Title: Sampling requirements and approaches to detect ecosystem shifts

Authors: Rosalie Bruel[1*] and Easton R. White[*†2,3]

Addresses: [1]Rubenstein Ecosystem Science Laboratory, University of Vermont, VT 05405, USA, [2]Department of Biology, University of Vermont, VT 05405, USA, [3]Gund Institute for Environment, University of Vermont, VT 05405, USA, * Both authors contributed contributed equally and order was chosen alphabetically, †Corresponding author

Keywords: time series, changepoint, monitoring, optimal sampling, Cladocera

Number of words: 3400

Number of references: 40

Number of figures: 6

Submission as: Article at *Ecological Indicators*

Preprint available at: https://www.biorxiv.org/content/10.1101/2020.04.07.030643v1

Sampling requirements and approaches to detect ecosystem shifts

Rosalie Bruel*[1] and Easton R. White*†[2,3]

[1]Rubenstein Ecosystem Science Laboratory, University of Vermont, VT, USA

[2]Department of Biology, University of Vermont, VT, USA

[3]Gund Institute for Environment, University of Vermont, VT, USA

Abstract

Environmental monitoring is a key component of understanding and managing ecosystems. Given that most monitoring efforts are still expensive and time-consuming, it is essential that monitoring programs are designed to be efficient and effective. In many situations, the expensive part of monitoring is not sample collection, but instead sample processing, which leads to only a subset of the samples being processed. For example, sediment or ice cores can be quickly obtained in the field, but they require weeks or months of processing in a laboratory setting. Standard sub-sampling approaches often involve equally-spaced sampling. We use simulations to show how many samples, and which types of sampling approaches, are the most effective in detecting ecosystem change. We test these ideas with a case study of Cladocera community assemblage indicators reconstructed from a sediment core. We demonstrate that standard approaches to sample processing are less efficient than an iterative approach. For our case study, using an optimal sampling approach would have resulted in savings of 195 person-hours—thousands of dollars in labor costs. We also show that, compared with these standard approaches, fewer samples are typically needed to achieve high statistical power. We explain how our approach can be applied to monitoring programs that rely on video records, eDNA, remote sensing, and other common tools that allow re-sampling.

Keywords: time series, changepoint, monitoring, optimal sampling, Cladocera, paleoecology

*The authors contributed equally to this work. †Corresponding author (eastonrwhite@gmail.com)

Introduction

Environmental monitoring is one of the core components to modern ecosystem research and management (McDonald-Madden *et al.* 2010, White 2019, Lindenmayer *et al.* 2020). Within an adaptive management framework, monitoring is needed for both learning about the system under study and assessing the effectiveness of management interventions (Lovett *et al.* 2007). Increasingly, long-term monitoring programs, like the Long Term Ecological Research (LTER) Network in the USA, are becoming available (Maguran *et al.* 2010). However, environmental monitoring is still often expensive and time-consuming, especially when further processing is needed following sample collection (Zhang & Zhang 2012). Thus, for many fields there is a disparity between the amount of data that can be acquired and stored, and the ultimate number of samples that can be processed. Therefore, monitoring programs need to be designed in such a way to address the question of interest while using limited resources efficiently (**???**, Legg & Nagy 2006, McDonald-Madden *et al.* 2010, Lindenmayer *et al.* 2020).

Monitoring programs characteristics must be tightly linked to the questions of interest. For example, White (2019) found that 72% of vertebrate populations required at least 10 years of monitoring to detect significant changes in the population size over time. The specific number of years required depended on the species biology and the detection method used (White 2019). Other work has focused on the frequency of monitoring (Wauchope *et al.* 2019), the impact of allocating monitoring resources spatially versus temporally (Rhodes & Jonzen 2011, Weiser *et al.* 2019), imperfect detection (**???**), and the costs and benefits of increasing sampling breadth relying on citizen science (Weiser *et al.* 2020). Lastly, both the ecological and economical costs of failing to detect a true trend (type II error) have to be weighed against the risks of false (type I error) detection (Mapstone 1995). Given limited budgets, monitoring programs need to be designed to be cost-effective (Caughlan 2001, Grantham *et al.* 2008, Bennett *et al.* 2016).

Because ecological systems are dynamic in both space and time, it is essential that sampling designs account for spatio-temporal dynamics (Williams *et al.* 2018). To estimate a particular parameter, e.g. population abundance, optimal spatial sampling strategies are based on spatially balanced sampling (Kermorvant *et al.* 2019), while optimal temporal sampling strategies are more cost efficient with a targeted sampling, e.g., around the period of reproduction (Jackson *et al.* 2008). In either case, sampling time and locations can be chosen in an iterative process to be cost-effective and reduce the uncertainty in the process (Hooten *et al.* 2009). With the ability to choose precisely when to sample, we can move beyond random, interval, or opportunistic sampling designs. This is particularly relevant in situations where a subset of samples already collected need to be analyzed.

Because of new technological advances, there are many data sources that can be derived long after the actual processes occurred. For example, sediment cores can be retrieved from aquatic ecosystems with little sediment disturbances, such as lakes or lagoon, allowing reconstruction of past ecological communities or conditions (Cohen 2003). Similarly, environmental samples (e.g. water, soil) can be saved and processed later for composition, including eDNA (Bohmann *et al.* 2014). Likewise, photo- or video-based monitoring can record snapshots of a system and be analyzed later (O'Connell *et al.* 2011, Mallet & Pelletier 2014). In each of these cases, decisions have to made about how much data to extract from the previously collected samples (Zhang & Zhang 2012). Should the paleoecological core be analyzed at every centimeter? Should the video be assessed once per minute? As long as processing samples is expensive, these trade-offs will remain.

Building on ideas from dynamic survey design and optimal monitoring research, we develop a set of tools to determine the appropriate number of samples and sampling approach when dealing with data sources where only a subset of samples are analyzed. We tailor our analysis to the detection of a changepoint in a time series, but our approach is applicable to other questions as well. Changepoints are an important characteristic of a time series as they can indicate a underlying change in ecosystem processes (James & Matteson 2014). We focus on paleoecological core samples as one example of this type of data. We examine the situation where the goal is to detect the time at which a significant change in an ecological community occurs, i.e. a changepoint. However, our approaches are widely applicable to other questions and data types. We first investigate these tools using a simulation-based approach. We then test the tools on a case study from a paleosequence of Cladocera community assemblage from Lake Varese located in the subalpine region of north-western Italy (Bruel *et al.* 2018).

Sampling approaches and changepoint detection

For both our simulations and case study, we investigate the effect of different sampling strategies on our ability to detect a changepoint. We begin by either creating simulated time series or using an actual paleoecological time series (Fig. 1). We then subsampled each time series to test the effect of three different sampling approaches along with varying the sample size (Fig. 1c-f). We compared the estimated changepoint from the subsampled time series to that of the full time series as a measure of the effectiveness.

The random sampling approach involves taking a set number of random points throughout the time series (Fig. 1c). In the context of sediment cores, this would mean analyzing community composition at random locations along the core. Random sampling is recommended in designs aimed at quantifying the average size of a population (spatial approach) (Nad'o & Kaňuch 2018). We hypothesize that random sampling will perform the worst in estimating the changepoint. Regular sampling is commonly used (e.g., pigments in Milan *et al.* (2015)) and requires that a set number of samples be taken at regular intervals (Fig. 1d). Lastly, iterative sampling involves first taking a set number of samples and then iteratively adding samples until a pre-determined level of precision is achieved (Fig. 1e-g). For each scenario, we begin by sampling the first and last sample to ensure coverage of the whole time period. We describe each approach in more detail in the supplementary material and provide code.

We detect changepoints with the function *e.divisive* in the R package *ecp* (James *et al.* 2019). There are several methods available to detect changepoint (reviewed in James & Matteson (2014)); *e.divisive* is a divisive hierarchical estimation algorithm for multiple change point analysis. We chose this method because it is able to perform multiple change point analysis for both uni- and multi-variate time series, without a priori knowledge of the number of changepoints. Herein, we focus on detecting the most important changepoint (i.e. the one of largest magnitude), although we tested the method on a time-series that would have multiple changepoints (Fig. S3). In order to test the performance, we detected the "true" changepoint on the whole time-series, and compare the changepoint found on the sub-sample with the "true" one. The distance to true changepoint served as the performance diagnostic.

Simulation approach

Simulation model

We began with a theoretical exploration of the sampling requirements to detect a changepoint. We modeled a simple first order autoregressive (AR-1) process (the discrete-time version of the Ornstein–Uhlenbeck process) with a response variable (X_t) that represents either population size, biodiversity, or some other unidimensional metric of community composition at time t. The model includes temporal autocorrelation (ϕ), the mean of the process (μ_X), and a white noise term (w_t). The white noise term is a normal distribution with mean (μ_w) and variance (σ^2):

$$\begin{aligned} X_t &= \mu_X + \phi(X_{t-1} - \mu_X) + w_t \\ w_t &\sim Norm(\mu_w, \sigma^2). \end{aligned}$$

We included a changepoint by shifting μ_X at time τ given a specific shift size (δ).

We explored how each of these model parameters affected our ability to detect a change point. We simulate an entire time series to serve as the "true" data for comparison (White & Bahlai 2020). We specifically study how the number of samples and the type of sampling affects the detection probability. For simulations, statistical power is the fraction of simulations that were able to detect a changepoint. We define an accurate changepont detection if an estimate is within five time points (given a time series of 100 time points) of the true changepoint. The minimum number of samples required is the number needed for 0.8 statistical power.

Simulation results

In line with theory on optimal monitoring, we found that the probability of correcting identifying a changepoint decreased with smaller levels of population variability (σ) and autocorrelation (ϕ) (Fig. 2). We also found that the probability of correct changepoint detection increases with larger shift sizes, which is essentially the effect size (Fig. 2). There were interaction effects between the variables. For example, autocorrelation was only important if population variability was high (Fig. 2). Thus, the number of samples required to obtain high statistical power (above 0.8) increased with larger population variability, lower autocorrelation, and smaller shift sizes (Fig. 3). As predicted, iterative sampling performed best, followed by regular and random sampling (Fig. 3). The distance to the true changepoint, and consequently the minimum number of samples required, was lower for iterative sampling. (Fig. 3)

Case study

Case study background

We examined the performance of our approach to detect changepoints in a paleosequence. Paleolimnology allows to reconstruct past environments over long periods of time, under the premise that sedimentation was not perturbed (low mixing and disturbances). A sediment core is typically subsampled to narrow down periods of time to be compared, either at regular intervals (e.g., Milan *et al.* (2015)), or continuously (e.g., Perga *et al.* (2015)).

We tested different sampling methods on a real community time series from Lake Varese (IT), with the objective to detect the main changepoint in zooplankton Cladocera assemblage. Lake Varese is a small (14.8 km^2), deep (z_{max}= 26m), monomictic lake, in the subalpine region of north-western Italy (238 m asl). It underwent hyper-eutrophication over the 20th century due to increase in nutrient loads from the watershed. Nutrient status was responsible for restructuration of the lake communities across trophic levels (Crosta 1999, Bruel *et al.* 2018). Air temperature is now driving changes in plankton communities (Bruel *et al.* 2018).

In a previous study, Cladoceran assemblage was reconstructed continuously with a 1-cm subsampling resolution (2-3 years resolution), from a 74-cm sediment core covering the 1816(±26)—2010 time period (Bruel *et al.* 2018). Our objective was to evaluate whether the same changepoints could be identified using less samples. In this previous study, the variability in the community was summarized into independent axis using Detrended Component Analysis (Hill & Gauch 1980). Changepoints were then detected on the first component (46% of the total variability) in years 1926, 1946, and 1983. We defined these as the "true" changepoints given they came from an analysis of the complete data. The 1983 changepoint was the largest in magnitude, hence the changepoint we sought to find with our method. We also identified second and third changepoints (Fig. S3).

In line with our simulation approach, we subsampled the full record (74 observations) using the three methods described earlier (random, regular, iterative). These subsamples were from the initial community dataset (Fig. 6a). We reduced the dimensionality of the assemblage-level data to an ordination axis using the same method than in the original study, and detected the changepoint on the first component (univariate vector). In the case of the iterative method, a new sample was added and the ordination was run again (Fig. 6d). For each of the three methods, we examined the error (difference between the true changepoint and the detected changepoint) when using different numbers of samples.

Case study results

We found that random sampling performed the worst, as changepoint analysis was left to chance (Fig. 5). Regular sampling provided good estimates from 8 samples, but detecting the true changepoint depended on the interval falling close to the true changepoint (i.e., also left to chance). Iterative sampling performed the best, as no more than 9 samples were ever necessary to get the true changepoint (Fig. 5c). We show how iterative sampling slightly changes the scores on the first component but not the overall ordination, as more samples are added (Fig. 6)

We also tested how the three methods performed at detecting other changepoints of lower magnitude (as three changepoints were detected in the initial study, Bruel *et al.* (2018)). Iterative sampling still performed best, especially if an higher number of initial samples (7) was chosen (Fig. S3).

As another application of our approach, we examined the same sediment core data, but examined total abundance as opposed to community composition (Fig. S4). We tested the three subsampling methods, and it took 11 samples to find the “true” changepoint (Fig. S5). The initial 5 subsamples analyzed were the same than the subsamples analyzed to answer the question of the change in community (Fig. 5c). The implication is that a very limited number of processed sample can rapidly and efficiently be used to narrow down different questions on a same dataset.

Discussion

Due to time or funding limitations, there is often a difference between the number of samples collected and the total number of samples that can later be processed. When the processing time is disproportionately higher than the collection time especially, a subsampling can be done prior to processing. A decision must then be made as to which subsamples to analyze. To address this question in the context of detecting changepoints, we tested three subsampling methods: subsampling random points, regular intervals, and an iterative sampling approach (Fig. 1). We found that the iterative method was systematically better at detecting changes than the two other methods, random subsampling being the least efficient (Figs. 4, 5, S1, S2). Autocorrelation, variance, and shift size, had an effect on how many samples were needed to detect the shift, but did not change which approach was optimal (Fig. 3).

Multiple subsampling strategies can be chosen (Fig. 1), but only iterative sampling detected the true changepoint with a limited number of samples (Fig. 4c). Analyzing 11% of the sample was enough in most cases to approach the "true" changepoint. Applied to the real case study, the iterative method allowed us to find the main changepoint with only 9 samples analyzed (Fig. 5). The method also worked well to detect other changepoints of lower magnitude (Fig. S3). Bruel *et al.* (2018) processed one sample at each centimeter in a 74-centimeter sediment core. Each sample took an average of 3 hours to process. We found that using an iterative approach would have eliminated 195 hours of sample processing, or about 24 days, which is just a little over a month of work. This correspond to several thousands of US dollars depending on labor costs.

Our approach goes beyond just paleoecological analyses. Running simulations or using past data to understand the amount of sampling effort required is important in many systems where sample collection or processing is expensive (White 2019, White & Bahlai 2020). The specific sampling techniques can also be compared to determine the optimal strategy in terms of accuracy and cost. Our specific approach applies to situations where more subsamples can be added, or processed, after the dynamics occurred (Zhang & Zhang 2012). It corresponds very well to paleoecological data: samples are taken long after the phenomenon of interest occurred, and allows subsampling at finer or rougher intervals (Wingard *et al.* 2017). However, both different types of data and different questions than those used here can be addressed with the same approach. Suppose instead that the goal was to detect a change in relative abundance over time with video-based approaches. It is often not practical to watch entire videos, so it can be useful to choose strategic time-points that would address a specific question of interest. Using an interval sampling approach, one could take a fixed number of samples to start. The trend over time from simple linear regression could be taken. Then, samples can be taken at random locations one-by-one and to see which samples have the largest effect on the trend estimates. If a particular sample has a large effect on the trend, then it would be best to choose another nearby sample. Sampling would continue until the trend estimate reached an equilibrium. Thus, the iterative sampling approach is particularly relevant to data sources where additional samples can be taken long after the initial dynamics. These approaches would also be appropriate for environmental samples, such as water or soil, that can be analyzed later or eDNA that can be extracted from previously-collected samples (Bohmann *et al.* 2014).

Our approach is applicable to a wide range of systems and questions, but it does have limitations. When less resources are needed for sample analysis, as opposed to collection, investigators will likely be able to process every sample, and analyzing all samples to obtain a whole picture may be preferred. We note that if resources need to be saved by collecting less samples in the first place, then regular sampling performs better than random sampling (Figs. 4, S1, S2). Another example where our method is less useful is when addressing questions that require a continuous time series, or at least a regular sampling interval. For example, continuous, high-resolution subsampling of a time-series is generally required to detect critical slowing-down or early warning of shifts (Frossard *et al.* 2015, Doncaster *et al.* 2016).

However, recent work suggest that combining indicators (in the specific study, trait dynamics and abundance-based early warning signals) allows forecasting population collapses even with at lower resolution and time-series length (Arkilanian *et al.* 2020). Critical slowing down does not necessarily result in a shift, and a shift can occur without critical slowing down (Spears *et al.* 2017). Signs of critical slowing downs are important to understand and recognize because they provide potential early warnings (Doncaster *et al.* 2016), but in terms of management, knowing the timing of a shift can have larger implications in addressing the underlying driver. Thus, selecting a set number of samples or specific approach may also limit what future questions can be asked.

Conclusions

Analyzing a subsample of a time series as opposed to the whole time series will inevitably leads to a lesser understanding of the phenomenon observed (White 2019). We show here that an informed subsampling can still allow detection of critical information, such as a changepoint in a time series. Monitoring programs have to be able to address our questions of interest with sufficient statistical power. In addition, optimizing sampling efforts is valuable given the high costs of many monitoring programs (Caughlan 2001, Bennett *et al.* 2014). Thus, costs of monitoring have to weighed against the value gained from monitoring—a value of information approach (Lovett *et al.* 2007, Bennett *et al.* 2018). Monitoring programs should try to anticipate the potential questions of tomorrow, and reducing the data collected, or analyzed, must be done with the best foresight possible on how these data may be necessary to manage ecosystems in the future. If only a subsample of the samples can be analyzed, it may be better to choose samples strategically as opposed to random or regular sampling.

This can improve the accuracy of the results and reduce costs overall.

Author contributions

ERW and RB conceived the ideas and designed methodology; RB collected the data; ERW designed the simulations; RB wrote the R code that simulates the three sampling approaches; ERW and RB analysed the data; Both authors contributed critically to the drafts and gave final approval for publication

Data availability

Data and code for all the figures can be found at https://github.com/rosalieb/temporal-sampling.

References

Arkilanian, A.A., Clements, C.F., Ozgul, A. & Baruah, G. (2020). Effect of time series length and resolution on abundance- and trait-based early warning signals of population declines. *Ecology*, n/a, e03040.

Bennett, J.R., Maxwell, S.L., Martin, A.E., Chadès, I., Fahrig, L. & Gilbert, B. (2018). When to monitor and when to act: Value of information theory for multiple management units and limited budgets. *Journal of Applied Ecology*, 55, 2102–2113.

Bennett, J.R., Rühland, K.M. & Smol, J.P. (2016). No magic number: Determining cost-effective sample size and enumeration effort for diatom-based environmental assessment analyses. *Canadian Journal of Fisheries and Aquatic Sciences*, 74, 208–215.

Bennett, J.R., Sisson, D.R., Smol, J.P., Cumming, B.F., Possingham, H.P. & Buckley, Y.M. (2014). Optimizing taxonomic resolution and sampling effort to design cost-effective ecological models for environmental assessment. *Journal of Applied Ecology*, 51, 1722–1732.

Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S. & Knapp, M. *et al.* (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution*, 29, 358–367.

Bruel, R., Marchetto, A., Bernard, A., Lami, A., Sabatier, P. & Frossard, V. *et al.* (2018). Seeking alternative stable states in a deep lake. *Freshwater Biology*.

Caughlan, L. (2001). Cost considerations for long-term ecological monitoring. *Ecological Indicators*, 1, 123–134.

Cohen, A.S. (2003). *Paleolimnology: The History and Evolution of Lake Systems*. OUP USA, Oxford.

Crosta, M. (1999). Il Lago di Varese: evoluzione trofica negli ultimi quarant'anni e stato attuale. PhD thesis. University of Insubria.

Doncaster, C.P., Alonso Chávez, V., Viguier, C., Wang, R., Zhang, E. & Dong, X. *et al.* (2016). Early warning of critical transitions in biodiversity from compositional disorder. *Ecology*, 97, 3079–3090.

Frossard, V., Saussereau, B., Perasso, A. & Gillet, F. (2015). What is the robustness of early warning signals to temporal aggregation? *Frontiers in Ecology and Evolution*, 3.

Grantham, H.S., Moilanen, A., Wilson, K.A., Pressey, R.L., Rebelo, T.G. & Possingham, H.P. (2008). Diminishing return on investment for biodiversity data in conservation planning. *Conservation Letters*, 1, 190–198.

Hill, M. & Gauch, H. (1980). Detrended correspondance analysis: An improved ordination technique. *Vegetatio*, 42, 47–58.

Hooten, M.B., Wikle, C.K., Sheriff, S.L. & Rushin, J.W. (2009). Optimal spatio-temporal hybrid sampling designs for ecological monitoring. *Journal of Vegetation Science*, 20, 639–649.

Jackson, A.L., Broderick, A.C., Fuller, W.J., Glen, F., Ruxton, G.D. & Godley, B.J. (2008). Sampling design and its effect on population monitoring: How much monitoring do turtles really need? *Biological Conservation*, 141, 2932–2941.

James, N.A. & Matteson, D.S. (2014). Ecp: An R Package for Nonparametric Multiple Change Point Analysis of Multivariate Data. *Journal of Statistical Software*, 62, 1–25.

James, N.A., Zhang, W. & Matteson, D.S. (2019). Ecp: Non-Parametric Multiple Change-Point Analysis of Multivariate Data.

Kermorvant, C., Caill-Milly, N., Bru, N. & D'Amico, F. (2019). Optimizing cost-efficiency of long term monitoring programs by using spatially balanced sampling designs: The case of manila clams in Arcachon bay. *Ecological Informatics*, 49, 32–39.

Legg, C.J. & Nagy, L. (2006). Why most conservation monitoring is, but need not be, a waste of time. *Journal of Environmental Management*, 78, 194–199.

Lindenmayer, D., Woinarski, J., Legge, S., Southwell, D., Lavery, T. & Robinson, N. *et al.* (2020). A checklist of attributes for effective monitoring of threatened species and threatened ecosystems. *Journal of Environmental Management*, 262, 110312.

Lovett, G.M., Burns, D.A., Driscoll, C.T., Jenkins, J.C., Mitchell, M.J. & Rustad, L. *et al.* (2007). Who needs environmental monitoring. *Frontiers in Ecology and the Environment*, 5, 253–260.

Maguran, A.E., Baillie, S.R., Buckland, S.T., Dick, J.M., Elston, D.A. & Scott, E.M. *et al.* (2010). Long-term datasets in biodiversity research and monitoring: Assessing change in ecological communities through time. *Trends in Ecology and Evolution*, 25, 574–582.

Mallet, D. & Pelletier, D. (2014). Underwater video techniques for observing coastal marine biodiversity: A review of sixty years of publications (19522012). *Fisheries Research*, 154, 44–62.

Mapstone, B.D. (1995). Scalable decision rules for environmental impact studies : Effect Size , type I , and type II errors. *Ecological Applications*, 5, 401–410.

McDonald-Madden, E., Baxter, P.W.J., Fuller, R.A., Martin, T.G., Game, E.T. & Montambault, J. *et al.* (2010). Monitoring does not always count. *Trends in Ecology and Evolution*, 25, 547–550.

Milan, M., Bigler, C., Salmaso, N., Guella, G. & Tolotti, M. (2015). Multiproxy reconstruction of a large and deep subalpine lake's ecological history since the Middle Ages. *Journal of Great Lakes research*, 41, 982–994.

Nad'o, L. & Kaňuch, P. (2018). Why sampling ratio matters: Logistic regression and studies of habitat use. *PLoS ONE*, 13.

O'Connell, A.F., Nichols, J.D. & Karanth, K.U. (Eds.). (2011). *Camera Traps in Animal Ecology*. Springer Japan, Tokyo.

Perga, M.-E., Frossard, V., Jenny, J.-P., Alric, B., Arnaud, F. & Berthon, V. *et al.* (2015). High-resolution paleolimnology opens new management perspectives for lakes adaptation to climate warming. *Frontiers in Ecology and Evolution*, 3.

Rhodes, J.R. & Jonzen, N. (2011). Monitoring temporal trends in spatially structured populations: How should sampling effort be allocated between space and time? *Ecography*, 34, 1040–1048.

Spears, B.M., Futter, M.N., Jeppesen, E., Huser, B.J., Ives, S. & Davidson, T.A. *et al.* (2017). Ecological resilience in lakes and the conjunction fallacy. *Nature Ecology & Evolution*, 1.

Wauchope, H.S., Johnston, A., Amano, T. & Sutherland, W.J. (2019). When can we trust population trends? A method for quantifying the effects of sampling interval and duration.

Methods in Ecology and Evolution, 498170–498170.

Weiser, E.L., Diffendorfer, J.E., Grundel, R., López-Hoffman, L., Pecoraro, S. & Semmens, D. *et al.* (2019). Balancing sampling intensity against spatial coverage for a community science monitoring programme. *Journal of Applied Ecology*, 56, 2252–2263.

Weiser, E.L., Diffendorfer, J.E., Lopez-Hoffman, L., Semmens, D. & Thogmartin, W.E. (2020). Challenges for leveraging citizen science to support statistically robust monitoring programs. *Biological Conservation*, 242, 108411.

White, E.R. (2019). Minimum time required to detect population trends: The need for long-term monitoring programs. *BioScience*, 69, 40–46.

White, E.R. & Bahlai, C.A. (2020). *Experimenting with the past to improve environmental monitoring programs* (Preprint). EcoEvoRxiv.

Williams, P.J., Hooten, M.B., Womble, J.N., Esslinger, G.G. & Bower, M.R. (2018). Monitoring dynamic spatio-temporal ecological processes optimally. *Ecology*, 99, 524–535.

Wingard, G.L., Bernhardt, C.E. & Wachnicka, A.H. (2017). The Role of Paleoecology in Restoration and Resource ManagementThe Past As a Guide to Future Decision-Making: Review and Example from the Greater Everglades Ecosystem, U.S.A. *Frontiers in Ecology and Evolution*, 5.

Zhang, J. & Zhang, C. (2012). Sampling and sampling strategies for environmental analysis. *International Journal of Environmental Analytical Chemistry*, 92, 466–478.

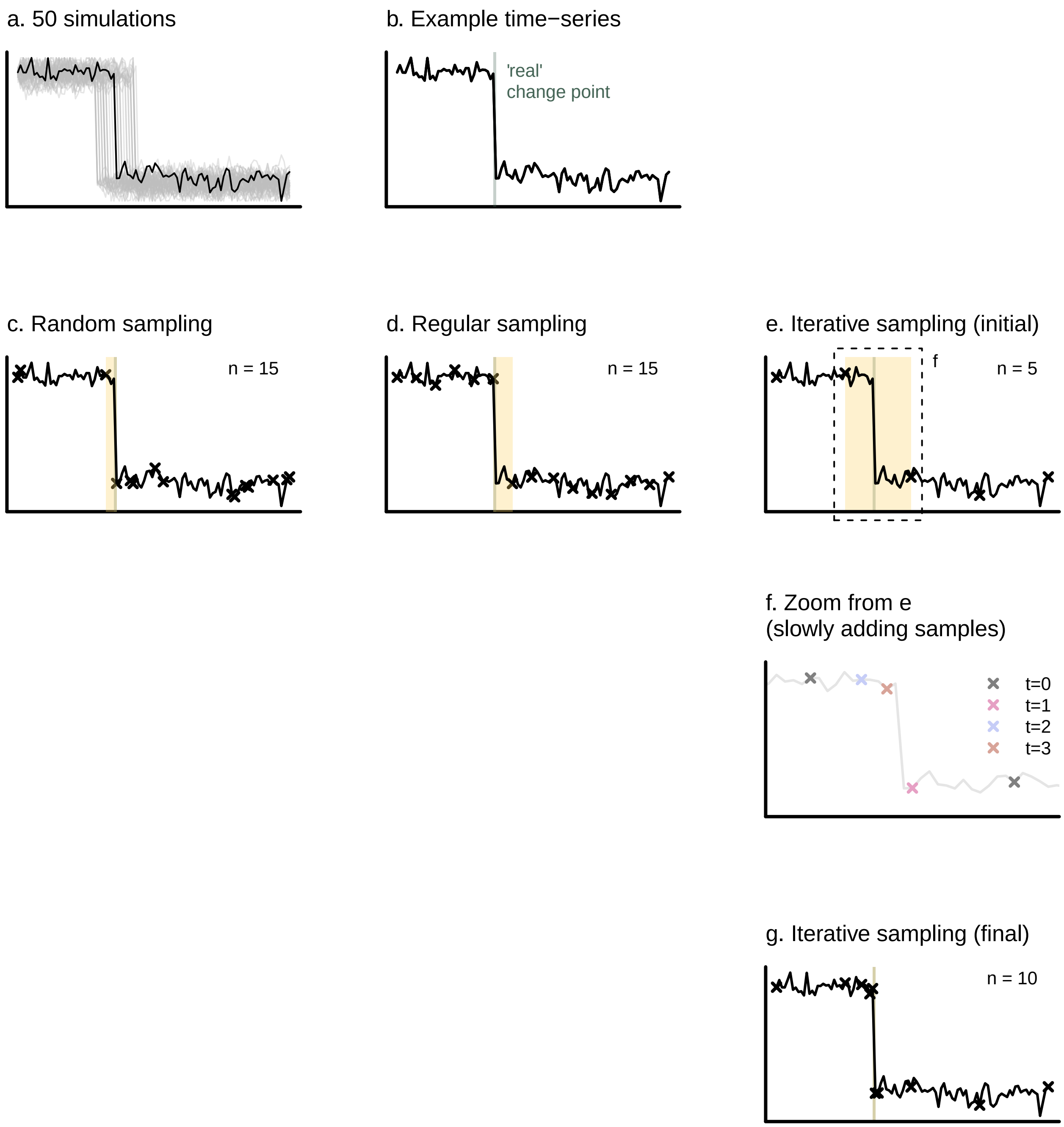

Figure 1: Conceptual diagram illustrating the process of taking (a) simulations of a time series and (b) selecting a single simulation to analyze with three different sampling approaches: (c) random, (d) regular, and (e) iterative. The iterative sampling approach requires (f) adding samples around a detected changepoint until (g) a certain level of accuracy is achieved.

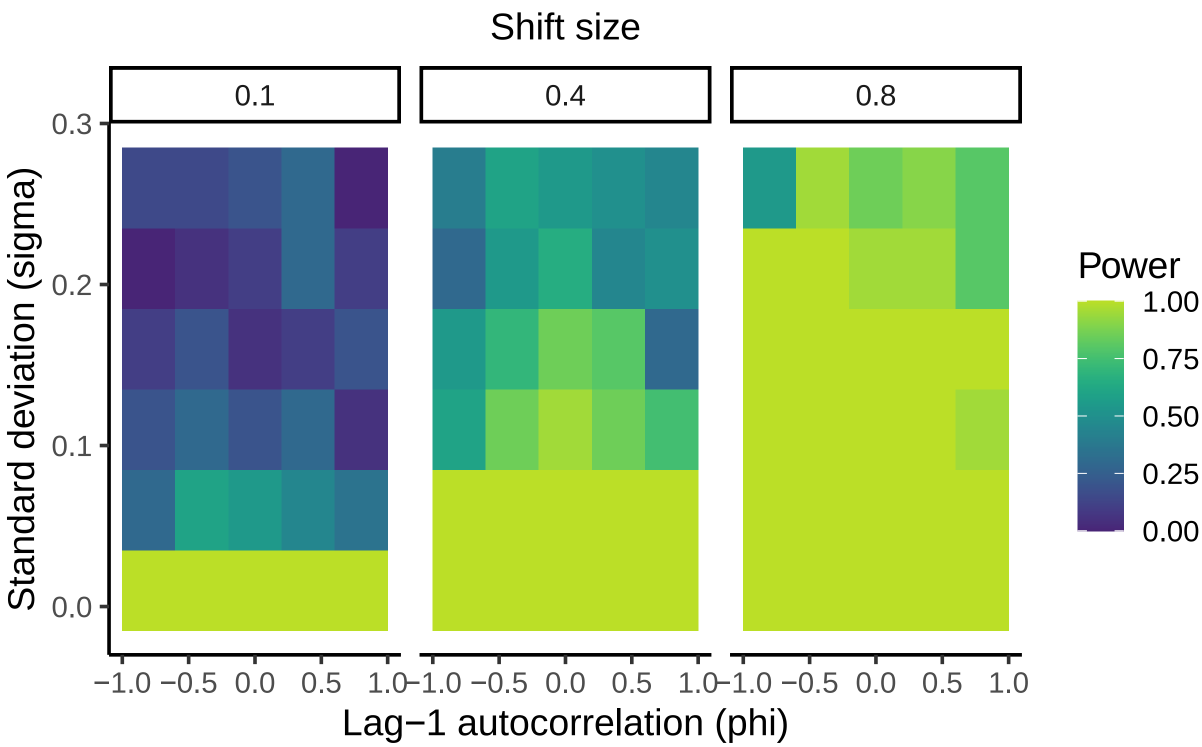

Figure 2: Regular sampling statistical power (fraction of 100 simulations which detected a changepoint within five time points of the true changepoint) for different levels of standard deviation (σ), lag-1 autocorrelation (ϕ), and shift size (δ). For each parameter combination, 20 samples were used. An increase in samples would increase the statistical power across this graph.

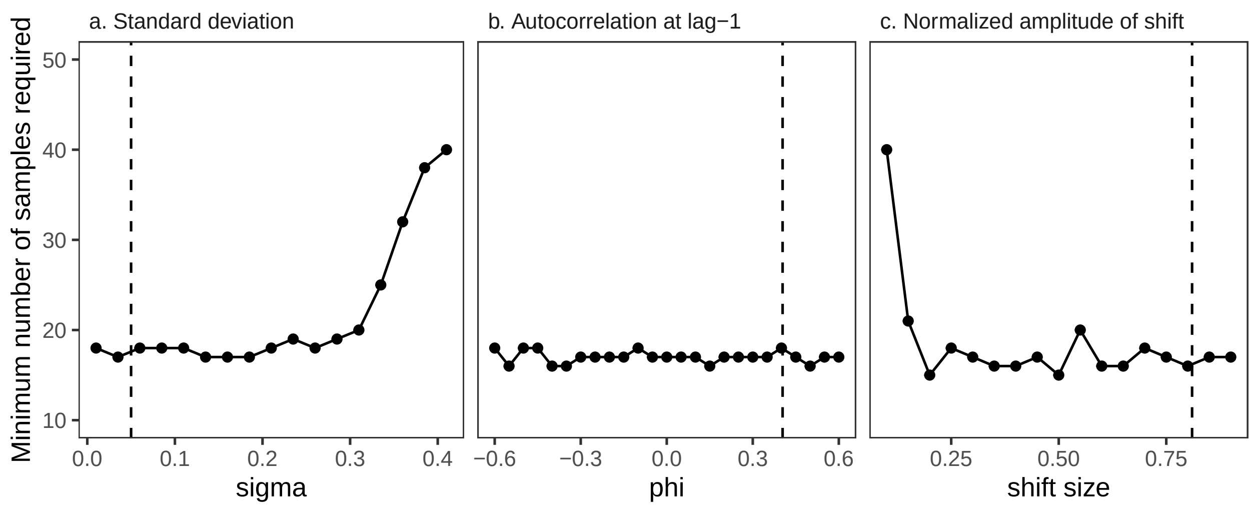

Figure 3: Minimum number of samples required for 0.8 statistical power given different levels of (a) temporal variability (σ), (b) temporal autocorrelation (ϕ), and (c) shift size (δ). Regular sampling was used with the default parameters: $\sigma = 0.53$, $\phi = 0.404$, and a shift size $= 0.81$ to match the case study. The exact timing of the true changepoint varied for each simulation between time steps 30 and 70. Vertical line indicate respective parameter calculated from the case study time series. The same analyses for random and iterative sampling is in Figs. S1,S2.

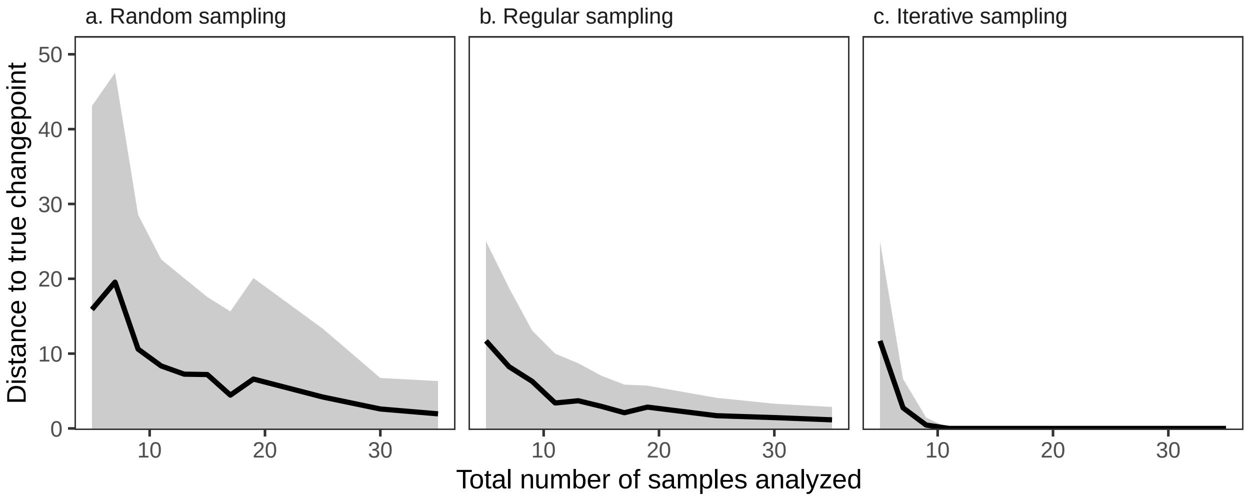

Figure 4: Distance to true changepoint per total number of samples analyzed for simulations, following (a) random sampling, (b) regular sampling, and (c) iterative sampling. Each sample size and approach combination was simulated 50 times with the same parameters as Figure 3. The error bars represent the middle 95% of the simlations.

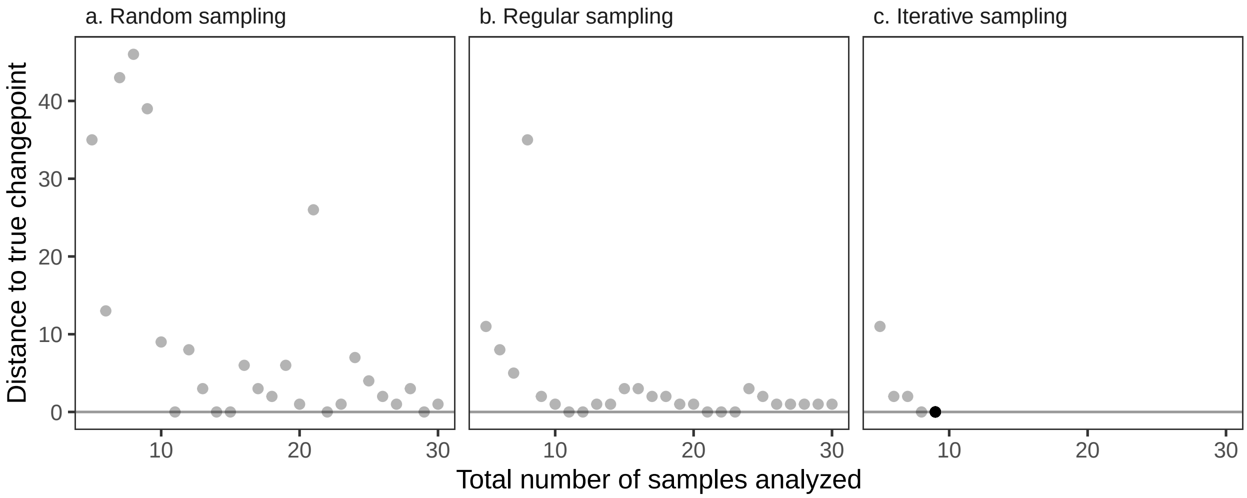

Figure 5: Distance to true changepoint for total number of samples analyzed, following (a) random sampling, (b) regular sampling, and (c) iterative sampling. Total number of samples was set between 5 and 30, out of the 74 initial time series.

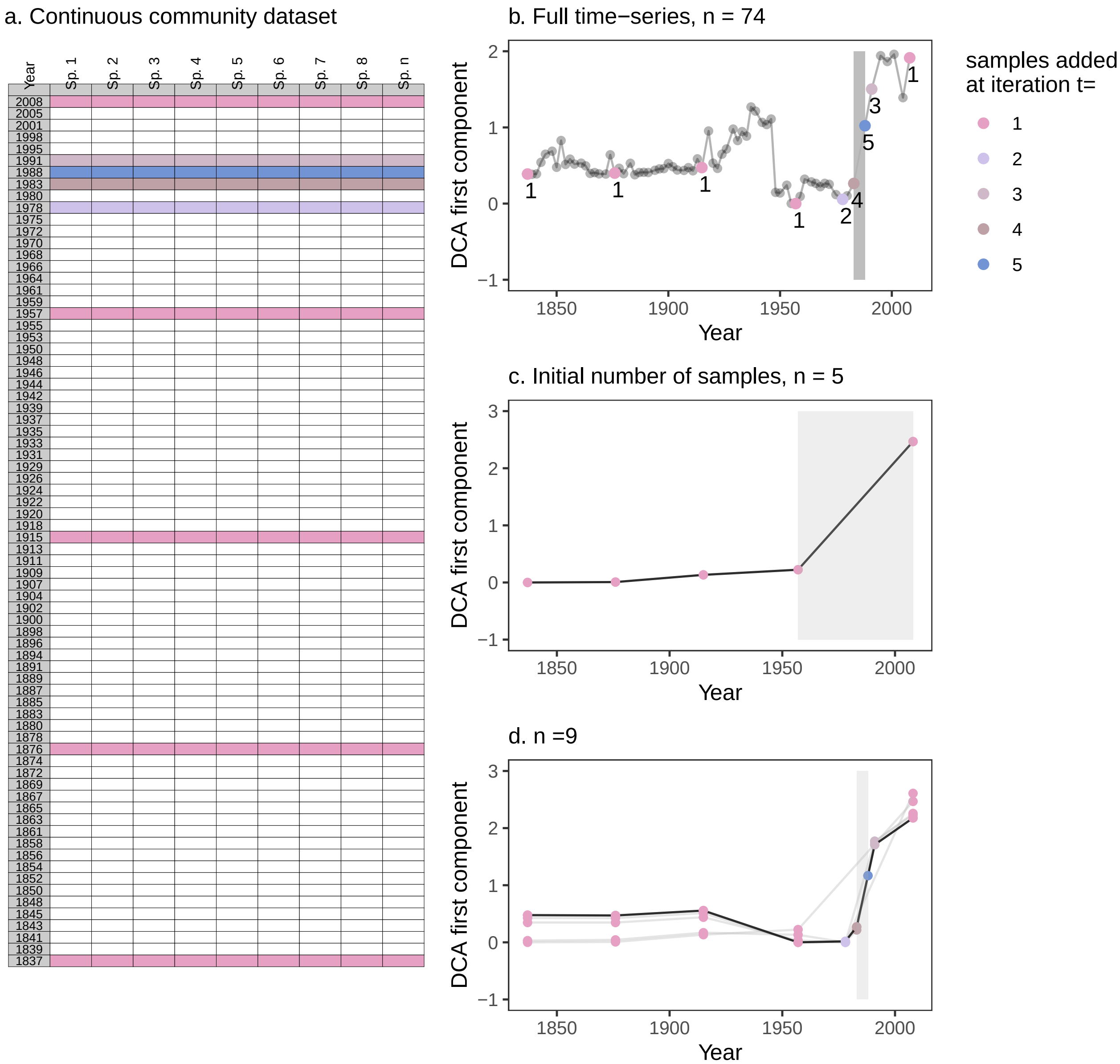

Figure 6: Iterative sampling and impact on ordination. (a) Initial community dataset. Samples are initially sampled at regular interval. The multivariate data is converted to univariate vector by Detrended Component Analysis. Samples were iteratively added, after computation of changepoint at each step. (b) Full time series, on which "true" changepoint was detected. (c) First step of iterative sampling, n = 5. (d) Final step of iterative samples: by adding 4 samples, the true changepoint was detected.