AGU ADVANCING EARTH AND SPACE SCIENCES

Water Resources Research

RESEARCH ARTICLE
10.1029/2023WR034484

The Music of Rivers: The Mathematics of Waves Reveals Global Structure and Drivers of Streamflow Regime

Brian C. Brown[1,2], Aimee H. Fulerton[3], Darin Kopp[4], Flavia Tromboni[5,6], Arial J. Shogren[7,8], J. Angus Webb[9], Claire Ruffing[10], Matthew Heaton[11], Lenka Kuglerová[12], Daniel C. Allen[13], Lillian McGill[14], Jay P. Zarnetske[15], Matt R. Whiles[16], Jeremy B. Jones Jr.[17], and Benjamin W. Abbott[1]

[1]Department of Plant and Wildlife Sciences, Brigham Young University, Provo, UT, USA, [2]Department of Computer Science, Brigham Young University, Provo, UT, USA, [3]Fish Ecology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, USA, [4]Oak Ridge Institute for Science and Education (ORISE), Corvallis, OR, USA, [5]Rheinland-PfälzischeTechnische Universität Kaiserslautern Landau, Landau, Germany, [6]Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany, [7]Earth and Environmental Sciences Department, Michigan State University, East Lansing, MI, USA, [8]Department of Biological Sciences, University of Alabama, Tuscaloosa, AL, USA, [9]Water, Environment and Agriculture Program, Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC, Australia, [10]The Nature Conservancy in Oregon, Portland, OR, USA, [11]Department of Statistics, Brigham Young University, Provo, UT, USA, [12]Department of Forest Ecology and Management, Swedish University of Agricultural Sciences, Umeå, Sweden, [13]Department of Ecosystem Science and Management, Penn State, University Park, PA, USA, [14]Center for Quantitative Science, University of Washington, Seattle, WA, USA, [15]Department of Earth and Environmental Sciences, Michigan State University, East Lansing, MI, USA, [16]Soil and Water Sciences Department, University of Florida, Gainesville, FL, USA, [17]Institute for Arctic Biology, University of Alaska Fairbanks, Fairbanks, AK, USA

Abstract River flows change on timescales ranging from minutes to millennia. These vibrations in flow are tuned by diverse factors globally, for example, by dams suppressing multi-day variability or vegetation attenuating flood peaks in some ecosystems. The relative importance of the physical, biological, and human factors influencing flow is an active area of research, as is the related question of finding a common language for describing overall flow regime. Here, we addressed both topics using a daily river discharge data set for over 3,000 stations across the globe from 1988 to 2016. We first studied similarities between common flow regime quantification methods, including traditional flow metrics, wavelets, and Fourier analysis. Across all these methods, the flow data showed low-dimensional structure (i.e., simple and consistent patterns), suggesting that fundamental mechanisms are constraining flow regime. One such pattern was that day-to-day variability was negatively correlated with year-to-year variability. Additionally, the low-dimensional structure in river flow data correlated closely with only a small number of catchment characteristics, including catchment area, precipitation, and temperature—but notably not biome, dam surface area, or number of dams. We discuss these findings in a framework intended to be accessible to the many communities engaged in river research and management, while stressing the importance of letting structure in data guide both mechanistic inference and interdisciplinary discussion.

Key Points:

- We compared tools for describing streamflow timeseries, including streamflow metrics, wavelet, and Fourier analysis
- Each method indicated streamflow data are structured: variability at short timescales is negatively correlated with long timescales
- Globally, dams were less correlated with streamflow regime than catchment size and climate were

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

B. C. Brown,
bcbrown365@gmail.com

Citation:

Brown, B. C., Fulerton, A. H., Kopp, D., Tromboni, F., Shogren, A. J., Webb, J. A., et al. (2023). The music of rivers: The mathematics of waves reveals global structure and drivers of streamflow regime. *Water Resources Research*, *59*, e2023WR034484. https://doi.org/10.1029/2023WR034484

Received 24 JAN 2023
Accepted 2 JUL 2023

1. Introduction

River flow drives the structure and function of aquatic systems on sub-daily to decadal timescales, and sculpts landscapes on geological timescales from months to millennia (Fisher et al., 1998; Pinay et al., 2018; Tucker & Hancock, 2010). For people, variability in river flow regulates access to freshwater, with extreme flow events such as floods and droughts imposing immense personal and societal costs (Abbott et al., 2019; Van Loon et al., 2016; Vörösmarty et al., 2010). For ecosystems, water flow through soils, aquifers, and surface-water networks mediates aquatic and riparian biodiversity (Bochet et al., 2020; Hain et al., 2018; Poff & Zimmerman, 2010; Poff et al., 1997). Additionally, the direction, volume, and timing of flow define terrestrial-aquatic connectivity, and thereby mediate the delivery of biogeochemical substituents, including pollutants, to aquatic and marine ecosystems, including human pathogens, excess nutrients, and novel entities (Bernhardt et al., 2017; Frei et al., 2020; Gorski & Zimmer, 2021; S. Liu et al., 2022; Moatar et al., 2017; Raymond et al., 2016; Zarnetske et al., 2018). From the various viewpoints of human society, biogeochemical fluxes, and aquatic habitat, no single timescale stands out as singularly important regarding flow regime (Figure 1).

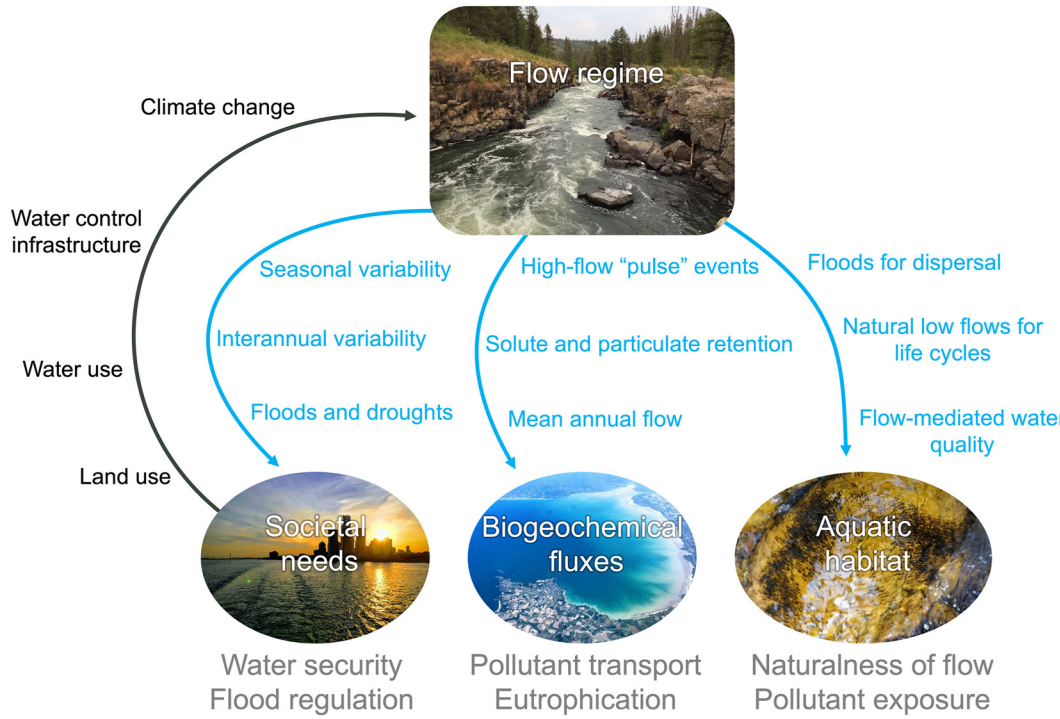

Figure 1. Conceptual diagram representing the societal, biogeochemical, and ecological importance of river flow regime. The relevant dimensions of flow regime are represented in blue, the consequences of flow regime are in gray, and the human influences on flow regime are in black.

In the Anthropocene, human interference with climate, land, and water is altering flow regimes, with direct impacts from dams and levees being better understood than indirect impacts imposed by land use and climatological changes (Chalise et al., 2021, 2023; Goeking & Tarboton, 2020; Zhou et al., 2015). These changes in turn impact aquatic ecosystems, human water security, and biogeochemical cycles at planetary scales (Abbott et al., 2019; Döll & Schmied, 2012; Gleeson et al., 2020; Hogeboom et al., 2020; Lin et al., 2019; Zipper et al., 2020), creating a pressing scientific challenge and opportunity to identify how climate and catchment parameters interact with direct human modifications of rivers such as dams and levees to influence river flow, and thereby shape the hydrological resilience of socioecological communities (Abbott et al., 2018; Berghuijs et al., 2019; Bunn & Arthington, 2002; Díaz et al., 2019; Harrison et al., 2018; Teixeira et al., 2019). As human modifications of land, water, and the atmosphere increase (Ascott et al., 2021; Minaudo et al., 2017; Zhou et al., 2015), understanding how to describe and predict river flow in the context of human involvement is becoming increasingly important.

Though global data sets of river flow observations and modeled natural discharge rates are now available (Alfieri et al., 2020; Gerten et al., 2008; Hales et al., 2022; Hannah et al., 2011; J. Liu et al., 2018; Masaki et al., 2017; McMahon et al., 2007), a unified framework for describing and interpreting river flow across multiple relevant timescales has not been widely adopted (McMillan, 2021). Efforts to quantify flow regime (e.g., variation in river discharge, including magnitude, frequency, duration, timing, and rate of change of flow) have resulted in the development of over 600 metrics (George et al., 2021; Gnann et al., 2021; Jones et al., 2014; Poff et al., 1997). Many of these metrics are designed to describe key features of flow pertinent to society and ecosystems, such as interannual variability of low flows and the seasonal timing of flooding. Other metrics are designed to quantify hydrological processes such as the rate of increase and decrease of flow following rain, or baseflow conditions between storm events (Archfield et al., 2014; Carlisle et al., 2010; McMillan, 2021). While all these metrics are useful for individual studies and management, their sheer range and redundancy creates a problem of comparability at regional to global scales (Olden & Poff, 2003). In addition to these "traditional" flow metrics, the hydrological literature has widely used the spectral properties of flow regime obtained via wavelet decompositions and Fourier analysis, two related analytical techniques which leverage the concise mathematics of waves to describe variability at multiple timescales simultaneously. These analyses effectively identify which timescales are most important in a timeseries (Carey et al., 2013; Labat, 2010; Sang, 2013; Smith et al., 1998; White et al., 2005), and

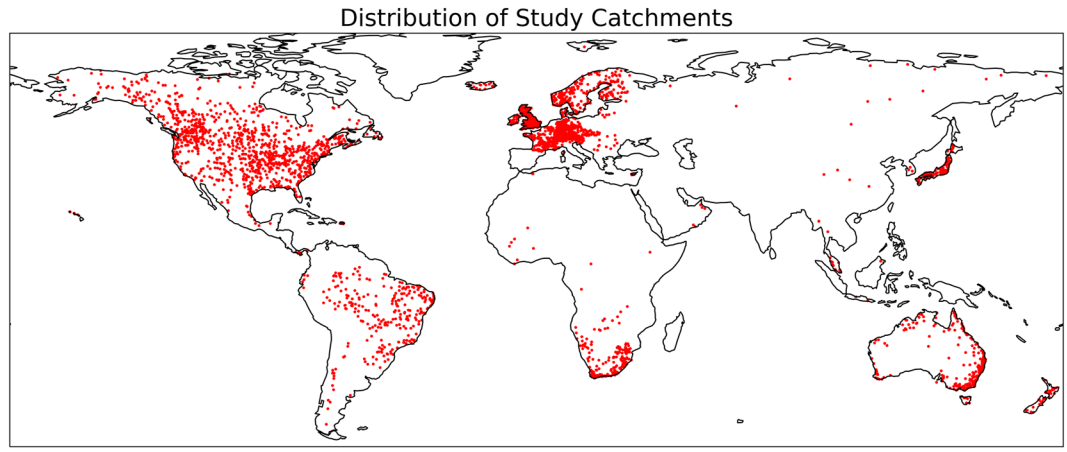

Figure 2. Distribution of catchments used in this study. Each dot represents a flow gage with nine or more years of daily flow data during the study period and minimal gaps. While the uneven spatial coverage of the stations precludes a representative global sample, we note that an extensive variety of climatic and socioecological conditions are represented.

have recently been used to quantify the impact of human water infrastructure on natural flow regimes (Ashraf et al., 2022; Chalise et al., 2023; F.-C. Wu et al., 2015). However, wavelet decompositions and the traditional flow metrics are rarely used in concert, and similarities and differences between the two approaches have not been quantified.

The complexity of measuring and characterizing flow regime likely contributes to the persistent difficulty in understanding the factors influencing river flow. Even when constraining the discussion to specific timescales or metrics such as annual flow or runoff ratios during storm events, the physical, biological, and human controls on flow at the catchment scale are still being debated (Lane et al., 2017; Lin et al., 2019; Reaver et al., 2020; Savenije, 2018; Sivapalan, 2006; Tetzlaff et al., 2008; Zhou et al., 2015). Climatic, surface, and subsurface parameters have been proposed as primary controls on the timing and magnitude of river flow across sites, including the amount of soil and aquifer water storage, the relative availability of energy and water, the configuration and size of the surface water network, and the extent and type of vegetation (Carlisle et al., 2010; Lane et al., 2017; Oldfield, 2016; Ryo et al., 2015; Sanborn & Bledsoe, 2006; Zhou et al., 2015). Understanding variation and similarity in flow regimes across biomes and ecoregions could reveal drivers of aquatic ecology and explain differences in success of water management and ecosystem protections in different conditions (Berghuijs et al., 2019; Bunn & Arthington, 2002; Zhou et al., 2015).

In this context, we analyzed a globally-sampled data set of river flow to compare methods for characterizing flow regime and to identify flow relationships with climatic and catchment features. We combined traditional flow metrics with wavelet and Fourier analysis to describe 3,120 time series of river flow, each with over 9 years of continuous data between 1988 and 2016 (Figure 2). In addition to quantifying the relationship between streamflow metrics and frequency analyses, we sought to identify which climatic, geomorphological, and human attributes are most important for determining variability in flow at timescales ranging from days to a decade. These flow behaviors across timescales are rarely analyzed in concert (McMillan, 2021; Olden & Poff, 2003), but we further hypothesized that variability in flow at different timescales acts as an interacting set of variables, meaning that changes in flow volume that last only a few days are fundamentally linked to changes in flow volume that last several years. If present, these linkages would imply low-dimensional structure in streamflow data, which we believe would be fundamental to developing a concise vocabulary for describing streamflow regime and understanding its controls. Because the same climatic and catchment attributes influence flow on multiple timescales, considering potential interactions across timescales could open new pathways towards understanding and predicting flow regimes. For example, because the relative abundance of energy and water influence vegetation and soil development (Malone et al., 2018; Tank et al., 2020), hot and dry catchments could simultaneously exhibit high seasonal variability in flow and greater extractive human water use, causing long-term reductions in the water table. Likewise, because larger catchments integrate heterogenous subcatchments over larger and longer spatiotemporal scales (Chezik et al., 2017; Dupas et al., 2019; Levia et al., 2020), we predict they will show less short-term variability but greater sensitivity to long-term changes in water balance.

2. Materials and Methods

2.1. River Flow and Catchment Characteristics Data

We obtained daily river discharge time series from the Global Runoff Data Center (GRDC; https://www.bafg.de/GRDC). We used several criteria to select from the 6,544 stations with discharge data from a recent 30-year period of interest (1988–2016). Because continuous time series are required for the calculation of many flow metrics, we first removed stations that had less than nine complete water years over the period of interest. This left us with 4,762 candidate stations (2,399 without any gaps and 2,363 with some gaps). For all stations, we removed records for partial water years, that is, those before the first complete water year or after the last complete water year. For those time series with gaps, we computed the number of days in each missing period and the total number of missing periods. We summarized the number of missing days (e.g., minimum, mean, maximum, and percentiles), and calculated the proportion of days in the record for which data were available. We filled gaps via linear interpolation for stations that met the following criteria: <25% missing data, the longest data gap was less than 2 years, and the 75th percentile of consecutive days of missing data was less than 3 months. For stations that passed this test (1,163 of the 2,363), we visually inspected the result of interpolation to ensure that obvious peaks or troughs in each station's data record were not omitted. We discarded 104 stations that showed anomalous effects during interpolation, leaving 1,059 stations. For the stations with gaps that did not meet our criteria, 509 were located more than 1 km from an included station, and many were in data-sparse regions with relatively few observations. Despite their gaps, some of these stations had long data records within the period of interest. Therefore, we determined which stations had sufficiently long (>9 years) intact stretches that could be extracted from the longer time series. We were able to salvage an additional 227 stations using an automated approach followed by visual inspection. Therefore, our final set of stations included those with complete records (2,399), those with interpolation that met our inclusion criteria (1,059), and additional salvaged stations (227), for a total of 3,685 stations—56% of the original GRDC stations. We chose to adjust southern hemisphere flow data forward by 4 months to account for the opposite seasonality between hemispheres (see Text S1 in Supporting Information S1 for a more complete treatment of this decision).

The GRDC streamflow data set reports the upstream catchment area associated with each station but does not directly reference them to the hydrography we used in this study. As such, differences in data sources could have created mismatches between the location of a GRDC station and the upstream catchment we delineated from the integrated Shuttle Radar Topography Mission Digital Elevation Model (DEM) and the GTOPO30 DEM (http://files.ntsg.umt.edu/data/DRT). Following Barbarossa et al. (2018), we geo-referenced each station to the pixel that was most similar in catchment area and within 5 km from its original location. We designated stations as high, medium, or low quality if the difference in catchment area was <5%, 5%–10%, or 10%–50%, respectively (Barbarossa et al., 2018).

After delineating each watershed, we extracted 117 variables obtained from a variety of geospatial data sources (Table S1). These variables capture the stream network structure, climate, landcover (including lakes and soils), and anthropogenic impacts (including population density and reservoirs) upstream of each GRDC location. Depending on the parameter, we calculated cumulative values (e.g., total precipitation) or catchment means (e.g., mean annual temperature). Because the configuration and density of stream networks can influence propagation of water and solutes (Godsey & Kirchner, 2014; Helton et al., 2011), we quantified stream network structure with TauDem (Terrain Analysis Using Digital Elevation Models, https://hydrology.usu.edu/taudem/taudem5/) using blended HydroSHEDS and HYDRO1K data (with HydroSHEDS data used preferentially) parsed by the DRT method prepared by (H. Wu et al., 2011). TauDem is an open-source software which implements highly parallelized algorithms that can efficiently process large data sets (Barbarossa et al., 2018). We used the AreaD8 function to calculate the number of pixels upslope from a station (i.e., the flow accumulation grid) and the GridNet function to calculate stream network attributes (e.g., stream order and total network length). Alternatively, the MERIT-Hydro hydrography data set (Yamazaki et al., 2019), which has more complete global coverage and is based on more recent elevation data which is generally more accurate for small rivers in forested areas, could have been used. However, at the time this study began this resource was not yet available. We also obtained data describing dam number and surface area from Lehner et al. (2011), climate data from Fick and Hijmans (2017), land use data from Amatulli et al. (2018), and human impact data from the "last of the wild v2" database (https://sedac.ciesin.columbia.edu/data/collection/wildareas-v2). Because of missing geospatial data, as few as 3,120

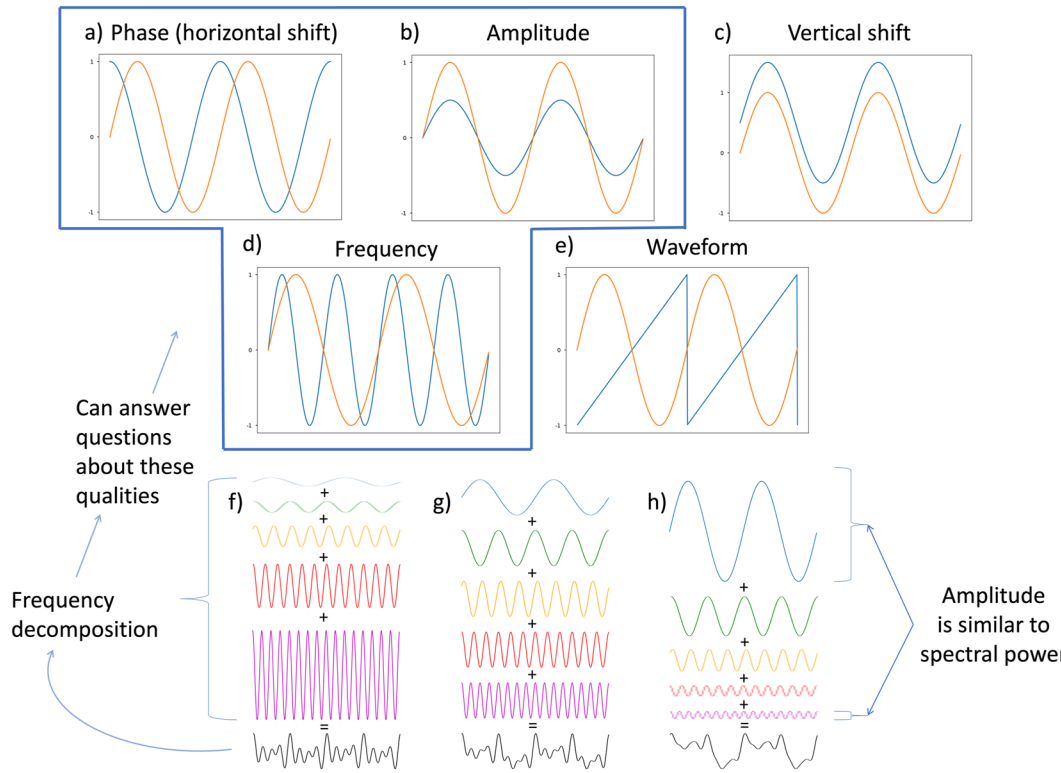

Figure 3. (a–e) Five fundamental components of flow regime (or any time series): Many of the behaviors in streamflow timeseries relate back to these five fundamental principles. The lower portion of the figure represents frequency decompositions of three timeseries: (f) a timeseries dominated by high-frequency variability, (g) a timeseries with equal variability across all timescales, and (h) a timeseries dominated by low-frequency variability. In each example, adding together the five colored waves produces the complex curve shown in black at the bottom.

streams were used in some subsequent analyses, and up to 14 of the derived catchment characteristics were ignored, including redundant measures of catchment size as well as measures of human population density.

2.2. Characterizing Flow Regime—Conceptual Introductions

Frequency decompositions rely on the fact that timeseries are fundamentally related to waves. Waves are phenomena that repeat through time, and can occur in any number of dimensions, though in this scope we consider one-dimensional waves that represent a single variable changing through time. Waves can be described with five fundamental descriptors: (a) *phase*: horizontal shift or timing of the oscillation; (b) *amplitude*: magnitude of variation around the mean, or equilibrium point; (c) *vertical shift*: changes in the level of the mean; (d) *frequnecy*: the number of oscillations that occur within a given timeframe; and (e) *waveform*: differences in the shape of the repeating pattern (e.g., a typical sinusoid curve, or a more unusual shape such as a square, triangle, or saw-tooth shape). Together, phase, amplitude, vertical shift, frequency, and waveform describe essentially any difference between any two waves (Figure 3).

The terms phase, amplitude, vertical shift, frequency, and waveform have familiar analogs in hydrology. Consider an imaginary catchment with a hydrograph that follows a perfect sinusoidal curve that goes up and down over the course of a year. The amplitude of this wave plus any vertical shift relates closely to the familiar concept of peak annual flow, and vertical shift minus the amplitude relates to baseflow, or minimum flow. In this catchment, the frequency of one cycle per *year* relates to the timescale containing the most variance. The phase of the wave indicates the time of year in which snowmelt or monsoon rains occur and would be opposite for a northern versus southern hemisphere catchment. The waveform relates to the rate of rise or fall of the year-long increase and decrease in flow. Now imagine a second catchment whose flow follows another perfect sinusoidal wave, but which oscillates at one cycle per 2 weeks. This "flashy" catchment neither accrues nor loses long-term storage and might hypothetically

occur in a warm climate with no snow but with identical rain storms every 2 weeks. Both catchments exhibit variability in flow, but in the first, the variance is maximized at the timescale of 1 year, and in the second at 2 weeks.

However, most catchments exhibit both flashiness and some seasonal variability. Adding together the perfect sine wave from the first catchment with the perfect sine wave from the second catchment would produce a complex curve that can no longer be described with amplitude, phase, vertical shift, waveform, and frequency, but which more closely resembles a real-world catchment. This process of adding new catchments that epitomize behavior on different timescales could be repeated infinitely many times, producing an ever more complex, and hence more realistic hydrograph, but which could always be decomposed back into a collection of simple waves that can individually be described by the same few, succinct variables. Mathematical tools exist to run this process *backwards*—decomposing a timeseries into a set of perfect sinusoids that together recreate the original timeseries. These are known as *frequency decompositions and* can be thought of as functioning similarly to a prism, which decomposes white light into a rainbow of colors ranging from high to low frequency, or to a computer program that takes in the sound recording of a symphony and outputs a musical score of notes representing air vibrations at particular frequencies. No matter the timeseries, the amplitudes of the resultant decomposed waves at different frequencies relate to the amount of variability in the data that occurs on those timescales, reported in a characteristic known as *spectral power*. Spectral power thus provides a unit for describing variability in streamflow across every timescale present in a hydrograph. Formally, the *power spectrum* is defined as follows: for a given time, τ (i.e., the day in the year), and a given timescale, s (i.e., the number of days over which the signal varies), and a *wavelet function* of those two variables, wavelet(τ,s), the spectral power is given by:

$$\text{power}(\tau, s) = \frac{1}{s} \cdot |\text{wavelet}(\tau, s)|^2$$

The wavelet function belongs to a family of similarly-shaped waves defined by a "mother wavelet." When this family of functions comprises non-decaying sinusoids, the frequency decomposition is known as a Fourier transform. However, in most modern applications, the mother wavelet has some decay across time, allowing for a blended time-frequency decomposition known as a wavelet analysis. The waveform of the mother wavelet (e.g., Figure 3 panel e) impacts which types of patterns will be most easily identified in the data. The Morlet wavelet is commonly used in many climate-linked timeseries and is a standard choice for hydrology timeseries (Ashraf et al., 2022; Torrence & Compo, 1998; White et al., 2005), but other common mother wavelets include the Haar, Ricker, Meyer, and Daubechies wavelets. Many of these wavelets have asymmetric shapes that may be helpful in identifying irregular wave patterns in data, such as a rise rate or a fall rate of a river following storm events.

Decision trees are a primal machine learning model that are foundational to many more complex models, such as random forests and gradient boosting forests. Conceptually, decision trees take in an array of prediction features and step-by-step combine multiple points of data along the feature array. Using relatively simple logic, they distill information further and further until a single prediction is made (Myles et al., 2004). Decision trees are generally known to have high bias (typically viewed as undesirable) with low variance, though they are still occasionally used because of their inherent interpretability.

Random forests are called "forests" because they comprise many individual decision trees, usually of significant depth, whose collective predictions are averaged to produce an output that is generally less biased and more accurate than individual decision tree regressors (Biau & Scornet, 2016). The "random" aspect comes from an innovation in 2001 where successive trees are trained on independent random samples with replacement from the larger data set (Breiman, 2001).

Gradient boosting regressors are similar to random forest regressors, but they differ in that new trees are added in a way that minimizes error in a targeted, rather than a random fashion. This targeted approach is achieved by adding new trees according to the gradient of a user-defined loss function, which is simply a function which characterizes the error of the model (Elith et al., 2008).

Principal Components Analysis, or *PCA*, projects high dimensional data onto a lower dimensional space where each axis is a linear combination of the original variables in the high dimensional space, and where the number of dimensions projected onto is the user's choice. As an intuitive example, imagine a "high-dimensional" data set with two variables, x and y. If, for every step in the x direction, data tend to take two steps in the y direction, the two variables are redundant and linearly related; total least squares linear regression would draw a line through the two axes with a slope of ~2. PCA on these two axes would project data points onto that regression

line. That is, instead of listing data points by their x and y coordinates, the PCA projection would list data points by their location on a new axis, z, which is two parts y, and one part x. The "two parts" and "one part" that describe how much each original axis contributes to the new projected axis are referred to as the *loadings matrix*. The loadings matrix effectively describes how correlated (positive or negative) each of the original axes in the high-dimensional space is with the low-dimensional axes PCA projects the data onto. Thus, like wavelet decompositions, PCA identifies variability in data. But instead of identifying variability at different timescales in a timeseries, PCA identifies variables (or combination of variables) in tabular data along which the data vary most. If a group of original variables (columns) have high magnitude loadings for a given principal component, then that principal component can be thought of as a combination of those original variables. In other words, the resulting components from PCA describe low-dimensional linear structure in data which in turn corresponds to simple, high-level concepts. Examining the loadings matrix is one of the best methods for adding interpretability to the abstract components that result from a PCA projection.

2.3. Streamflow Analysis

2.3.1. Quantifying Streamflow Regime

Traditional methods for describing streamflow regime include over 600 flow regime metrics available in the literature that describe concepts such as variability in monthly flow, annual maximum of 90-day moving average of flow, low flood pulse count, etc., and are collectively both diverse and in many cases redundant (Olden & Poff, 2003). We calculated a subset of these metrics that are commonly used in hydrology, based on the availability of statistical packages and recent flow regime papers. First, we calculated the "Magnificent 7" (Archfield et al., 2014). Second, we calculated 171 metrics from the Hydrological Index Tool (Henriksen et al., 2006), reimplemented in the EflowStats package (Archfield et al., 2014). Finally, we calculated the 11 metrics from Sabo and Post (2008), for a total of 189 metrics. Given previously identified redundancy in streamflow metrics (Olden & Poff, 2003), we are confident that this set covers the full range of hydrological variability.

To identify the amount of redundancy in the selected metrics we applied PCA using the R package FRK and the NNGP method (Zammit-Mangion & Cressie, 2017). We retained seven dimensions for further analysis and which we hereafter refer to as "PCA Metrics." These seven dimensions collectively explained 68% of the variability in the 189 streamflow metrics. We summarized the top correlates suggested by the loadings matrix (see Section 2.2) to provide qualitative descriptors of the resulting metrics. Separately, we quantified streamflow regime using two related methods of frequency decompositions: Fourier and wavelet analysis. Classically, frequency decompositions are performed using the discrete-time Fourier transform, yielding an output that quantifies the variability in the signal at different timescales using a unit called "spectral power" (see Section 2.2, Unpingco 2014). Recently it has become more common to use a related analysis called a Wavelet transform (Carey et al., 2013; Labat, 2010; Sang, 2013; Smith et al., 1998; White et al., 2005), which generates a blended time-frequency decomposition of the input. For the scope of this contribution, we opted to not analyze time-varying components of frequency decompositions, but we were still interested in comparing similarities between flow metrics, wavelet analysis, and Fourier analysis. We therefore calculated a Fourier decomposition of each timeseries in addition to a time-averaged wavelet decomposition of each timeseries, meaning we obtained two semi-redundant frequency representations of our data. However, because the wavelet transform uses a differently-shaped "mother wavelet," and because it is not perfectly information-preserving when averaged across time, it will have some differences (Schmidbauer & Roesch, 2018; Torrence & Compo, 1998). We found that some information loss was advantageous because it reduced noise that otherwise obscured patterns in the data. Most of the results presented in the main manuscript therefore report time-averaged wavelet decomposition results, while Fourier-related results (which, even with more noise, were very similar) are presented in the supplementary material. We calculated the time-averaged wavelet decomposition using the default settings of the WaveletComp R package (Rösch & Schmidbauer, 2018), specifically using the Morlet wavelet, which is preferred for hydrology (Ashraf et al., 2022; Torrence & Compo, 1998). Fourier transforms were calculated using NumPy and SciPy with a Blackman window to prevent spectral leakage (Harris et al., 2020; Virtanen et al., 2020).

2.3.2. Similarities Between Streamflow Metrics and Frequency Decomposition

We calculated Spearman correlations between each frequency's spectral power (wavelet only) and each of the 189 flow metrics across all catchments in the data set. Seeking to confirm the results of the correla-

tion analysis through an alternate technique, we also trained machine learning models to predict each of the streamflow metrics using the frequency decompositions (wavelet only) as inputs. To account for variability between models and divisions of data, 18 models were trained on each of the 189 streamflow metrics. For each metric, 9 were random forest regressors and 9 were gradient boosting regressors. Data were divided with an 80:20 training to testing ratio, with the divisions done randomly and independently for each model. Models were then validated on the 20% portion reserved for testing and an r-squared was calculated between model output and the actual values of the given streamflow metric for the 20% testing data. Finally, the "feature importances" were extracted from each model to determine which input features were most important in the models' decision-making processes (Frei et al., 2021). Models were implemented in Python using the Sci-kit Learn library and feature importances were extracted using the "feature_importance_" method (Pedregosa et al., 2011).

To connect the previously-calculated PCA axes to frequency analyses (wavelet only), we ran a Spearman correlation analyses between each of these PCA metrics and the spectral power of each frequency. Similar to each of the 189 flow metrics, we also trained 360 machine learning models, with an even split between random forest regressors and gradient boosting regressor models, to predict each PCA metric using the frequency domain, again with a unique, random 80:20 split between training and testing data.

Structure in the outputs of these three analyses suggested that variability at shorter timescales was linked to variability at longer timescales. To isolate and quantify this phenomenon, we calculated the pairwise spearman rank correlation between the spectral powers at each frequency and the spectral powers at all other frequencies (with both Fourier and wavelet decompositions).

2.3.3. Identifying Controls on Streamflow Regime

Whereas in the previous section we sought to quantify similarities between methods for describing streamflow regime, in the following section we describe analyses in which we sought to understand which catchment characteristics are the best predictors (and therefore likely controls) of flow regime. Consequently, we trained three separate machine learning model classes, decision tree regressors, random forest regressors, and gradient boosting regressors, to predict each of the PCA metrics (which we consider concise surrogates for the full 189 flow metrics we calculated) from the 117 catchment characteristic input features. We used the *k*-folds validation process with a *k* of 10, meaning that we trained 10 separate models on different 90:10 splits of data and validated each model on the unique 10% of the data not used for training that model. Validation was done by calculating the model *r*-squared between predictions and ground truth. Prior to training, data were normalized using min/max normalization. As before, feature importances were extracted to understand which input features (i.e., catchment characteristics) were most important in determining flow regime. To confirm these results, we also ran a Spearman correlation analysis between the 117 streamflow metrics and the spectral power for each frequency. All correlation analyses in this paper were implemented using the SciPy library in python (Virtanen et al., 2020).

Additionally, we trained two classes of machine learning models to predict the spectral power of streamflow timeseries at different frequencies (wavelet only). Similar to the machine learning analysis predicting streamflow metrics from wavelet analyses, for each of the 1,101 frequencies identified by the wavelet analysis, we trained 27 random forest regressors and 27 gradient boosting regressors to predict spectral power using catchment characteristics, for a total of 59,454 models. Prior to training, 14 columns were removed because they were too sparse. These included measures of human population density and measures of catchment area. We divided data with an 80:20 training to testing ratio, with the divisions done randomly and independently for each model. Models were then validated on the 20% portion reserved for testing and an r-squared was calculated between model output and the actual values of the given streamflow metric for the 20% testing data. Data were normalized to be mean zero and standard deviation of 1. The importance of each prediction feature was then extracted from the models and features were grouped into categories to determine which categories of features were most important for predicting streamflow regime. These results were also confirmed by calculating the Spearman correlation between each of the 117 catchment characteristics and the spectral power for each frequency (using both Fourier and wavelet decompositions).

Table 1
List of Top Seven Principal Components Derived From 189 Flow Metrics Calculated for 3,685 River Flow Time Series

PCA (% variance explained)	Name	Description of correlates	Hypothesized cause(s)
1, (26%)	Magnitude	High total amount of flow, high minimum flows (rarely dry), and low flow variation in high flows	Big rivers
2, (16%)	High-frequency stability	Long-lasting but infrequent high flows, large portion of flux occurs at high flows, few reversals or short-term changes in direction, few low flow events, red or black noise in the daily discharge data, and strong and skewed seasonal signal	Big rivers (surface-dominated or unduly influenced by high-flow tributaries)
3, (9%)	Low-frequency stability	High interannual flow stability, low event flashiness, predictable interannual high flows, low flood frequency, high base flow	High overall storage, low synchrony among sub-catchments, groundwater dominated
4, (6%)	Interannual variability	Low interannual stability in high flow magnitude and duration, low stability in annual flow, low seasonality, low annual flow (specific and absolute), variable timing of annual min and max flow, frequent floods, skewed annual flows, variable event response, short-lived flow events	Arid or semi-arid sites
5, (6%)	High and stable baseflow	High baseflow (rarely dry), high skewness, low exceedance flows, frequent floods of moderate magnitude, variable flow, variable moderate flows, variable event response	Near-surface groundwater
6, (3%)	Variable baseflow	Variability in number of no-flow days, very few and short baseflow pulses, high flow constancy and predictability (same timing of variation), more zero-flow months, little range in daily flows, little autocorrelation, higher minimum annual flow, later arrival of minimum flow (freshet pattern), high skewness, more no-flow days	Snowmelt, intermittency, semi-arid, flashy
7, (2%)	Daily variability	High spread in daily flows, low magnitude of interannual high flows, consistently rapid changes in flow, low variability in no-flow days, short and small pulses, more no-flow months, seasonally variable flooding, high signal to noise, variable monthly flows, later arrival of max flows (monsoonal), high interannual variability, frequent floods	Arid, small headwaters, Mediterranean

3. Results

3.1. Similarities Between Streamflow Metrics and Frequency Decompositions

Several lines of evidence suggested that streamflow metrics and frequency decompositions carry a substantial amount of similar information. For example, the average maximum Spearman correlation coefficient between the 189 flow metrics and any frequency in the frequency decompositions was 0.46 (Figures S1 and S2 in Supporting Information S1). Similarly, the average r-squared for machine learning models trained to predict the 189 flow metrics exclusively using the frequency decomposition was 0.33 (Figures S3 and S4 in Supporting Information S1). And finally, the average r-squared for machine learning models that were trained to predict the seven PCA flow metrics exclusively using the frequency decomposition was 0.42. Together, these results indicate that frequency decompositions such as the wavelet transform describe between 30% and 45% of the same information as streamflow metrics (or alternatively that both approaches describe phenomena that are highly correlated).

3.2. Low-Dimensional Structure in Streamflow Timeseries

PCA analysis of streamflow metrics suggested that low-dimensional linear structure exists alongside a nontrivial amount of nonlinear structure in streamflow data: 68% of the variance in the original 189 flow metrics could be explained in 7 PCA axes, each capturing increasingly less variability in the data (Figure S5 in Supporting Information S1). PCA metrics that explained more variance in the original 189 metrics tended to correlate more strongly to the frequency domain (e.g., metrics 1–4), while those that explained less variance in the original metrics tended to relate less strongly to the frequency domain (e.g., metrics 5–7, see Figure S6 in Supporting Information S1). A summary of the loading matrices of each metric are found in Table 1, and more extensive descriptions of the matrices are given in Tables S2–S8 in Supporting Information S1. The spatial distributions of the metrics across the globe are plotted in Figure S7 in Supporting Information S1.

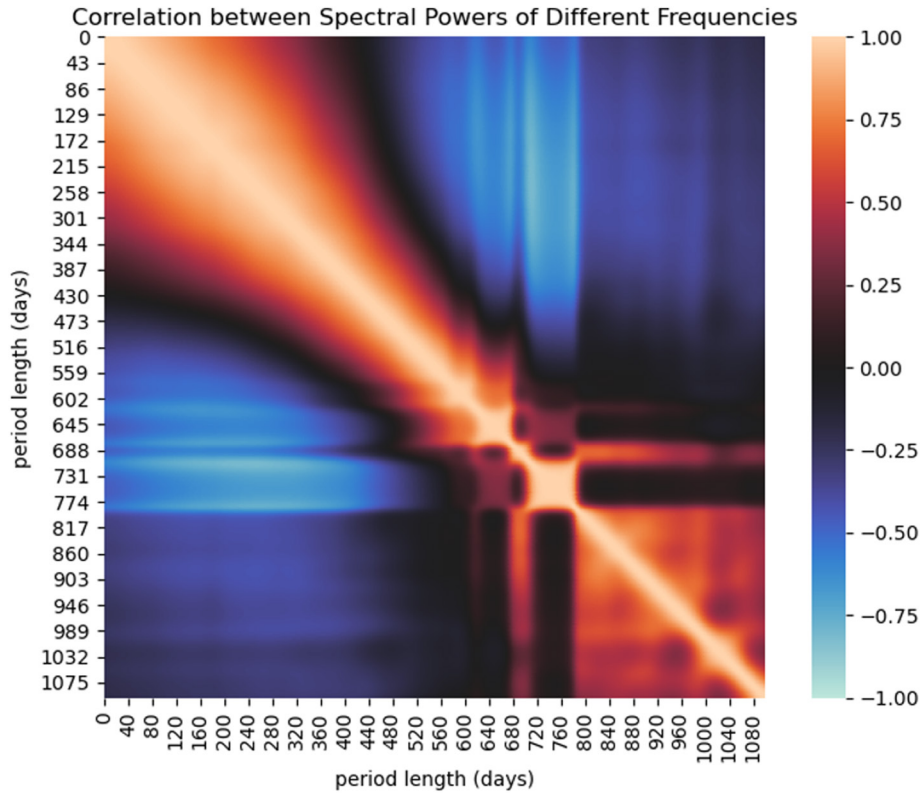

Figure 4. Pairwise correlations between spectral power for each period length. The coefficient of correlation from the spearman correlation is represented as color, with brighter orange representing a stronger positive monotonic relationship and brighter blue representing a stronger negative monotonic relationship. This unexpectedly simple structure in the data suggests that even non-redundant streamflow metrics may be correlated with each other because inherent correlations exist between variability at short timescales and variability at long timescales. The coherent patterns imply that streamflow data are low-dimensional and easily compressed (described). This low-dimensionality also suggests that a relatively small number of mechanisms may govern streamflow variability across multiple timescales.

Wavelet analysis of streamflow timeseries also suggested that streamflow data are highly compressible (and therefore easily summarized). Spectral power of high frequencies was negatively correlated with spectral power of low frequencies (Figure 4, Figure S8 in Supporting Information S1). This indicates that a tradeoff exists between changes in flow that occur over several days and changes in flow that occur over several months or years. This structure also indicates that streamflow data are fairly low-dimensional when represented in the frequency domain. Figure 5 anecdotally demonstrates the tradeoff between long and short-term variability in flow data using example hydrographs and their associated frequency decompositions from our data set.

We also found that on average, variability in flow occurs at four distinct timescales (Figure 6). These are multi-day variations, multi-month variations, annual variations, and multi-annual variations. Annual variation was the strongest, followed by multi-month variation and multi-day.

3.3. Identifying Controls on Streamflow Regime With PCA Metrics

Three types of machine learning models corroboratively suggested that just a few catchment characteristics are sufficient to accurately predict flow regime (as measured by the PCA metrics, Figures S9 and S10 in Supporting Information S1). These included dominant contributions of cumulative precipitation for PCA metrics 1, 5, and 7, catchment area for metric 2, climate variables for metrics 3 and 4, and land cover for metrics 3, 5, and 6. In addition, the length of the timeseries was an important feature for several metrics. The *r*-squared values across models decreased from the higher variance-explaining metrics to the lower variance explaining metrics. Specifically, model accuracy decreased from a maximum of ~0.85 for metric 1 to a maximum of ~0.45 for metric 7 (Figure S10 in Supporting Information S1). To visualize the relationship

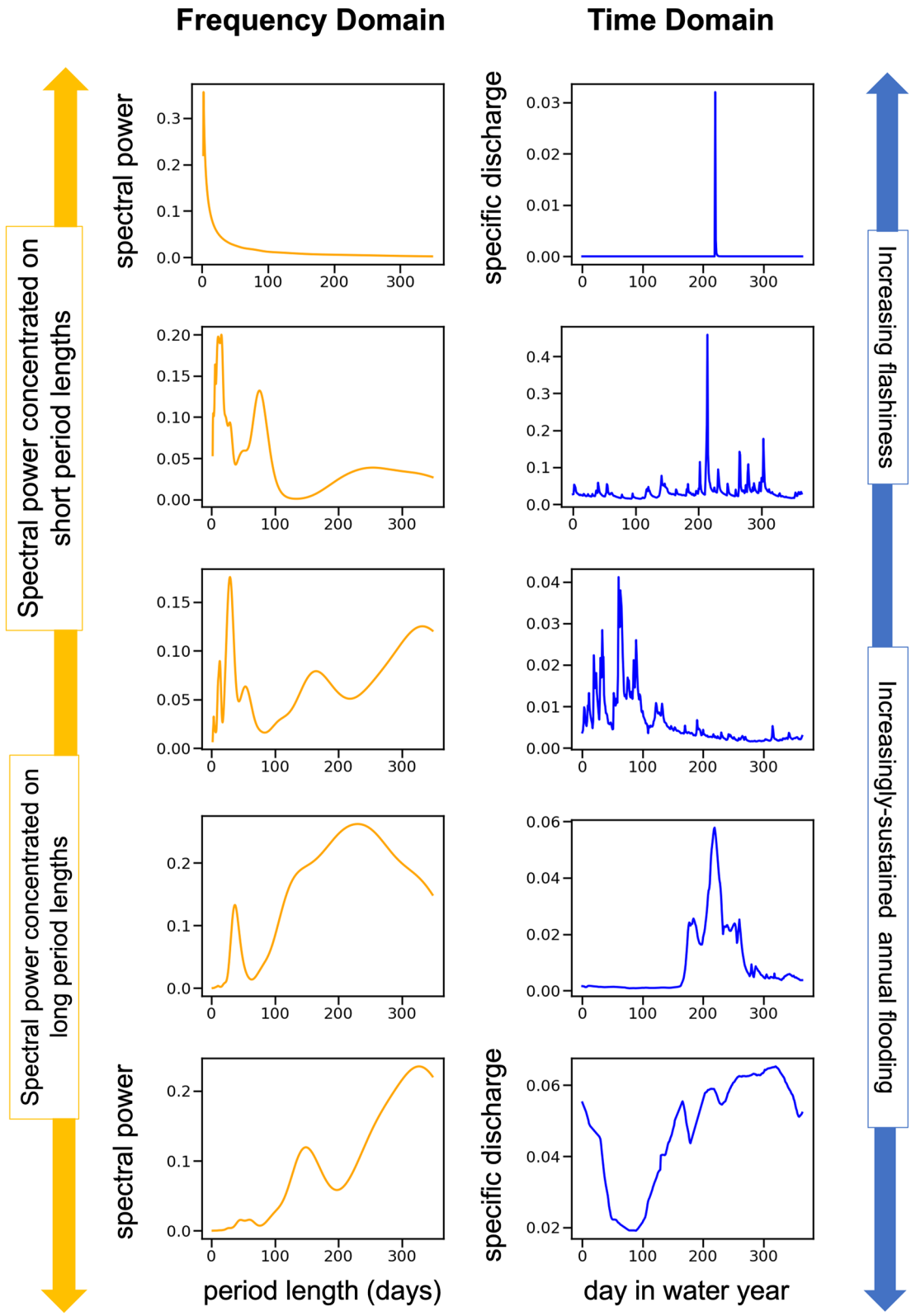

Figure 5. Comparison between frequency domain and time domain representations of hydrographs. Frequency domain representations (pictured on the left in yellow) show how much variability in the data occurs along a particular time scale, while their corresponding time-domain representations (pictured on the right in blue) show the raw time series measured by streamflow gauges. The frequency domain representation allows for the quantification of many qualitative attributes of flow regime properties that might otherwise take dozens of metrics to fully describe. Note that these example hydrographs anecdotally demonstrate the global phenomenon that spectral power at short period lengths is negatively correlated with spectral power at long period lengths.

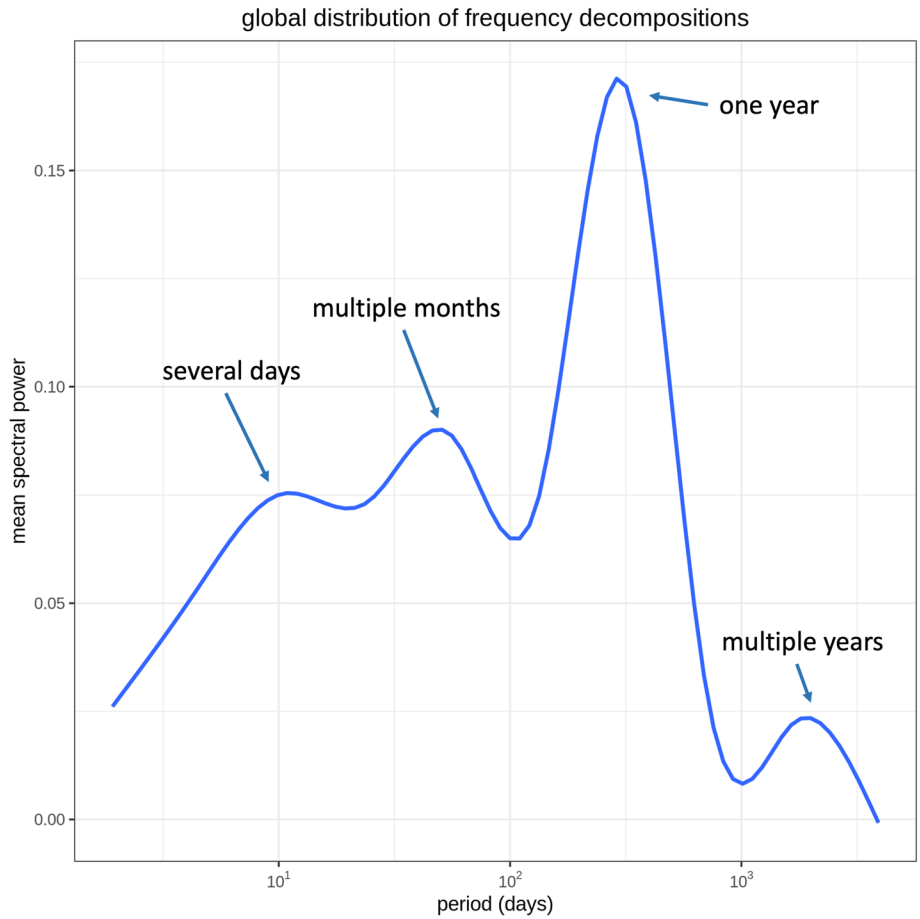

Figure 6. Mean global spectral decomposition of streamflow timeseries. The horizontal axis represents the period length of oscillations in streamflow timeseries on a logarithmic scale, while the vertical axis represents the spectral power, a unit that can be intuitively understood as how much a given timescale contributes to the variance in the data.

between flow metrics and the subset of catchment characteristics that the machine learning analysis suggested were important, as well as catchment characteristics suggested to be important by hydrological theory, we plotted the relationships between the PCA metrics and selected catchment characteristics (Figures 7 and 8, Figures S11–14 in Supporting Information S1). The dominant role of catchment size was quite clear, including non-linear relationships between catchment size and metrics 3, 5, and 7, in which the largest streams tended to behave similarly to the smallest streams. Several notable null results were also apparent: relationships between biomes were surprisingly ambiguous given that biome delineations are defined by temperature and precipitation. In addition, the relationship between flow regime and dam count and reservoir surface area was relatively weak. This was true whether flow regime was quantified via PCA metrics (Figure 8), or as detailed below, wavelet analysis (Figure S15 in Supporting Information S1), or Fourier analysis (Figure S16 in Supporting Information S1).

3.4. Identifying Controls on Streamflow Regime With Frequency Decompositions

Many of the drivers of flow regime suggested by the PCA flow metrics were also highlighted by the wavelet analysis (Figure 9) and Fourier analysis (Figure S17 in Supporting Information S1). These frequency-based analyses also demonstrated that single catchment features impacted multiple timescales. For example, catchment size was negatively correlated with high frequency (short-term) phenomena but positively correlated with low frequency (long-term) phenomena. Temperature followed a more complex relationship where high winter temperatures were positively correlated with multi-day phenomena and negatively correlated with multi-month to year-long phenomena. In contrast, summer temperatures most strongly correlated with multi-year phenomena.

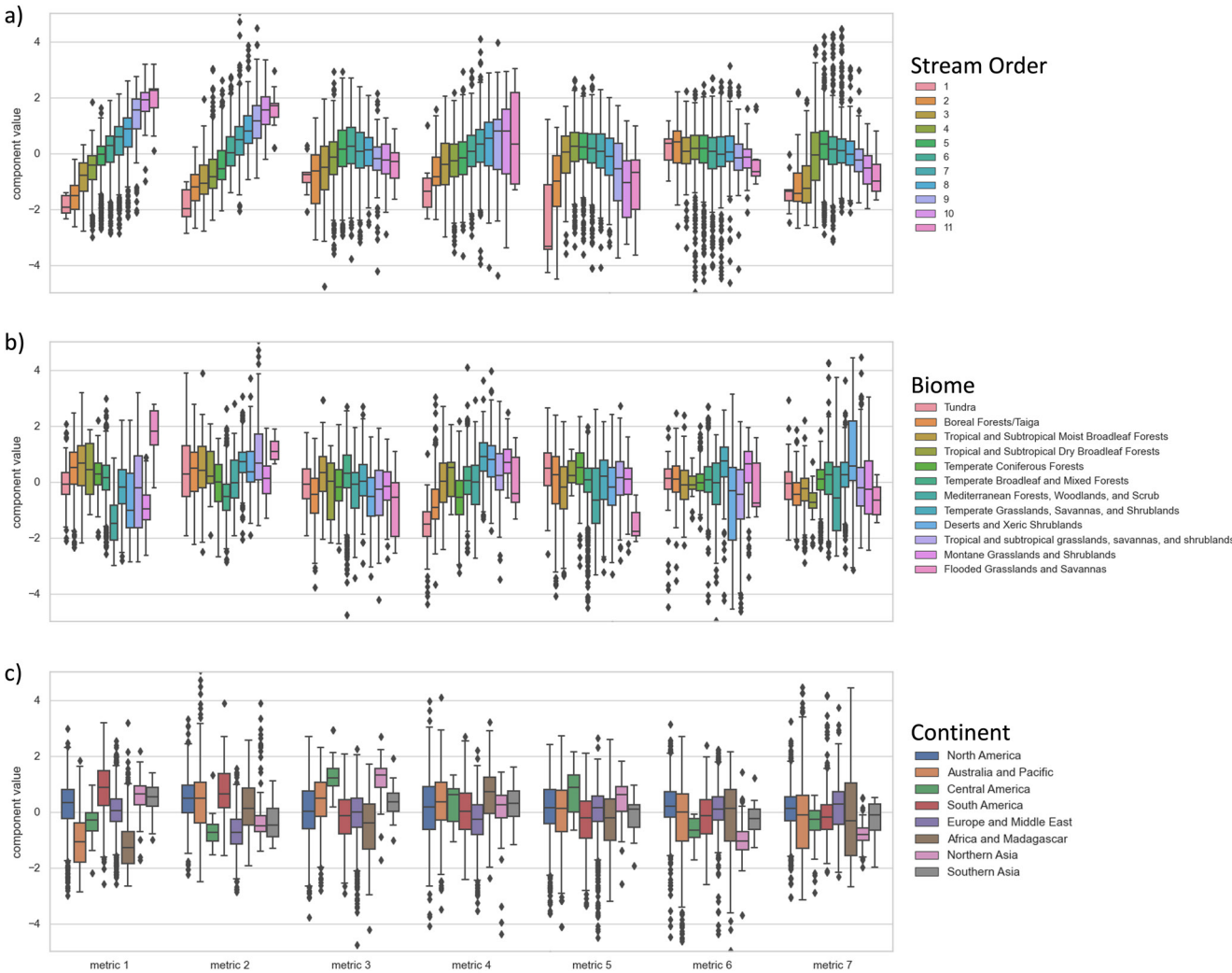

Figure 7. The seven Principal Components Analysis (PCA) flow metrics divided according to (a) stream order, (b) biome, and (c) continental region. Box plots are constructed using standard conventions: boxes are the range of the first through third quartiles, lines represent the range between the minimum and maximum values when these values are within 1.5 times the inter-quartile range (IQR), and dots represent outliers beyond 1.5 IQR. Note that the relationship between stream order and flow regime is much stronger than the relationship between biome and flow regime across most metrics, and that continental region was an even poorer predictor of flow regime. Also note the non-linear relationship between stream order and several of the PCA metrics.

Many land-use characteristics followed similar complex relationships across multiple timescales (Figures S15 and S16 in Supporting Information S1).

Machine learning models trained to predict spectral power using catchment characteristics consistently suggested that climate was the most important predictor, followed by land cover and catchment area, with human impact becoming increasingly important at longer timescales (Figure 10). Again, contrary to expectations, dams were not particularly important predictors of flow regime (Figure 10, Figure S18 in Supporting Information S1). The number of dams in the upstream catchment and reservoir surface area loosely correlated with several flow metrics, PCA metrics, and certain period lengths, with coefficients of correlation between about −0.15 and 0.2 (Figure S15 in Supporting Information S1). Dam count and surface area were somewhat related to PCA metrics 1 and 2, which represented total flow magnitude and high-frequency stability in big rivers, respectively (Table 1), but these correlations were small, and variance was large (Figure 8). Additionally, variability at shorter timescales was easier to predict than variability at longer timescales (Figure S19 in Supporting Information S1), possibly due to the inherent loss in sample size for longer-lasting phenomena.

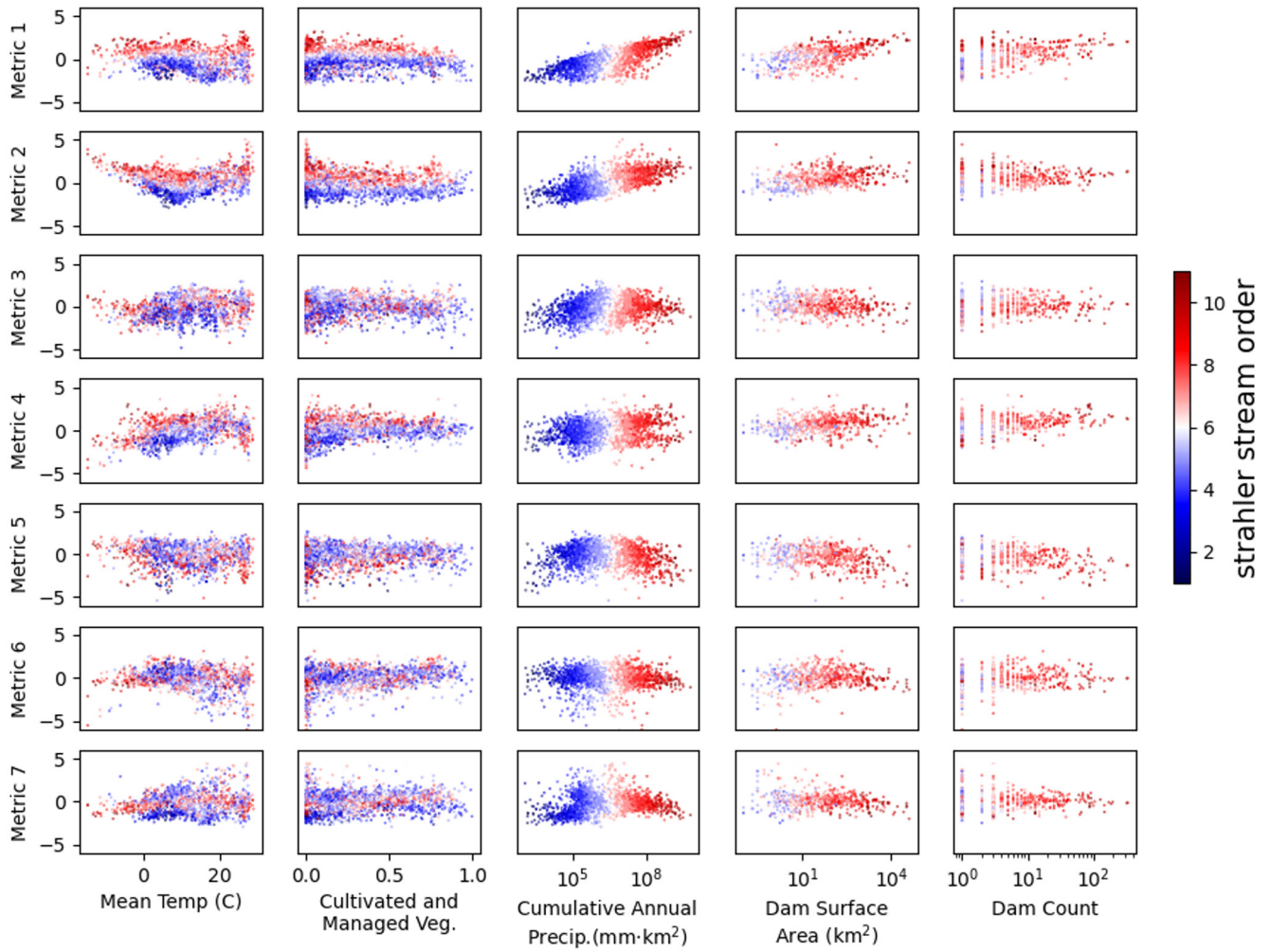

Figure 8. Continuous relationships between seven Principal Components Analysis flow metrics and catchment properties of mean annual temperature, mean annual precipitation (normalized for catchment size), catchment size, percent forest cover, and percent human influence. Streams have been colored according to stream order (with group 4 representing the largest catchments). The plots demonstrate the coherence of the relationship between catchment size and streamflow regime, while highlighting the comparatively weak relationship between human and forest cover and flow regime. Also of note are the complex, non-linear relationships between temperature and precipitation and the seven flow metrics.

4. Discussion

River networks connect and unite much of life on Earth, including human societies (Figure 1). Like a constellation of linked ecological heartbeats, river flows rise and fall across myriad timescales, sculpting aquatic habitat, driving biogeochemical flux, and quenching human water needs. In an increasingly human-dominated world of dams, agricultural water use, and changing climate, it is critical to understand patterns in hydrological processes at planetary scales, and to identify which climate, land cover, and water-use factors drive those patterns. One of the necessary milestones needed to achieve this understanding has been the development of a quantitative language for describing streamflow regime that is both concise enough to favor meaningful insight, yet broad enough to capture the wide range of behaviors seen in streams around the world. Therefore, our primary purpose in this paper was to explore possible methods for describing streamflow regime, and then to leverage those methods to identify patterns in and drivers of flow regime. Given the complexity that is traditionally attributed to streamflow regime (Dey & Mujumdar, 2022; Sivapalan, 2006; Tetzlaff et al., 2008), the large number of different hydrological models (Horton et al., 2022), and the number of parameters these models usually take, we were surprised by the low-dimensionality (i.e., simplicity) that global streamflow data consistently exhibited through a variety of analyses. At its core, this low dimensionality was driven by linkages of streamflow properties between timescales, and was not as closely correlated with dams as we expected. Below, we discuss our findings in

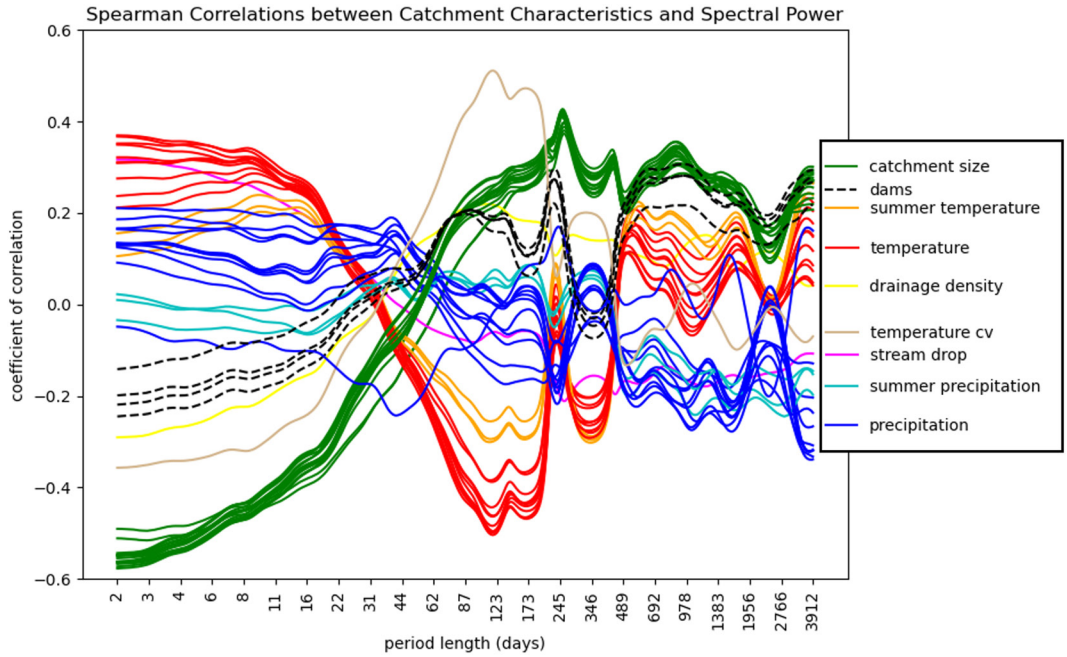

Figure 9. Correlations between catchment characteristics and spectral power across period lengths ranging from 2 days to almost 10 years.

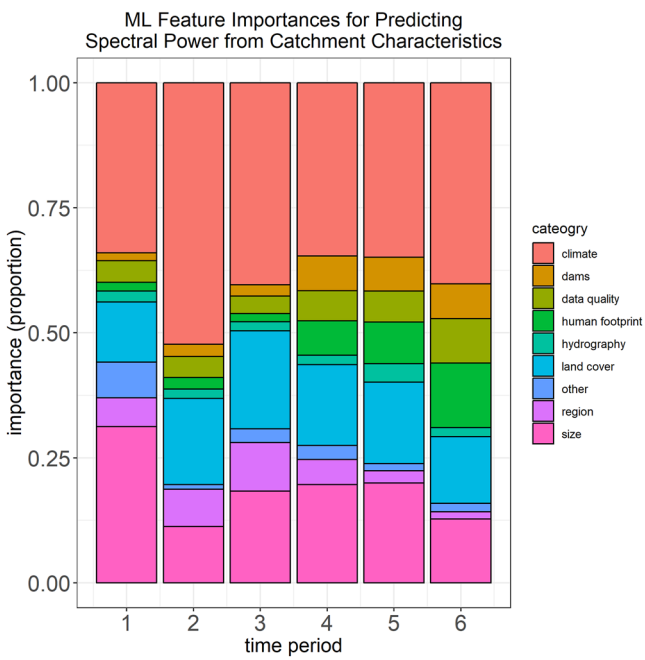

Figure 10. Mean feature importances of over 59,000 machine learning models trained to predict spectral power across different intervals (1: 2–40 days, 2: 41–180 days, 3: 181–365 days, 4: 366–1,095 days, 5: 1,096–2,190 days, and 6: 2,191–10,000 days) using catchment characteristics data. For simplicity, catchment characteristics have been grouped into eight categories. The relative contribution of each category to the predictive power of the models is represented by the height of each bar. Measures of urban density and agricultural spread are grouped in "human footprint" separately from measures of dam size and abundance, while cumulative measures of climate such as cumulative precipitation are grouped under "size."

light of current ecological challenges and hydrological theory, with particular emphasis on the importance of understanding timescales as interacting units with low effective dimensionality and simple driving mechanisms.

4.1. Are Streamflow Metrics or Frequency Decompositions Better?

Streamflow metrics and frequency decompositions such as wavelet analyses facilitate different, albeit related insights into streamflow regime. Streamflow metrics are not limited by a strict mathematical framework and therefore describe a wide range of phenomena, including variability, timing, and volume of flow with precise, albeit poorly organized, detail. Data-driven techniques such as PCA can counter this disorganization by identifying latent low-dimensional structure in streamflow metrics. One of the key contributions of this work was to apply data-driven structure identification techniques to unmodeled streamflow data at a global scale. Indeed, PCA analysis suggested that globally, streamflow metrics are inherently compressible along linear manifolds, with 68% of the variability in flow data explained by seven linear principal components. However, our analysis also showed that a substantial portion of the information provided by streamflow metrics (the remaining 32%) is not well described by linear structures. In other words, streamflow metrics, and by extensions streamflow data, contain an inherently information rich component, and thus the large number of metrics used to describe streamflow is well-justified. We suggest that streamflow is both a simple and complex phenomenon, with minor, complex (high-dimensional) structures emerging on top of the dominant, simple (low-dimensional) patterns that are consistent at global scales. This dominant compressibility has previously been attributed to redundancy in streamflow metrics (Olden & Poff, 2003)—not an unlikely outcome given the sheer number of metrics available. However, the disorganization inherent within this approach also belies that the dominant low-dimensional structure is in-part a manifestation

of linkages in flow properties among timescales. Identifying these linkages (e.g., Figure 4) is another key contribution of this work. Our results suggest that these linkages arise from a small number of hydrological phenomena that are tuned by relatively few catchment properties such as drainage basin size, mean annual temperature, precipitation, and land use. Streamflow metrics fail to identify these linkages because they have not traditionally been organized by timescale and analyzed as an interacting set of variables. This is one of the great advantages of frequency decompositions—they organize phenomena in a timeseries by their duration, from days to decades. Complex dynamics are quantified with a concise vocabulary: the amplitude, phase, waveform, and vertical shift of waves of varying frequency. This vocabulary resides at a level of abstraction that is perhaps uncomfortably distant from real-world reservoir management, flood preparedness, and climate change assessment but is nonetheless remarkably useful for organizing structure in data, which we hold to be essential for identifying the key processes and interactions in a dynamic system.

We further suggest that as data sets increase in temporal and spatial scales, data-driven descriptors based on low-dimensional structure are key for several reasons: (a) they provide sanity checks of intuitive notions or concepts common in sub-disciplines that are otherwise quantitatively unfalsifiable (Kipper, 2021). For example, in hydrology the notions of "semi-arid watersheds" and "snow-driven watersheds" are commonly employed in the literature (Arheimer et al., 2017; Cosh et al., 2008; Manning et al., 2022; Poon & Kinoshita, 2018). We suggest that these intuitive concepts represent informal, expert-driven versions of dimensionality reduction (Bates, 2020; Wolff, 2019). Confirming that our expert-derived vernacular corresponds to patterns in data is critical as policy decisions are made regarding restoration efforts and global climate-change action. (b) Correlating low-dimensional structure with system characteristics is a first step toward developing understandable, causal mathematical models that can more reliably be used to predict system behavior under novel conditions such as climate change. We distinguish data-driven descriptors (low-dimensional structure) from data-driven models, which tend to be black-box models whose generalizability to novel conditions is harder to verify (Rudin, 2019). (c) Low-dimensional structure forms a concise vocabulary for communicating major issues to non-experts (see Eckmann & Tlusty, 2021; Lum et al., 2013; Nicolau et al., 2011 for examples from other fields), significantly aiding interdisciplinary discussions. In the realm of ecohydrology where so many organisms, processes, and societal communities are involved, succinct communication is key for progress to be made in the face of increasing environmental degradation (Abbott et al., 2019; Frei et al., 2021).

4.2. Streamflow Metrics Can Be Predicted From Temperature, Precipitation, and Catchment Size

Multiple analyses independently suggested that only a few catchment properties (temperature, precipitation, and catchment size) were necessary for predicting the dominant structures in streamflow regime data, consistent with the parsimonious modeling framework frequently used by hydrological modelers (Coutu et al., 2012; Mei et al., 2014; Perrin et al., 2003; Roux et al., 2011). This was true regardless of the method used for quantifying flow regime. For example, three types of machine learning models suggested that PCA flow metrics could be predicted almost exclusively from temperature, precipitation, and catchment area. When plotted together, the relationships between PCA flow metrics and these variables were visually obvious, while the relationships between PCA flow metrics and variables identified as less important by the machine learning models (e.g., drainage density) were markedly less clear (Figure 8, Figures S11–14 in Supporting Information S1). Similar patterns emerged when wavelet and Fourier analyses were used to quantify flow regime: correlational and machine learning analyses consistently identified climate and catchment size as important predictors of flow variability at all temporal scales considered. However, there was one important difference between features that were important for predicting PCA flow metrics and frequency decompositions: the influence of land use. This category grouped several variables, including percent forest cover, percent shrub cover, percent snow cover, etc. under a single label. And while individual land use characteristics were not particularly important in isolation, machine learning models consistently ranked this group to be as important as catchment area for predicting variability in flow at all timescales longer than a few months (Figure 10). We speculate that precipitation, temperature, and catchment size may regulate near-universal hydrological processes that are responsible for the highly compressible components of streamflow regime (those captured by the seven PCA metrics; Giano, 2021; Poff et al., 1997), and that the more complicated ecohydrological interactions introduced by the myriad possible land use regimes and geological factors are responsible for the streamflow properties that were harder to identify with PCA analysis (Bladon et al., 2014; Manning et al., 2022; Tague & Grant, 2004; S. Wu et al., 2021), corroborating the need for a large number of non-generalizable parameters that frequently occurs in modeling scenarios (Horton et al., 2022). Said

differently, our results imply that a simple, emergent physics may exist at the catchment scale, where a handful of mean catchment properties accurately predict flashiness, timing, and volume of flow at the basin's outlet, to the extent that biological interactions remain simple (Sposito, 2017; Zhou et al., 2015).

4.3. What About Dams?

Surprisingly, dams were unimpressive predictors of flow regime according to every analysis we performed. At local and even regional scales, dams can and do fundamentally alter flow regimes (Arheimer et al., 2017; Chalise et al., 2023). However, our results suggest that at global scales, and for the majority of flow regime properties, these impacts appear to be small or difficult to measure relative to other catchment features' impacts.

Many factors mediate the impact of dams on hydrology, including dam size, river size, impounded area, and dam operations (i.e., the volume, frequency, and timing of releases). Likewise, dam location is not randomly distributed—many of the flow regime attributes and ecological factors that influence flow regime also affect dam construction and operation, including precipitation, topography, catchment size, and temperature (Lehner et al., 2011; Liermann et al., 2012). Together, this means that the impacts of a particular dam on flow regime may vary dramatically. This could explain the lack of a consistent pattern in our analysis because our data set comprised large and small dams on rivers from multiple continents. For example, dams with large reservoirs used for irrigation reduce flow variability downstream, while hydroelectric dams may increase it (Graf, 2006; Kennedy et al., 2016), reducing the overall effects at global scales.

We also note that dam number and surface area correlated strongly with catchment area (Figure S18 in Supporting Information S1). Given that there are very few large rivers that are not dammed (Maavara et al., 2020), and that large rivers are more likely to be managed for human uses, the relationship between dams, catchment area, and land use may obscure many of the direct effects of dams on flow regime. Furthermore, the signal dampening that occurs in large catchments may dominate the dynamics introduced by human dam management such as episodic releases to meet hydroelectric production, flood control, and water storage (Chezik et al., 2017).

One of our major goals was to use empirical data to characterize the drivers of streamflow regime in the context of human domination of the water cycle (Abbott et al., 2019; Chalise et al., 2021; Palmer & Ruhi, 2019). However, in this context of collinearity and qualitative differences across scales, we think that separating cause from correlation may be very difficult (Thomas et al., 2016). We recommend caution when interpreting global patterns in the interplay of infrastructure, land use, and climate in determining flow regime. However, our results suggest that at global scales, human alterations to earth's climate and land surface have the potential to impact river flow at least as much as the construction and operation of dams (Nijssen et al., 2001; Schneider et al., 2013; Wenger et al., 2011; Xenopoulos & Lodge, 2006).

5. Conclusions

In closing, river flow is a critical component of ecosystem and societal functioning (Palmer & Ruhi, 2019). Given the massive scale of human alterations to the water cycle, it has never been more important to understand how climatic, geomorphological, biological, and industrial factors interact to mediate the rise and fall of rivers. We propose that river flow can only be understood as a phenomenon occurring across many *interacting* timescales. These interactions are visible through low-dimensional structure in streamflow data that correlates closely with a small number of catchment characteristics. Together, these results suggest that global river flow dynamics are controlled by just a few dominant hydrological mechanisms that are locally tuned by land use, geology, climate, and human infrastructure. The implications of organizing low-dimensional structure within streamflow data for hydrology are broad and far reaching, inasmuch as simplicity provides a *lingua franca* for the diverse academic and societal communities whose livelihoods pulse to the rhythm of earth's rivers and streams.

Data Availability Statement

The data used in this study are available on https://www.researchgate.net at https://doi.org/10.13140/RG.2.2.24985.95842 and https://doi.org/10.13140/RG.2.2.31696.84487 under CC BY 4.0 licenses. Code for the analyses in this paper can be found at https://doi.org/10.5281/zenodo.7820994, or alternatively release v1.1.0 at https://github.com/Populustremuloides/TheMusicOfRivers.git.

Acknowledgments
This project was funded by the U.S. National Science Foundation (Grant DEB-1354867, EAR-2011439, EAR-2012123) and the Utah Division of Natural Resources Watershed Restoration Initiative. We thank the Stream Resiliency Research Coordination Network for initiating and coordinating this collaboration.

References

Abbott, B. W., Bishop, K., Zarnetske, J. P., Hannah, D. M., Frei, R. J., Minaudo, C., et al. (2019). A water cycle for the Anthropocene. *Hydrological Processes*, *33*(23), 3046–3052. https://doi.org/10.1002/hyp.13544

Abbott, B. W., Bishop, K., Zarnetske, J. P., Minaudo, C., Chapin, F. S., Krause, S., et al. (2019). Human domination of the global water cycle absent from depictions and perceptions. *Nature Geoscience*, *12*(7), 533–540. https://doi.org/10.1038/s41561-019-0374-y

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

Alfieri, L., Lorini, V., Hirpa, F. A., Harrigan, S., Zsoter, E., Prudhomme, C., & Salamon, P. (2020). A global streamflow reanalysis for 1980–2018. *Journal of Hydrology X*, *6*, 100049. https://doi.org/10.1016/j.hydroa.2019.100049

Amatulli, G., Domisch, S., Tuanmu, M.-N., Parmentier, B., Ranipeta, A., Malczyk, J., & Jetz, W. (2018). A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific Data*, *5*(1), 180040. https://doi.org/10.1038/sdata.2018.40

Archfield, S. A., Kennen, J. G., Carlisle, D. M., & Wolock, D. M. (2014). An objective and parsimonious approach for classifying natural flow regimes at a continental scale. *River Research and Applications*, *30*(9), 1166–1183. https://doi.org/10.1002/rra.2710

Arheimer, B., Donnelly, C., & Lindström, G. (2017). Regulation of snow-fed rivers affects flow regimes more than climate change. *Nature Communications*, *8*(1), 62. https://doi.org/10.1038/s41467-017-00092-8

Ascott, M. J., Gooddy, D. C., Fenton, O., Vero, S., Ward, R. S., Basu, N. B., et al. (2021). The need to integrate legacy nitrogen storage dynamics and time lags into policy and practice. *Science of the Total Environment*, *781*, 146698. https://doi.org/10.1016/j.scitotenv.2021.146698

Ashraf, F. B., Haghighi, A. T., Riml, J., Mathias Kondolf, G., Kløve, B., & Marttila, H. (2022). A method for assessment of sub-daily flow alterations using wavelet analysis for regulated rivers. *Water Resources Research*, *58*(1), e2021WR030421. https://doi.org/10.1029/2021WR030421

Barbarossa, V., Huijbregts, M. A. J., Beusen, A. H. W., Beck, H. E., King, H., & Schipper, A. M. (2018). FLO1K, global maps of mean, maximum and minimum annual streamflow at 1 km resolution from 1960 through 2015. *Scientific Data*, *5*(1), 180052. https://doi.org/10.1038/sdata.2018.52

Bates, C. (2020). Efficient data compression in human perception. Retrieved from https://urresearch.rochester.edu/institutionalPublicationPublicView.action?institutionalItemId=35450

Berghuijs, W. R., Harrigan, S., Molnar, P., Slater, L. J., & Kirchner, J. W. (2019). The relative importance of different flood-generating mechanisms across Europe. *Water Resources Research*, *55*(6), 4582–4593. https://doi.org/10.1029/2019WR024841

Bernhardt, E. S., Rosi, E. J., & Gessner, M. O. (2017). Synthetic chemicals as agents of global change. *Frontiers in Ecology and the Environment*, *15*(2), 84–90. https://doi.org/10.1002/fee.1450

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Bladon, K. D., Emelko, M. B., Silins, U., & Stone, M. (2014). Wildfire and the future of water supply. *Environmental Science & Technology*, *48*(16), 8936–8943. https://doi.org/10.1021/es500130g

Bochet, O., Bethencourt, L., Dufresne, A., Farasin, J., Pédrot, M., Labasque, T., et al. (2020). Iron-oxidizer hotspots formed by intermittent oxic–anoxic fluid mixing in fractured rocks. *Nature Geoscience*, *13*(2), 1–7. https://doi.org/10.1038/s41561-019-0509-1

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

Bunn, S. E., & Arthington, A. H. (2002). Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. *Environmental Management*, *30*(4), 492–507. https://doi.org/10.1007/s00267-002-2737-0

Carey, S. K., Tetzlaff, D., Buttle, J., Laudon, H., McDonnell, J., McGuire, K., et al. (2013). Use of color maps and wavelet coherence to discern seasonal and interannual climate influences on streamflow variability in northern catchments. *Water Resources Research*, *49*(10), 6194–6207. https://doi.org/10.1002/wrcr.20469

Carlisle, D. M., Falcone, J., Wolock, D. M., Meador, M. R., & Norris, R. H. (2010). Predicting the natural flow regime: Models for assessing hydrological alteration in streams. *River Research and Applications*, *26*(2), 118–136. https://doi.org/10.1002/rra.1247

Chalise, D. R., Sankarasubramanian, A., Olden, J. D., & Ruhi, A. (2023). Spectral signatures of flow regime alteration by dams across the United States. *Earth's Future*, *11*(2), e2022EF003078. https://doi.org/10.1029/2022EF003078

Chalise, D. R., Sankarasubramanian, A., & Ruhi, A. (2021). Dams and climate interact to alter river flow regimes across the United States. *Earth's Future*, *9*(4), e2020EF001816. https://doi.org/10.1029/2020EF001816

Chezik, K. A., Anderson, S. C., & Moore, J. W. (2017). River networks dampen long-term hydrological signals of climate change. *Geophysical Research Letters*, *44*(14), 7256–7264. https://doi.org/10.1002/2017GL074376

Cosh, M. H., Jackson, T. J., Moran, S., & Bindlish, R. (2008). Temporal persistence and stability of surface soil moisture in a semi-arid watershed. *Remote Sensing of Environment*, *112*(2), 304–313. https://doi.org/10.1016/j.rse.2007.07.001

Coutu, S., Del Giudice, D., Rossi, L., & Barry, D. A. (2012). Parsimonious hydrological modeling of urban sewer and river catchments. *Journal of Hydrology*, *464*(465), 477–484. https://doi.org/10.1016/j.jhydrol.2012.07.039

Dey, P., & Mujumdar, P. (2022). On the statistical complexity of streamflow. *Hydrological Sciences Journal*, *67*(1), 40–53. https://doi.org/10.1080/02626667.2021.2000991

Díaz, S., Settele, J., Brondízio, E. S., Ngo, H. T., Agard, J., Arneth, A., et al. (2019). Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science*, *366*(6471). https://doi.org/10.1126/science.aax3100

Döll, P., & Schmied, H. M. (2012). How is the impact of climate change on river flow regimes related to the impact on mean annual runoff? A global-scale analysis. *Environmental Research Letters*, *7*(1), 014037. https://doi.org/10.1088/1748-9326/7/1/014037

Dupas, R., Abbott, B. W., Minaudo, C., & Fovet, O. (2019). Distribution of landscape units within catchments influences nutrient export dynamics. *Frontiers in Environmental Science*, *7*. https://doi.org/10.3389/fenvs.2019.00043

Eckmann, J.-P., & Tlusty, T. (2021). Dimensional reduction in complex living systems: Where, why, and how. *BioEssays*, *43*(9), 2100062. https://doi.org/10.1002/bies.202100062

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Fisher, S. G., Grimm, N. B., Martí, E., Holmes, R. M., & Jones, J. B., Jr. (1998). Material spiraling in stream corridors: A telescoping ecosystem model. *Ecosystems*, *1*(1), 19–34. https://doi.org/10.1007/s100219900003

Frei, R. J., Abbott, B. W., Dupas, R., Gu, S., Gruau, G., Thomas, Z., et al. (2020). Predicting nutrient incontinence in the Anthropocene at watershed scales. *Frontiers in Environmental Science*, *7*, 200. https://doi.org/10.3389/fenvs.2019.00200

Frei, R. J., Lawson, G. M., Norris, A. J., Cano, G., Vargas, M. C., Kujanpää, E., et al. (2021). Limited progress in nutrient pollution in the U.S. caused by spatially persistent nutrient sources. *PLoS One*, *16*(11), e0258952. https://doi.org/10.1371/journal.pone.0258952

George, R., McManamay, R., Perry, D., Sabo, J., & Ruddell, B. L. (2021). Indicators of hydro-ecological alteration for the rivers of the United States. *Ecological Indicators*, *120*, 106908. https://doi.org/10.1016/j.ecolind.2020.106908

Gerten, D., Rost, S., von Bloh, W., & Lucht, W. (2008). Causes of change in 20th century global river discharge. *Geophysical Research Letters*, *35*(20), L20405. https://doi.org/10.1029/2008GL035258

Giano, S. I. (2021). Fluvial geomorphology and river management. *Water*, *13*(11), 1608. https://doi.org/10.3390/w13111608

Gleeson, T., Wang-Erlandsson, L., Porkka, M., Zipper, S. C., Jaramillo, F., Gerten, D., et al. (2020). Illuminating water cycle modifications and Earth system resilience in the Anthropocene. *Water Resources Research*, *56*(4), e2019WR024957. https://doi.org/10.1029/2019WR024957

Gnann, S. J., Coxon, G., Woods, R. A., Howden, N. J. K., & McMillan, H. K. (2021). TOSSH: A toolbox for streamflow signatures in hydrology. *Environmental Modelling & Software*, *138*, 104983. https://doi.org/10.1016/j.envsoft.2021.104983

Godsey, S. E., & Kirchner, J. W. (2014). Dynamic, discontinuous stream networks: Hydrologically driven variations in active drainage density, flowing channels and stream order. *Hydrological Processes*, *28*(23), 5791–5803. https://doi.org/10.1002/hyp.10310

Goeking, S. A., & Tarboton, D. G. (2020). Forests and water yield: A synthesis of disturbance effects on streamflow and snowpack in western coniferous forests. *Journal of Forestry*, *118*(2), 172–192. https://doi.org/10.1093/jofore/fvz069

Gorski, G., & Zimmer, M. A. (2021). Hydrologic regimes drive nitrate export behavior in human-impacted watersheds. *Hydrology and Earth System Sciences*, *25*(3), 1333–1345. https://doi.org/10.5194/hess-25-1333-2021

Graf, W. L. (2006). Downstream hydrologic and geomorphic effects of large dams on American rivers. *Geomorphology*, *79*(3), 336–360. https://doi.org/10.1016/j.geomorph.2006.06.022

Hain, E. F., Kennen, J. G., Caldwell, P. V., Nelson, S. A. C., Sun, G., & McNulty, S. G. (2018). Using regional scale flow–ecology modeling to identify catchments where fish assemblages are most vulnerable to changes in water availability. *Freshwater Biology*, *63*(8), 928–945. https://doi.org/10.1111/fwb.13048

Hales, R. C., Nelson, E. J., Souffront, M., Gutierrez, A. L., Prudhomme, C., Kopp, S., et al. (2022). Advancing global hydrologic modeling with the GEOGloWS ECMWF streamflow service. *Journal of Flood Risk Management*, e12859. https://doi.org/10.1111/jfr3.12859

Hannah, D. M., Demuth, S., Lanen, H. A. J. V., Looser, U., Prudhomme, C., Rees, G., et al. (2011). Large-scale river flow archives: Importance, current status and future needs. *Hydrological Processes*, *25*(7), 1191–1200. https://doi.org/10.1002/hyp.7794

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., et al. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Harrison, I., Abell, R., Darwall, W., Thieme, M. L., Tickner, D., & Timboe, I. (2018). The freshwater biodiversity crisis. *Science*, *362*(6421), 1369. https://doi.org/10.1126/science.aav9242

Helton, A. M., Poole, G. C., Meyer, J. L., Wollheim, W. M., Peterson, B. J., Mulholland, P. J., et al. (2011). Thinking outside the channel: Modeling nitrogen cycling in networked river ecosystems. *Frontiers in Ecology and the Environment*, *9*(4), 229–238. https://doi.org/10.1890/080211

Henriksen, J., Heasley, J., Kennen, J., & Nieswand, S. (2006). Users' manual for the hydroecological integrity assessment process software (including the New Jersey assessment tools).

Hogeboom, R. J., de Bruin, D., Schyns, J. F., Krol, M. S., & Hoekstra, A. Y. (2020). Capping human water footprints in the world's river basins. *Earth's Future*, *8*(2), e2019EF001363. https://doi.org/10.1029/2019EF001363

Horton, P., Schaefli, B., & Kauzlaric, M. (2022). Why do we have so many different hydrological models? A review based on the case of Switzerland. *WIREs Water*, *9*(1), e1574. https://doi.org/10.1002/wat2.1574

Jones, N. E., Schmidt, B. J., & Melles, S. J. (2014). Characteristics and distribution of natural flow regimes in Canada: A habitat template approach. *Canadian Journal of Fisheries and Aquatic Sciences*, *71*(11), 1616–1624. https://doi.org/10.1139/cjfas-2014-0040

Kennedy, T. A., Muehlbauer, J. D., Yackulic, C. B., Lytle, D. A., Miller, S. W., Dibble, K. L., et al. (2016). Flow management for hydropower extirpates aquatic insects, undermining river food webs. *BioScience*, *66*(7), 561–575. https://doi.org/10.1093/biosci/biw059

Kipper, J. (2021). Intuition, intelligence, data compression. *Synthese*, *198*(27), 6469–6489. https://doi.org/10.1007/s11229-019-02118-8

Labat, D. (2010). Cross wavelet analyses of annual continental freshwater discharge and selected climate indices. *Journal of Hydrology*, *385*(1), 269–278. https://doi.org/10.1016/j.jhydrol.2010.02.029

Lane, B. A., Dahlke, H. E., Pasternack, G. B., & Sandoval-Solis, S. (2017). Revealing the diversity of natural hydrologic regimes in California with relevance for environmental flows applications. *JAWRA Journal of the American Water Resources Association*, *53*(2), 411–430. https://doi.org/10.1111/1752-1688.12504

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., et al. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, *9*(9), 494–502. https://doi.org/10.1890/100125

Levia, D. F., Creed, I. F., Hannah, D. M., Nanko, K., Boyer, E. W., Carlyle-Moses, D. E., et al. (2020). Homogenization of the terrestrial water cycle. *Nature Geoscience*, *13*(10), 656–658. https://doi.org/10.1038/s41561-020-0641-y

Liermann, C. R., Nilsson, C., Robertson, J., & Ng, R. Y. (2012). Implications of dam obstruction for global freshwater fish diversity. *BioScience*, *62*(6), 539–548. https://doi.org/10.1525/bio.2012.62.6.5

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., et al. (2019). Global reconstruction of naturalized river flows at 2.94 million reaches. *Water Resources Research*, *55*(8), 6499–6516. https://doi.org/10.1029/2019WR025287

Liu, J., Zhang, Q., Singh, V. P., Song, C., Zhang, Y., Sun, P., & Gu, X. (2018). Hydrological effects of climate variability and vegetation dynamics on annual fluvial water balance in global large river basins. *Hydrology and Earth System Sciences*, *22*(7), 4047–4060. https://doi.org/10.5194/hess-22-4047-2018

Liu, S., Kuhn, C., Amatulli, G., Aho, K., Butman, D. E., Allen, G. H., et al. (2022). The importance of hydrology in routing terrestrial carbon to the atmosphere via global streams and rivers. *Proceedings of the National Academy of Sciences*, *119*(11), e2106322119. https://doi.org/10.1073/pnas.2106322119

Lum, P. Y., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., et al. (2013). Extracting insights from the shape of complex data using topology. *Scientific Reports*, *3*(1), 1236. https://doi.org/10.1038/srep01236

Maavara, T., Chen, Q., Van Meter, K., Brown, L. E., Zhang, J., Ni, J., & Zarfl, C. (2020). River dam impacts on biogeochemical cycling. *Nature Reviews Earth & Environment*, *1*(2), 103–116. https://doi.org/10.1038/s43017-019-0019-0

Malone, E. T., Abbott, B. W., Klaar, M. J., Kidd, C., Sebilo, M., Milner, A. M., & Pinay, G. (2018). Decline in ecosystem δ13C and mid-successional nitrogen loss in a two-century postglacial chronosequence. *Ecosystems*, *21*(8), 1659–1675. https://doi.org/10.1007/s10021-018-0245-1

Manning, A. L., Harpold, A., & Csank, A. (2022). Spruce beetle outbreak increases streamflow from snow-dominated basins in southwest Colorado, USA. *Water Resources Research*, *58*(5), e2021WR029964. https://doi.org/10.1029/2021WR029964

Masaki, Y., Hanasaki, N., Biemans, H., Schmied, H. M., Tang, Q., Wada, Y., et al. (2017). Intercomparison of global river discharge simulations focusing on dam operation—Multiple models analysis in two case-study river basins, Missouri–Mississippi and green–Colorado. *Environmental Research Letters*, *12*(5), 055002. https://doi.org/10.1088/1748-9326/aa57a8

McMahon, T. A., Vogel, R. M., Peel, M. C., & Pegram, G. G. S. (2007). Global streamflows—Part 1: Characteristics of annual streamflows. *Journal of Hydrology*, *347*(3), 243–259. https://doi.org/10.1016/j.jhydrol.2007.09.002

McMillan, H. (2021). A review of hydrologic signatures and their applications. *WIREs Water*, *8*(1), e1499. https://doi.org/10.1002/wat2.1499

Mei, Y., Anagnostou, E. N., Stampoulis, D., Nikolopoulos, E. I., Borga, M., & Vegara, H. J. (2014). Rainfall organization control on the flood response of mild-slope basins. *Journal of Hydrology*, *510*, 565–577. https://doi.org/10.1016/j.jhydrol.2013.12.013

Minaudo, C., Dupas, R., Gascuel-Odoux, C., Fovet, O., Mellander, P.-E., Jordan, P., et al. (2017). Nonlinear empirical modeling to estimate phosphorus exports using continuous records of turbidity and discharge. *Water Resources Research*, *53*(9), 7590–7606. https://doi.org/10.1002/2017WR020590

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

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics*, *18*(6), 275–285. https://doi.org/10.1002/cem.873

Nicolau, M., Levine, A. J., & Carlsson, G. (2011). Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences*, *108*(17), 7265–7270. https://doi.org/10.1073/pnas.1102826108

Nijssen, B., O'Donnell, G. M., Hamlet, A. F., & Lettenmaier, D. P. (2001). Hydrologic sensitivity of global rivers to climate change. *Climatic Change*, *50*(1), 143–175. https://doi.org/10.1023/A:1010616428763

Olden, J. D., & Poff, N. L. (2003). Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*, *19*(2), 101–121. https://doi.org/10.1002/rra.700

Oldfield, J. D. (2016). Mikhail budyko's (1920–2001) contributions to global climate science: From heat balances to climate change and global ecology. *Wiley Interdisciplinary Reviews: Climate Change*, *7*(5), 682–692. https://doi.org/10.1002/wcc.412

Palmer, M., & Ruhi, A. (2019). Linkages between flow regime, biota, and ecosystem processes: Implications for river restoration. *Science*, *365*(6459). https://doi.org/10.1126/science.aaw2087

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 6.

Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, *279*(1), 275–289. https://doi.org/10.1016/S0022-1694(03)00225-7

Pinay, G., Bernal, S., Abbott, B. W., Lupon, A., Marti, E., Sabater, F., & Krause, S. (2018). Riparian corridors: A new conceptual framework for assessing nitrogen buffering across biomes. *Frontiers in Environmental Science*, *6*, 47. https://doi.org/10.3389/fenvs.2018.00047

Poff, N. L., Allan, J. D., Bain, M. B., Karr, J. R., Prestegaard, K. L., Richter, B. D., et al. (1997). The natural flow regime. *BioScience*, *47*(11), 769–784. https://doi.org/10.2307/1313099

Poff, N. L., & Zimmerman, J. K. H. (2010). Ecological responses to altered flow regimes: A literature review to inform the science and management of environmental flows. *Freshwater Biology*, *55*(1), 194–205. https://doi.org/10.1111/j.1365-2427.2009.02272.x

Poon, P. K., & Kinoshita, A. M. (2018). Spatial and temporal evapotranspiration trends after wildfire in semi-arid landscapes. *Journal of Hydrology*, *559*, 71–83. https://doi.org/10.1016/j.jhydrol.2018.02.023

Raymond, P. A., Saiers, J. E., & Sobczak, W. V. (2016). Hydrological and biogeochemical controls on watershed dissolved organic matter transport: Pulse-shunt concept. *Ecology*, *97*(1), 5–16. https://doi.org/10.1890/14-1684.1

Reaver, N. G. F., Kaplan, D. A., Klammler, H., & Jawitz, J. W. (2020). Reinterpreting the budyko framework. *Hydrology and Earth System Sciences Discussions*, 1–31. https://doi.org/10.5194/hess-2020-584

Rösch, A., & Schmidbauer, H. W. (2018). WaveletComp: Computational wavelet analysis. R package version 1.1.

Roux, H., Labat, D., Garambois, P.-A., Maubourguet, M.-M., Chorda, J., & Dartus, D. (2011). A physically-based parsimonious hydrological model for flash floods in Mediterranean catchments. *Natural Hazards and Earth System Sciences*, *11*(9), 2567–2582. https://doi.org/10.5194/nhess-11-2567-2011

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Ryo, M., Iwasaki, Y., Yoshimura, C., & Saavedra, V. O. C. (2015). Evaluation of spatial pattern of altered flow regimes on a river network using a distributed hydrological model. *PLoS One*, *10*(7), e0133833. https://doi.org/10.1371/journal.pone.0133833

Sabo, J. L., & Post, D. M. (2008). Quantifying periodic, stochastic, and catastrophic environmental variation. *Ecological Monographs*, *78*(1), 19–40. https://doi.org/10.1890/06-1340.1

Sanborn, S. C., & Bledsoe, B. P. (2006). Predicting streamflow regime metrics for ungauged streamsin Colorado, Washington, and Oregon. *Journal of Hydrology*, *325*(1), 241–261. https://doi.org/10.1016/j.jhydrol.2005.10.018

Sang, Y.-F. (2013). A review on the applications of wavelet transform in hydrology time series analysis. *Atmospheric Research*, *122*(Supplement C), 8–15. https://doi.org/10.1016/j.atmosres.2012.11.003

Savenije, H. H. G. (2018). HESS Opinions: Linking Darcy's equation to the linear reservoir. *Hydrology and Earth System Sciences*, *22*(3), 1911–1916. https://doi.org/10.5194/hess-22-1911-2018

Schmidbauer, H., & Roesch, A. (2018). WaveletComp 1.1: A guided tour through the R package.

Schneider, C., Laizé, C. L. R., Acreman, M. C., & Flörke, M. (2013). How will climate change modify river flow regimes in Europe? *Hydrology and Earth System Sciences*, *17*(1), 325–339. https://doi.org/10.5194/hess-17-325-2013

Sivapalan, M. (2006). Pattern, process and function: Elements of a unified theory of hydrology at the catchment scale. In *Encyclopedia of hydrological Sciences*. John Wiley & Sons, Ltd. https://doi.org/10.1002/0470848944.hsa012

Smith, L. C., Turcotte, D. L., & Isacks, B. L. (1998). Stream flow characterization and feature detection using a discrete wavelet transform. *Hydrological Processes*, *12*(2), 233–249. https://doi.org/10.1002/(SICI)1099-1085(199802)12:2<233::AID-HYP573>3.0.CO;2-3

Sposito, G. (2017). Understanding the budyko equation. *Water*, *9*(4), 236. https://doi.org/10.3390/w9040236

Tague, C., & Grant, G. E. (2004). A geological framework for interpreting the low-flow regimes of Cascade streams, Willamette River Basin, Oregon. *Water Resources Research*, *40*(4), 14. https://doi.org/10.1029/2003WR002629

Tank, S. E., Vonk, J. E., Walvoord, M. A., McClelland, J. W., Laurion, I., & Abbott, B. W. (2020). Landscape matters: Predicting the biogeochemical effects of permafrost thaw on aquatic networks with a state factor approach. *Permafrost and Periglacial Processes*, *31*(3), 358–370. https://doi.org/10.1002/ppp.2057

Teixeira, H., Lillebø, A. I., Culhane, F., Robinson, L., Trauner, D., Borgwardt, F., et al. (2019). Linking biodiversity to ecosystem services supply: Patterns across aquatic ecosystems. *Science of the Total Environment*, *657*, 517–534. https://doi.org/10.1016/j.scitotenv.2018.11.440

Tetzlaff, D., McDonnell, J., Uhlenbrook, S., McGuire, K., Bogaart, P., Naef, F., et al. (2008). Conceptualizing catchment processes: Simply too complex? *Hydrological Processes*, *22*, 11–1730. https://doi.org/10.1002/hyp.7069

Thomas, Z., Abbott, B. W., Troccaz, O., Baudry, J., & Pinay, G. (2016). Proximate and ultimate controls on carbon and nutrient dynamics of small agricultural catchments. *Biogeosciences*, *13*(6), 1863–1875. https://doi.org/10.5194/bg-13-1863-2016

Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, *79*(1), 61–78. https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2

Tucker, G. E., & Hancock, G. R. (2010). Modelling landscape evolution. *Earth Surface Processes and Landforms*, *35*(1), 28–50. https://doi.org/10.1002/esp.1952

Unpingco, J. (2014). Discrete-time Fourier transform. In J. Unpingco (Ed.) *Python for signal processing: Featuring IPython notebooks* (pp. 45–55). Springer International Publishing. https://doi.org/10.1007/978-3-319-01342-8_3

Van Loon, A. F., Gleeson, T., Clark, J., Van Dijk, A. I. J. M., Stahl, K., Hannaford, J., et al. (2016). Drought in the Anthropocene. *Nature Geoscience*, *9*(2), 89–91. https://doi.org/10.1038/ngeo2646

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., et al. (2010). Global threats to human water security and river biodiversity. *Nature*, *467*(7315), 555–561. https://doi.org/10.1038/nature09440

Wenger, S. J., Isaak, D. J., Luce, C. H., Neville, H. M., Fausch, K. D., Dunham, J. B., et al. (2011). Flow regime, temperature, and biotic interactions drive differential declines of trout species under climate change. *Proceedings of the National Academy of Sciences*, *108*(34), 14175–14180. https://doi.org/10.1073/pnas.1103097108

White, M. A., Schmidt, J. C., & Topping, D. J. (2005). Application of wavelet analysis for monitoring the hydrologic effects of dam operation: Glen Canyon Dam and the Colorado River at Lees Ferry, Arizona. *River Research and Applications*, *21*(5), 551–565. https://doi.org/10.1002/rra.827

Wolff, J. G. (2019). Information compression as a unifying principle in human learning, perception, and cognition. *Complexity*, *2019*, e18797466. https://doi.org/10.1155/2019/1879746

Wu, F.-C., Chang, C.-F., & Shiau, J.-T. (2015). Assessment of flow regime alterations over a spectrum of temporal scales using wavelet-based approaches. *Water Resources Research*, *51*(5), 3317–3338. https://doi.org/10.1002/2014WR016595

Wu, H., Kimball, J. S., Mantua, N., & Stanford, J. (2011). Automated upscaling of river networks for macroscale hydrological modeling. *Water Resources Research*, *47*(3), 2011-03. https://doi.org/10.1029/2009WR008871

Wu, S., Zhao, J., Wang, H., & Sivapalan, M. (2021). Regional patterns and physical controls of streamflow generation across the conterminous United States. *Water Resources Research*, *57*(6), e2020WR028086. https://doi.org/10.1029/2020WR028086

Xenopoulos, M. A., & Lodge, D. M. (2006). Going with the flow: Using species–discharge relationships to forecast losses in fish biodiversity. *Ecology*, *87*(8), 1907–1914. https://doi.org/10.1890/0012-9658(2006)87[1907:GWTFUS]2.0.CO;2

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M. (2019). MERIT hydro: A high-resolution global hydrography map based on latest topography dataset. *Water Resources Research*, *55*(6), 5053–5073. https://doi.org/10.1029/2019WR024873

Zammit-Mangion, A., & Cressie, N. (2017). FRK: An R package for spatial and spatio-temporal prediction with large datasets.

Zarnetske, J. P., Bouda, M., Abbott, B. W., Saiers, J., & Raymond, P. A. (2018). Generality of hydrologic transport limitation of watershed organic carbon flux across ecoregions of the United States. *Geophysical Research Letters*, *45*(21), 11702–11711. https://doi.org/10.1029/2018GL080005

Zhou, G., Wei, X., Chen, X., Zhou, P., Liu, X., Xiao, Y., et al. (2015). Global pattern for the effect of climate and land cover on water yield. *Nature Communications*, *6*(1), 5918. https://doi.org/10.1038/ncomms6918

Zipper, S. C., Jaramillo, F., Wang-Erlandsson, L., Cornell, S. E., Gleeson, T., Porkka, M., et al. (2020). Integrating the water planetary boundary with water management from local to global scales. *Earth's Future*, *8*(2), e2019EF001377. https://doi.org/10.1029/2019EF001377