

Can we detect ecosystem critical transitions and signals of changing resilience from paleo-ecological records?

ZOFIA E. TARANU,^{1,†} STEPHEN R. CARPENTER,² VICTOR FROSSARD,³ JEAN-PHILIPPE JENNY,^{3,4} ZOË THOMAS,⁵ JESSE C. VERMAIRE,⁶ AND MARIE-ELODIE PERGA⁷

¹Department of Biology, University of Ottawa, Ottawa, Ontario K1N 6N5 Canada

²Center for Limnology, University of Wisconsin-Madison, Madison, Wisconsin 53706 USA

³UMR CARTEL, Centre Alpin de Recherche sur les Réseaux Trophiques et Ecosystèmes Limniques, Université Savoie Mont Blanc-INRA, Le Bourget-du-Lac 73370 France

⁴Max-Planck-Institute for Biogeochemistry, Jena 07745 Germany

⁵ARC Centre of Excellence in Australian Biodiversity and Heritage, and Palaeontology, Geobiology and Earth Archives Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Kensington, Sydney, New South Wales 2052 Australia

⁶Department of Geography and Environmental Studies, Institute of Environmental Science, Carleton University, Ottawa, Ontario K1S 5B6 Canada

⁷Institute of Earth Surface Dynamics, University of Lausanne, Lausanne 1015 Switzerland

Citation: Taranu, Z. E., S. R. Carpenter, V. Frossard, J.-P. Jenny, Z. Thomas, J. C. Vermaire, and M.-E. Perga. 2018. Can we detect ecosystem critical transitions and signals of changing resilience from paleo-ecological records? *Ecosphere* 9(10): e02438. 10.1002/ecs2.2438

Abstract. Nonlinear responses to changing external pressures are increasingly studied in real-world ecosystems. However, as many of the changes observed by ecologists extend beyond the monitoring record, the occurrence of critical transitions, where the system is pushed from one equilibrium state to another, remains difficult to detect. Paleo-ecological records thus represent a unique opportunity to expand our temporal perspective to consider regime shifts and critical transitions, and whether such events are the exception rather than the rule. Yet, sediment core records can be affected by their own biases, such as sediment mixing or compression, with unknown consequences for the statistics commonly used to assess regime shifts, resilience, or critical transitions. To address this shortcoming, we developed a protocol to simulate paleolimnological records undergoing regime shifts or critical transitions to alternate states and tested, using both simulated and real core records, how mixing and compression affected our ability to detect past abrupt shifts. The smoothing that is built into paleolimnological data sets apparently interfered with the signal of rolling window indicators, especially autocorrelation. We thus turned to time-varying autoregressions (online dynamic linear models, DLMs; and time-varying autoregressive state-space models, TVARSS) to evaluate the possibility of detecting regime shifts and critical transitions in simulated and real core records. For the real cores, we examined both varved (annually laminated sediments) and non-varved cores, as the former have limited mixing issues. Our results show that state-space models can be used to detect regime shifts and critical transitions in some paleolimnological data, especially when the signal-to-noise ratio is strong. However, if the records are noisy, the online DLM and TVARSS have limitations for detecting critical transitions in sediment records.

Key words: aquatic transitions; eutrophication; online dynamic linear model; paleolimnology; regime shifts; resilience indicators; time-varying autoregressions; time-varying autoregressive state-space.

Received 26 April 2018; **revised** 6 August 2018; **accepted** 17 August 2018. Corresponding Editor: Debra P. C. Peters.

Copyright: © 2018 The Authors. 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.

† E-mail: zofia.taranu@gmail.com

INTRODUCTION

The observation that ecosystems can respond discontinuously to changing external pressures has shed some light on their complex nonlinear dynamics. The concepts of critical transitions, thresholds, and alternative stable states spread throughout the ecological and environmental management literature, from the initial proposal by Holling (1973) to field evidence for ecosystem regime shifts (Scheffer et al. 2001, Carpenter 2003). Most recently, however, reviews have raised some skepticism about whether nonlinear ecosystem responses to an environmental press are the exception rather than the rule (Capon et al. 2015). The use of resilience indicators as tools for detecting the possible onset of critical transitions has also been questioned (Kéfi et al. 2013, Burthe et al. 2015, Gsell et al. 2016).

Although a regime shift, defined as a large change with prolonged consequences, is often obvious when it occurs, whether it qualifies as a catastrophic transition that pushes the system from one equilibrium state to another is usually difficult to prove (Carpenter 2003). A critical transition occurs when the system operates a bifurcation leading to loss or gain of a new attractor (Scheffer 2009). Most methods used to detect critical transitions depend on demonstrating that the dominant eigenvalue (i.e., recovery rate) of the variance–covariance matrix of a plausible dynamical model crosses from inside to outside the unit circle (in discrete time) or passes through zero from below (in continuous time). That is, mathematically, a critical transition is defined as an unstable equilibrium point where the rate with which the system returns to this equilibrium (dominant eigenvalue) approaches zero, with the consequence that the system becomes increasingly slow in recovering from small perturbations (loss in resilience) and thus vulnerable to major changes caused by small perturbations. When approaching a critical transition, the system's intrinsic rates will differ less from previous time points (increase memory or autocorrelation; Ives et al. 2003) while its behavior becomes more like a random walk (rise in variance; Scheffer et al. 2015).

In general, a deeper understanding of the causes of regime shifts and whether these are critical transitions requires high-resolution and

lengthy times series; often monitoring records do not meet these requirements. Paleo-ecological records thus represent a valuable opportunity to extend our temporal perspective and enhance our ability to detect critical transitions retrospectively (Dakos et al. 2008, Lenton et al. 2012, Spanbauer et al. 2014, Belle et al. 2017, Beck et al. 2018). However, to the exception of varved sediments which provide annual integration, sediment core records have their own biases and shortcomings, such as sediment focusing and irregular temporal integration (Carstensen et al. 2013, Frossard et al. 2015). As such, sediment core time series are not strictly analogous to instrumental data sets, and there is a growing need for the development of quantitative methods to identify regime shifts and critical transitions in paleolimnological records (Spears et al. 2016).

In this work, we examined three sediment core archives displaying either a gradual change in trophic state, an abrupt state change, or plausible flickering between two alternate states, to (1) show that regime shifts are discernible in real sediment cores and (2) evaluate whether we could use these cores to test for evidence of critical transitions between alternate states. To address the latter, we built a model of core formation (simulated core records) to help define the conditions under which resilience indicators of critical transitions can be detected with paleolimnological data. Our simulations showed that factors encountered in core records (integration of core intervals, sediment compaction with time or uneven degradation) weaken our ability to detect past critical transitions using generic statistics (i.e., rise in rolling window variance and lag-1 autocorrelation). In particular, mixing among core intervals increased the autocorrelation at lag-1, whereas greater compaction of sediment layers with time eroded the temporal autocorrelation signal. This suggests that the ability to detect transition using generic resilience indicators will depend on the core sampling resolution (e.g., different results are expected if using high-resolution varved cores derived from micro-XRF scanning vs. low-resolution acquisition from extruding cores at 1-cm intervals). In light of the shortfall of the rolling window approach in paleo-records, we employed time-varying autoregressions (Ives and Dakos 2012) computed using dynamic linear models (DLMs,

Pole et al. 1994) and time-varying autoregressive state-space models (TVARSS; Ives and Dakos 2012) to evaluate the possibility of detecting critical transitions in simulated and real sediment core records. Our results showed that regime shifts and indicators of past critical transitions in core records can be detected in paleolimnological data if noise is moderate compared to the magnitude of change.

METHODS

Observed lake sediment records

To provide a range of regime shift time series, we cross-examined the published literature and identified core records with (1) an abrupt onset of a eutrophic state (Roxton Pond; echinenone pigment used as a proxy of total cyanobacteria abundance; Vermaire et al. 2017); (2) a noisier eutrophication signal (Lake Anarry; myxoxanthophyll pigment used as a proxy of colonial cyanobacteria abundance; Stevenson et al. 2016); and (3) possible flickering between alternate states (Lake Geneva; a composite proxy reconstructed from multiple intercorrelated varved cores tracking the inferred volume of hypolimnetic hypoxia; Jenny et al. 2014). Cyanobacterial dominance and hypolimnetic hypoxia are well-established symptoms of anthropogenic disturbances and increased nutrient loading to lakes (Edmondson 1961, Reckhow 1977); thus, both paleolimnological proxies are used here as indicators of a shift from oligotrophic to eutrophic conditions (Taranu et al. 2015, Jenny et al. 2016). All proxies were standardized (z-scores) prior to analysis, though we refrain from comparing the magnitude of change among these three example cores as we used three very different lake types (i.e., shallow vs. very deep) which we would expect to behave and respond in different ways to the same forcing.

Roxton Pond is a eutrophic, shallow lake (maximum depth 6.2 m) in the Eastern Townships of Quebec, Canada, that experienced a pronounced increase in nutrient loading during the 1970s as a result of agricultural activities (Vermaire et al. 2017). According to the paleolimnological record, this was followed by a change in the zooplankton and phytoplankton community and an increase in cyanobacterial pigment concentrations in the 1990s. Lake Anarry is an upland, shallow lake

(maximum depth 7.8 m) situated in northwest Ireland, whose catchment was afforested in the 1960s (Stevenson et al. 2016). Forest plantations accounted for >50% of the lake's catchment area. Increases in cyanobacteria abundance reflect a combination of mineral and nutrient enrichment associated with the forest fertilization and organic matter influx. In particular, there was a twofold increase in cyanobacterial pigment concentrations following the 1963 forest planting. Lake Geneva is a subalpine, deep lake (maximum depth 310 m) at the Swiss-France border, in the foothills of the western European Alps that shifted from oligotrophy to eutrophy in the 1960–1980s and then restored gradually. Geochemical and visual analyses of this high-resolution archive (temporal resolution on the order of ± 2 yr) indicate that bottom-water hypoxia appeared in the 1950s with further changes in the volume of hypoxic waters (Jenny et al. 2014).

Simulating the eutrophication process

To simulate sediment core archives (see Fig. 1 for flowchart of our protocol), we first generated a time series of lake dynamics using a one-dimensional version of the phosphorus load and lake eutrophication model presented by Carpenter and Brock (2006). The dynamics of phosphorus in the lake water follow

$$\begin{aligned} dx &= [L_t - sx_t - hx_t + f(x_t)]dt + \sigma x_t dW \\ f(x) &= \frac{rMx^q}{m^q + x^q} \end{aligned} \quad (1)$$

In Eq. 1, x is phosphorus in the lake water (mass/area), L is the loading rate (mass/(area \times time)), s and h are the sedimentation and hydraulic washout coefficients, respectively (1/time), $f(x)$ is the recycling function (mass/(area \times time)), r is a recycling coefficient (dimensioned so that $f(x)$ has units of mass/(area \times time)), M is mass/area of phosphorus in sediments vulnerable to recycling, m is the half-maximum coefficient for phosphorus recycling (mass/area), and q is an exponent that determines the rate of recycling at the half-maximum point (dimensionless). The standard deviation of environmental noise is σ (i.e., coefficient of variation of environmental noise multiplied by the mean to obtain a standard deviation), and dW is a Wiener stochastic process. Using the Euler-Maruyama method (Higham

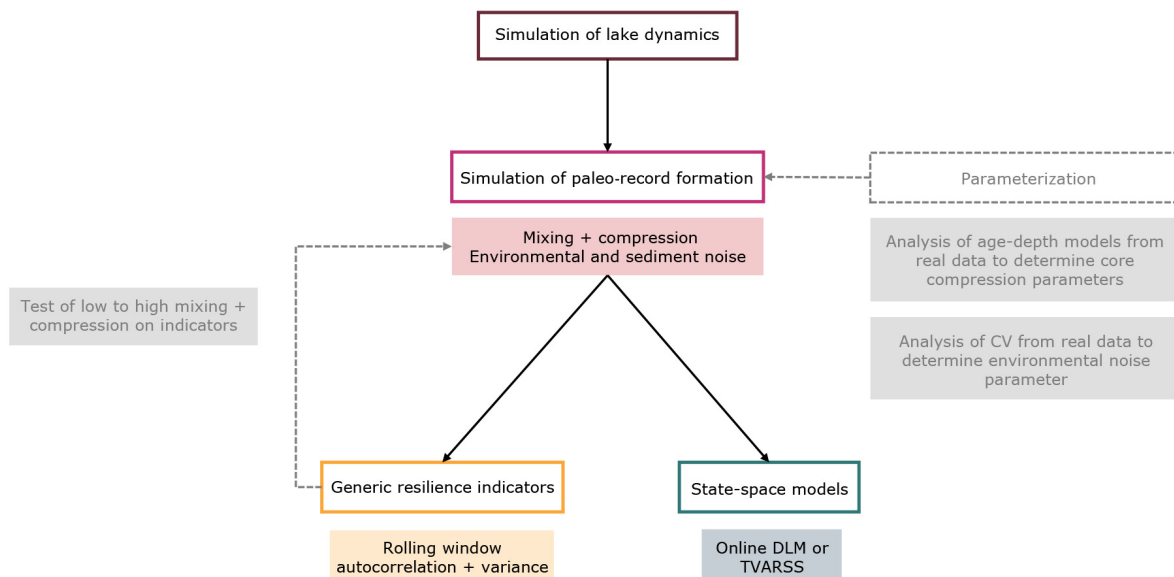


Fig. 1. Flowchart of simulations and analyses conducted in this study.

2001), we numerically solved Eq. 1 over short finite time increments dt as

$$x_{t+1} = x_t + [L_t - sx_t - hx_t + f(x_t)]dt + \sigma x_t z_t \sqrt{dt} \quad (2)$$

In Eq. 2, z_t is a series of independent normally distributed random numbers with mean 0 and unit standard deviation. Simulations presented here used the following parameter values: $s = 0.71$, $h = 0.29$, $r = 0.019$, $m = 4$, $q = 4$. These values are within the range of uncertainty for Lake Mendota, Wisconsin, USA (Carpenter and Lathrop 2008). Values of σ and M were set appropriately for the numerical experiments described below. Time series of load L_t were gradually increased through each model run to simulate eutrophication. We used a time step of $dt = 0.1$ yr.

Modifying simulations to mimic sediment core records

The next step of our protocol consisted in modifying the ecosystem simulations to more accurately reflect sediment core archives. Since Eq. 2 simulates ecosystem dynamics but fails to account for additional sources of variance that occur in the sediment record (e.g., variance in sedimentation, decay, and preservation processes), we included an additional source of variance Eq. 3 that affects sediment dynamics but does not affect ecosystem dynamics:

$$x_{s,t} = x_t + v_t z_{s,t} \quad (3)$$

The concentration of tracer that is included in the simulated core, $x_{s,t}$, is equal to the sum of x_t plus a standard deviation for core formation, v_t , times a series of independent normally distributed random numbers with mean 0 and unit standard deviation, $z_{s,t}$. The series z_t (Eq. 2) and $z_{s,t}$ are also uncorrelated.

Using Eqs. 2, 3 with selected σ , v , and M values (see *Numerical experiments* section), we generated high-frequency time series of deposition (10,000 time steps of length 0.1, corresponding to 1000 yr) over a range of loading rates, L , from very low phosphorus to highly eutrophic phosphorus levels. Since we were interested in the concentration of a tracer per unit mass of this deposited material for the pigment stratigraphies, we assumed that each sediment layer contained the annually deposited mass of total organic plus inorganic material per unit of lake bottom (i.e., that the sediment marker of eutrophication was directly proportional to phosphorus density [mass/area] in the lake). The same assumption would apply to other paleolimnological proxies (e.g., isotope of an element, frustules of a diatom species, and carapaces of a zooplankton species). For the varved stratigraphies, we assumed that the increase in the yearly volume

of hypolimnetic hypoxia, reconstructed from the multiple cores, was directly related to the eutrophication process

To represent a 400-yr paleo-archive of a lake that was eutrophied ~100 yr ago, we used the Self Exciting Threshold Autoregressive model (SETAR; package *tsDyn* in R; Stigler 2010, Di Narzo, Aznarte and Stigler 2014) to estimate the time point when the system shifted to a eutrophic regime in the 1000-yr time series and selected a window of 300 yr prior to and 100 yr post this SETAR estimate of the switchpoint. All SETARs were modeled with only one threshold, which was estimated by a search over a reasonable grid.

The last step in converting the simulated time series to realistic sediment core archives was to include temporal averaging (annual resolution is rarely available in paleolimnology), changes in temporal resolution (resulting from changes in sedimentation rates due to eutrophication and core compression), and sediment mixing (due to bioturbation). To do so, we converted the 400-yr time series to archives of consecutive 1-cm slices (determined from an age–depth model) by mixing (average of localized time steps) and compressing (making the averaging window increasingly large as we move downcore) the time series.

Core mixing

We assumed that mixing occurs among proximate layers of the sediment core over a window of length -2υ to $+2\upsilon$ centered at the midpoint of a given layer t , where υ is the standard deviation of mixing in units of annually deposited layers. We then computed a vector of weights as areas of the unit normal distribution centered on t for a sequence of small intervals from $t - 2\upsilon$ to $t + 2\upsilon$. The weights were relativized so that they sum to one to deal with rounding error and the fact that the normal integral between $\pm 2\upsilon$ is <1 . The weights were then multiplied by the tracer concentrations in the annually deposited layers around the target layer. This step was then repeated for all possible target layers t . To assess the effect of mixed vs. unmixed cores over a wide range of kernels, we tested scenarios of no mixing (which represented a varved core), medium mixing (normal kernel with a standard deviation of $\upsilon = 2.5$), and high mixing (normal kernel with a standard deviation of $\upsilon = 5.5$).

Empirical model for core compression

To model core compression, we cross-examined cores from Taranu et al. (2015) and related the years per centimeter of core (y) to the age at the midpoint of a core section between two age measurements (x). In most cases, y and x were related approximately linearly (i.e., $y = a_0 + a_1x$) by the coefficients a_0 and a_1 . We fit the inverse of this relationship (which was necessary for modeling core compression) by a nonlinear regression, providing us centimeters per year of core ($1/y$) as a function of age:

$$\frac{1}{y} = \frac{1}{a_0 + a_1x} \quad (4)$$

Comparison of the linear and nonlinear fits showed that estimates of a_0 and a_1 were comparable between the two methods.

Fitting the nonlinear regression (Eq. 4) to the suite of real core archives indicated that sedimentation rates were quite variable among lakes and regions (Appendix S1: Fig. S1a, b). Although this variability was related to lake eutrophication (i.e., sedimentation rates increased with concentrations of total phosphorus in the water column), over longer time scales, the contribution of terrigenous supplies to the sedimentation rates may be equal to or higher than eutrophication. We therefore used this data set to define parameter estimates of low ($a_0 = 1.28$, $a_1 = 0.015$), intermediate ($a_0 = 2.24$, $a_1 = 0.034$), and high ($a_0 = 4.35$, $a_1 = 0.088$) compression in natural lakes within the context of paleo-studies conducted over the past 200 yr. In the case of no compression, the a_1 slope was 0 (i.e., age and depth are linearly related, and the plot of y vs. x has no slope that is statistically significant).

Once we obtained a realistic range of a_0 and a_1 values, the next step was to use these estimates to parametrize simulated cores with low, intermediate, and high compression. That is using the vector of annual layers (indexed 0 to T , either with or without mixing applied), we created a core of 1 cm thick slices (indexed 1 to S , where S is the length of the core) by integrating Eq. 4 over time (from $T = 0$ –400 yr). Although the thickness at which cores are extruded varies among paleolimnological studies, and many do sample sequences at <1 cm resolution, for the present exercise we assumed that only complete 1-cm

slices were analyzed and set the integral of Eq. 4 from t_0 to t_1 to be equal to 1 cm throughout the core. Solving the integral of Eq. 4 from the time at the top of the core, $t_0 = 0$, to the time at the bottom of the first slice, t_1 , with the left side of the equation equal to 1, gives:

$$\begin{aligned}
 1 &= \frac{1}{a_1} [\ln(a_1 t_1 + a_0) - \ln(a_1 t_0 + a_0)] \\
 &= \frac{1}{a_1} \ln \left[\frac{a_1 t_1 + a_0}{a_1 t_0 + a_0} \right] \\
 e^{a_1} &= \frac{a_1 t_1 + a_0}{a_1 t_0 + a_0} \\
 t_1 &= \frac{1}{a_1} [(a_1 t_0 + a_0)e^{a_1} - a_0]
 \end{aligned} \tag{5}$$

The compression protocol thus starts with the top of the core ($t_0 = 0$) and solves for the time at the bottom of the first 1-cm slice, t_1 . The tracer concentrations are then averaged over the annual layers between t_0 and t_1 to get the tracer concentration in the first slice. Any fractions of annual layers are weighted by the fraction in computing the average tracer concentration. Then, t_1 from the first slice becomes t_0 for the second slice. The t_1 for the second slice is then computed using the above equation, and the average tracer concentration is computed in the second slice. These steps are repeated for deeper slices until the core is completed.

Numerical experiments

To test under which condition we could expect to detect resilience indicators in sediment archives, we simulated series with different combinations of environmental and sediment noise (σ and v) and sediment P mass values (M ; i.e., series with or without deterministic regime shifts and critical transitions) and tested how well we could retrieve the known signals.

Reasonable estimates of environmental noise, σ , for real eutrophication indicators were determined by cross-examining sedimentary cyanobacterial pigment profiles synthesized in Taranu et al. (2015), which provided a range of coefficients of variation (CV) in oligotrophic to hyper-eutrophic lakes. Lakes where the concentration of cyanobacterial pigments remained at 0 throughout the time series (CV = 0) were removed. The baseline CV was based on time series that did not undergo a substantial change

in pigment concentrations over the last ~200 yr. In general, the CV in these cores was well below 0.5 (Appendix S1: Fig. S2a). Similarly, the range of sedimentation noise, v , was determined by comparing our simulated series to the same set of pigment profiles.

To determine the critical P mass, we explored a suite of different M values (Appendix S1: Fig. S2b). For the values of s , h , r , m , and q stated above, the simulation with $M = 50$ had no regime shift and no critical transition, though the system became more eutrophic. An increase to $M = 200$ produced a regime shift that was non-critical (i.e., the lake suddenly became eutrophic, but the shift did not imply any bifurcation and thereby hysteresis). Finally, transitions were critical for $M > 210.5$.

Based on the above, we simulated five types of sediment core records: (1) a control with low noise and a non-critical eutrophication signal ($M = 125$, $\sigma = 0.05$, and $v = 0$) for which the regime shift is not reached over the range of the time series; (2) a record with a regime shift that is not a critical transition ($M = 200$, $\sigma = 0.05$, and $v = 0$); (3) a record with a regime shift that is a critical transition ($M = 225$, $\sigma = 0.15$, and $v = 0.15$) which represents a critical slowing down scenario for which resilience indicators are expected to be clear and occur in a rather narrow bifurcation window; (4) a core record with large ecosystem noise ($\sigma = 0.45$, keeping $M = 225$ and $v = 0.15$) that flickered between two alternate states before settling into the alternate, eutrophic state at the bifurcation point; and (5) a record with both large environmental ($\sigma = 0.45$) and sedimentation ($v = 0.25$) noise (keeping M equal to 225).

For the non-critical simulations (scenarios 1 and 2), we expected a gradual increase in the P mass with eutrophic conditions. For the critical slowing down simulation (scenario 3), a transition was expected midway through the series, where the time series would show a shift in the mean as well as a rise in variance and autocorrelation. With flickering (scenarios 4 and 5), we expected the ecosystem to reach the eutrophic state for brief periods of time before P loading reached a level where the system remained in the eutrophic state. Thus, to obtain an index of the time of regime shift in systems that flickered, we used SETAR to estimate the first time when the ecosystem spent a significant amount of time in the eutrophic state.

The cumulative number of crossings of the SETAR threshold was calculated and plotted against time. To construct cumulative plots, we subtracted the state value at the SETAR threshold from the data, creating positive or negative values in the time series depending on whether the state is above or below the threshold. Positive values were assigned a value of 1, and negative values were assigned a value of 0. The scores were then tallied cumulatively for each year. If the cumulative sum followed a sigmoidal shape, the steepest part of the curve should represent the time window of frequent flickering between oligotrophic and eutrophic states. In addition, we plotted the number of times that the sign (difference between value and SETAR threshold) changed between time t and time $t + 1$. Specifically, a 0 is assigned (i.e., remained in same state) if the sign is the same, whereas if the sign changes (i.e., change in state), a value of 1 is assigned. It is expected that, at the start and end of the series, more 0s will be assigned (same state), whereas during flickers more 1s will be assigned (switching back and forth). We compared both cumulative plots to the deterministic critical point to evaluate SETAR as an indicator of the time of transition to eutrophy.

Rolling window resilience indicators in paleo-records

Critical transitions in monitoring records have typically been quantified using generic resilience indicators (e.g., rolling window variance and lag-1 autocorrelation); however, with core records factors such as sediment mixing and compaction with time may hinder the use of such statistics. Using the above scenarios, we thus tested whether mixing or compaction resulted in false-positive signals of critical transition and/or eroded the signals of true-positive transitions. For the former, if core compression increases with depth, the number of years per centimeter will increase toward the bottom of the core. This effect could mask the lower deposition rates in the older, more oligotrophic phases of the core (i.e., lower nutrient concentrations during oligotrophic phase would result in relatively reduced sediment accumulation). In contrast, the years per cm would decrease in more recent eutrophic years (greater primary production, decomposition, and sedimentation) resulting in much less averaging to dampen the variance. High

compression alone may thus produce a gradual increase in variance toward the top of the core and a false-positive detection of a critical transition in recent years.

Test for false positive

Making use of simulations with no critical transition (scenarios 1 and 2), we examined the effect of mixing and compression alone on the rolling window variance statistics (using the *rollapply* function from the *zoo* package in R; Zeileis and Grothendieck 2005, with a window width that was half the length of the time series, and a window that moved by one time step; Dakos et al. 2008). We crossed no mixing (mimicking a varved core), medium mixing, and high mixing with low, medium, and high core compression and, for each combination, tested whether the rolling window variance and autocorrelation increased over time.

Test for true positive

Conversely, to determine how mixing and compression jointly affected the signals of a critical transition (scenarios 3–5), we crossed no mixing to high mixing with low to high compression and examined their effect on the rolling window variance and temporal autocorrelation prior to the deterministic switchpoints.

State-space modeling of paleo-records

Simulation studies show that the rolling window method often works, despite its obvious shortcomings: It cannot account for higher order AR(p) processes, the choice of window length is arbitrary, and it may not account for rapid changes in stability close to critical points. To provide an alternate approach in light of these factors, Ives and Dakos (2012) introduced the idea of time-varying autoregressions to evaluate changes in stability of observed time series.

By definition, a critical transition in discrete time occurs if the dominant eigenvalue of the Jacobian matrix passes from inside to outside the unit circle on the complex plane (i.e., the modulus passes above 1 from below; Biggs et al. 2009, Scheffer et al. 2015). When working with observed time series, it is common to approximate the eigenvalue by fitting a stationary AR(1) model to sequential rolling windows of observations as described above (Dakos et al. 2012). In that case, an AR(1)

coefficient near zero implies that the state variable returns rapidly toward the mean, but as the AR(1) coefficient reaches 1, this tendency to return toward the mean disappears and we thus have evidence of a critical transition (Ives et al. 2003).

In the case of our simulated time series, we started with a eutrophication model for which the rolling window AR(1) method is known to be a reliable indicator of the critical transition. However, sediment mixing and compression within paleo-records causes complications that could cause the linear rolling window AR(1) approximation to fail. We thus turned to the time-varying autoregressions to fit higher order and time-varying parameters of AR(p) (Ives and Dakos 2012) to allow greater complexity in approximating the underlying processes that create paleo-records.

We fitted time-varying autoregressions in the form of DLMs (Pole et al. 1994). The AR(p) model in DLM form is as follows:

$$\begin{aligned} y_t &= F'_{t-1} \phi_{t-1} + \varepsilon_t \\ \phi_t &= I \phi_{t-1} + \omega_t \end{aligned} \quad (6)$$

The first equation of (6) relates the observation at time t (y_t) to previous observations F_t and current values of the coefficients ϕ_t (intercept and autoregressive coefficients) and error ε_t (assumed to be normally distributed with mean zero and variance R). For an AR(p) process, F_t and ϕ_t are defined as follows:

$$\begin{aligned} F_t &= [1 y_{t-1} \dots y_{t-p}]' \\ \phi_t &= [\phi_{0,t} \phi_{1,t} \dots \phi_{p,t}]' \end{aligned} \quad (7)$$

The time-varying intercept $\phi_{0,t}$ tracks the baseline of the time series, and the other coefficients estimate the autoregressive process. The second equation of (6) allows the autoregression coefficients to evolve through a random walk with shocks ω_t (distributed as a multivariate normal with means 0 and covariance matrix Q). These shocks ω_t are independent of the autoregression process errors ε_t .

We used an online method to estimate the coefficients ϕ_t , error variance R , and covariances Q . The method uses sequential Bayesian updating to compute estimates of the parameters at each time step (Pole et al. 1994). The user specifies a discount parameter δ ($0 < \delta < 1$) used to adjust the variability of the random walk. The weight of past data points is multiplied by δ at each time step.

Thus, the influence of past data declines over time as δ^k where k is the number of time steps between two observations. As the total number of time steps approaches infinity, the total weight of all data points contributing to the estimates at any particular time step approaches $1/(1 - \delta)$. Lower values of δ thus assign lower weight to past data points and thereby allow more rapid adaptation of the DLM to shifts in the data.

The adaptability of the online method to non-stationary data is advantageous, but the arbitrary choice of δ is a disadvantage of the online method. In addition, the DLM does not account for errors in the observations (elements of the vector F). Therefore, we also estimated the AR parameters in time-varying autoregressive state-space form by maximum likelihood estimation using TVARSS (Ives and Dakos 2012).

The TVARSS method estimates the time-varying autoregression coefficients by fitting the model

$$\begin{aligned} y_t &= F'_{t-1} \phi_{t-1} + \varepsilon_t \\ \phi_t &= I \phi_{t-1} + \omega_t \\ x_t &= y_t + v_t \end{aligned} \quad (8)$$

by maximum likelihood. The first two equations of (8) are analogous to equations (6) for the DLM. The third equation models the errors in the observations, assumed to be normally distributed with individual errors v_t having mean 0 and standard deviation P . Thus, TVARSS accounts for the dynamics of the estimated system, changes in the autoregression parameters, and observation errors in the data. The advantages and disadvantages of the online DLM and TVARSS methods are complementary in some ways, and therefore, we used both methods.

For both the online DLM and TVARSS, we chose the optimal AR order p using Akaike's information criterion (AIC). We calculated eigenvalues from the AR parameters according to Box et al. (2008). R functions for the online DLM, example data sets, and R programs demonstrating calculations are downloadable from https://github.com/SRCarpen/Demo_DynamicLinearModels.

RESULTS

Core simulations

The simulations showed that annualizing alone (averaging 10,000 step time series by blocks of 10

to produce a 1000-yr time series) decreased the variability in the simulated time series (Appendix S1: Fig. S3a; example shown for critical slowing down, scenario 3). Once the series were annualized, SETAR was used to estimate the switchpoint between the two states. For the selected simulations with low to moderate environmental noise (scenarios 1–3), this occurred at time steps 782, 802, and 635, respectively. The records were thus clipped to 400 annual data points (300 time steps before the SETAR estimate of the switchpoint, one time step during the switch, and 99 time steps after the switch; e.g., Appendix S1: Fig. S3b). When a critical transition occurred (critical slowing down, scenario 3), our simulations showed that annualizing the record did not bias the resilience indicators and a discernible rise in rolling window variance was detected near the switchpoint (Appendix S1: Fig. S3c). The autocorrelation signal was also discernible close to the SETAR switchpoint (Appendix S1: Fig. S3d). Nonetheless, both indicators peaked after the SETAR switchpoint.

For the flickering simulations with either high environmental noise (scenario 4: $M = 225$, $\sigma = 0.45$, $v = 0.15$) or both high environmental and sediment noise (scenario 5: $M = 225$, $\sigma = 0.45$, $v = 0.2$), we expected the first state switch to occur before the threshold had been reached (Appendix S1: Fig. S4a; example shown for flickering with high environmental noise, scenario 4). Indeed, the first SETAR switchpoints of the annualized time series occurred well before the deterministic switchpoints (Appendix S1: Fig. S4b). If the ecosystem reaches the alternate stable state (e.g., eutrophic at all times), then the center of the distribution of critical transition points should occur between the first flicker and the time of staying in the alternate state. The cumulative plots suggested that the peak of the distribution of critical thresholds is close to the time step when approximately half of the maximum number of crossings has occurred (Appendix S1: Fig. S4c, d). Note that detection of this pattern requires a sufficiently large sample size. As with the low to moderate environmental noise simulations (scenarios 1–3), we clipped the flickering records to 400 annual time points and rescaled the x -axes so that the critical switchpoint was at year 300.

Once the 400-yr core series were obtained for all five scenarios, the last step in our simulation

protocol was to apply intermediate mixing and compression to all simulated series to mimic sediment cores extruded at 1-cm intervals (Fig. 2a; example shown for critical slowing, scenario 3).

Rolling window resilience indicators

To examine whether generic resilience statistics could be applied to paleolimnological records where mixing and/or core compression occur, we tested for false- vs. true-positive detection of critical transitions using the simulated scenarios described above.

Test for false positive

To evaluate whether core compression alone caused an increase in variance during the more recent, eutrophic years (less dampening when sediment accumulation is high), we tested for false-positive transition indicators using the control scenario with low to high compression combined with medium mixing ($v = 2.5$; Appendix S1: Fig. S5a). Our simulations showed that core compression did not amplify the variance signal (Appendix S1: Fig. S5b, c). Instead, the rolling window variance decreased slightly toward the eutrophic years. The false-positive test for the slightly higher P load scenario with a possible regime shift but no criticality (scenario 2) was comparable to the control scenario (not shown).

Test for true positive

To test the effect of core mixing and compression on the detection of a true deterministic critical transition, we examined the resilience indicators for the critical slowing down scenario with low compression combined with either no mixing (varved core), medium mixing ($v = 2.5$) or high mixing ($v = 5.5$), or combining intermediate mixing ($v = 2.5$) with low to high compression (Fig. 2b–e). Although mixing and compression did not affect the rise in variance as a resilience indicator (Fig. 2b, c), the effect on the rise in autocorrelation was more noticeable (Fig. 2d, e). Mixing among neighboring core intervals effectively increased the AR(1) autocorrelation signal, whereas core compression eroded the autocorrelation signal, creating an AR process that was higher order than 1 (Fig. 2e). In addition, the peak in autocorrelation often occurred after the SETAR switchpoint, illustrating that choosing the peak as the endpoint, which is customary in many field

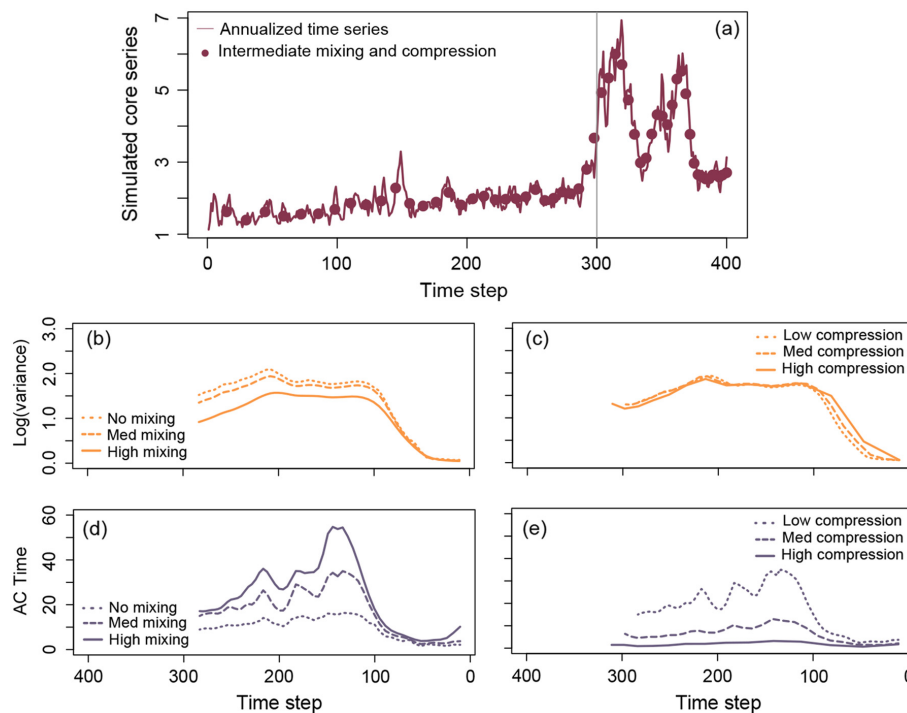


Fig. 2. Effect of mixing and compression on the annualized critical slowing down time series (scenario 3: $M = 225$, $\sigma = 0.15$, $v = 0.15$). (a) Annualized time series overlaid with intermediate mixing and compression. Effect of mixing: rolling window ($n/2$ time steps) (b) variance and (d) autocorrelation for low compression combined with no (dotted line), medium (dashed line), and high (solid line) mixing ($n = 115$ for all three scenarios). Effect of compression: rolling window (c) variance and (e) autocorrelation for low (dotted line; $n = 115$), intermediate (dashed line; $n = 57$), and high (solid line; $n = 25$) compression combined with intermediate mixing. Given the short time series of these rolling window statistics, we reversed the time series and ran the rolling window statistics in reverse to provide information about baseline values.

applications, biases the Kendall tau trend coefficient toward a significant outcome. To test this result further, we ran 100 simulations of the critical slowing down scenario with intermediate mixing and core compression, retaining the Kendall tau statistic for the rolling window autocorrelation and variance leading up to the SETAR switchpoint each time (Appendix S1: Fig. S6). The exercise showed that the Kendall tau trend test on rolling window autocorrelation and variance did not consistently increase prior to a critical transition. The selected example shown (Fig. 2) fell close to the median of the simulated runs (Appendix S1: Fig. S6). Even with low core compression where the autocorrelation signal should be stronger (Fig. 2e), we failed to identify a consistent rise in the autocorrelation across 100 runs (Appendix S1: Fig. S7).

With the flickering simulations (scenarios 4 and 5), the rolling window resilience indicators showed a discernible rise in variance prior to the switchpoints, but the autocorrelation signal was once again largely eroded (not shown).

Our simulations illustrate the limitations of the rolling window approach to detect indicators of a critical transition in sediment core records. We thus turned to time-varying regressions as an alternate approach to rolling window statistics and compared the fit of models with higher order AR processes using AIC.

State-space modeling

For the state-space model simulations presented, we used intermediate mixing and core compression throughout. For the control (scenario 1; Appendix S1: Fig. S8a–f) and regime shift only

(scenario 2; Appendix S1: Fig. S8g–l), simulations were as expected, where a sharp increase in the intercept (regime shift) was only detected in the latter and eigenvalues of both series hovered near one, without a distinctive transition.

For the critical slowing down simulation (scenario 3; Fig. 3a–c), the online DLM with an AR(1) process fit the data, eigenvalue, and process variance well, except for slightly over-predicting the response values immediately after the switchpoint. The change in the intercept tracked the shift in the mean (Fig. 3b) and thus acted as an indicator of the regime shift. The change in the eigenvalue correctly located the switchpoint (crossing of unit circle) and was thus indicative of critical slowing down at a critical transition (Fig. 3c). When using the TVARSS model, we detected the regime shift (i.e., the time-varying intercept tracked the change in the mean) and the eigenvalue crossed the unit circle at the regime shift (Appendix S1: Fig. S9a–c; best-fit AR(2) process shown). The TVARSS model also detected a small increase in the eigenvalue above 1 early on in the series (near time step 75). To

further examine these results, we ran 100 simulations of this critical slowing down scenario with intermediate mixing and core compression, which showed that in general the TVARSS model was more sensitive than the online DLM, whereby the eigenvalue crossed 1 long before the deterministic critical transition on numerous runs (Fig. 4a, b). However, as seen in the selected example (Appendix S1: Fig. S9c), this first crossing rarely coincided with the local maxima and the maximum eigenvalue tended to be close to the deterministic SETAR switchpoint for both the online DLM and TVARSS (Fig. 4c, d).

With higher environmental noise (flickering between alternate states, scenario 4; Fig. 3d–f, Appendix S1: Fig. S4), the best-fit online DLM was the AR(1) model and detected both the regime shift (change in intercept; Fig. 3e) and evidence of a critical transition, although the latter was inaccurate (i.e., the eigenvalue peaked and crossed 1 after the deterministic critical transition; Fig. 3f). The best-fit TVARSS (AR(2) process) detected the regime shift but lacked precision in detecting the deterministic switchpoint (Appendix S1: Fig. S9d–f).

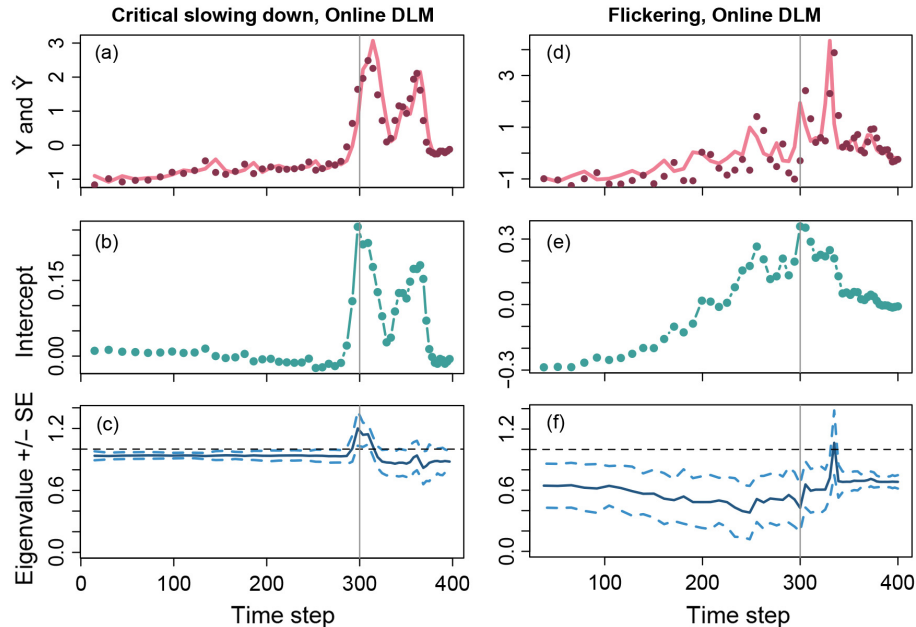


Fig. 3. Online dynamic linear model (DLM) results for the (a–c) critical slowing down (scenario 3: $M = 225$, $\sigma = 0.15$, $v = 0.15$) and (d–f) flickering ($M = 225$, $\sigma = 0.45$, $v = 0.15$) sediment core simulations with intermediate mixing and compression ($n = 57$). The 400-yr time series (dark red points), online DLM goodness of fit (light red line), intercept (green line), and eigenvalue (blue line) are shown for each scenario.

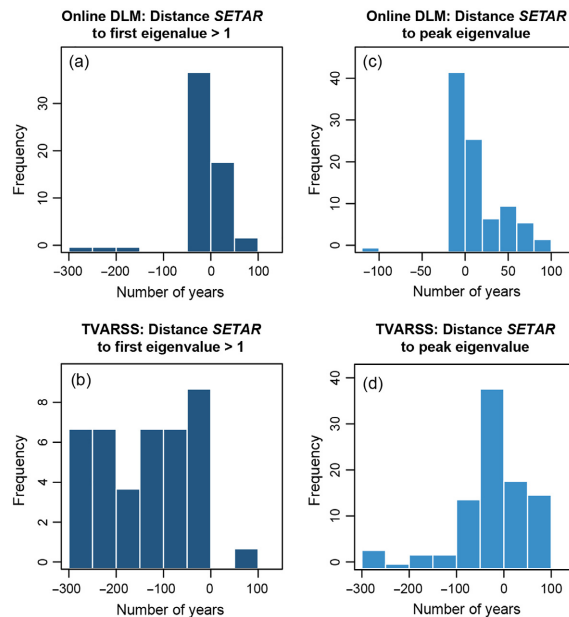


Fig. 4. Summary histograms of the online dynamic linear model (DLM) and time-varying autoregressive state-space (TVARSS) for 100 runs of the critical slowing down scenario ($M = 225$, $\sigma = 0.15$, $v = 0.15$). (a, b) Distance (in years) between the Self Exciting Threshold Autoregressive (SETAR) switchpoint and the first time the eigenvalue exceeds 1. (c, d) Distance (in years) between the SETAR switchpoint and the peak eigenvalue. AR(1) fit is shown for both the online DLM and TVARSS.

For fifth scenario, with both high environmental and sediment noise, the eigenvalue of the best-fit online DLM (AR(3)) always exceeded 1, with a peak near the regime shift (not shown). Although sub-optimal ($\Delta AIC = 13$), the AR(1) online DLM behaved as expected, with the eigenvalue crossing the unit circle at the regime shift. The best-fit TVARSS was of order 3, and the eigenvalue peaked shortly after the year 300.

Overall, for the simulation with low to moderate environmental and sediment noise (scenarios 1–3), the online DLM and TVARSS correctly identified whether regime shifts and critical transitions were absent or present at the onset. As environmental noise increased (scenarios 4–5), both approaches had greater difficulty in providing the precise and accurate timing of the critical transitions.

Lake data analysis

To compare the simulations to observed lake data, we applied the online DLM in lakes where sedimentary proxies indicated either an abrupt change (the shallow Roxton Pond), a more gradual, noisier change to the eutrophic state (the shallow, upland Lake Anarry), or possible flickering between states prior to switching to the eutrophic state (the deep, subalpine Lake Geneva) over the past ~200 yr.

For the Roxton Pond pigment record (Fig. 5a–c), the AR(1) process provided the optimal fit for the online DLM. The intercept detected the recent regime shift (~1990; Fig. 5b), at which point the eigenvalue approached 1 (Fig. 5c), suggesting a nearby critical transition. The Lake Anarry pigment record was much noisier, but it did not behave like flickering (Fig. 5d). Instead,

the eutrophic state seemed to have a higher variance and to display a high-order AR process that was not well constrained by the data. The AR(2), AR(3), and AR(4) fit the pigment time series equally well; the results for AR(3) are shown here. The intercept indicated a regime shift ca. 1950 (Fig. 5e) and the eigenvalue hovered near 1 (Fig. 5f), suggesting instability throughout most of the record. The Anarry core is consistent with an unstable, noise-dominated system. For the Lake Geneva record (Fig. 5g), we found that an AR(1) process was optimal and that the regime shift was picked by the intercept (Fig. 5h). The eigenvalue indicated instability consistent with critical transition around 1970 (Fig. 5i).

The best-fitting TVARSS detected the critical transitions in the time series with moderate environmental noise and indication of critical slowing down (Roxton Pond; Appendix S1: Fig. S10a–c), and the time series with flickering (Lake Geneva; Appendix S1: Fig. S10g–i). For the noisier Lake Anarry time series (Appendix S1: Fig. S10d), the best-fit AR(1) model did not detect the critical transition. However, the higher order AR(3) TVARSS did and is presented here (Appendix S1: Fig. S10e, f). As with the simulations, these results suggest that a lack of

consistency among modeling approaches may arise when the time series are noisier.

DISCUSSION

Resilience indicators of critical transitions, such as rising variance and autocorrelation, have been shown to perform well in theoretical models (Scheffer et al. 2015), laboratory experiments (Drake and Griffen 2010, Dai et al. 2012, Veraart et al. 2012), as well as field experiments (Carpenter et al. 2011, Wilkinson et al. 2018). When using observational time series data, the detection of impending, abrupt shifts to an alternative state has been more difficult to detect, even in systems with long monitoring programs of response metrics and drivers (Spears et al. 2016). In such cases, generic resilience indicators either fail to consistently forewarn a catastrophic shift (Gsell et al. 2016) or provide false positives (i.e., increasing variance and autocorrelation in systems experiencing a continuous transition without bistability; Kéfi et al. 2013, Burthe et al. 2015). A better understanding of why certain resilience indicators failed to detect critical transition is needed, as are methods specific to the system and indicators under study (Kéfi et al.

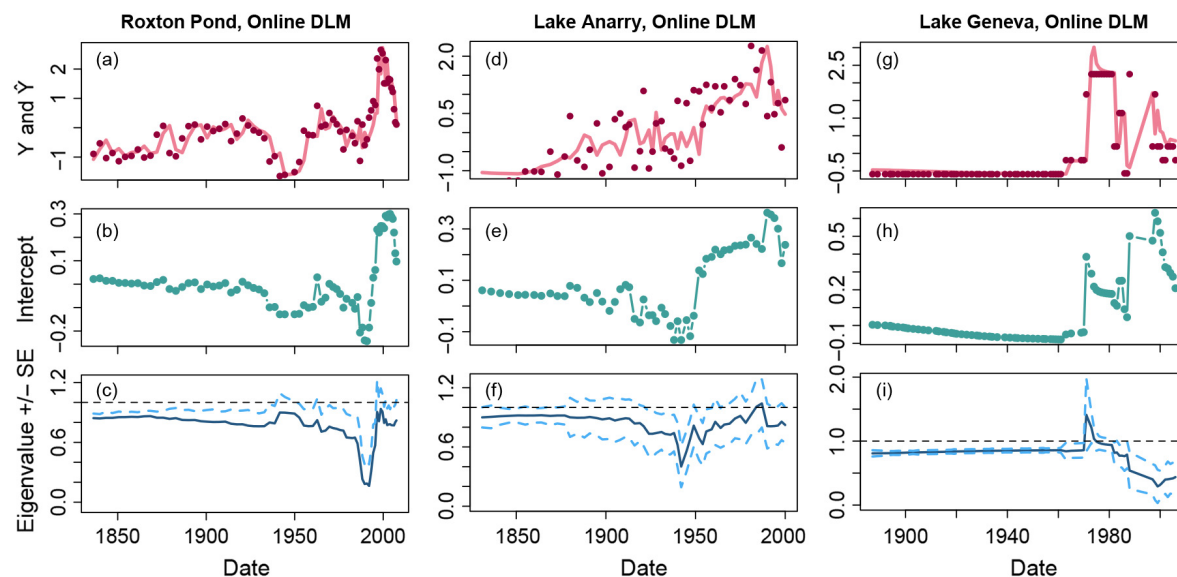


Fig. 5. Summary of online dynamic linear model (DLM) results for (a–c) Roxton Pond ($n = 73$), (d–f) Lake Anarry ($n = 60$), and (g–i) Lake Geneva ($n = 77$). The time series (dark red points), goodness of fit (light red line), intercept (green line), and eigenvalue (blue line) are shown for each lake.

2013). Here, we used paleolimnological records to extend the temporal scale under which regime shifts and critical transitions could be evaluated in aquatic ecosystems. We developed a protocol to mimic paleolimnological records undergoing critical transitions (Fig. 1) to test how certain features of these time series may bias the results of generic resilience indicators (Carstensen et al. 2013). Further, we tested whether our new application of time-varying regressions was better equipped at detecting shifts in this specific field.

Rolling window statistics and paleo-records

Rolling window variance and AR(1) are common methods used to indicate critical transitions. However, our analysis of simulated and real cores showed that rolling window AR(1) was not a reliable indicator for paleolimnological archives (Fig. 2). Rolling window variance was more consistent, but typically both indicators must be used in tandem in rolling window analyses (Scheffer et al. 2015). Instead, we noted rising variance with inconsistent AR(1) responses, even in simulations where we knew the underlying dynamics. These findings are consistent with prior work examining the effect of different ranges of temporal aggregation (from 1 to 10 time units, typical of high-resolution paleolimnological data) on rolling window indicators (Frossard et al. 2015). Similarly, Ditlevsen and Johnsen (2010) found inconsistent signals from autocorrelation and variance in paleo-climate interpretations, while Thomas (2016) found using a simple bifurcation model to mimic paleo-climate data that autocorrelation displayed a better performance for detecting critical slowing down than variance, possibly due to increased sensitivity to additional processes and mechanisms affecting this parameter (Verbesselt et al. 2016). However, degrading the data length, data resolution and data density diminished both the autocorrelation and variance signals (Thomas 2016). Bruel et al. (2018) recently noted an increased variance and autocorrelation preceding non-critical transitions from paleo-records, but that the values and significance of resilience indicators strongly depend on the rolling window size. Lastly, an analysis of forest–grassland transitions using a simple model as well as observed paleo-records found that rolling window variance, but

not autocorrelation, could reliably detect the transition (M. Stegner et al., *unpublished manuscript*). Together, these findings suggest that rolling window analyses are questionable for paleo-data. Therefore, alternatives such as the time-varying regressions with DLM or other methods as introduced by Ives and Dakos (2012) should be sought.

Time-varying regression as a new approach

Our simulations showed that both the online DLM and TVARSS could detect regime shifts and critical transitions in sediment core records even when severe core compression and higher order lags affected the AR process. We noted some differences, however, between the online DLM and the TVARSS methods in their sensitivity to detect critical transitions in time series experiencing a one-way transition (critical slowing down). Similarly, with flickering dynamics the precision and accuracy of the critical transition differed between methods. Although the detection of regime shifts is fairly reliable, critical transitions impose the additional criterion that the eigenvalue approaches one. This did not always occur at the deterministic switchpoint for TVARSS (Fig. 4b), and consequently, our simulations showed that regime shifts were more accurately detected than critical transitions. Nonetheless, the peak in the eigenvalue tended to be more accurate than the first-time crossing of the unit circle (Fig. 4d). If we can extrapolate from the simulations to the real core records, we suggest that both the time of the regime shift (intercept) and the time of pronounced increase in the eigenvalue be used to identify when a critical transition occurs.

Core vs. time domain

It is important to note that our study was conducted in the core domain, such that events that took place in the time domain were transformed to a record in the core domain by the processes of core formation and paleolimnological sampling. The core domain is thus a transformed representation of events in the time domain. Here, we tested whether regime shifts and critical transitions that occur in the time domain can be detected in the core domain without back-transforming to the time domain. We learned that (1) regime shifts and critical transitions can

be detected in the core domain if mixing and compression are sufficiently small. (2) Some cores show evidence of regime shifts, and some of these cores also show evidence of critical transition. Working in the core domain may also provide an opportunity to improve the way we sample sediment cores, by re-sampling according to the age models close to the transition and testing whether this improved detections.

A follow-up question to study is whether we can transform back to the time domain. For instance, given mixing and compression parameters, can we estimate regime shifts and critical transition indicators in the time domain? This follow-up question has at least three parts: (1) Given the core data (time, depth, and indicator, such as pigment concentrations), can we reconstruct the time domain series? (2) Given the time domain series, what are the regime shift and critical transition measures? and (3) How do reconstructed time domain estimates of regimes shifts and critical transitions compare to core domain estimates and to the original (true) time domain measures? Weighing by number of years per core interval, where models would be altered to include missing values to control the influence of data points when there is a large time gap, could be one solution (Einarsson et al. 2016). We tested the effect of pre-processing the simulated and real cores (weighed by time gaps) on the resilience indicators and state-space model results and found that in some instances (e.g., low environmental noise) time-gap weighting created strong trends that could be corrected by detrending (not shown). With higher environmental noise, however, the effect of time-gap weighting, detrending, or both was not consistent and would need to be explored further with a large population of paleo-records (not shown). That said, our test of false positives suggests that working in the core domain did not bias the rolling window variance statistic. Had the rise in variance in the control been large enough to interfere with the inference of whether a core experienced a state shift, the appropriate adjustment would be to weight each core sample by the number of years it represents. This would inflate the deeper, low-nutrient variances and have little effect on the core surface resulting in a flatter (and more accurate) trend of variance vs. core depth.

CONCLUSIONS

The case studies selected here showed that, across different proxies, regime shifts and critical transitions could be detected in sediment core archives experiencing both critical slowing down (Roxton Pond and Lake Anarry) and possible flickering (Lake Geneva). We worked both with indicators that are very integrative of the overall functioning of the lake (hypolimnetic hypoxia) and slightly less integrative (algal community), each contributing a complementarity to the eutrophication diagnosis. However, future broad-scale syntheses of sediment core archives are needed to better understand the conditions under which time-varying regressions provide consistent findings, how the model outcomes of critical slowing down and flickering time series vary among lake types (i.e., deep vs. shallow lakes), and if certain proxies are better indicators of critical transitions. Different proxies correspond to species with different life cycles, where one year would cover many generations for some and would thus present inherently different temporal mixing effects. Degradation is also likely to vary among proxies (e.g., diatom frustules vs. genetic markers) or lake types (deep, oligotrophic vs. shallow, eutrophic with slow vs. fast sedimentation rate; Kyle et al. 2015) and can be a limitation for the detection of regime shifts. Although artifacts due to degradation are not considered in this paper, this merits further evaluation. To better capture and explain the degradation effect, some significant additional work is needed, namely a synthesis of many studies where multiple proxies are examined in the same lake, that these study lakes be situated along a gradient of environmental factors, and that new simulations (modeling lake depth, eutrophication, light attenuation, anoxia, etc.) be conducted. Multi-proxy studies would also be fruitful as examining changes in the functional composition of the community as a whole may provide better diagnostic information of an approaching critical transition than any one proxy (Doncaster et al. 2016, Beck et al. 2018). This would also allow to test whether regime shifts and critical transitions could be upscaled to the whole aquatic ecosystem, that is, whether resilience indicators at one trophic level are detectable in other components of a food web (Carpenter et al. 2008). Lastly, the drivers, whether extrinsic

or intrinsic, varied among study locations and the links between proxy and drivers would likewise need to be evaluated (*sensu* Seddon et al. 2014, Randsalu-Wendrup et al. 2016). Indeed, there is relatively little work on how ecosystems respond to different or multiple interacting drivers (Andersen et al. 2009). Interestingly, this could be examined further with TVARSS which can handle covariates and potentially help tease apart the role of each driver.

In general, with field data, the inference of critical transitions depends on convergence of several sources of information. For example: Is there a sensible ecosystem model that can generate critical transitions? Is this model consistent with diverse kinds of data (mechanistic field studies, experiments, comparative studies, etc.)? Does the model generate data that resemble the observed time series using various statistical diagnostics? Does the observed time series exhibit characteristics consistent with critical transitions? In the case of lakes, are there other, similar, lakes that seem to exhibit the same critical transitions and do their time series have similar characteristics? For paleolimnological studies, we offer the following advice: (1) More samples (closer-interval sampling) are better, and (2) time series with high signal to noise is especially valuable for identifying a threshold. Most importantly, this study indicates that paleolimnological data are sometimes capable of discerning critical transitions, despite the problems with mixing, compression, and various kinds of noise that may occur in these archives.

ACKNOWLEDGMENTS

A significant portion of this paper was made possible thanks to the valuable input, statistical modeling, and R code provided by Dr. Anthony R Ives, to whom we are most grateful. We are very grateful to Drs. Mark Stevenson and Suzanne McGowan for providing data on Lake Anarry. We also thank Dr. Irene Gregory-Eaves for fruitful discussions and comments on earlier drafts of the manuscript. We are also very thankful to Drs. Peter R. Leavitt, Lynda Bunting, Teresa Buchaca, Jordi Catalan, Isabelle Domaizon, Piero Guilizzoni, Andrea Lami, Heather Moorhouse, Frances R. Pick, Patrick L. Thompson, Rolf D. Vinebrooke, and the late Giuseppe Morabito for providing sediment core archive data that helped parametrize the compression model. This paper is a contribution for the PAGES Aquatic Transition working group, from which the

original idea emerged. ZET is funded through an NSERC postdoctoral grant. SRC is supported by the North Temperate Lakes Long-Term Ecological Research Program (NSF DEB-1440297) and NSF grants (DEB-1455461 and DEB-1754712) from the Ecosystems Studies Program.

LITERATURE CITED

- Andersen, T., J. Carstensen, E. Hernández-García, and C. M. Duarte. 2009. Ecological thresholds and regime shifts: approaches to identification. *Trends in Ecology & Evolution* 24:49–57.
- Beck, K. K., M.-S. Fletcher, P. S. Gadd, H. Heijnis, K. M. Saunders, G. L. Simpson, and A. Zawadzki. 2018. Variance and rate-of-change as early warning signals for a critical transition in an aquatic ecosystem state: a test case from Tasmania, Australia. *Journal of Geophysical Research: Biogeosciences* 123:495–508.
- Belle, S., et al. 2017. Rising variance and abrupt shifts of subfossil chironomids due to eutrophication in a deep sub-alpine lake. *Aquatic Ecology* 51:307–319.
- Biggs, R., S. R. Carpenter, and W. A. Brock. 2009. Turning back from the brink: detecting an impending regime shift in time to avert it. *Proceedings of the National Academy of Sciences USA* 106:826–831.
- Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 2008. *Time series analysis: forecasting and control*. Fourth edition. John Wiley & Sons, New York, New York, USA.
- Bruel, R., A. Marchetto, A. Bernard, A. Lami, P. Sabatier, V. Frossard, and M.-E. Perga. 2018. Seeking alternative stable states in a deep lake. *Freshwater Biology*. <https://doi.org/10.1111/fwb.13093>
- Burthe, S. J., et al. 2015. Do early warning indicators consistently predict nonlinear change in long-term ecological data? *Journal of Applied Ecology* 53:666–676.
- Capon, S. J., et al. 2015. Regime shifts, thresholds and multiple stable states in freshwater ecosystems; a critical appraisal of the evidence. *Science of the Total Environment* 534:122–130.
- Carpenter, S. R. 2003. Regime shifts in lake ecosystems: pattern and variation. International Ecology Institute, Oldendorf/Luhe, Germany.
- Carpenter, S. R., and W. A. Brock. 2006. Rising variance: a leading indicator of ecological transition. *Ecology Letter* 9:311–318.
- Carpenter, S. R., W. A. Brock, J. J. Cole, J. F. Kitchell, and M. L. Pace. 2008. Leading indicators of trophic cascades. *Ecology Letter* 11:128–138.
- Carpenter, S. R., and R. Lathrop. 2008. Probabilistic estimate of a threshold for eutrophication. *Ecosystems* 11:601–613.

- Carpenter, S. R., et al. 2011. Early warnings of regime shifts: a whole-ecosystem experiment. *Science* 332:1079–1082.
- Carstensen, J., R. J. Telford, and H. J. B. Birks. 2013. Diatom flickering prior to regime shift. *Nature* 498: E11–E12.
- Dai, L., D. Vorselen, K. S. Korolev, and J. Gore. 2012. Generic indicators for loss of resilience before a tipping point leading to population collapse. *Science* 336:1175–1177.
- Dakos, V., et al. 2008. Slowing down as an early warning signal for abrupt climate change. *Proceedings of the National Academy of Sciences USA* 105:14308–14312.
- Dakos, V., et al. 2012. Early warnings of critical transitions: methods for time series. *PLoS ONE* 7:e0041010.
- Di Narzo, A. F., J. L. Aznarte, and M. Stigler. 2009. tsDyn: time series analysis based on dynamical systems theory. <http://stat.ethz.ch/CRAN/web/packages/tsDyn/vignettes/tsDyn.pdf>
- Ditlevsen, P. D., and S. J. Johnsen. 2010. Tipping points: early warning and wishful thinking. *Geophysical Research Letters* 37:L19703.
- Doncaster, C. P., et al. 2016. Early warning of critical transitions in biodiversity from compositional disorder. *Ecology* 97:3079–3090.
- Drake, J. M., and B. D. Griffen. 2010. Early warning signals of extinction in deteriorating environments. *Nature* 467:456–459.
- Edmondson, W. T. 1961. Changes in Lake Washington following an increase in the nutrient income. *Proceedings of the International Association of Theoretical and Applied Limnology* 16:153–158.
- Einarsson, Á., U. Hauptfleisch, P. R. Leavitt, and A. R. Ives. 2016. Identifying consumer–resource population dynamics using paleoecological data. *Ecology* 97:361–371.
- Frossard, V., B. Saussereau, A. Perasso, and F. Gillet. 2015. What is the robustness of early warning signals to temporal aggregation? *Frontiers in Ecology and Evolution* 3:1–9.
- Gsell, A. S., et al. 2016. Evaluating early-warning indicators of critical transitions in natural aquatic ecosystems. *Proceedings of the National Academy of Sciences USA* 113:E8089–E8095.
- Higham, D. 2001. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review* 43:525–546.
- Holling, C. S. 1973. Resilience and stability of ecological systems. *Annual Review of Ecology Systematics* 4:1–23.
- Ives, A. R., and V. Dakos. 2012. Detecting dynamical changes in nonlinear time series using locally linear state-space models. *Ecosphere* 3:58.
- Ives, A. R., B. Dennis, K. L. Cottingham, and S. R. Carpenter. 2003. Estimating community stability and ecological interactions from time-series data. *Ecological Monographs* 73:301–330.
- Jenny, J. P., et al. 2014. Inherited hypoxia: a new challenge for reoligotrophic lakes under global warming. *Global Biogeochemical Cycles* 28:1413–1424.
- Jenny, J. P., et al. 2016. Urban point sources of nutrients were the leading cause for the historical spread of hypoxia across European lakes. *Proceedings of the National Academy of Sciences USA* 113:12655–12660.
- Kéfi, S., et al. 2013. Early warning signals also precede non-catastrophic transitions. *Oikos* 122:641–648.
- Kyle, M., S. Haande, J. Sønstebo, and T. Rohrlack. 2015. Amplification of DNA in sediment cores to detect historic Planktothrix occurrence in three Norwegian lakes. *Journal of Paleolimnology* 53:61–72.
- Lenton, T. M., V. N. Livina, V. Dakos, E. H. van Nes, and M. Scheffer. 2012. Early warning of climate tipping points from critical slowing down: comparing methods to improve robustness. *Philosophical Transactions of the Royal Society A* 370:1185–1204.
- Pole, A., M. West, and J. Harrison. 1994. *Applied Bayesian forecasting and time series analysis*. Chapman & Hall, New York, New York, USA.
- Randsalu-Wendrup, L., D. J. Conley, J. Carstensen, and S. C. Fritz. 2016. Paleolimnological records of regime shifts in lakes in response to climate change and anthropogenic activities. *Journal of Paleolimnology* 56:1–14.
- Reckhow, K. H. 1977. Empirical lake models for phosphorus: development, applications, limitations and uncertainty. Pages 193–221 *in* D. Scavia and A. Robertson, editors. *Lake ecosystem modeling*. Ann Arbor Science, Ann Arbor, Michigan, USA.
- Scheffer, M. 2009. *Critical transitions in nature and society*. Princeton University Press, Princeton, New Jersey, USA.
- Scheffer, M., S. R. Carpenter, V. Dakos, and E. H. van Nes. 2015. Generic indicators of ecological resilience: inferring the chance of a critical transition. *Annual Review of Ecology, Evolution, and Systematics* 46:145–167.
- Scheffer, M., S. R. Carpenter, J. A. Foley, C. Folke, and B. Walker. 2001. Catastrophic shifts in ecosystems. *Nature* 413:591–596.
- Seddon, A. W. R., C. A. Froyd, A. Witkowski, and K. J. Willis. 2014. A quantitative framework for analysis of regime shifts in a Galápagos coastal lagoon. *Ecology* 95:3046–3055.
- Spanbauer, T. L., et al. 2014. Prolonged instability prior to a regime shift. *PLoS ONE* 9:e108936.

- Spears, B. M., et al. 2016. Ecological instability in lakes: A predictable condition? *Environmental Science & Technology* 50:3285–3286.
- Stevenson, M. A., S. McGowan, N. J. Anderson, R. H. Foy, P. R. Leavitt, Y. R. McElarney, D. R. Engstrom, and S. Pla-Rabés. 2016. Impacts of forestry planting on primary production in upland lakes from north-west Ireland. *Global Change Biology* 22:1490–1504.
- Stigler, M. 2010. tsDyn: threshold cointegration: overview and implementation in R. <http://stat.ethz.ch/CRAN/web/packages/tsDyn/vignettes/ThCointOverview.pdf>
- Taranu, Z. E., et al. 2015. Acceleration of cyanobacterial dominance in north temperate subarctic lakes during the Anthropocene. *Ecology Letters* 18:375–384.
- Thomas, Z. A. 2016. Using natural archives to detect climate and environmental tipping points in the Earth system. *Quaternary Science Reviews* 152:60–71.
- Veraart, A. J., E. J. Faassen, V. Dakos, E. H. van Nes, M. Lürling, and M. Scheffer. 2012. Recovery rates reflect distance to a tipping point in a living system. *Nature* 481:357–360.
- Verbesselt, J., N. Umlauf, M. Hirota, M. Holmgren, E. H. van Nes, M. Herold, A. Zeileis, and M. Scheffer. 2016. Remotely sensed resilience of tropical forests. *Nature Climate Change* 6:1028–1031.
- Vermaire, J. C., Z. E. Taranu, G. K. MacDonald, K. Velghe, E. M. Bennett, and I. Gregory-Eaves. 2017. Extrinsic versus intrinsic regimes shifts in shallow lakes: long-term response of cyanobacterial blooms to historical catchment phosphorus loading and climate warming. *Frontiers in Ecology and Evolution* 5:146.
- Wilkinson, G. M., S. R. Carpenter, J. J. Cole, M. L. Pace, R. D. Batt, C. D. Buelo, and J. T. Kurtzweil. 2018. Early warning signals precede cyanobacterial blooms in multiple whole-lake experiments. *Ecological Monographs*. <https://doi.org/10.1002/ecm.1286>
- Zeileis, A., and G. Grothendieck. 2005. zoo: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software* 14:1–27.

SUPPORTING INFORMATION

Additional Supporting Information may be found online at: <http://onlinelibrary.wiley.com/doi/10.1002/ecs2.2438/full>

Ecosphere. **Can we detect ecosystem critical transitions and signals of changing resilience from paleo-ecological records?** Taranu Z.E., S.R. Carpenter, V. Frossard, J.-P. Jenny, Z. Thomas, J.C. Vermaire, and M.-E. Perga

Appendix 1. Supplemental figures

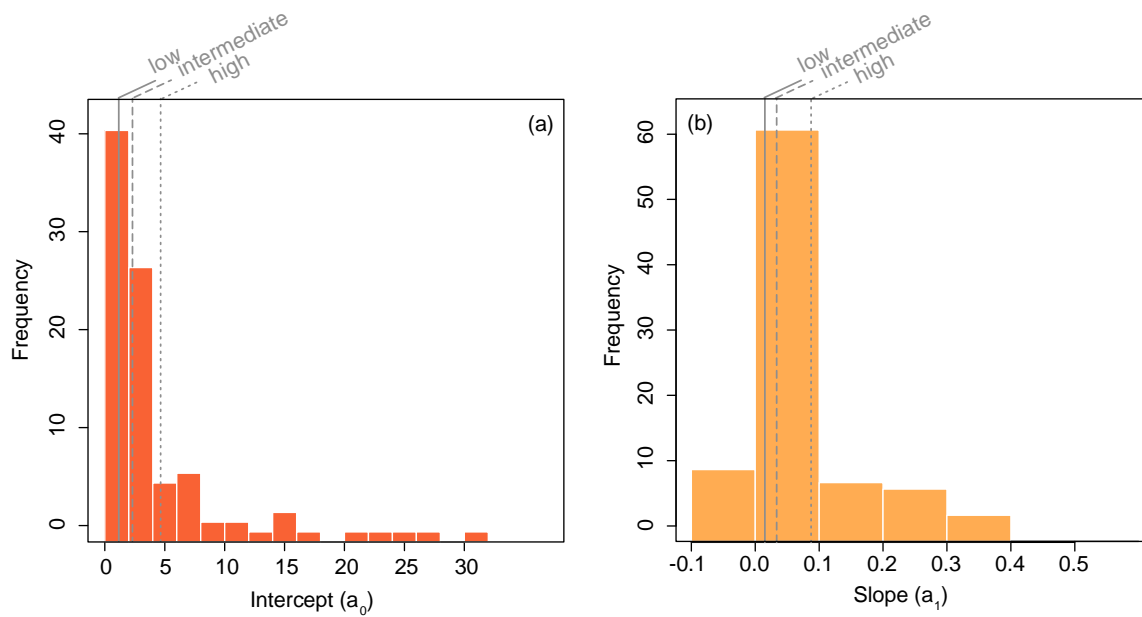


Figure S1. Histogram of (a) intercept, a_0 , and (b) slope, a_1 , of the sedimentation rate model from cores synthesized by Taranu et al. (2015). Sedimentation rate coefficients of example cores with relatively low (1st quartile: $a_0 = 1.28$, $a_1 = 0.015$), intermediate (median: $a_0 = 2.24$, $a_1 = 0.034$), and high (3rd quartile: $a_0 = 4.35$, $a_1 = 0.088$) compression are shown by grey solid, dashed, and dotted lines, respectively.

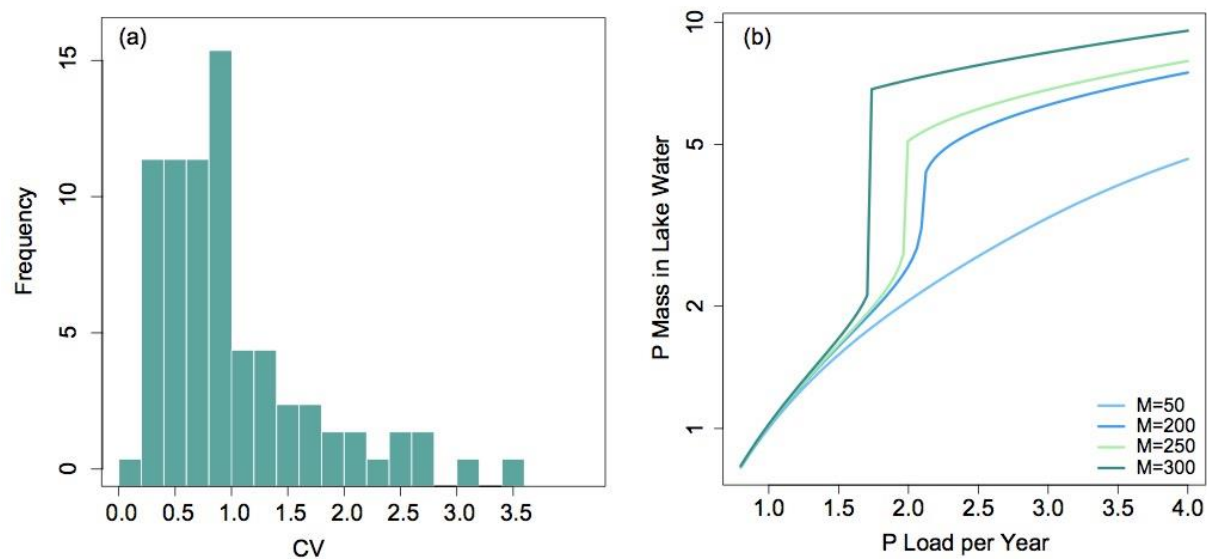


Figure S2. (a) Histogram of the coefficient of variation, CV, of cyanobacterial pigment concentrations from cores synthesized by Taranu et al. (2015). (b) Examples of deterministic eutrophication models with incremental sediment phosphorus mass (i.e. from non-critical to critical transitions).

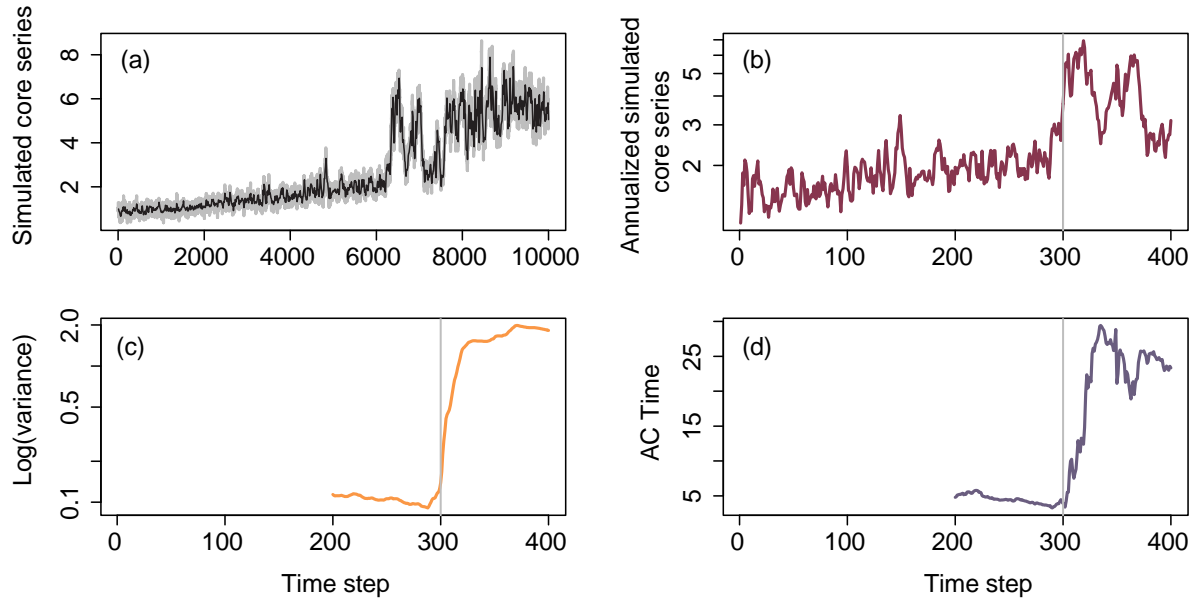


Figure S3. Simulated time series of critical slowing down (scenario 3: $M = 225$, $\sigma = 0.15$, $v = 0.15$) from the P load and lake eutrophication model. (a) Annualized time series (black) plotted over the 0.1-year series (grey). (b) Annualized time series truncated to 300 years prior and 100 post the SETAR threshold (re-scaled axis is shown). (c) Rolling window ($n/2$ time steps) variance and (d) lag-1 autocorrelation, presented as $[1/\log(\text{lag-1 autocorrelation coefficient})]$, of the truncated 400 year series. The SETAR switchpoint is shown by the grey line in panels (b-d).

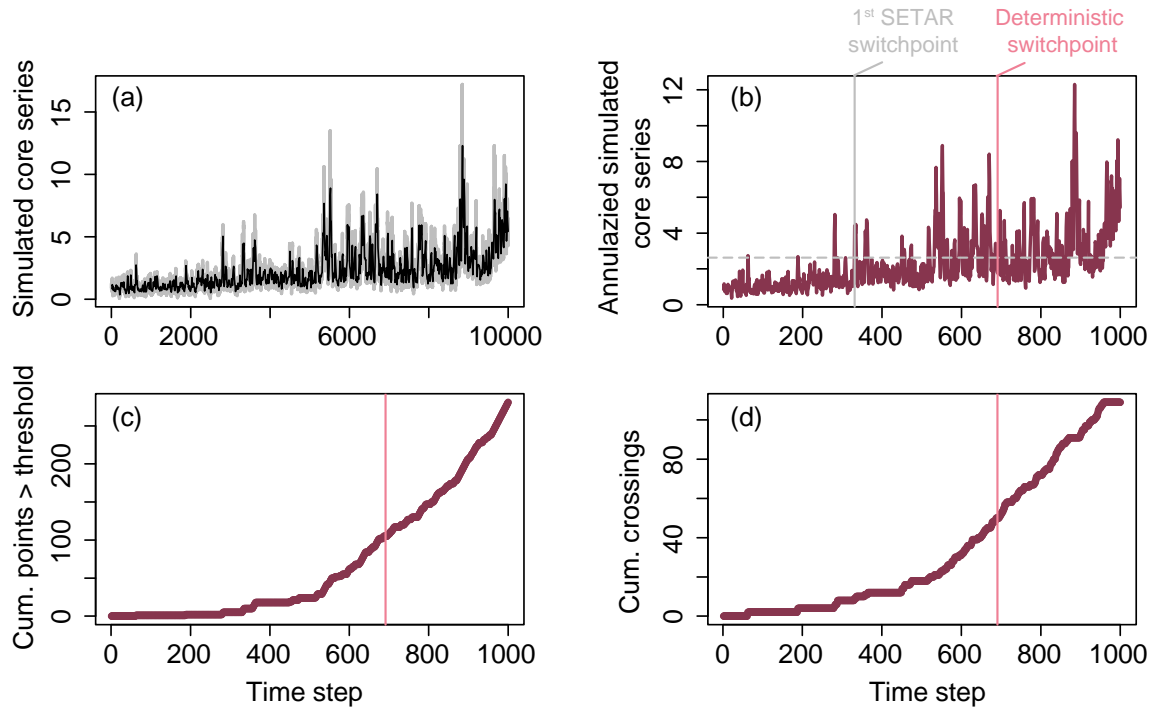


Figure S4. Simulated time series of flickering with high ecosystem noise ($M = 225$, $\sigma = 0.45$, $v = 0.15$). (a) Annualized time series (black) plotted over the 0.1-year series (grey). (b) Comparison of the true deterministic switchpoint (light red line) and the first SETAR crossing (grey line) for the annualized time series. (c) Cumulative sum of simulated values above SETAR threshold, where values above the threshold were assigned a value of 1, and those below were assigned a value of 0. (d) Cumulative number of times a state change occurs, where a 0 was assigned if the value either remained in an oligotrophic or eutrophic state from one time step to the next, otherwise a 1 was assigned.

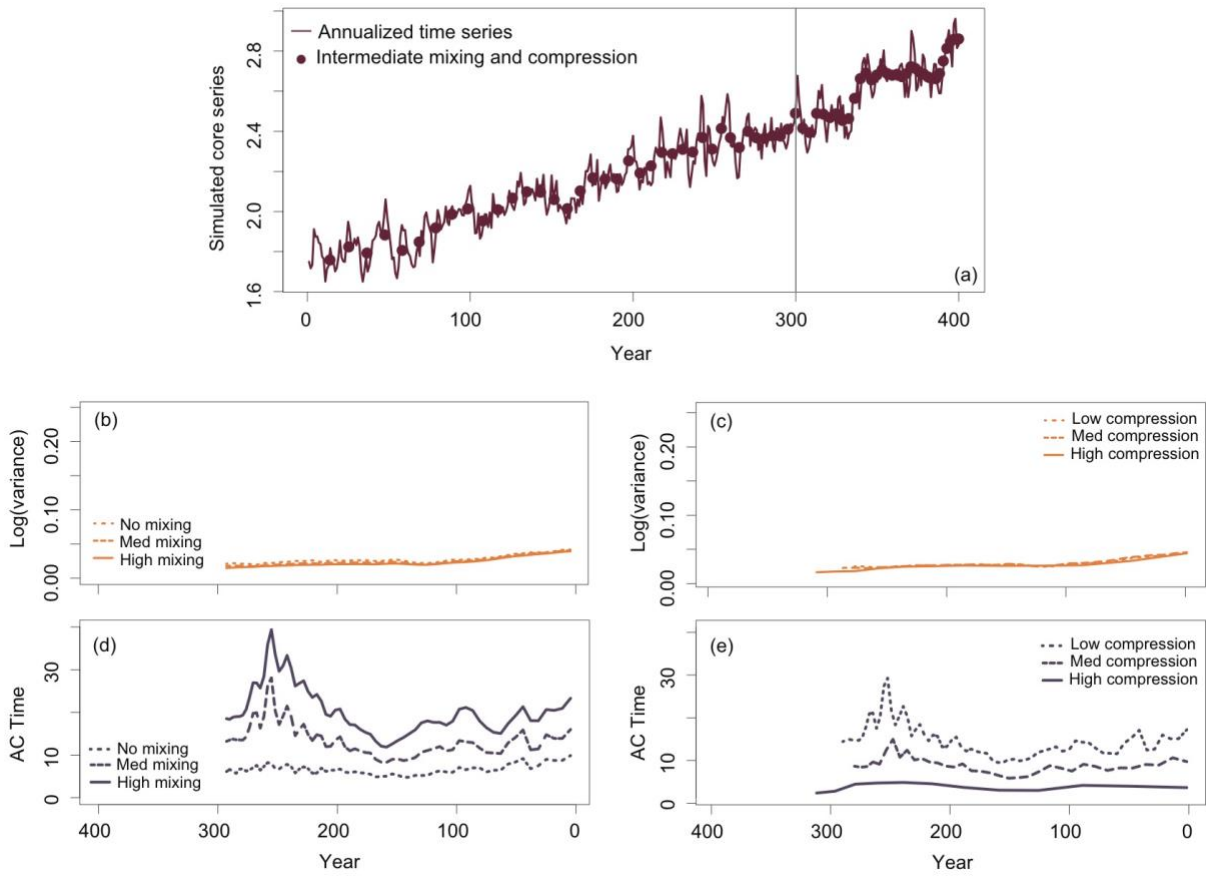


Figure S5. Effect of mixing and compression on the annualized control time series (scenario 1). See Fig. 2 for details.

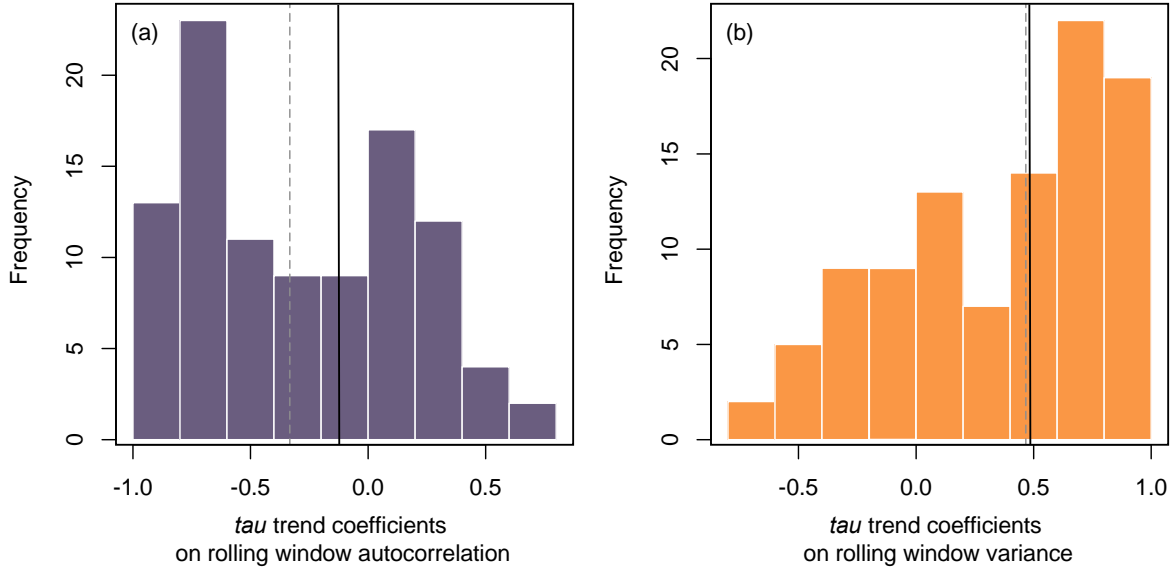


Figure S6. Summary histograms of the Kendall τ statistics on the rolling window (a) autocorrelation and (b) variance for 100 runs of the critical slowing down scenario ($M = 225$, $\sigma = 0.15$, $v = 0.15$) with intermediate mixing and intermediate compression. The rolling window length was decreased from 28.5 ($n/2$ time steps) to 21 to provide at least 10 data points for the Kendall τ statistic, which was calculated from the start of the rolling window series (time step 240) to the SETAR switchpoint (time step 300). Results from the selected example (Fig. 2) are shown by the black solid lines and the median of the 100 runs are shown by the grey dashed lines.

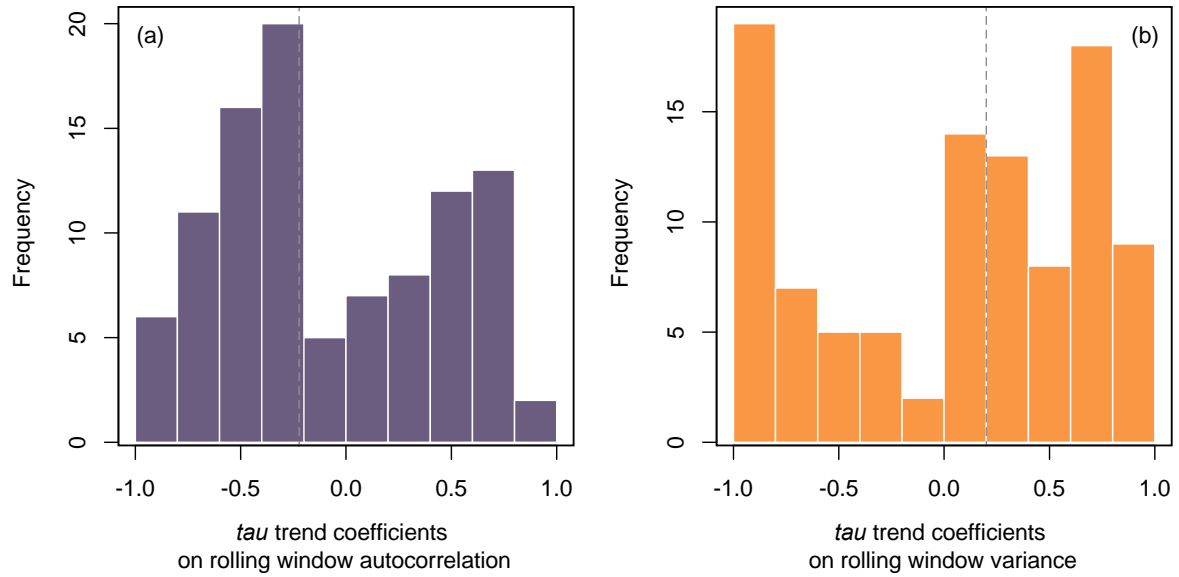


Figure S7. Summary histograms of the Kendall τ statistics on the rolling window (a) autocorrelation and (b) variance for 100 runs of the critical slowing down scenario with intermediate mixing and low compression. The rolling window length was decreased from 57.5 ($n/2$) to 45 to provide a series that covered the same time frame as in Fig. S6 (i.e. from time step 240 to 300). The median of the 100 runs are shown by the grey dashed lines.

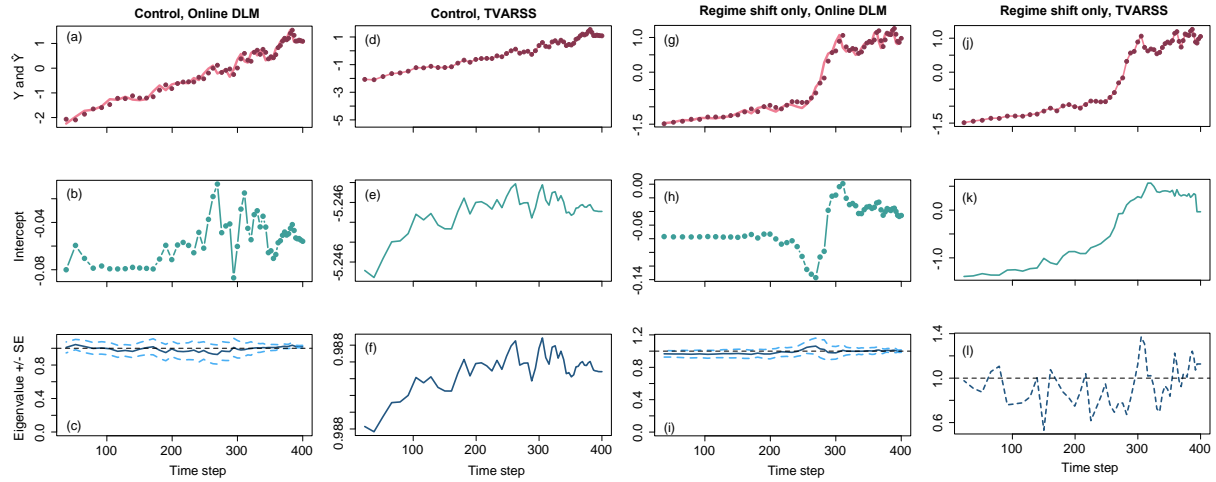


Figure S8. Online DLM and TVARSS results for the (a-f) control (scenario 1: $M = 125$, $\sigma = 0.05$, $v = 0$; optimal online DLM: AR(1); optimal TVARSS: AR(2)), and (d-f) regime shift only (scenario 2: $M = 200$, $\sigma = 0.05$, $v = 0$; optimal online DLM: AR(1); optimal TVARSS: AR(3)) sediment core simulations with intermediate core mixing and compaction ($n = 57$). The 400-year time series (dark red points), goodness of fit (light red line), intercept (green line), and eigenvalue (blue line) are shown for each scenario and modeling approach.

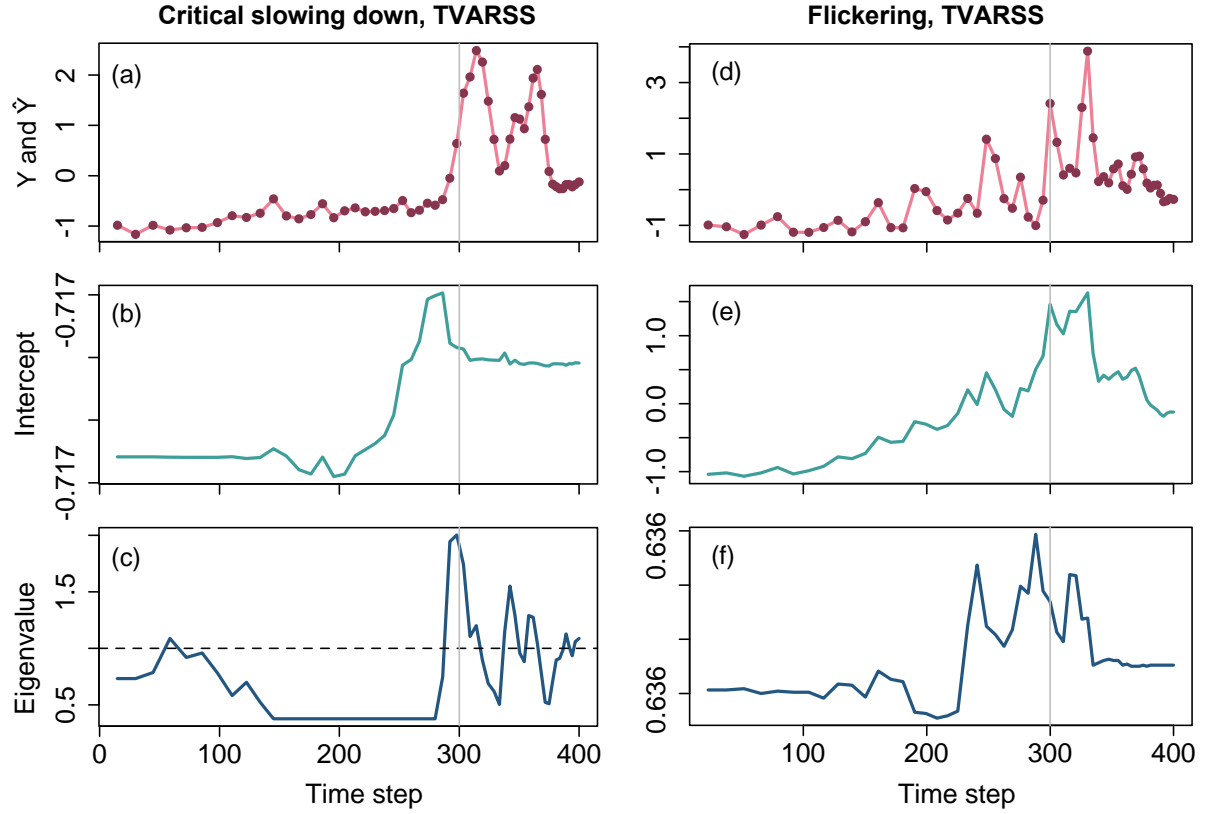


Figure S9. Summary of TVARSS for the (a-c) critical slowing down (scenario 3: $M = 225$, $\sigma = 0.15$, $v = 0.15$) and (d-f) flickering (scenario 4: $M = 225$, $\sigma = 0.45$, $v = 0.15$) sediment core simulations with intermediate mixing and compression. The time series (dark red points), goodness of fit (light red line), intercept (green line), and eigenvalues (blue line) are shown for each scenario.

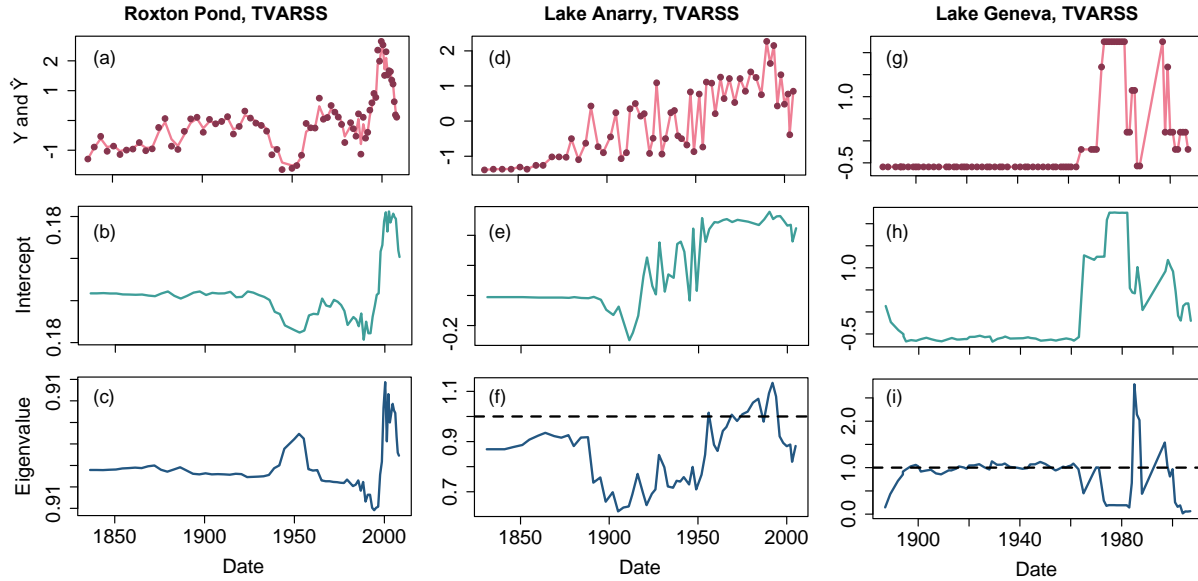


Figure S10. Summary of TVARSS for (a-c) Roxton Pond (optimal model: AR(1)), (d-f) Lake Anarry (optimal model: AR(3)) and (g-i) Lake Geneva (optimal model: AR(1)). The time series (dark red points), goodness of fit (light red line), intercept (green line), and eigenvalue (blue line) are shown for each lake. Note that for Lake Geneva, we included winter temperature as a covariate to control for the sharp increase in the response; that is, the increase in 1970 followed by 10 years of no change caused a challenge for TVARSS as the AR parameters follow a random walk and are limited in their rates of change. The covariate was thus used to capture the big jump.