PNAS RESEARCH ARTICLE | ENVIRONMENTAL SCIENCES

Long-range dependence and extreme values of precipitation, phosphorus load, and Cyanobacteria

Stephen R. Carpenter[a,1], Mark R. Gahler[a], Christopher J. Kucharik[b], and Emily H. Stanley[a]

Contributed by Stephen R. Carpenter; received August 23, 2022; accepted October 14, 2022; reviewed by Todd V. Royer and Sarah M. Stackpoole

Extreme daily values of precipitation (1939–2021), discharge (1991–2021), phosphorus (P) load (1994–2021), and phycocyanin, a pigment of Cyanobacteria (June 1–September 15 of 2008–2021) are clustered as multi-day events for Lake Mendota, Wisconsin. Long-range dependence, or memory, is the shortest for precipitation and the longest for phycocyanin. Extremes are clustered for all variates and those of P load and phycocyanin are most strongly clustered. Extremes of P load are predictable from extremes of precipitation, and precipitation and P load are correlated with later concentrations of phycocyanin. However, time delays from 1 to 60 d were found between P load extremes and the next extreme phycocyanin event within the same year of observation. Although most of the lake's P enters in extreme events, blooms of Cyanobacteria may be sustained by recycling and food web processes.

Cyanobacteria | extreme values | long-range dependence | phosphorus | precipitation

Significance

The rising frequency of extreme precipitation is associated with the rising frequency of extreme nutrient flows from land to lakes, but direct links of load extremes to blooms of Cyanobacteria are unresolved. In Lake Mendota, Wisconsin, daily phosphorus load is correlated with high concentrations of Cyanobacteria 2–3 wk later. Extreme load events provide large reservoirs of nutrients that accumulate to support future blooms of Cyanobacteria. However, responses of Cyanobacteria to extreme storms and phosphorus loads have long and variable time delays.

Weather stations around the world report rising frequency of extreme precipitation (1, 2). Extreme precipitation events are expected to increase as the climate warms (3, 4), thereby increasing flood risk (5) with adverse effects on communities, ecosystem processes, and species' capacities to adapt to change (6). Extreme precipitation drives hydrological extremes (7) including drainage to lakes. In agricultural watersheds with highly nutrient-enriched soils (8, 9), extreme precipitation events cause erosion of soils and sediments, driving nutrient inputs to lakes (10–12) that support blooms of Cyanobacteria (13–15).

High concentrations of Cyanobacteria, or blooms, in lakes or reservoirs are a serious and expanding environmental problem (16). Consequences include hypoxia, mass mortality of fishes, and adverse effects on human health (17–19). Limnologists define a bloom as a period of net growth to high concentrations associated with several causal patterns related to mixing, temperature, nutrients, and grazing (19).

Extreme values, defined as time series peaks over a threshold (20), are rare, occur irregularly, and are an inherent feature of complex stochastic systems (7, 21) such as enriched lakes with Cyanobacteria (22). Warm summer temperatures are a necessary precondition for Cyanobacteria blooms (23, 24). Blooms are associated with calm winds and stable stratification (25). Extreme nutrient pulses can be followed by extreme concentrations of Cyanobacteria (26). Associations of extreme events in time can be identified by extreme value models with covariates, including Pareto models used here (27, 28).

Precipitation and hydrologic flows may show persistent autocorrelations or long-range dependence or memory (7). Long-range dependent (also known as long-range memory) processes have autocorrelation functions that decay slowly over time (7, 29). The terms of the autocovariance function $\rho(\tau)$, where τ is the time lag, decay as $\tau^{-\gamma}$ with correlation exponent γ, $0 < \gamma < 1$. For a Poisson (independent) process $\gamma = 1$. Slow decay, γ near 0, or long memory, is associated with aggregation or clustering of extremes in hydrologic data (30), although short-range autocorrelations can cause clumping of extremes in some stochastic systems (21). Accumulation of nutrients in soils, sediments, or lentic waters may slow transport and further increase long-range memory and time lags in nutrient flows (31–33).

Lake Mendota, Wisconsin, USA, is a eutrophic lake that has experienced blooms of Cyanobacteria since the 1880s (34–36). Long-term records of precipitation, discharge, phosphorus (P) load, and Cyanobacteria present extreme values in recent decades (11, 12, 37–39). On average 29 d per year of extreme inputs account for 74% of the annual P load to the lake (11). However, the relationships between extreme P loads and extreme concentrations of Cyanobacteria have not yet been analyzed.

Here we analyze extremes of precipitation, discharge, P load, and the Cyanobacterial pigment phycocyanin in long-term daily records from Lake Mendota. Specifically, we ask: (1) Does long-range dependence (memory) increase from precipitation, to discharge and P load, to phycocyanin? (2) Do cross-correlations and Pareto models suggest causal associations

Author affiliations: [a]Center for Limnology, University of Wisconsin-Madison, Madison WI 53706; and [b]Department of Agronomy, University of Wisconsin-Madison, Madison, WI 53706

Author contributions: S.R.C., C.J.K., and E.H.S. designed research; S.R.C. and M.R.G. performed research; S.R.C. analyzed data; and S.R.C., M.R.G., C.J.K., and E.H.S. wrote the paper.

Reviewers: T.V.R., Indiana University Bloomington; and S.M.S., United States Geological Survey.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: Steve.Carpenter@wisc.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2214343119/-/DCSupplemental.

Published November 21, 2022.

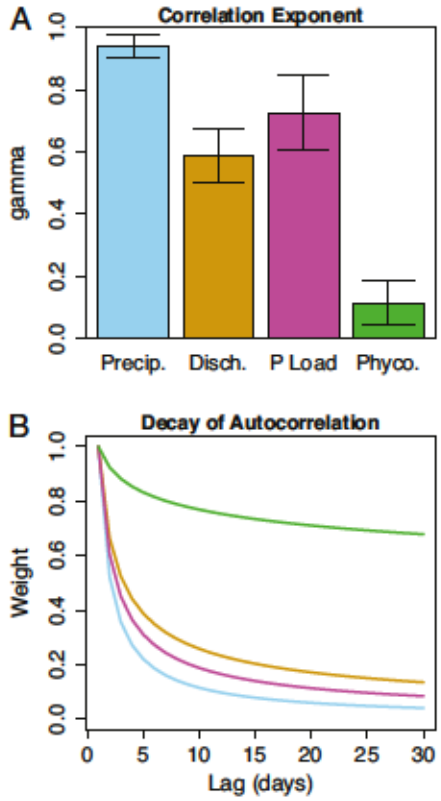


Fig. 1. (*A*) Correlation exponents $\gamma \pm 2$ SEs for concurrent daily observations of precipitation, discharge, P load, and phycocyanin. (*B*) Decay of autocorrelation function, weight = $\tau^{-\gamma}$, versus lag τ (days) for precipitation, discharge, P load, and phycocyanin. Line colors are the same as the bars in panel *A*.

of phycocyanin to precipitation or P load? (3) Are extreme values of precipitation, P load, and phycocyanin statistically independent, or are they clumped in time? (4) Are extreme values of P load followed by extreme concentrations of phycocyanin?

Results

The concurrent daily time series of precipitation, discharge of water to the lake, P load to the lake, and phycocyanin relative fluorescence units (RFU) (*SI Appendix*, Fig. S4) were analyzed for long-range dependence (Fig. 1). Precipitation is nearly independent (correlation exponent near 1), whereas phycocyanin (correlation exponent near 0) has the strongest long-range dependence of the four variates. Long-range dependence is intermediate for discharge and P load. Fig. 1*B* presents decay curves for autocorrelation functions showing rapid decay (low long-range dependence or short memory) for precipitation and slow decay (high long-range dependence or long memory) for phycocyanin, with intermediate rates for discharge and P load.

Connections of Precipitation, P, and Phycocyanin. Cross-correlations show immediate effects of precipitation on P load and delayed, variable effects of precipitation and P load on phycocyanin. (Fig. 2). Changes in precipitation are positively correlated with changes 1 d later in P load (Fig. 2*A*) and with changes 13–18 d later in phycocyanin (Fig. 2*B*). Changes in P load are positively correlated to changes 13–28 d later in phycocyanin (Fig. 2*C*).

Models for extreme values show that all of the time series' extremes fit Pareto distributions (40). The frequency and magnitude of extreme precipitation increased from 1940–2021. The magnitude (mm) of 2-, 5-, and 10-y precipitation events continues to increase (Fig. 3*A*). The return time of 100-mm events has declined from more than 5.5 y to less than 2.5 y over the period of record (Fig. 3*B*).

Extremes of precipitation are closely associated with extremes of P load (Fig. 3 *C* and *D*) through the fitted parameters of the Pareto equation (40). Return level (kg/d) of extreme P loads increases with daily precipitation more steeply as return time increases (Fig. 3*C*). Return level (kg/d) of extreme P loads increases with return interval of precipitation (Fig. 3*D*).

The best-fitting Pareto model for phycocyanin extremes had no significant effects of precipitation or P loads. Extreme phycocyanin concentrations from daily time series (*SI Appendix*, Fig. S4*D*) fit a stationary Pareto model (scale = 0.42 with standard error (s.e). = 0.079, shape = 0.30 with s.e. = 0.135). The stationary Pareto model had lower Akaike Information Criterion (AIC) than alternative Pareto models that represented scale or shape as functions of precipitation, P load, or discharge.

Return Intervals of Extreme Values. Return intervals of extreme values (i.e., the number of days between observations above the Pareto threshold) of precipitation, water discharge and P load to

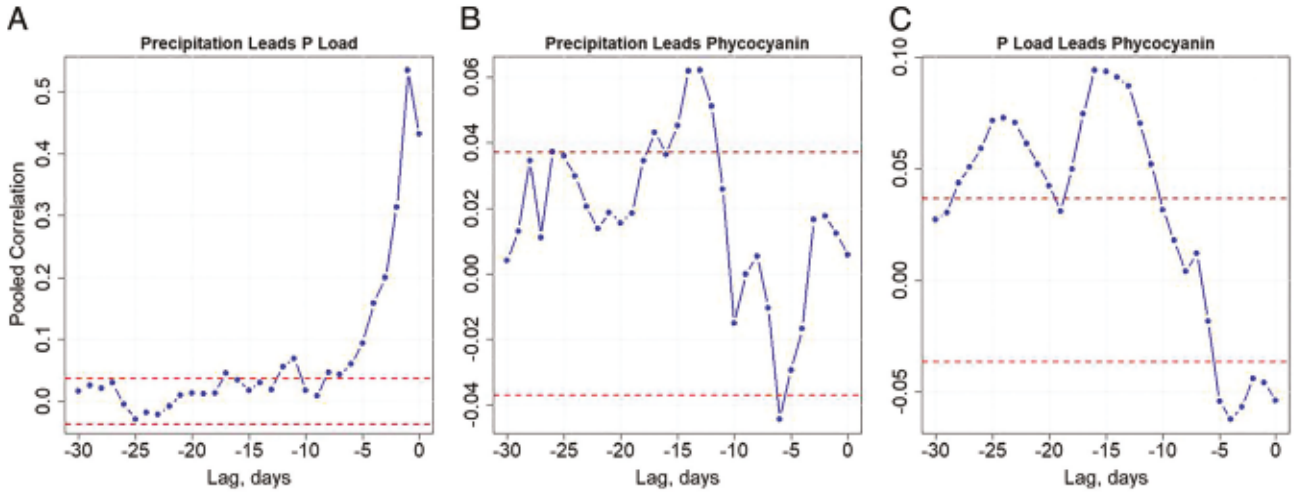


Fig. 2. Cross-correlations versus lag in days of (*A*) log10 precipitation before log10 P load, (*B*) log10 precipitation before log10 phycocyanin, and (*C*) log10 P load before log10 phycocyanin. Correlation coefficients are inverse-variance weighted means of annual values from 2008–2021. Horizontal lines show ± SD of the pooled cross-correlation function (CCF).

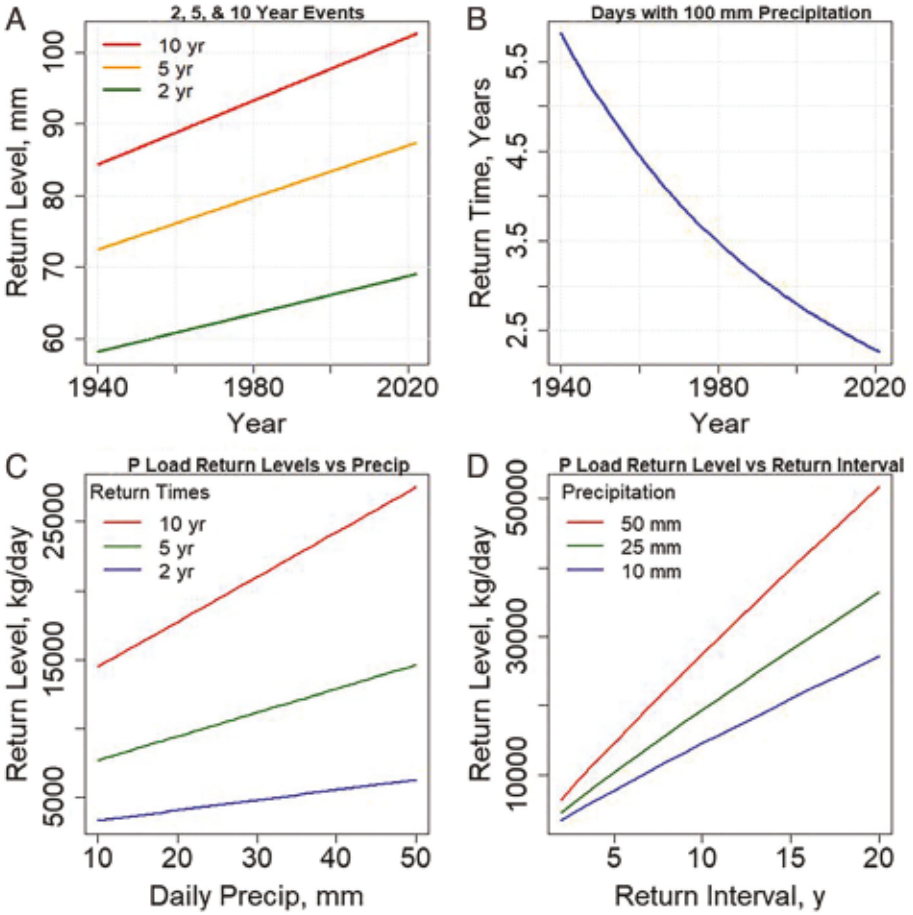


Fig. 3. Pareto models for precipitation and P load. The values selected for the examples show trends of the fitted Pareto models within the range of the data. (*A*) Precipitation return level versus year for 2-, 5-, and 10-y events. (*B*) Return time (years) of days with 100 mm precipitation versus year. (*C*) Return levels of P daily load (kg/d) versus daily precipitation (mm) for events with return times of 2, 5, and 10 y. (*D*) P load return levels (kg/d) versus return intervals (y) for daily precipitation events of 10, 25, and 50 mm. The Pareto threshold for precipitation is 20 mm/d.

the lake, and phycocyanin were compared to Poisson expectations to assess their independence (Fig. 4). For all variates, the indices of dispersion exceed 1, and plots of observed probabilities deviate from the Poisson distribution expected if events are independent in time (Kolmogorov–Smirnov test $P < 10^{-10}$). The deviations for precipitation are smaller than those of the other variates.

Extremes of precipitation (Fig. 4*A*) increase over time (Kendall rank correlation 0.307, $P < 0.0001$). Coefficients of dispersion exceed 1 for precipitation (Fig. 4*E*), consistent with clusters of extremes, but have no significant trend with time. Comparison to the Poisson distribution (Fig. 4*I*) shows a rather close fit. Nonetheless, short gaps between events are slightly less common than expected and long gaps between events are slightly more common than expected.

Annual extremes of water discharge into the lake (Fig. 4*B*) vary among years, but an upward trend is discernible (Kendall rank correlation = 0.338, $P = 0.007$). No extreme discharge events occurred in 2012 and the coefficient of dispersion cannot be calculated for that year (Fig. 4*F*). For the other years, the coefficient of dispersion is above 1 indicating clusters of extremes. Despite the inter-year variation, the coefficient of dispersion tends to decline over time (Kendall rank correlation = –0.274, $P = 0.0277$). Departures of discharge from the Poisson distribution (Fig. 4*C*) show that short gaps between extremes are less common than expected and long gaps between extremes are more common than expected.

Annual extremes of P load to the lake appear to increase over time (Fig. 4*C*), but the trend is not significant at the 5% level (Kendall tau = 0.238, $P = 0.087$). Dispersion coefficients exceed 1, indicating clusters of extremes, with no obvious trend through time (Fig. 4*G*). Departures from the Poisson distribution indicate that short gaps between extremes are too few and long gaps between extremes are too many (Fig. 4*K*).

Annual extremes of daily standardized log10 phycocyanin show no significant trend over the 14 y of data (Fig. 4*D*). No extremes were observed in 2016, and the coefficient of dispersion cannot be computed. In other years, the coefficient of dispersion exceeds 1, indicating aggregation of blooms in time (Fig. 4*H*). The apparent decline of dispersion over time is not significant. Comparison with the Poisson expectation shows that short gaps are too rare and long gaps are too common (Fig. 4*L*).

Of the 169 P load extremes observed during 2008–2021, 124 (73%) were followed, after a time delay of 1–60 d, by an extreme value of phycocyanin in the same year of observations (Fig. 5).

Discussion

The trend of rising average precipitation with increasingly frequent extreme values over a threshold is noted for many ecosystems (6, 7), and the Yahara watershed of Lake Mendota is among them (12).

In the Yahara watershed, extremes of precipitation, P load, and phycocyanin RFU are non-independent and tend to occur in clumps, such as multi-day periods of extreme rain, P load, and phycocyanin (Fig. 4). Daily mean precipitation and P load are lag-correlated with concentrations of phycocyanin. However,

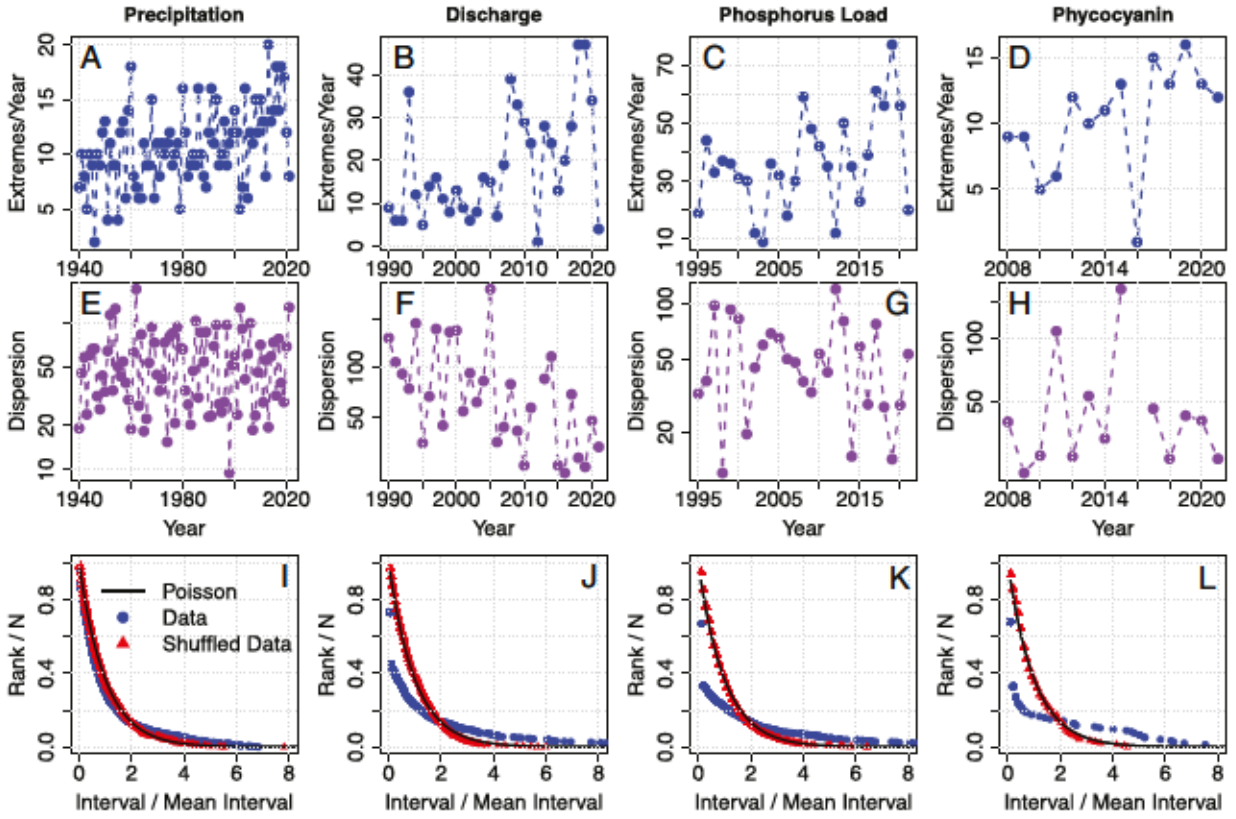


Fig. 4. Characteristics of extreme daily events for (*A, E, I*) precipitation, (*B, F, J*) discharge, (*C, G, K*) P load, and (*D, H, L*) phycocyanin. For each variate we present (*A–D*) number of extremes per year, (*E–H*) dispersion index, and (*I–L*) time intervals (days) between extremes as rank of interval length/number of intervals versus interval length/mean interval length for the observed data, randomly shuffled data, and the Poisson distribution.

Pareto models for phycocyanin extremes are stationary, with no significant trends over 14 y nor discernible effects of precipitation, discharge, or P load.

It is important to recognize that a high proportion of the annual P load arrives during extreme runoff events (11, 12). Once the P is in the lake water or sediments it can be recycled for many years (41, 42) and potentially support Cyanobacteria until it is flushed from the lake or added to permanent sediments. Extremely dry years have the opposite effect: P and pigment concentrations are unusually low, and the water is clear (37). However, excess nutrients can build up in soil during dry years and contribute to "weather whiplash" of unusually high nutrient loads in wet years (43).

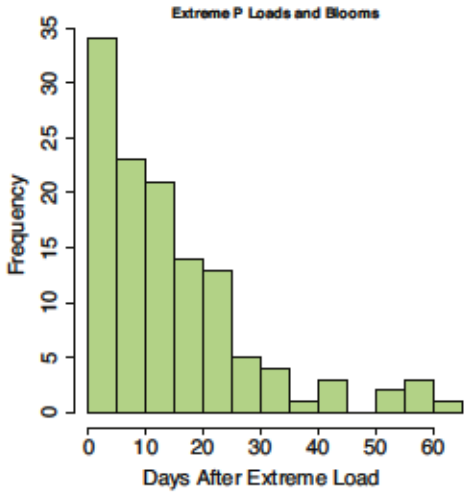


Fig. 5. Frequency distribution of time lags (days) from an extreme value of P load to the next extreme value of phycocyanin within the same year of observations. Forty-six of the 170 extreme P load events were not followed by an extreme value of phycocyanin within the same year and are omitted from the plot.

Extreme precipitation events can persist over several days (44) causing statistical dependencies (Fig. 1) that propagate from watershed to lake mixing, water transparency, and metabolism (45). Soil moisture can build up over successive rain events thereby priming the soil for runoff, drainage, and erosion over time. Consequently, clusters of extremes are more common than expected if events are independent (21, 30). Alternative statistical models for non-independent phenomena include the negative binomial for spatial patterns in ecology (46), the beta-binomial for spatial-temporal dependencies in river discharge (47), or the stretched Poisson for return intervals with correlation exponents below 1 (30). Such models could be investigated for systems of precipitation, P load, and Cyanobacteria.

Watershed processes amplify the dependencies over time. Lags and hysteretic shifts in nitrogen loading through nested subwatersheds showed decadal dependencies (33). Legacies of P accumulation generate long-term memory in P flows at continental scales (32). In the Yahara basin, multi-day precipitation events expand into longer series of days with extreme P loading to the lake. Extremes of precipitation translate rapidly into extremes of P load (Figs. 2*A* and 3 *C* and *D*), and the P residence time is longer than the water residence time of 4.4 y (37, 48).

It would be reasonable to expect that periods of high nutrient load lead directly to blooms of phytoplankton as found in other lakes (26). However, we found that cross-correlations of phycocyanin lagged P loads by 13–28 d (Fig. 2*C*), and time delays between a P load extreme and the next phycocyanin extreme

ranged up to 60 d (Fig. 5). Moreover, a general pattern of low and high alternate states of phycocyanin persists over a wide range of P load rates (49). Alternative stable states of phycocyanin appear as large shifts between low levels and bloom states (49), with high stochasticity and potentially rapid flickering between alternative states (50).

As water moves from the atmosphere over land to the lake, terrestrial processes increase long-range dependence for discharge and P load, and lake processes further increase long-range dependence of phycocyanin blooms (Fig. 1*B*). Phycocyanin blooms may be sustained by accumulated nutrients and endogenous lake dynamics (such as recycling from sediments) rather than a single extreme loading event (51). The long-term memory of phycocyanin, despite rapid daily fluctuations, is consistent with other syntheses. For example, decades of enriched soils and long-term nutrient flow to lakes (52), internal loading by sediment-to-water nutrient flux (41, 42, 53), and food web processes including grazing and nutrient regeneration (19, 54, 55) may enable blooms of Cyanobacteria year after year regardless of fluctuations in P load. The current water quality of the lake derives from climate trends, intensification of agriculture, urban expansion, and invasions of non-native plants and animals over recent decades (56). Together these generate long time lags, extreme events, and management surprises (57, 58). Projects to improve future water quality must adjust soil nutrients to match crop needs (59, 60) while coping with enriched sediments, increasing frequency of extreme precipitation, floods, and nutrient loads, unexpected assemblages of species, novel ecosystem behaviors, and rising demand for land and water resources. Impacts on food, water, and human well-being call for new approaches and collective action (61). The challenges of lakes in agricultural landscapes exemplify current global challenges that evoke transformative approaches for stewardship of land, water, and nature (62, 63).

Methods Summary

Yahara Watershed and Lake Mendota. The Lake Mendota watershed in southern Wisconsin, United States (43.2°N, 89.4°W) drains a 604 km^2 area of land devoted to dairy, corn, and soybean production, part of the Madison metropolitan area, and remnants of native vegetation (35, 64). The region has been significantly altered by intensification of dairy agriculture, expanding demand for biofuels, urban development, and climate change (65, 66). The watershed of Lake Mendota exemplifies Upper Midwestern watersheds, with an urbanizing agricultural landscape upstream of four large lakes with relatively poor water quality and frequent outbreaks of harmful algal blooms (49, 66, 67).

Hydrological Data. We analyzed daily discharge and P load data for the two tributaries of Lake Mendota, Pheasant Branch at Middleton (United States Geological Survey (USGS) No. 05427948; 43.10336 N, -89.5117 W; 1994–2021), and the Yahara River at Windsor (USGS No. 05427718; 43.20885 N, −89.3528 W; 1991–2021). Among the tributaries to Lake Mendota, Pheasant Branch and the Yahara River have the longest available time series for discharge and P load. These two tributaries account for more than 29% of the annual total P load to Lake Mendota and statistically they explain 96% of the variance in total annual P load to Lake Mendota from all tributaries combined (37). Thus these tributaries are a surrogate estimate of loading rates from the entire watershed. We combined the discharges and P loads from these two tributaries.

Discharge and P load were measured by USGS and downloaded from their website (http://waterdata.usgs.gov/nwis/sw) using the dataRetrieval() package (68, 69).

Precipitation Data. Continuous daily precipitation data were obtained for Madison Dane County Regional Airport (Cooperative Observer Network ID 474961; Global Historical Climatology Network ID USW00014837; coordinates 43.133 N, -89.349 W; elevation 264 m; Daily Data Range: October 1, 1939 to December 31, 2021). We define extreme precipitation values and return intervals based on calendar day totals and not for events that span a 24-h period over two consecutive calendar days.

Phycocyanin Data. Concentrations of phycocyanin, a characteristic pigment of Cyanobacteria, were monitored every minute at a central station (43.0995 N, -89.4045 W) in Lake Mendota during the ice-free seasons of 2008–2021 (70). Phycocyanin and chlorophyll data were recorded every minute using Turner Designs Cyclops 7 sensors suspended 1.0 m below the buoy until 2019. From 2019 to 2021, the sensor was a YSI EXO2 sonde. Water temperature at the sensor array and meteorology data for the atmosphere at the buoy, including wind speed and direction, were also recorded every minute. A complete list of sensors used over time is posted with the data (70). The duration of the measurements is around 175 d each year. We used data from June 1–September 15 (days of year 152–258) when blooms are common (71). Phycocyanin sensor readings were expressed as RFU. Log(RFU) is directly related to the log of phycocyanin concentration (72). We calculated daily means of log-transformed RFU using data from 00:00–04:00 and 22:00–24:00 each day to minimize quenching during daylight hours (71). Different sensors and maintenance regimes were used in different years, and these factors could affect the mean and SD of sensor readings within a year. To facilitate comparisons among years, daily means of the log(RFU) within each year were standardized as *z*-scores $z = (x - \hat{\mu}) / \hat{\sigma}$ where mean μ and SD σ are estimated over all days each year (49).

Statistical Analysis. All calculations were conducted in R 4.2.0 (73). Time series plots and descriptive statistics were calculated with R scripts and data presented online (69).

Cross-correlations for concurrent daily series were computed for each year and then pooled among years as weighted averages with inverse variance weights. The variance of a product-moment correlation coefficient *r* for a given lag in a given year is $(1 - r^2)^2/(n - 2)$ where n is the number of days used to calculate the correlation (74).

Extreme values, trends in extremes, and dependencies among extremes were assessed by fitting generalized Pareto distributions to each of the daily time series using the fevd() function of the extRemes() package (28). This procedure provides maximum likelihood estimates of location μ, scale σ, and shape ξ parameters (12). The threshold u for each time series was estimated according to Coles (40) as implemented by Carpenter et al. (12). These threshold values were used for peak-over threshold analyses of return intervals described below. R scripts and data for Pareto analyses are presented online (75).

Using threshold estimates from fits of Pareto models, we identified dates of threshold exceedances (extreme values) for precipitation, discharge, P load, and phycocyanin. For each variate, we calculated return intervals (durations of gaps between exceedances). For each variate's return intervals in each year, we calculated the index of dispersion (variance/mean) to assess independence or dependence of events from year to year. Trends in the number of extremes per year and the index of dispersion were assessed by Kendall rank correlation using the cor() function of *R* (76).

If extreme events are independent, then the return intervals (gaps between extreme events) are also independent and distributed according to the Poisson distribution (77). For a large sample of return intervals for a given threshold *q* with mean return interval R_q (i.e. the mean rate of extreme events is $\lambda = 1 / R_q$), the probability of one extreme in a return interval of length *r* is (77).

$$P_q(r) = r\lambda e^{-r\lambda}. \quad [1]$$

If extreme events are dependent, or clumped, in time then observed probabilities will deviate from [**1**] (30) as shown below. For a Poisson distribution the variance and mean are equal, hence the ratio variance/mean, called the index of dispersion, is 1 (46). For events that are dependent, the variance exceeds the mean and the index of dispersion is greater than 1 (46, 78).

For observed return intervals, we calculated cumulative rank distributions of r/R_q following Bunde et al. (30). For a given variate, we calculated 1 minus the relative rank of r/R_q (rank divided by the total number of events) and then plotted this quantity versus r/R_q. We compared these distributions with the Poisson distribution using Kolmogorov–Smirnov tests of the null hypothesis that data and Poisson expectation are drawn from the same distribution (function ks.test() with two-sided

option in R). As an additional visual test, we compared return intervals for observed time series of each variate with those of the randomly shuffled time series. Shuffled data were generated by the shuffle() function of the permute library in R (76, 79).

The correlation exponent γ was estimated by Detrended Fluctuation Analysis (DFA) (29). We fitted DFA models to daily time series of precipitation, P load, and phycocyanin using the dfa() function of the nonlinearTseries library (76, 80). The log–log plot of the resulting fluctuation function should be linear over the relevant time scales to estimate the correlation exponent (29, 81). We evaluated linearity by comparing a linear model to higher-order polynomials using AIC. When the linear model fits the data the slope α measures "self-affinity" of the time series, also called "long-term memory" since the correlation exponent γ approaches 0 as α approaches 1. The correlation exponent and self-affinity are related by the equation: $\gamma = 2(1 - \alpha)$. Since $\gamma = 1$ for independent or Poisson-distributed extremes, which have no temporal dependence or memory, the difference 1 - γ indicates dependence of extremes. Estimates of γ near zero indicate high levels of long-term dependence or memory (7).

Within each year, we recorded the number of days between an extreme value of P load and the next extreme value of phycocyanin. When summarized for all years the data show the time lags between extreme P loads and extreme phycocyanin concentrations (76).

Data, Materials, and Software Availability. Original Measurements data have been deposited in Environmental Data Initiative (https://doi.org/10.6073/pasta/fc8bd96677405945024ad708003be1fc).

ACKNOWLEDGMENTS. The North Temperate Lakes Long-Term Ecological Research program (NSF under Cooperative Agreement #DEB-2025982) supported this research. We are grateful to National Oceanic and Atmospheric Administration (NOAA) for providing precipitation data, USGS for providing hydrological and phosphorus load data, and David Buoy for the in-lake data. We thank E.G. Booth, R. Lathrop, T. Stuntebeck, and M.G. Turner for conversations about climate, land use, and changes of Lake Mendota.

1. S. Westra, L. V. Alexander, F. W. Zwiers, Global increasing trends in annual maximum daily precipitation. *J. Clim.* **26**, 3904–3918 (2013).
2. E. M. Fischer, R. Knutti, Observed heavy precipitation increase confirms theory and early models. *Nat. Clim. Change* **6**, 986–991 (2016).
3. S. Westra *et al.*, Future changes to the intensity and frequency of short-duration extreme rainfall. *Rev. Geophys.* **52**, 522–555 (2014).
4. A. F. Prein *et al.*, The future intensification of hourly precipitation extremes. *Nat. Clim. Change* **7**, 48–52 (2017).
5. X. Huang, S. Stevenson, A. D. Hall, Future warming and intensification of precipitation extremes: A "double whammy" leading to increasing flood risk in California. *Geophys. Res. Lett.* **47**, e2020GL088679 (2020).
6. C. C. Ummenhofer, G. A. Meehl, Extreme weather and climate events with ecological relevance: A review. *Philos. Trans. R. Soc. B: Biol. Sci.* **372** (2017), 10.1098/rstb.2016.0135.
7. M. Ghil *et al.*, Extreme events: Dynamics, statistics and prediction. *Nonlinear Process. Geophys.* **18**, 295–350 (2011).
8. G. K. MacDonald, E. M. Bennett, P. A. Potter, N. Ramankutty, Agronomic phosphorus imbalances across the world's croplands. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3086–3091 (2011).
9. A. Sharpley *et al.*, Phosphorus legacy: Overcoming the effects of past management practices to mitigate future water quality impairment. *J. Environ. Qual.* **42**, 1308–1326 (2013).
10. M. Motew, E. G. Booth, S. R. Carpenter, X. Chen, C. J. Kucharik, The synergistic effect of manure supply and extreme precipitation on surface water quality. *Environ. Res. Lett.* **13**, 044016. (2018).
11. S. R. Carpenter, E. G. Booth, C. J. Kucharik, R. C. Lathrop, Extreme daily loads: Role in annual phosphorus input to a north temperate lake. *Aquat Sci* **77**, 71–79 (2014).
12. S. R. Carpenter, E. G. Booth, C. J. Kucharik, Extreme precipitation and phosphorus loads from two agricultural watersheds. *Limnol. Oceanogr.* **63**, 1221–1233 (2018).
13. C. A. Stow, S. R. Carpenter, R. C. Lathrop, A Bayesian observation error model to predict cyanobacterial biovolume from spring total phosphorus in Lake Mendota, Wisconsin. *Can. J. Fish. Aquat. Sci.* **54**, 464–473 (1997).
14. R. C. Lathrop, S. R. Carpenter, C. A. Stow, P. A. Soranno, J. C. Panuska, Phosphorus loading reductions needed to control blue-green algal blooms in Lake Mendota. *Can. J. Fish. Aquat. Sci.* **55**, 1169–1178 (1998).
15. A. M. Michalak *et al.*, Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6448–6452 (2013).
16. D. M. Leech, A. I. Pollard, S. G. Labou, S. E. Hampton, Fewer blue lakes and more murky lakes across the continental US: Implications for planktonic food webs. *Limnol. Oceanogr.* **63**, 2661–2680 (2018).
17. W. W. Carmichael, G. L. Boyer, Health impacts from cyanobacteria harmful algal blooms: Implications for the North American Great Lakes. *Harmful Algae* **54**, 194–212 (2016).
18. J. Huisman *et al.*, Cyanobacterial blooms. *Nat. Rev. Microbiol.* **16**, 471–483 (2018).
19. P. D. Isles, F. Pomati, An operational framework for defining and forecasting phytoplankton blooms. *Front. Ecol. Environ.* **19**, 443–450 (2021).
20. R. W. Katz, G. S. Brush, M. B. Parlange, Statistics of extremes: Modeling ecological disturbances. *Ecology* **86**, 1124–1134 (2005).
21. C. L. E. Franzke, Extremes in dynamic-stochastic systems. *Chaos* **27**, 012101. (2017).
22. M. Scheffer, *Critical Transitions in Nature and Society* (Princeton University Press, Princeton, New Jersey, USA, 2009).
23. D. Trolle *et al.*, Projecting the future ecological state of lakes in Denmark in a 6 degree warming scenario. *Clim. Res.* **64**, 55–72 (2015).
24. R. I. Woolway, B. M. Kraemer, J. Zscheischler, C. Albergel, Compound hot temperature and high chlorophyll extreme events in global lakes. *Environ. Res. Lett.* **16**, (2021).
25. P. A. Soranno, Factors affecting the timing of surface scums and epilimnetic blooms of blue-green algae in a eutrophic lake. *Can. J. Fish. Aquat. Sci.* **54**, 1965–1975 (1997).
26. M. L. Larsen *et al.*, Extreme rainfall drives early onset cyanobacterial bloom. *Facets* **5**, 899–920 (2020).
27. R. W. Katz, "Statistical methods for nonstationary extremes" in *Extremes in a Changing Climate: Detection, Analysis and Uncertainty*, A. AghaKouchak, D. Easterling, K. Hsu, Eds. (Springer, Dordrecht, 2013), chap. 2. 10.1007/978-94-007-4479-02.
28. E. Gilleland, R. W., Katz, extRemes 2.0: An extreme value analysis package in R. *J. Stat. Softw.* **72**, 1–39 (2016).
29. J. W. Kantelhardt, E. Koscielny-Bunde, H. H. A. Rego, S. Havlin, A. Bunde, Detecting long-range correlations with detrended fluctuation analysis. *Physica A: Stat. Mech. Appl.* **295**, 441–454 (2001).
30. A. Bunde, J. F. Eichner, J. W. Kantelhardt, S. Havlin, Long-term memory: A natural mechanism for the clustering of extreme events and anomalous residual times in climate records. *Phys. Rev. Lett.* **94**, 048701. (2005).
31. T. V. Royer, M. B. David, L. E. Gentry, Timing of riverine export of nitrate and phosphorus from agricultural watersheds in Illinois: Implications for reducing nutrient loading to the Mississippi River. *Environ. Sci. Technol.* **40**, 4126–4131 (2006).
32. S. M. Stackpoole, E. G. Stets, L. A. Sprague, Variable impacts of contemporary versus legacy agricultural phosphorus on US river water quality. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 20562–20567 (2019).
33. K. J. Van Meter, N. B. Basu, Time lags in watershed-scale nutrient transport: An exploration of dominant controls. *Environ. Res. Lett.* **12**, 084017. (2017).
34. T. D. Brock, *A Eutrophic Lake: Lake Mendota, Wisconsin* (Springer-Verlag, New York, 1985).
35. R. C. Lathrop, "Lake Mendota and the Yahara River Chain" in *Food Web Management: A Case Study of Lake Mendota*, J. F. Kitchell, Ed. (Springer-Verlag, New York, NY, 1992), chap. 3, pp. 17–30.
36. S. R. Carpenter *et al.*, The ongoing experiment: Restoration of Lake Mendota and its watershed in *Long-Term Dynamics of Lakes in the Landscape*, J. J. Magnuson, T. K. Kratz, B. J. Benson, Eds. (Oxford University Press, London, England, 2006), pp. 236–256.
37. R. C. Lathrop, S. R. Carpenter, Water quality implications from three decades of phosphorus loads and trophic dynamics in the Yahara chain of lakes. *Inland Waters* **4**, 1–14 (2014).
38. R. C. Lathrop, S. R. Carpenter, "Phytoplankton and their relationship to nutrients" in *Food Web Management: A Case Study of Lake Mendota*, J. F. Wisconsin, Kitchell, Eds. (Springer-Verlag, New York, 1992), pp. 99–128.
39. R. C. Lathrop, S. R. Carpenter, L. G. Rudstam, Water clarity in Lake Mendota since 1900: Responses to differing levels of nutrients and herbivory. *Can. J. Fish. Aquat. Sci.* **53**, 2250–2261 (1996).
40. S. G. Coles, *An Introduction to Statistical Modeling of Extreme Values* (Springer, New York, 2001).
41. P. A. Soranno, S. R. Carpenter, R. C. Lathrop, Internal phosphorus loading in Lake Mendota: Response to external loads and weather. *Can. J. Fish. Aquat. Sci.* **54**, 1883–1893 (1997).
42. A. M. Kamarainen, H. Yuan, C. H. Wu, S. R. Carpenter, Estimates of phosphorus entrainment in Lake Mendota: A comparison of one-dimensional and three-dimensional approaches. *Limnol. Oceanogr.* **7**, 553–567 (2009).
43. T. D. Loecke *et al.*, Weather whiplash in agricultural regions drives deterioration of water quality. *Biogeochemistry* **133**, 7–15 (2017).
44. H. Du *et al.*, Precipitation from persistent extremes is increasing in most regions and globally. *Geophys. Res. Lett.* **46**, 6041–6049 (2019).
45. E. Jennings *et al.*, Effects of weather-related episodic events in lakes: An analysis based on high-frequency data. *Freshwater Biol.* **57**, 589–601 (2012).
46. E. C. Pielou, *An Introduction to Mathematical Ecology* (Wiley-Interscience, New York, pp. 294. 1969).
47. F. Serinaldi, C. G. Kilsby, Unsurprising surprises: The frequency of record-breaking and overthreshold hydrological extremes under spatial and temporal dependence. *Water Res. Res.* **54**, 6460–6487 (2018).
48. S. Carpenter, R. Lathrop, Phosphorus loading, transport and concentrations in a lake chain: A probabilistic model to compare management options. *Aquat. Sci.* **76**, 145–154 (2014).
49. S. R. Carpenter *et al.*, Stochastic dynamics of Cyanobacteria in long-term high-frequency observations of a eutrophic lake. *Limnol. Oceanogr. Lett.* **5**, 331–336 (2020).
50. B. M. S. Arani, S. R. Carpenter, L. Lahti, E. H. van Nes, M. Scheffer, Exit time as a measure of ecological resilience. *Science* **372**, eaay4895 (2021).
51. J. C. Vermaire *et al.*, Extrinsic vs. intrinsic regimes shifts in shallow lakes: Long-term response of cyanobacterial blooms to historical catchment phosphorus loading and climate Warming. *Front. Ecol. Evol.* **5**, (2017).
52. H. A. Ewing *et al.*, "New" cyanobacterial blooms are not new: Two centuries of lake production are related to ice cover and land use. *Ecosphere* **11**, e03170. (2020).
53. I. Markelov, R.-M. Couture, R. Fischer, S. Haande, P. Van Cappellen, Coupling water column and sediment biogeochemical dynamics: Modeling internal phosphorus loading, climate change responses, and mitigation measures in Lake Vansjø, Norway. *J. Geophys. Res.* **124**, 3847–3866 (2019).
54. J. F. Kitchell, Ed., *Food Web Management: A Case Study of Lake Mendota* (Springer, New York, 1992).
55. P. Urrutia-Cordero, M. K. Ekvall, L.-A. Hansson, Responses of cyanobacteria to herbivorous zooplankton across predator regimes: Who mows the bloom?. *Freshwater Biol.* **60**, 960–972 (2015).
56. J. R. Walsh, R. C. Lathrop, M. J. Vander Zanden, Invasive invertebrate predator, Bythotrephes longimanus, reverses trophic cascade in a north-temperate lake. *Limnol. Oceanogr.* **62**, 2498–2509 (2017).
57. S. R. Carpenter *et al.*, Plausible futures of a social-ecological system: Yahara watershed, Wisconsin, USA. *Ecol. Soc.* **20**, (2015).

58. J. W. Williams, S. T. Jackson, Novel climates, no-analog communities, and ecological surprises. *Front. Ecol. Environ.* **5**, 475–482 (2007).
59. P. M. Vitousek *et al.*, Nutrient imbalances in agricultural development. *Science* **324**, 1519–1520 (2009).
60. S. R. Carpenter, E. M. Bennett, Reconsideration of the planetary boundary for phosphorus. *Environ. Res. Lett.* **6**, 014009 (2011).
61. S. A. Levin *et al.*, Governance in the face of extreme events: Lessons from evolutionary processes for structuring interventions, and the need to go beyond. *Ecosystems* **25**, 697–711 (2022).
62. C. Folke *et al.*, Our future in the Anthropocene biosphere. *Ambio* **50**, 834–869 (2021).
63. National Academies of Sciences, Medicine. *2021 Nobel Prize Summit: Our Planet, Our Future: Proceedings of a Summit*, F. Carrero-Martínez, N. Sobhani, E. Kameyama, P. Whitacre, Eds. (The National Academies Press, Washington, DC, 2021), p. 88, DOI: 10.17226/26310.
64. J. Qiu, M. G. Turner, Spatial interactions among ecosystem services in an urbanizing agricultural watershed. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12149–12154 (2013).
65. J. Schatz, C. J. Kucharik, Seasonality of the urban heat island effect in Madison, Wisconsin. *J. Appl. Meteorol. Clim.* **53**, 2371–2386 (2014).
66. S. Gillon, E. G. Booth, A. R. Rissman, Shifting drivers and static baselines in environmental governance: Challenges for improving and proving water quality outcomes. *Reg. Environ. Change* **16**, 759–775 (2016).
67. R. C. Lathrop, Perspectives on the eutrophication of the Yahara lakes. *Lake Reserv. Manag.* **23**, 345–365 (2007).
68. L. A. DeCicco, D. Lorenz, R. M. Hirsch, W. Watkins, M. Johnson, *dataRetrieval: R Packages for Discovering and Retrieving Water Data Available from U.S. Federal Hydrologic Web Services* (U.S. Geological Survey, Reston, Virginia, U.S.A., 2021).
69. S. R. Carpenter, Precipitation, discharge, phosphorus load and phycocyanin: Time series plots and cross-correlations. Zenodo. https://doi.org/10.5281/zenodo.6672217. Accessed 28 April 2022.
70. J. J. Magnuson, S. R. Carpenter, E. H. Stanley, North Temperate Lakes, LTER: High frequency data: Meteorological, dissolved oxygen, chlorophyll, phycocyanin - Lake Mendota Buoy 2006-current. Environmental Data Initiative, 10.6073/pasta/fc8bd96677405945024ad708003be1fc. Accessed 3 January 2022.
71. B. Z. Rousso, E. Bertone, R. A. Stewart, K. Rinke, D. P. Hamilton, Light-induced fluorescence quenching leads to errors in sensor measurements of phytoplankton chlorophyll and phycocyanin. *Water Res.* **198**, 117133 (2021).
72. M. L. Pace *et al.*, Reversal of a cyanobacterial bloom in response to early warnings. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 352–357 (2017).
73. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2021).
74. A. L. Bowley, The standard deviation of the correlation coefficient. *J. Am. Stat. Assoc.* **23**, 31–34 (1928).
75. S. R. Carpenter Precipitation, discharge, phosphorus load and phycocyanin: Pareto models. Zenodo. https://doi.org/10.5281/zenodo.6672277. Accessed 28 April 2022.
76. S. R. Carpenter Precipitation, discharge, phosphorus load and phycocyanin: Long-range dependence and models for return intervals. Zenodo. https://doi.org/10.5281/zenodo.6914469. Accessed 28 April 2022.
77. H. v. Storch, F. W. Zwiers, *Statistical Analysis in Climate Research* (Cambridge University Press, Cambridge, New York, 1999).
78. D. Cox, P. A. W. Lewis, *The Statistical Analysis of Series of Events* (Methuen, UK, 1966), 10.1007/978-94-011-7801-3.
79. G. L. Simpson permute: Functions for Generating Restricted Permutations of Data. R package version 0.9-7. https://CRAN.R-project.org/package=permute. (2022).
80. A. G. Constantino nonlinearTseries: Nonlinear Time Series Analysis. R package version 0.2.11. https://CRAN.R-project.org/package=nonlinearTseries. (2021).
81. D. Maraun, H. W. Rust, J. Timmer, Tempting long-memory - on the interpretation of DFA results. *Nonlin. Process. Geophys.* **11**, 495–503 (2004).

Supporting Information for

Long-Range Dependence and Extreme Values of Precipitation, Phosphorus Load and Cyanobacteria

By

S.R. Carpenter, M. R. Gahler, C. Kucharik and E.H. Stanley

Supporting Information

Table of Contents

Time Series: Precipitation, Phosphorus Load, and Phycocyanin

Figure S1. Daily precipitation 1940-2021

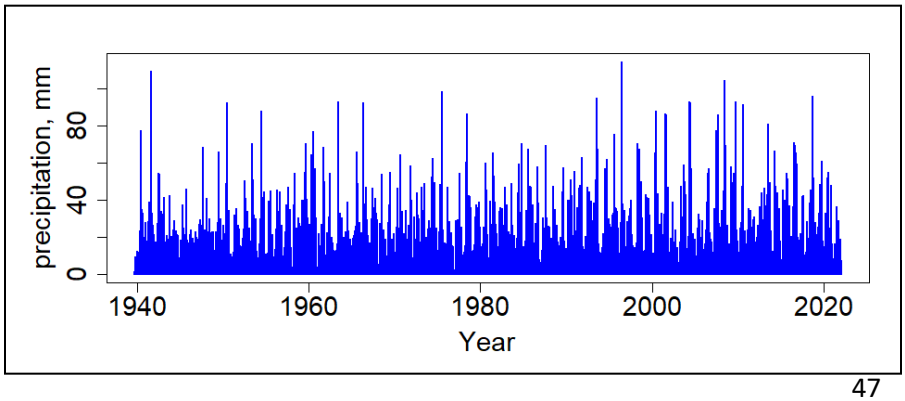


Figure S2. Daily discharge as m^3d^{-1} to Lake Mendota from Yahara River and Pheasant Branch combined, 1990-2021. Note log y-axis.

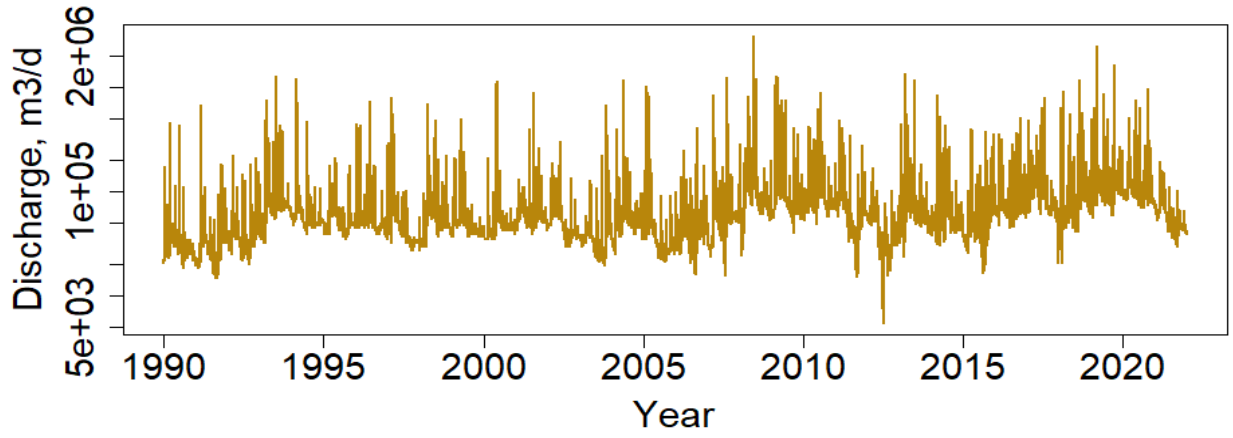
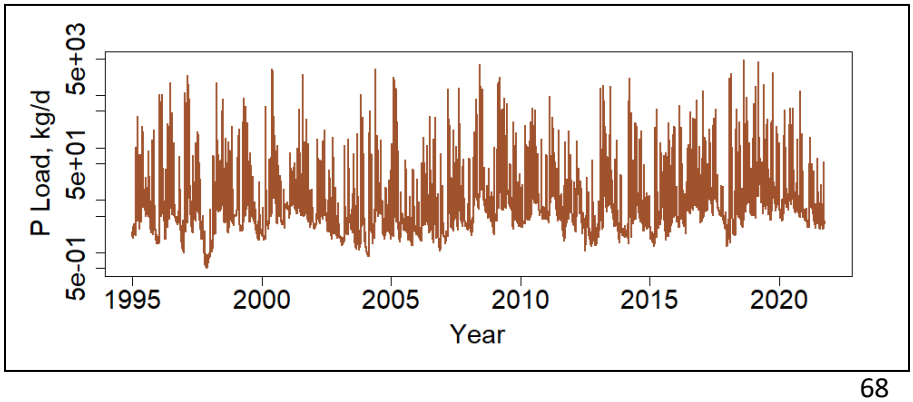


Figure S3. Daily P load to Lake Mendota from Yahara River and Pheasant Branch combined, 1995-2021. Note log y axis. These two tributaries combined account for 30% of the P load to Lake Mendota from all sources and together explain 96% of the variance in P load from all sources (1)



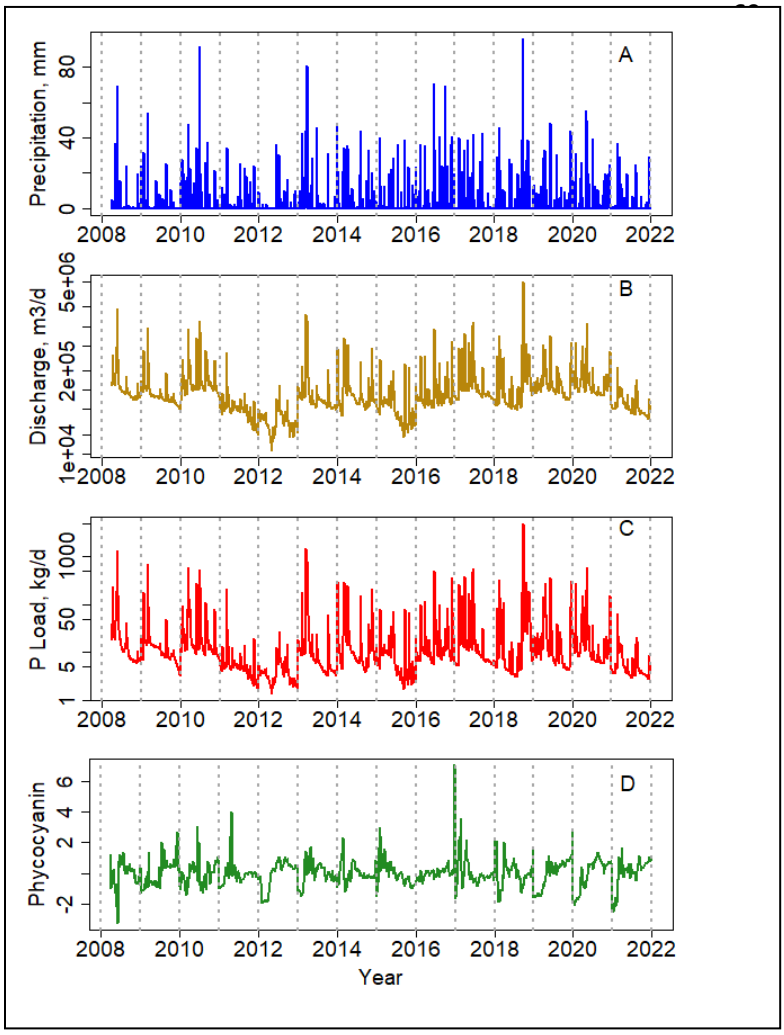


Figure S4. Time series from the days when precipitation, P load, and phycocyanin were all available for Lake Mendota. (A) precipitation (mm d^{-1}), (B) discharge from Yahara River and Pheasant Branch to Lake Mendota (m^3 d^{-1}; note log y axis), (C) P load from Yahara River and Pheasant Branch combined to Lake Mendota (kg d-1; note log y axis), and (C) phycocyanin log10 relative fluorescence units (RFU) measured in Lake Mendota (Z scores, units of standard deviation).

Long-Range Dependence

Long-range dependence, or memory, is indicated by the correlation exponent γ estimated by detrended fluctuation analysis (DFA) (2). The main text presents correlation exponents for concurrent days when all four variates were measured. Here we present DFA results using all available data for each variate. Data and R scripts for DFA are presented online (3).

For a Poisson distribution with minimal long-term dependence or memory γ = 1. Analyses of all available data show that departures from a Poisson pattern increase from precipitation to discharge to phosphorus load, with phycocyanin presenting slightly less long-term dependence than phosphorus load (Table S1). The fluctuation plot for phycocyanin was quadratic, despite detrending within each year to remove seasonality, and the estimate of γ is therefore biased (4). For the other three time series the fluctuation plot was best-fit by a linear model according to AIC, consistent with assumptions of DFA. Thus the phycocyanin data may not meet the assumptions for estimating the correlation exponent. Other studies demonstrate alternate states with rapid transitions between low and high levels of phycocyanin (5). Annual repetition of alternate states suggests long-term memory of nonlinear stochastic dynamics that may be difficult to measure with correlation exponents.

Table S1. Detrended fluctuation analysis results using all available data for daily time series of precipitation, P load, and phycocyanin. Phycocyanin data were detrended within each year to remove seasonal effects.

Time series	Days of data	Self-Affinity α (s.e.)	Correlation Exponent γ (s.e.)
Precipitation	30042	0.531 (0.0023)	0.939 (0.00465)
Log(Discharge)	11684	0.895 (0.010)	0.203 (0.0114)
Log(P load)	9770	0.962 (0.013)	0.0769 (0.0130)
Log Phycocyanin	1438	0.943 (0.018)	0.114 (0.036)

References

1. R. C. Lathrop, S. R. Carpenter, Water quality implications from three decades of phosphorus loads and trophic dynamics in the Yahara chain of lakes. *Inland Waters* **4**, 1-14 (2014).
2. M. Ghil *et al.*, Extreme events: dynamics, statistics and prediction. *Nonlin. Processes Geophys.* **18**, 295-350 (2011).
3. S. R. Carpenter, Precipitation, discharge, phosphorus load and phycocyanin: long-range dependence and models for return intervals. Zenodo.
4. D. Maraun, H. W. Rust, J. Timmer, Tempting long-memory - on the interpretation of DFA results. *Nonlin. Processes Geophys.* **11**, 495-503 (2004).
5. S. R. Carpenter *et al.*, Stochastic dynamics of Cyanobacteria in long-term high-frequency observations of a eutrophic lake. *Limnology and Oceanography Letters* **5**, 331-336 (2020).