Title: Rethinking discretization to advance limnology in a rapidly changing world

Author: B.M. Kraemer

1) IGB Leibniz Institute for Freshwater Ecology and Inland Fisheries, Berlin, Germany,

Correspondence to: ben.m.kraemer@gmail.com
Abstract:

Limnologists often adhere to a discretized view of waterbodies—they classify them, divide them into zones, promote discrete management targets, and use research tools, experimental designs, and statistical analyses focused on discretization. This approach to limnology has profoundly benefited the way we understand, manage, and communicate about waterbodies. But the research questions and the research tools in limnology are changing rapidly with consequences for the relevance of our current discretization schemes. Here, I examine how and why we discretize and argue that selectively rethinking the extent to which we must discretize gives us an exceptional chance to advance limnology in new ways.

Keywords: Classification, management, big data, computing, statistics, trophic state, zonation

Main Text:

Rethinking discretization can elicit limnological progress

Limnology was founded on the premise that lake ecosystems are distinct from the terrestrial ecosystems around them. Lake ecosystems were considered superb habitats for ecological experiments because their biological, chemical, and physical processes were considered discrete. In 1887, Stephen Forbes, a founder of limnology, went so far
as to doubt whether annihilating all terrestrial animals would have any important effect on lake ecosystems at all (Forbes, 1925).

But, we know today that lake ecosystems are intricately connected to surrounding ecosystems and vice versa. The boundary where lake ecosystems end and the next ecosystem begins has become increasingly blurred in recent decades. During the research boom on “ecosystem subsidies” beginning in the 2000’s, it was shown that a large proportion of lake ecosystem carbon can be derived from terrestrial production (Pace et al., 2004). And it is recognized today that lakes emit greenhouse gases that have widespread effects beyond their boundaries (Raymond et al., 2013). In recognition that lakes are not discrete ecosystems, we have substantially advanced our understanding of their role in global biogeochemical cycles (Tranvik et al., 2018). The emergent recognition that lake ecosystems are not discrete has been a substantial limnological advance (Tranvik et al., 2018).

Limnologists have begun to critically reflect on many other aspects of discretization, leading to a variety of other limnological advances. For example, in recognition of the subjectivity of defining discrete mixed layers, some recently argued that an idealised concept of a well-defined mixed layer does not necessarily reflect the reality of aquatic physics (Gray et al., 2019). In recognition of the substantial amount of subsurface flow across catchments, “discrete” hydrogeological units are increasingly recognized as having no real boundaries to water movement (Fan, 2019). By moving beyond research focused on single, discrete waterbody types, we have also begun to find the remarkable
commonalities among seemingly disparate aquatic ecosystem types leading to more
general theories for how waterbodies function. For instance, waterbodies across size
and flow gradients have been shown to have similar controls on their nutrient limitation
(Elser et al., 2007), metabolism (Yvon-Durocher et al., 2012), trophic cascades (Shurin
et al., 2002), and responses to human activity (King et al., 2019).

The benefits of rethinking discretization are also demonstrated by major scientific
advances in fields outside limnology. In biology, Charles Darwin and Alfred Russel
Wallace challenged the discretized worldview that all organisms belong to discrete,
unrelated species. In his famous “theory of relativity,” Albert Einstein powerfully
demonstrated that time and space are not discrete, further arguing that “physical reality
must be described in terms of continuous functions” (Einstein 1979). In social
psychology, Judith Butler radically disrupted the binary view of sex, gender, and
sexuality (Butler, 1990). Over the course of the 20th century, and especially since the
rapid growth of human genome sequencing, it has become clear that standard racial
categories do not reflect actual genetic structure in humans (Yudell et al., 2016). These
key advances show how rethinking our discretization systems can promote inspiring
scientific progress.

In limnology, opportunities for rethinking discretizations are as widespread as the
discretizations themselves. We classify waterbodies based on their size (pond, lake,
Great Lake, or Ocean), flow rates (creek, stream, river, or large river), trophic status
(oligotrophic, mesotrophic, or eutrophic), mixing (stratified or unstratified), salinity
(freshwater or saline), latitude (tropical, temperate, or arctic), and human hydrological influence (lake or reservoir). We also divide waterbodies into different zones according to their mixing (epilimnion, metalimnion, or hypolimnion), light climate (photic or aphotic), distance from shore (littoral or limnetic), and distance from the bottom (benthic or pelagic). The publications which codify these waterbody discretizations are often very well-cited (e.g. Lewis Jr. 1983).

But any calls for casting aside specific waterbody discretizations are likely to be met with fierce resistance because limnological progress to date has relied heavily on them. For instance, early in limnology’s history, lakes were divided into low, medium, and high trophic classes based on their productivity (Carlson, 1977). So-called “trophic state classifications” expanded rapidly following this early work, with new limnologists frequently reinventing the scale. But in the 1970’s, limnologists began to find the contradictions among trophic state classification systems problematic because the class boundaries were scientifically un-testable, preventing the development of a single general classification system (Carlson, 1977). Robert E. Carlson argued for replacing a classification-based system with a less arbitrary continuous gradient from 0 to 100 called the “trophic state index” (Carlson, 1977). While Carlson’s paper receives many citations, the discretized, classification-based systems are still widely-used in limnology today. Similar calls for replacing other waterbody classifications or zonations are likely to be met with similar resistance, and often with good reason.
Limnologists rely on trophic state classifications and other waterbody discretizations because they offer extraordinarily useful shortcuts that can facilitate limnological progress, management, and communication. In a hypothetical research context, high-resolution sampling for total phosphorus might be unnecessary if a discrete, 3-sample approximation from the epilimnion, metalimnion, and hypolimnion would suffice to answer the research question at hand. In a management context, the ecological integrity of a waterbody could be painstakingly described using the complete nucleotide sequences for all organisms that occupy it. But, simply classifying the waterbody’s ecological integrity as “good” or “bad” based on the presence of a few indicator species may be adequate, depending on the management goal. Discretization can also simplify communicating with the public. In the case of public swimming advisories, a dichotomous “safe” or “not safe” advisory may facilitate swimmer decisions about whether or not to get in the water. But an advisory stating the exact quantitative probabilities that swimmers will be exposed to all toxins may complicate rather than facilitate swimmer decision-making. Thus, the appropriateness of any specific discretization system depends on the objectives and the tools at hand to meet them.

But our objectives and our tools in limnology are changing rapidly. Recently deployed remote sensing platforms have higher resolutions that capture more waterbodies at higher frequencies. Continuously profiling cameras can collect underwater images of microscopic organisms, process those images using artificial intelligence, and generate real-time biodiversity profiles. Automated sensors for measuring dissolved nutrients are being improved every day, inspiring new questions about the drivers of high-frequency
variability. Numerous computing resources are putting advanced statistical tools at our finger tips for free. National, international, and global databases containing data from many thousands of waterbodies are inspiring new questions and becoming an increasingly important tool to understand aquatic ecosystems (King et al., 2019). And mobile technologies, social media, and open access philosophies are reshaping the ways limnologists communicate with each other and with the public.

These rapid changes could have consequences for how and when limnologists find certain discretizations appropriate. Just as modern genomics has strengthened calls to move beyond the biological concept of race as a scientific categorization (Yudell et al., 2016), changes in limnology may also require rethinking limnological discretizations. Arguments about the appropriateness of any specific discretization scheme in limnology may seem trivial. But in aggregate, resolving these arguments could substantially influence how the ongoing information deluge improves our science. Fully casting aside any specific discretization system may be reckless and extremely impractical, but questioning the extent to which we must rely on them is an appropriate and timely pursuit given the ongoing changes to the field. Importantly, the process of selectively rethinking discretization should be informed by a keen awareness of discretization’s advantages and disadvantages discussed here.

2. Advantages and disadvantages of discretization
Waterbody discretization is widespread, in part, because it can facilitate communication among experts by offering useful linguistic shortcuts—jargon. Single words of limnological jargon can stand for whole paragraphs of text in plain English, so jargon can save time and lead to the development of a unifying scholarly identity. But just as the jargon associated with discretization can facilitate communication among experts, it can also hamper communication with non-experts. Jargon is widely denounced in the public sphere by scientific communication specialists who view it as a key boundary to public understanding of science. Jargon associated with waterbody discretization can even hamper communication among experts from closely related fields as different fields can have different definitions for the same jargon (e.g. the ‘littoral zone’ refers to the ‘shallow illuminated zone’ in freshwater ecology but to the “intertidal zone” in marine ecology). Thus, waterbody discretization and its associated jargon can be profoundly useful by expediting communication among experts, but it can hamper communication with non-experts and experts from closely related fields.

Discretizing waterbodies can benefit limnology by guiding expectations for how waterbodies function leading to important research and collaborations. For instance, many limnologists inherently expect “discrete” classes and zones to behave in distinct ways. Reservoirs are thought to function in fundamentally different ways compared to lakes (Hayes et al., 2017). Some limnologists encourage developing a unique limnology for very small ponds (Hoverman and Johnson, 2012). And large rivers, it is argued, should be modelled separately from other rivers (Puckridge et al., 1998). These expectations can lead limnologists to form discrete, collaborative teams focused on
illuminating exceptionally important research topics even if they are predominantly relevant only for specific waterbody types and zones. As a result, ponds, lakes, wetlands, streams, rivers, and oceans are typically studied in isolation at both fine and broad scales due to their perceived differences (Chaloner and Wotton 2011; King et al. 2019).

But expectations that waterbody classes and zones represent real structure in nature can also be a disadvantage. The compartmentalization of limnology is problematic because it counteracts the formulation of general ecological theory and hypotheses founded on waterbody relatedness. It has been argued that this isolation has slowed the development of a common mechanistic understanding of the drivers of carbon (Hotchkiss et al. 2018), nutrient (Elser et al. 2007) and energy (Chaloner and Wotton 2011) dynamics in aquatic ecosystems. When limnologists study different waterbody types and zones concurrently along continuous gradients, the basic commonalities among all waterbodies can be more apparent, which promotes synthesis and general theory formulation (Chaloner and Wotton 2011; Hotchkiss et al. 2018). Trans-disciplinarity is widely touted in the scientific literature (Chaloner and Wotton 2011), and may be key to merging the understanding generated from studying specific waterbody types.

Like pixelating an image, waterbody discretization also partially masks variability within groups, causing a loss of signal in the gradient (Gray et al., 2019). For example, limnologists often divide the continuous gradient of human influence on waterbody
hydrology into the categories, “lake” and “reservoir.” But these terms mask the wide variety of waterbodies that fall along the continuous gradient that underlies them. The term, “reservoir” can be used to describe waterbodies ranging from those that have been created by humans de novo outside a water network to those that are pre-existing with slight modifications in their water levels due to a dam. Discretization errors associated with the term “reservoir” are partly reduced by adding more classes (e.g. “run-of-river” reservoir). Adding enough classes to sufficiently reduce discretization errors can take the limnological lexicon down a path toward overbearing complexity that impedes communication rather than enhancing it (e.g. semi-lacustrine-oligotrophic-tropical-run-of-river reservoir).

Discretization in limnology goes far beyond waterbody classification and zonation—discretization is also widely relied on by limnologists when promoting certain resource management targets. For example, discrete pollution limits are promoted as a waterbody management tool which allows polluters to pollute up to a specific threshold without having to pay. Threshold-based management is widely encouraged in the scientific literature (Liu et al., 2015) and adopted by local, national, and international management authorities with considerable resource benefits (e.g. Total Maximum Daily Loads). Discrete management targets have been relied on for decades to control waterbody stressors with some success.

But in some contexts, management approaches based on discrete management targets can be suboptimal. The effectiveness of discrete management targets partially depends
on whether waterbodies have a predictable, well-defined, discrete capacity to withstand stress. There is an abundant limnological literature on thresholds in stressor-response relationships, but this literature shows that strong thresholds in these relationships are rare, uncertain, and difficult to predict (Groffman et al., 2006; Gsell et al., 2016). Furthermore, potential thresholds may be dynamic through time, making them an unrealistic management target even with multitudinous data from a specific system. Discrete stressor targets can also cause defeatism—the thinking that stressor reductions are valuable only if they take stressors below a threshold. Defeatism may prevent incremental stressor reductions which are beneficial but don’t meet the threshold. Conversely, discrete stressor targets can cause complacency—the thinking that stressor reductions that occur below a threshold are worthless, when further reductions would still elicit resource benefits.

Limnologists’ discretized approach to collecting and analysing data has been the core of limnological progress for decades. Limnologists sample waterbodies at discrete depths and sampling locations; design experiments using discrete, replicated treatments (Johnson et al., 2009); test for statistical significance using discrete p-values; cluster data using k-means and other clustering approaches (Savoy et al., 2019); and use classification trees to explain variability in data by dividing it into classes based on discrete cut-offs in predictor variables (O’Reilly et al., 2015). These discretizations can save time, resources, and can simplify the interpretation of statistical findings.
Despite the widespread dependence of limnological progress on data collection and analysis using discretization, overdependence on the practice can lead limnologists to design limnological studies with less statistical power that are more difficult to incorporate into ecological theory. For instance, analysis of variance (ANOVA)—a widely used statistical technique in limnology—has less statistical power than linear regressions when both tests’ assumptions are met, yet we often design experiments with a discrete, replicated ANOVA design (Cottingham et al., 2005). Furthermore, linear regression can provide quantitative output with fewer parameters that can be more effectively incorporated into ecological models than ANOVA output (Cottingham et al., 2005). Simple classification tree analysis is also commonly used in limnology, but its discretized simplicity belies its many disadvantages. Simple classification trees make highly approximated representations of continuous functions, their output can be extremely unstable when fit with new data, and other methods vastly outperform them according to widely ranging performance metrics (Prasad et al., 2006). Limnologists also use discrete thresholds in p-values to test for statistical significance. Yet statisticians have made widespread calls to end dichotomous significance testing because the practice often leads to misunderstandings and misinterpretation of results (Amrhein et al., 2019; Johnson, 2007). Just as dichotomous significance testing has been widely criticized, so too have the various statistical clustering approaches commonly used in limnology for easily finding structure in unstructured data (Cormack, 1971).

3. The big picture, recommendations, and the future of discretization
Regardless of the advantages and disadvantages, discretization is partially unassailable because it is essential to science, language, and modern computing. After all, science itself advances through discrete datasets, analyses, and publications. Publications are written using language which is constructed with discrete letters and words. As you read this, your brain needs discretization to group the individual letters that make up this sentence into something understandable. Moreover, computers are essentially discretization machines—turning precise information into 1's and 0's so that it can be represented by on-off transistors in the computer's hardware. Casting aside discretization as a whole would mean casting aside all of modern science, language, and computing. Thus, discretization is perpetually engrained in limnology as it is in virtually all human pursuits.

As humans, the tendency to discretize has a potent evolutionary basis. We often discretize to make decisions on how to react (e.g. was that snake that just bit me poisonous, or not?). When facing threats or challenges, we also discretize ourselves into cooperative groups that may be better positioned to overcome them. This evolutionary human tendency to discretize spills over into many aspects of our research leading to the plethora of ways in which we discretize, for better and worse. Aside from our general human reliance on discretization for science, language, and computing, discretization may simply be unavoidable in specific contexts. Our data may have been pre-discretized, we may lack sufficient funds, sampling power, or computational power to fully capture continuous gradients, or implement regression-based experimental
designs. Or, we may be temporarily required by law to discretize (e.g. section 314 of the United States Clean Water Act requires that lakes be classified according to their "eutrophic" character). In these cases where discretization is the only option, the best we can do is to avoid discretization’s disadvantages.

But when limnologists do have a choice, I suggest that we carefully examine the value of our discretizations and rigorously compare them to the alternatives. The value of any discretization system should be judged based on the extent to which it satisfies three criteria: objectivity (independent researchers make the same conclusion about the number and definitions of discrete boundaries), stability (doesn’t change when new observations contribute to the system), and predictivity (performs well when predicting independent variables). The objectivity, stability, and predictivity of discretizations should always be compared to alternatives such as analogous continuous gradients when they are available. For example, the HydroLakes database, containing information on 1.4 million lakes worldwide, now includes data on the proportion of the waterbody’s volume which has been impounded (continuous gradient) in addition to the discretized categories, “lake” and “reservoir” (Messager et al., 2016). Tests of the relative predictive power of these two variables could be illuminating. Similarly, discrete pollution targets should be rigorously compared against pollution taxes or pollution trading that affect all levels of pollution regardless of whether it is above or below a discrete threshold. Pollution taxes and trading may lead to more optimal outcomes in resource management in some contexts (Muller and Mendelsohn, 2009), but the relative effectiveness of these two management approaches needs to be tested for waterbodies...
at broad scales. Rigorous testing of discretization schemes according to their objectivity, stability, and predictivity may even lead us toward scientifically testable discretizations that avoid potentially unproductive arguments about the appropriate number and definitions of boundaries between groups.

Discretization is inevitable in limnology because it is at the core of science, language, and computing. Discretization primarily supports limnological progress by providing shortcuts and tools to meet research and management objectives. But the numerous advantages and disadvantages call for careful consideration of the ways we discretize. More than 100 years since Stephen Forbes’ publication of “The Lake as a Microcosm,” we have moved beyond the idea that lake ecosystems are discrete. Limnologists may be poised to further transform their reliance on discretization in the next 100 years. Radical innovations in sensing technology could largely supplant the need for discrete water samples. Unforeseen statistical and data visualization approaches could allow us to better capture and communicate the full signal in continuous gradients. And quantum computing could even move us one big step past binary computing’s limitations (Arute et al., 2019). The extent to which we benefit from these future changes will depend on our mindfulness of discretization and the role it plays in our research. Selectively rethinking discretization could open up new interactions, new questions, and could even lead to the next major advance. So for now, let’s carefully weigh the advantages against the disadvantages and compare to the alternatives when they are available. Doing so will improve our science.
4. Conclusions:

Discretization has profoundly benefitted the way we study, understand, manage, and communicate about aquatic ecosystems. But, limnologists sometimes have a choice over the extent to which they rely on discretization. In these cases, discretization’s advantages should be carefully weighed against their disadvantages.

The core advantage of discretization is that it can provide extraordinarily useful shortcuts, especially when faced with limited resources. These shortcuts can facilitate data collection, data analysis, data interpretation, communication, decision-making, management, and can be used to guide expectations for how ecosystems function. But, discretization also has several key disadvantages in specific contexts which are sometimes overlooked. Discretization can inhibit communication, distract from general theory formation, introduce unnecessary subjectivity, mask the relatedness between discrete groups, mask the variability within discrete groups, lead to suboptimal management approaches, and lead to suboptimal research designs.

Discretization is partially unassailable because it is the foundation of modern science, language, and computing. But recent changes to the field of limnology including big data and high resolution sensors have challenged specific aspects of the way we discretize, leading to substantial limnological advances. In light of the rapid and ongoing changes to the field of limnology, we encourage the careful and selective examination of various aspects of limnological discretization. This examination may help limnology stay relevant in a rapidly changing world.
Acknowledgements:
The author would like to thank Catalina Munteanu and Karan Kakouei for their feedback on previous versions of this manuscript.

Funding:
The author was supported by funding from the Leibniz Institute for Freshwater Ecology and Inland Fisheries’ International Postdoctoral Research Fellowship and from the German Research Foundation through the LimnoScenES project (AD 91/22-1).

References:

Sank, D., Satzinger, K.J., Smelyanskiy, V., Sung, K.J., Trevithick, M.D.,


https://doi.org/10.1016/j.watres.2019.115136


https://doi.org/10.1007/s10021-003-0142-z


https://doi.org/10.1073/PNAS.1608242113


https://doi.org/10.1002/lol2.10036


Naumenko, M., Noges, P., Noges, T., North, R.P., Plisnier, P.-D.D., Rigosi, A.,
Rimmer, A., Rogora, M., Rudstam, L.G., Rusak, J.A., Salmaso, N., Samal, N.R.,
Teubner, K., Verburg, P., Voutilainen, A., Watkinson, A., Williamson, C.E., Zhang,
G., 2015. Rapid and highly variable warming of lake surface waters around the
https://doi.org/10.1002/2015GL066235

Pace, M.L., Cole, J.J., Carpenter, S.R., Kitchell, J.F., Hodgson, J.R., Van De Bogert,
https://doi.org/10.1038/nature02227

Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer Classification and Regression Tree
Techniques: Bagging and Random Forests for Ecological Prediction. Ecosystems
9, 181–199. https://doi.org/10.1007/s10021-005-0054-1

https://doi.org/10.1071/MF94161

Raymond, P.A., Hartmann, J., Lauerwald, R., Sobek, S., McDonald, C., Hoover, M.,
Butman, D., Striegl, R., Mayorga, E., Humborg, C., Kortelainen, P., Dürr, H.,
Meybeck, M., Ciais, P., Guth, P., 2013. Global carbon dioxide emissions from

Savoy, P., Appling, A.P., Heffernan, J.B., Stets, E.G., Read, J.S., Harvey, J.W.,
Bernhardt, E.S., 2019. Metabolic rhythms in flowing waters: An approach for
classifying river productivity regimes. Limnol. Oceanogr.

https://doi.org/10.1002/lno.11154


