

# An operational framework for defining and forecasting phytoplankton blooms

Peter DF Isles<sup>1,2\*</sup> and Francesco Pomati<sup>1,3</sup>

Phytoplankton blooms are complex ecological events that emerge from the dynamics of an entire ecosystem. Increasing efforts to forecast blooms are hampered by inconsistent definitions of what constitutes a bloom event, from both conceptual (mechanistic) and empirical (quantitative) perspectives. By clarifying definitions of blooms using temporal system dynamics, we propose to target modeling and forecasting methods to appropriate settings, and generate testable ecological hypotheses into the underlying processes fueling bloom development. Here, we present a general bloom definition that highlights both growth and loss processes, and identify quantitative metrics of time-series structure associated with several subclasses of blooms. We hypothesize ecological processes that are consistent with these time-series structures, and that suggest promising approaches for forecasting different types of blooms.

*Front Ecol Environ* 2021; doi:10.1002/fee.2376

Over the past 50 years, a considerable amount of research has focused on the deterioration of inland water quality due to phytoplankton blooms (Burford *et al.* 2019). Despite this attention, however, there is still no consensus in the scientific community as to what constitutes a “bloom” event. Blooms in different ecosystems may have different levels of intrinsic predictability and be driven by a variety of dynamics due to emergent properties of complex adaptive systems (Norberg 2004), and as such different types of blooms may be suited to different modeling and forecasting approaches. Drawing these distinctions is difficult in the absence of a clear framework for defining what a bloom event is, or for distinguishing whether there are distinct types of blooms, which is especially relevant today as attempts to forecast phytoplankton blooms become

increasingly more common (Rousso *et al.* 2020). This consideration underscores the need for a widely applicable conceptual framework that, on the basis of underlying system dynamics, can guide modeling and forecasting efforts, facilitate comparisons across systems, and ultimately further our understanding of the mechanisms driving phytoplankton blooms.

In this paper, we present a review of the literature, a summary of existing knowledge of phytoplankton bloom dynamics, and a synthesis of a set of conceptual bloom definitions. Categorization of ecological events is inherently difficult, and occasionally unjustified, because real-world ecological systems lay along continuums in multiple dimensions. The literature on bloom dynamics currently covers a wide range of conditions in which fundamentally different ecological mechanisms are likely to play roles. We therefore believe it is useful to break down the complexity of bloom events by delineating subcategories that facilitate useful comparisons, while acknowledging that the boundaries are gradients rather than hard divisions. The framework that we propose can guide modeling efforts, but also suggests testable ecological hypotheses to investigate phytoplankton bloom triggering factors and proposes data-driven approaches to forecasting suited to different types of blooms.

## In a nutshell:

- No widely accepted definition of what constitutes a phytoplankton bloom currently exists
- This lack of a clear definition complicates efforts to compare patterns and processes across systems, and to develop bloom forecasting guidelines
- We reviewed the literature on algal bloom definitions and dynamics, focusing primarily on factors relevant to defining a bloom event and forecasting blooms
- We then developed a framework of bloom categories that connects temporal patterns in bloom time-series with underlying theories of process consistent with the literature, and propose promising forecasting approaches for each category

## ■ A review of existing phytoplankton bloom definitions and forecasting approaches

### Threshold- and anomaly-based bloom definitions

Definitions based on thresholds of phytoplankton biomass or based on anomalies are common in the literature on “harmful algal blooms” (HABs). In such contexts, HABs are defined as having detrimental effects on one or more target species or communities (WebTable 1; Smayda 1997). HAB definitions are generally associated with specified densities of potentially toxic species, chlorophyll concentrations,

<sup>1</sup>Department of Aquatic Ecology, Swiss Federal Institute of Aquatic Science and Technology (Eawag), Dübendorf, Switzerland; <sup>2</sup>Vermont Department of Environmental Conservation, Watershed Management Division, Montpelier, VT \*(peter.isles@vermont.gov); <sup>3</sup>ETH-Zürich, Institute of Integrative Biology, Zürich, Switzerland

or toxin levels. Thresholds are useful because they are easy to apply widely, but choice of thresholds may be subjective, sensitive to sampling and analysis methods, and vary among monitoring organizations with respect to target variables and alert levels. Although helpful in studies of bloom impacts, threshold-based definitions may not be the most suitable from a predictive standpoint because multiple processes may result in conditions exceeding alert thresholds.

Some researchers have defined blooms as anomalies or extreme events, generally relative to a system-specific baseline (Miller *et al.* 2013) or as deviations from a presumed statistical distribution. Such anomaly-based definitions are useful in situations where blooms occur infrequently but are inadequate in situations where blooms occur regularly. In highly eutrophic lakes, for example, blooms may recur annually, and the normal range of observations may encompass situations characterized by high concentrations of cyanobacteria and potentially toxic conditions. In such systems, anomaly-based definitions may be inappropriate.

### Advective blooms

We define “advective blooms” as those in which the variability in a time-series of algal biomass is dominated by spatial transport of existing biomass rather than biological growth and loss processes. Advective processes can act both vertically and horizontally. Vertical transport of cells upward in the water column can drive surface blooms in buoyant and colonial cyanobacteria, fostered by stable water columns that allow floating colonies to overcome turbulence (Reynolds *et al.* 1987). Forecasting blooms in lakes with vertical heterogeneity and variable thermal structures requires vertically resolved data. Lateral advective processes can concentrate algal biomass in one part of a water body due to currents. These events are particularly common in marine environments and in large lakes. Three-dimensional hydrodynamic models driven by meteorological forecasts are best suited for predicting lateral advective blooms (Davidson *et al.* 2016), although models using convolutional neural networks and satellite images have also shown promising results and are likely to become more widely used (Barzegar *et al.* 2020).

### Ecological definitions of blooms

Several “ecological” definitions of blooms have been proposed that focus on processes governing phytoplankton growth and loss rates, as well as phytoplankton community structure. High *dominance* of a blooming taxon is one such metric. Cyanobacteria blooms, for example, are frequently characterized by the dominance of a single species or functional group (Molot *et al.* 2014), and can even be dominated by a single genotype (D’Alelio *et al.* 2011). However, there are also events with high concentrations of potentially toxic cyanobacteria in species-rich phytoplankton communities (Bukowska *et al.* 2017). Consequently, while diversity may be an important component of bloom definitions, it is insufficient to define a bloom.

Blooms can also be defined in terms of trophic dynamics, as events when *phytoplankton growth outpaces the control by grazers* (Kjørboe *et al.* 1996). This definition can encompass both the spring bloom and grazing-resistant summer blooms. In the spring bloom, rapid growth of relatively edible species, boosted by nutrient availability, leads to temporary accumulation until grazer populations increase in response to abundant food supplies (Sommer *et al.* 1986). In a summer bloom, some phytoplankton can accumulate large populations despite relatively low intrinsic growth rates because of defense traits limiting grazing pressure, such as large mucilaginous or filamentous colonies that are physical impediments to grazing (Reynolds 2007), poor food quality for grazers (Sterner and Hessen 1994; Brett and Müller-Navarra 1997), and toxin production (Janssen 2019). Phytoplankton can therefore form a bloom either by maximizing growth rates or by minimizing loss rates, or by a combination of the two.

Phytoplankton biomass accumulation may be linked to positive or negative feedbacks to top-down (grazing) or bottom-up (nutrient supply) mechanisms that affect the magnitude and temporal dynamics of blooms (Sunda *et al.* 2006). Positive and negative feedbacks may give rise to alternate basins of attraction that allow bloom conditions to persist for longer than they would in the absence of feedbacks; this has led some researchers to consider blooms to be *alternate stable states* (Wilkinson *et al.* 2018). However, not all events with high phytoplankton biomass exhibit alternate stable state dynamics (Spears *et al.* 2017), making the presence of feedbacks a useful but not a necessary consideration in defining blooms.

### Forecasting algal blooms

Models currently used to forecast algal blooms fall into several broad categories (see Rouso *et al.* 2020). Statistical models test specific hypotheses regarding cause and effect, making interpretation relatively simple. Parametric statistical models of phytoplankton biomass as functions of nutrient loading, nutrient concentrations, temperature, or other abiotic drivers have been used to try to predict phytoplankton biomass for decades (Vollenweider 1968; Stumpf *et al.* 2012), but they may be poorly suited to modeling nonlinear dynamics or threshold effects. Bayesian models provide explicit handling of probability distributions and estimation of uncertainty, a critical feature of forecasting from the perspective of ecosystem management (Dietze 2017). Bayesian models can also incorporate existing knowledge in the form of priors and are therefore suited to iterative forecasting approaches (Dietze *et al.* 2018). Classical and Bayesian hypothesis-based models may be difficult to apply if underlying processes are not well understood.

“Process-based” aquatic ecosystem models (AEMs) combine numerical simulations of physical and biogeochemical processes (Hipsey *et al.* 2015; Janssen *et al.* 2015). AEMs are effective tools for testing and generating hypotheses about lake

processes and for generating long-term projections under different climate or nutrient loading scenarios (Zia *et al.* 2016). Producing accurate short-term forecasts is difficult with AEMs, in part due to the challenges of real-time model–data integration. AEMs can be particularly useful in systems where limited data are available, because they incorporate a large body of ecological knowledge in the underlying model design.

In contrast to hypothesis-based models and AEMs, machine-learning (ML) models and other data-driven approaches show promise for strict forecasting applications where no a priori assumptions are made about ecosystem structure beyond the specification of input variables (Reichstein *et al.* 2019). There has recently been a proliferation of ML approaches (eg artificial neural networks, genetic programming, random forests; Recknagel *et al.* 2006; Thomas *et al.* 2018) and model-free methods (Ye *et al.* 2015) applied to ecological forecasting. ML models and model-free methods often perform well in strict forecasting applications (Bzdok *et al.* 2018) but generally have high data requirements. In many cases, it may be worthwhile to consider hybrid modeling approaches (Reichstein *et al.* 2019) that combine the flexibility of ML for poorly understood system variables with the interpretability of process-based models for processes that are more thoroughly understood.

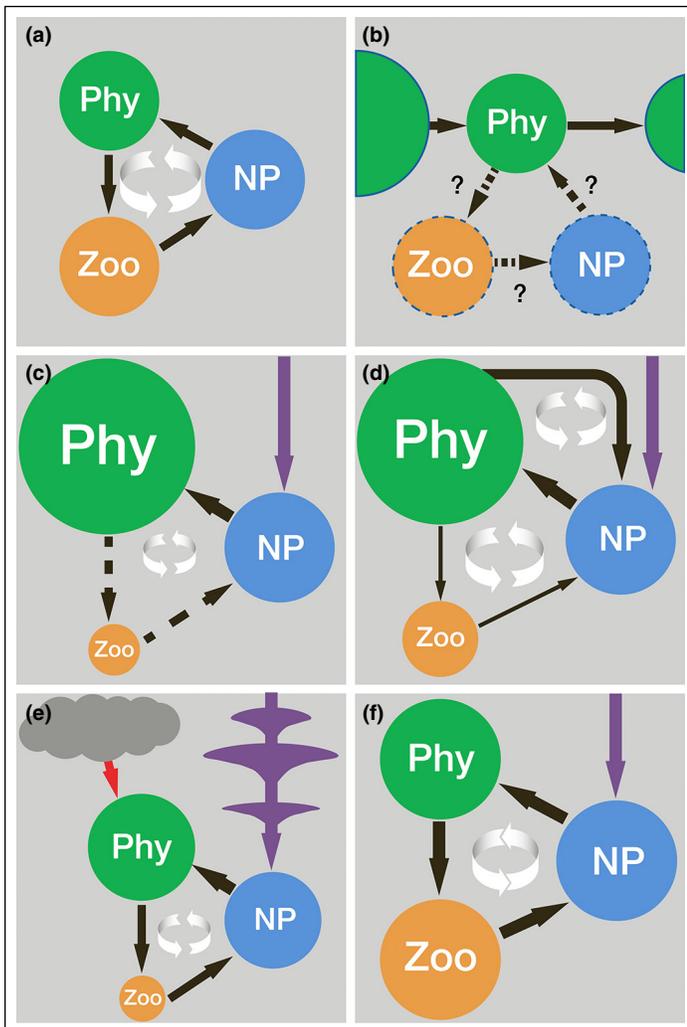
### ■ A set of operational phytoplankton bloom definitions for predictive modeling

We propose that a bloom event can be generally defined as a *period of net phytoplankton biomass accumulation within a defined area or volume, resulting from growth rates that exceed loss rates, followed by eventual decline to near baseline concentrations*. In contrast to other definitions, we do not assign a biomass threshold but instead focus on phytoplankton traits related to growth and loss rates. Our broad definition is inclusive enough to account for events that have a growth phase, a peak, and a decline phase, whether they occur in relatively oligotrophic or relatively eutrophic systems. On the basis of this general definition, we build on the conceptual approach proposed by Ryo *et al.* (2019) to describe subcategories of blooms based on temporal structures. Blooms can be conceived of as single events with several basic shapes, which may interact in various ways that result in emergent time-series structures.

	Basic structure	Sample time-series	Forecasting and diagnostic indicators
0: Seasonally determined bloom			Predictable based on seasonal patterns + 1, 3, 11 - 8
1: Sporadic bloom			Prediction is difficult. Hydrodynamic models may improve forecasts + 6, 8, 13, 14 - 1, 2
2: Self-terminating single bloom			Relatively predictable due to high autocorrelation. + 1, 2 - 5, 6, 14
3: Persistent single bloom			Relatively predictable, but nonlinear dynamics are likely + 1, 2, 7, 10, 12 - 5, 9
4: Single-type composite bloom			Predictability limited by stochastic drivers. + 1, 5 - 2, 14
5: Multiple-type composite bloom			Difficult to forecast due to possible chaotic dynamics. + 5, 6, 14 - 11, 15

**Figure 1.** Proposed framework of operational bloom categories. Left panels indicate the basic temporal structure of each bloom category (x-axis represents time over the course of a single bloom season, y-axis represents phytoplankton biomass). Central panels indicate synthetic time-series resembling real data that would fit in each bloom category (horizontal red lines indicate hypothetical alert thresholds). Inset panels in the center column show kernel density plots of phytoplankton biomass over the time-series. Numbers in the right column refer to diagnostic indicators of time-series structure from WebTable 2; + and – symbols indicate whether relatively high or low values of indicators are associated with that bloom category, respectively. For a more detailed version of this framework, see WebTable 3.

We identify six classes of blooms based upon time-series properties that are consistent with existing knowledge and theories of phytoplankton blooms (Figure 1). The proposed arrangement offers specific testable hypotheses concerning bloom driving mechanisms (Figure 2), as well as metrics of time-series structure that help to identify bloom events and provide clues to intrinsic predictability (WebTable 2). Based on these premises, different modeling and forecasting methodologies may be more or less applicable to each subcategory of bloom.



**Figure 2.** Hypothesized simple food webs showing processes and feedbacks that may explain differences between bloom categories. Phy = phytoplankton, Zoo = zooplankton, NP = nutrients and light. Black arrows represent flows between compartments, with larger arrows representing larger flows and smaller and dashed arrows representing smaller flows; white arrows represent cycling rates; purple arrows represent external nutrient loads. (a) A healthy food web, where modest nutrient inputs result in a stable food web with efficient recycling of nutrients and a Type 0 bloom. (b) A system where phytoplankton dynamics are dominated by advection, potentially resulting in a Type 1 bloom and obscuring underlying ecological dynamics. (c) The emergence of inedible phytoplankton disrupts the negative grazer feedback loop, reducing zooplankton populations and slowing nutrient regeneration, potentially resulting in a Type 2 bloom. (d) A positive feedback develops between phytoplankton and nutrient or light resources, resulting in a self-perpetuating bloom state where nutrient supply is no longer dependent on regeneration by grazers, potentially resulting in a Type 3 bloom. (e) Example of a bloom in a stochastically dominated system, where external nutrient inputs (pulses in purple arrow), storm impacts on phytoplankton populations (cloud), or other stochastic processes (eg disease outbreaks) may result in multiple peaks of a single dominant species and a Type 4 bloom. (f) A eutrophic system where high grazer densities are maintained. Large flows between all system compartments and rapid cycling rates may result in the emergence of chaos and a Type 5 bloom.

### Time-series features for bloom identification

An important goal of our approach is to develop quantitative time-series metrics to distinguish between bloom events and enable comparisons across systems. These metrics are particularly useful when comparing blooms across different lakes or monitoring programs, where clustering raw time-series is complicated by variations in sample frequencies, seasonality or duration of blooms, and bloom magnitudes. Several time-series features have been proposed as signals of underlying ecological processes, particularly relating to early warning indicators (Dakos *et al.* 2012) and ecosystem predictability (Pennekamp *et al.* 2019) (WebTable 2). Many metrics describing time-series structure – beyond commonly used statistics such as mean, median, and variance – have been used in other fields to summarize information in time-series (Fulcher *et al.* 2013; Hyndman *et al.* 2016). Rather than focusing on any single metric, we propose that bloom time-series should be assigned to ecological categories using multiple features, which, when combined with modeling of biomass time-series, may result in more robust time-series classifications.

#### Type 0: baseline – seasonally determined blooms

In “healthy” temperate aquatic ecosystems, predictable seasonal blooms are likely to occur following well-described processes (Sommer *et al.* 1986). Autoregressive models with seasonal components should be well suited to prediction of Type 0 blooms, but seasonality may at times need to be adjusted relative to cardinal events like the onset of stratification or clear water phases in lakes. Mechanisms used to classify seasonally determined blooms are similar to those in other categories described below, but forecasting seasonally determined blooms should be relatively simple; therefore, it may be useful to test for seasonal predictability before attempting more complicated diagnostic and forecasting approaches.

#### Type 1: sporadic blooms

We define sporadic blooms as rare biomass spikes with very high rates of change and rapid rates of decline that deviate from seasonal baselines and appear as anomalies in a time-series (Figure 1). Sporadic blooms can be identified from time-series features associated with low autocorrelation and predictability and high nonlinearity (WebTable 2; Figure 1). Empirical distributions of biomass should have long right-side tails representing bloom events.

Many sporadic blooms may involve a combination of modest increases in total water column biomass coupled with physical processes that concentrate biomass at particular points in space and time (advective blooms). For example, moderate concentrations of cyanobacteria may accumulate at the surface during calm periods through vertical advection, and

horizontal advection along the water surface may concentrate these populations at lake margins (Ndong *et al.* 2017). As such, we might expect sporadic blooms to be commonly observed in shoreline monitoring data (Pobel *et al.* 2011). Advective processes may be inferred if observed growth rates are higher than would be expected based on known maximum growth rates. Forecasting sporadic bloom events is difficult, yet may be possible if the physical conditions that concentrate biomass are well understood. In this case, weather predictions or hydrodynamic model outputs could be used to refine the probabilities of bloom occurrence. Aggregating data from multiple sites may provide a clearer signal of the underlying population dynamics and improve forecasting accuracy.

### Type 2: self-terminating single blooms

We define Type 2 blooms as those with a single prominent biomass peak that reaches a maximum concentration and then declines to concentrations similar to a seasonal baseline (Figure 1). Statistically, these blooms should be characterized by metrics indicating high predictability and autocorrelation, with clear trajectories of growth and decline. Distributions of log-biomass during the bloom should be unimodal.

Type 2 blooms should feature negative feedbacks, the mechanisms of which may be related to the emergence of top-down controls or exhaustion of nutrient supplies, or both (Figures 1 and 2). Blooms with high growth rates may be more susceptible to grazing due to trade-offs between investments in defense (eg large colonies) and growth (eg nutrient uptake rates) (Van Donk *et al.* 2011); these blooms terminate when enemy populations (eg grazers, parasites) increase to a point where they match phytoplankton accumulation. In contrast, blooms with slow growth rates could emerge if selective grazing pressure promotes expansion of less-preferred food species (Mittra and Flynn 2006) that in turn suppress grazer populations and regeneration of nutrients (Figure 2c). This may lead to severe nutrient limitation at the peak of the bloom (Sunda *et al.* 2006), which can then trigger a negative feedback on phytoplankton growth rates to the point where they can no longer compensate for loss processes. Because planktonic nutrient regeneration is low, in such cases maximum biomass may be controlled by external loads of limiting nutrients.

If the bloom is fast growing and composed of edible species, autoregressive predator–prey models are likely to have high predictive ability; if blooms are slow growing and composed primarily of inedible species, nutrient loads – particularly in spring and early summer – will likely be strong predictors of maximum biomass (Stumpf *et al.* 2012). AEMs may perform particularly well for slow-growing Type 2 blooms, due to the focus on bottom-up eutrophication issues in the development of many of these models. Bayesian models are also well suited to Type 2 blooms, and may be especially effective because of the explicit quantification of uncertainty involved in these models (Dietze *et al.* 2018).

### Type 3: persistent single blooms

Type 3 blooms remain at near-peak concentrations for an extended period of time, and are characterized initially by a step rather than a pulse response (cf Ryo *et al.* 2019). These blooms should be characterized by statistical properties associated with multiple basins of attraction and critical transitions (Dakos *et al.* 2012), such as multimodal biomass distributions (Figure 1). Type 3 blooms are likely to be sustained by positive feedbacks to resource supply (eg Figure 2d), such as access to benthic or hypolimnetic nutrient pools through vertical migration (Reynolds *et al.* 1987; Cottingham *et al.* 2015), hypoxia-driven feedbacks resulting in ammonium ( $\text{NH}_4^+$ ) and phosphate ( $\text{PO}_4^{3-}$ ) fluxes from sediments (Hupfer and Lewandowski 2008; McCarthy *et al.* 2016), nitrogen (N) fixation, and monopolization of light resources through positive buoyancy (Reynolds *et al.* 1987). Type 3 blooms eventually end, but termination might occur as a result of changes in exogenous drivers (eg declining temperatures) rather than endogenous feedbacks.

Model-free methods, such as empirical dynamic modeling (Sugihara *et al.* 2012), are explicitly designed to be robust in systems with nonlinear dynamics, as with Type 3 blooms. ML approaches like random forests and neural networks also perform effectively (Thomas *et al.* 2018). Early warning indicators, including flickering (oscillating between two distinct levels of biomass) or critical slowing down (which describes the rate at which a system recovers from small perturbations), may provide advance warnings of bloom onset (Scheffer *et al.* 1993; Dakos *et al.* 2012). AEMs may also be effective if feedback processes are well represented. Some widely used methods such as regressions that predict cyanobacteria as a function of nutrients are likely to perform poorly because bloom-forming species are frequently drivers of nutrient concentrations in addition to being driven by nutrient concentrations (Figure 2d), and predictor variables are likely to have different relationships to cyanobacteria biomass before and after the change in system state, resulting in weak predictive relationships.

### ■ Composite blooms

Observed time-series of blooms often do not have trajectories as clear as the Type 2 and Type 3 blooms described above, and multiple peaks, which may or may not be considered a single event, are frequently observed. We categorize some multiple-peak time-series as “composite blooms”. Multiple individual peaks can occur in response to external drivers like nutrient pulses, heat waves, or mixing events, or they may also arise as a result of within-lake processes due to complex interactions within plankton communities. We divide composite blooms into two primary classes: single type (different peaks are composed of the same species) and multiple type (different peaks are composed of different species).

#### Type 4: single-type composite blooms

These blooms contain multiple biomass peaks and are composed of the same taxa or assemblage of taxa. Individual peaks may be conceived of as discrete bloom events, each of which is initiated by a specific driver, such as a nutrient pulse, a period of calm, hot weather (Jöhnk *et al.* 2008), or wind mixing events that induce light limitation (Huisman *et al.* 1999). Type 4 blooms can be distinguished by statistics indicating multiple time-series peaks of relatively short duration (Figure 1). If peaks overlap, they may merge and resemble other bloom categories; in this case, analysis of long-term data can provide an indication of the underlying dynamics, which may be important for the accurate identification of causal factors.

We hypothesize that stochastic drivers usually determine the timing of peaks (Figure 2e). As such, individual peaks are unlikely to be spaced evenly along the time-series, and may overlap. We further hypothesize that overlapping peaks in a single-species bloom should have a multiplicative effect on blooming populations, because drivers will have a greater effect on total biomass if populations are already elevated due to a previous peak. This is likely to make maximum biomass accumulation in single-type composite blooms particularly sensitive to the timing of drivers responsible for individual peaks (Figure 1).

Forecasting Type 4 blooms is likely to be challenging because stochastic drivers (such as wind or storm events) cannot be reliably estimated beyond the limits of weather forecast models and because the responses to these drivers may often be nonlinear (Carpenter *et al.* 2020). If drivers can be forecasted, the frequency and timing of stochastic events is important for accurate predictions (Isles *et al.* 2017). The most practical approach to prediction might be to first identify stochastic drivers (these can also be internal, like demographic stochasticity) and then develop multiple projection scenarios based on possible driver values drawn from realistic distributions, which can be used to develop estimates of bloom probability.

#### Type 5: multiple-type composite blooms

Determining when individual peaks with different species should be considered as distinct blooms or as a single composite bloom is not straightforward. However, there is substantial evidence of chaotic dynamics in plankton communities (Benincà *et al.* 2008; Telesh *et al.* 2019), and prediction of bloom conditions in such chaotic systems is highly dependent on earlier internal system states, including the dynamics of other species. It may be useful to consider multiple peaks of different taxa as composite blooms if they occur within a period of high and sustained productivity, are highly variable among years, and show evidence of chaotic dynamics (eg Lyapunov exponents greater than zero).

Multiple-type composite blooms may be more likely in systems with high and variable resource supply (nutrients and

light), resulting in generally high biomass and growth rates (Figure 2f) but a changing competitive landscape preventing the emergence of a single dominant group. Mathematical simulations and experimental evidence suggest that chaotic dynamics are most likely to emerge in systems with high growth rates, fluctuating environmental conditions, and/or high grazing rates (Benincà *et al.* 2008).

Because of these chaotic dynamics, multiple-type composite blooms may be particularly difficult to forecast using linear methods. Model-free methods (Sugihara *et al.* 2012, Ye *et al.* 2015) or ML approaches appropriate to time-series (eg long short-term memory recurrent neural networks; Kim *et al.* 2018) could provide short-term predictions if the effects of internal system dynamics outweigh the effects of external stochastic drivers. Aggregating the unit of prediction by taxonomic or functional group may result in models with acceptable forecast accuracy but lower specificity (Pennekamp *et al.* 2017).

#### ■ Advantages and challenges of this bloom identification and forecasting framework

We have developed a basic framework integrating (1) operational phytoplankton bloom classes that are connected to time-series structure (Ryo *et al.* 2019), (2) quantitative diagnostic metrics (WebTable 2; Figure 1), (3) hypotheses associated with each class reflecting current ecological theories of mechanisms dictating bloom dynamics (Figures 1 and 2), and (4) potentially applicable forecasting approaches for each class. We expect the categories and associated metrics proposed here to be further refined in the future. In some cases, it may be difficult to identify which category of bloom is occurring, and applying multiple forecasting methodologies may result in more robust predictions. Comparing features of phytoplankton time-series across multiple systems should make it possible to highlight common features of water bodies predisposed to different bloom types. This will provide managers with more targeted approaches to water resource monitoring, forecasting, and management in the face of ongoing eutrophication, warming, and other global change drivers, and will help to manage expectations for predictability.

Challenges remain when attempting to apply this framework, particularly in times of dramatically changing environmental conditions. Climate change and human impacts are altering the natural structure of ecological communities from local to global scales, driving the reorganization of food webs due to the addition or loss of interactions (Bartley *et al.* 2019). The disruption of ecological networks as a result of, for example, climate change will likely be abrupt (Trisos *et al.* 2020). This consideration has several consequences for modeling and forecasting phytoplankton blooms, which are becoming more frequent (Ho *et al.* 2019), because dominant mechanisms and ecological arrangements (eg Figure 2) may shift. As a result, it might become problematic to forecast

beyond the range of the historical data used to train the forecasting model, particularly for data-driven methods such as ML and EDM. Short- and medium-range data-driven forecasts (daily through annual timescales) may still be possible because year-to-year variability is usually greater than long-term trends at these timescales, but the most extreme events are likely to be overlooked. Long-term (decadal-scale) projections should rely on ensembles of mechanistic models incorporating projected changes in climate, and validated on training data from multiple trophic levels and across large spatial and temporal scales.

The distinctions between bloom categories described here highlight the need for advances in aquatic ecosystem monitoring technologies. Data collected at biweekly or monthly scales (as is the case for most public monitoring programs) may be insufficient for resolving the temporal dynamics described in our conceptual model. In addition, traditional monitoring data are usually not available in real-time, precluding the possibility of rapid short- and medium-range data-driven forecasts. It is now common practice to deploy sensors measuring bulk biomass (eg chlorophyll-*a* or phycocyanin fluorescence) that stream data in real-time, but bulk measures may be inadequate for identifying the individual taxon dynamics necessary for understanding certain underlying processes. Emerging technologies, such as in situ imaging systems (Lombard *et al.* 2019), in situ flow cytometers (Thomas *et al.* 2018), and multiple fluorescence probes, could be used to accurately distinguish blooms of different species or functional groups at high temporal resolution. Our conceptual framework will be particularly useful with these types of data, where substantial information about system-level processes may be contained in temporal structures that are not observable in data at coarser resolution.

## ■ Acknowledgements

This work was supported by the Swiss Federal Institute of Aquatic Science and Technology (Eawag) through internal discretionary funds. We thank V Dakos for useful discussions.

## ■ References

- Bartley TJ, McCann KS, Bieg C, *et al.* 2019. Food web rewiring in a changing world. *Nat Ecol Evol* **3**: 345–54.
- Barzegar R, Aalami MT, and Adamowski J. 2020. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stoch Env Res Risk A* **34**: 415–33.
- Benincà E, Huisman J, Heerkloss R, *et al.* 2008. Chaos in a long-term experiment with a plankton community. *Nature* **451**: 822–25.
- Brett MT and Müller-Navarra DC. 1997. The role of highly unsaturated fatty acids in aquatic foodweb processes. *Freshwater Biol* **38**: 483–99.
- Bukowska A, Kaliński T, Koper M, *et al.* 2017. Predicting blooms of toxic cyanobacteria in eutrophic lakes with diverse cyanobacterial communities. *Sci Rep-UK* **7**: 8342.
- Burford M, Carey C, Hamilton DP, *et al.* 2019. Perspective: advancing the research agenda for improving understanding of cyanobacteria in a future of global change. *Harmful Algae* **91**: 101601.
- Bzdok D, Altman N, and Krzywinski M. 2018. Points of significance: statistics versus machine learning. *Nat Methods* **15**: 233–34.
- Carpenter SR, Arani BMS, Hanson PC, *et al.* 2020. Stochastic dynamics of cyanobacteria in long-term high-frequency observations of a eutrophic lake. *Limnol Oceanogr Lett* **5**: 331–36.
- Cottingham KL, Ewing HA, Greer ML, *et al.* 2015. Cyanobacteria as biological drivers of lake nitrogen and phosphorus cycling. *Ecosphere* **6**: 1–19.
- D’Alelio D, Gandolfi A, Boscaini A, *et al.* 2011. *Planktothrix* populations in subalpine lakes: selection for strains with strong gas vesicles as a function of lake depth, morphometry and circulation. *Freshwater Biol* **56**: 1481–93.
- Dakos V, Carpenter SR, Brock WA, *et al.* 2012. Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PLoS ONE* **7**: e41010.
- Davidson K, Anderson DM, Mateus M, *et al.* 2016. Forecasting the risk of harmful algal blooms. *Harmful Algae* **53**: 1–7.
- Dietze MC, Fox A, Beck-Johnson LM, *et al.* 2018. Iterative near-term ecological forecasting: needs, opportunities, and challenges. *P Natl Acad Sci USA* **115**: 1424–32.
- Dietze MC. 2017. Ecological forecasting. Princeton, NJ: Princeton University Press.
- Fulcher BD, Little MA, and Jones NS. 2013. Highly comparative time-series analysis: the empirical structure of time series and their methods. *J R Soc Interface* **10**: 20130048.
- Hipsey MR, Hamilton DP, Hanson PC, *et al.* 2015. Predicting the resilience and recovery of aquatic systems: a framework for model evolution within environmental observatories. *Water Resour Res* **51**: 7023–43.
- Ho JC, Michalak AM, and Pahlevan N. 2019. Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature* **574**: 667–70.
- Huisman J, Jonker RR, Zonneveld C, and Weissing FJ. 1999. Competition for light between phytoplankton species: experimental tests of mechanistic theory. *Ecology* **80**: 211–22.
- Hupfer M and Lewandowski J. 2008. Oxygen controls the phosphorus release from lake sediments – a long-lasting paradigm in limnology. *Int Rev Hydrobiol* **93**: 415–32.
- Hyndman RJ, Wang E, and Laptev N. 2016. Large-scale unusual time series detection. In: Cui P, Dy J, and Aggarwal C, *et al.* (Eds). Proceedings of the 15th Institute of Electrical and Electronics Engineers (IEEE) International Conference on Data Mining Workshops; 14–17 Nov 2015; Atlantic City, NJ. New York, NY: IEEE.
- Isles PDF, Rizzo DM, Xu Y, and Schroth AW. 2017. Modeling the drivers of interannual variability in cyanobacterial bloom severity using self-organizing maps and high-frequency data. *Inland Waters* **7**: 333–47.
- Janssen ABG, Arhonditsis GB, Beusen A, *et al.* 2015. Exploring, exploiting and evolving diversity of aquatic ecosystem models: a community perspective. *Aquat Ecol* **49**: 513–48.
- Janssen EM-L. 2019. Cyanobacterial peptides beyond microcystins – a review on co-occurrence, toxicity, and challenges for risk assessment. *Water Res* **151**: 488–99.

- Jöhnk KD, Huisman J, Sharples J, *et al.* 2008. Summer heatwaves promote blooms of harmful cyanobacteria. *Glob Change Biol* **14**: 495–512.
- Kim K, Kim D-K, Noh J, and Kim M. 2018. Stable forecasting of environmental time series via long short term memory recurrent neural network. *IEEE Access* **6**: 75216–28.
- Kjørboe T, Hansen JLS, Alldredge AL, *et al.* 1996. Sedimentation of phytoplankton during a diatom bloom: rates and mechanisms. *J Mar Res* **54**: 1123–48.
- Lombard F, Boss E, Waite AM, *et al.* 2019. Globally consistent quantitative observations of planktonic ecosystems. *Front Mar Sci* **6**: 196.
- McCarthy MJ, Gardner WS, Lehmann MF, *et al.* 2016. Benthic nitrogen regeneration, fixation, and denitrification in a temperate, eutrophic lake: effects on the nitrogen budget and cyanobacteria blooms. *Limnol Oceanogr* **61**: 1406–23.
- Miller TR, Beversdorf L, Chaston SD, and McMahon KD. 2013. Spatiotemporal molecular analysis of cyanobacteria blooms reveals *Microcystis-Aphanizomenon* interactions. *PLoS ONE* **8**: e74933–15.
- Mitra A and Flynn KJ. 2006. Promotion of harmful algal blooms by zooplankton predatory activity. *Biol Lett-UK* **2**: 194–97.
- Molot LA, Watson SB, Creed IF, *et al.* 2014. A novel model for cyanobacteria bloom formation: the critical role of anoxia and ferrous iron. *Freshwater Biol* **59**: 1323–40.
- Ndong M, Bird D, Nguyen Quang T, *et al.* 2017. A novel Eulerian approach for modelling cyanobacteria movement: thin layer formation and recurrent risk to drinking water intakes. *Water Res* **127**: 191–203.
- Norberg J. 2004. Biodiversity and ecosystem functioning: a complex adaptive systems approach. *Limnol Oceanogr* **49**: 1269–77.
- Pennekamp F, Adamson MW, Petchey OL, *et al.* 2017. The practice of prediction: what can ecologists learn from applied, ecology-related fields? *Ecol Complex* **32**: 156–67.
- Pennekamp F, Iles AC, Garland J, *et al.* 2019. The intrinsic predictability of ecological time series and its potential to guide forecasting. *Ecol Monogr* **89**: e01359.
- Pobel D, Robin J, and Humbert J-F. 2011. Influence of sampling strategies on the monitoring of cyanobacteria in shallow lakes: lessons from a case study in France. *Water Res* **45**: 1005–14.
- Recknagel F, Cao H, Kim B, *et al.* 2006. Unravelling and forecasting algal population dynamics in two lakes different in morphometry and eutrophication by neural and evolutionary computation. *Ecol Inform* **1**: 133–51.
- Reichstein M, Camps-Valls G, Stevens B, *et al.* 2019. Deep learning and process understanding for data-driven Earth system science. *Nature* **566**: 195–204.
- Reynolds CS, Oliver RL, and Walsby AE. 1987. Cyanobacterial dominance: the role of buoyancy regulation in dynamic lake environments. *New Zeal J Mar Fresh* **21**: 379–90.
- Reynolds CS. 2007. Variability in the provision and function of mucilage in phytoplankton: facultative responses to the environment. *Hydrobiologia* **578**: 37–45.
- Rousso BZ, Bertone E, Stewart R, and Hamilton DP. 2020. A systematic literature review of forecasting and predictive models for cyanobacteria blooms in freshwater lakes. *Water Res* **182**: 115959.
- Ryo M, Aguilar-Trigueros CA, Pinek L, *et al.* 2019. Basic principles of temporal dynamics. *Trends Ecol Evol* **34**: 723–33.
- Scheffer M, Hosper SH, Meijer ML, *et al.* 1993. Alternative equilibria in shallow lakes. *Trends Ecol Evol* **8**: 275–79.
- Smayda TJ. 1997. Harmful algal blooms: their ecophysiology and general relevance to phytoplankton blooms in the sea. *Limnol Oceanogr* **42**: 1137–53.
- Sommer U, Gliwicz Z, Lampert W, and Duncan A. 1986. The PEG-model of seasonal succession of planktonic events in fresh waters. *Arch Hydrobiol* **106**: 433–71.
- Spears BM, Fitter MN, Jeppesen E, *et al.* 2017. Ecological resilience in lakes and the conjunction fallacy. *Nat Ecol Evol* **1**: 1616–24.
- Sterner RW and Hessen DO. 1994. Algal nutrient limitation and the nutrition of aquatic herbivores. *Annu Rev Ecol Syst* **25**: 1–29.
- Stumpf RP, Wynne TT, Baker DB, and Fahnenstiel GL. 2012. Interannual variability of cyanobacterial blooms in Lake Erie. *PLoS ONE* **7**: e42444.
- Sugihara G, May R, Ye H, *et al.* 2012. Detecting causality in complex ecosystems. *Science* **338**: 496–500.
- Sunda WG, Graneli E, and Gobler CJ. 2006. Positive feedback and the development and persistence of ecosystem disruptive algal blooms. *J Phycol* **42**: 963–74.
- Telesh IV, Schubert H, Joehnk KD, *et al.* 2019. Chaos theory discloses triggers and drivers of plankton dynamics in stable environment. *Sci Rep-UK* **9**: 1–12.
- Thomas MK, Fontana S, Reyes M, *et al.* 2018. The predictability of a lake phytoplankton community, over time-scales of hours to years. *Ecol Lett* **21**: 619–28.
- Trisos CH, Merow C, and Pigot AL. 2020. The projected timing of abrupt ecological disruption from climate change. *Nature* **580**: 496–501.
- Van Donk E, Ianora A, and Vos M. 2011. Induced defences in marine and freshwater phytoplankton: a review. *Hydrobiologia* **668**: 3–19.
- Vollenweider RA. 1968. Scientific fundamentals of the eutrophication of lakes and flowing waters, with particular reference to nitrogen and phosphorus as factors in eutrophication. Paris, France: OECD.
- Wilkinson GM, Carpenter SR, Cole JJ, *et al.* 2018. Early warning signals precede cyanobacterial blooms in multiple whole-lake experiments. *Ecol Monogr* **88**: 188–203.
- Ye H, Beamish RJ, Glaser SM, *et al.* 2015. Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. *P Natl Acad Sci USA* **112**: E1569–76.
- Zia A, Bomblies A, Schroth AW, *et al.* 2016. Coupled impacts of climate and land use change across a river–lake continuum: insights from an integrated assessment model of Lake Champlain’s Missisquoi Basin, 2000–2040. *Environ Res Lett* **11**: 1–13.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## Supporting Information

Additional, web-only material may be found in the online version of this article at <http://onlinelibrary.wiley.com/doi/10.1002/fee.2376/supinfo>