Beyond the Tipping Point: Predicting the Collapse of Complex Environmental Systems

Sam Whitehall *

*University of Southampton, School of Electronics and Computer Science

An Individual Research Project Final Report submitted for the award of MEng Computer Science.

A broad range of complex, dynamical systems have demonstrated a tendency towards sudden dynamical regime shifts in response to slow changes in external conditions, which are often difficult to reverse—so called tipping points or critical transitions. In this report, the theoretical basis for stability is examined, as well as two broad approaches to predicting critical transitions: explicit process-based modelling and statistical early warning signals. Published results are summarised for two environmental case studies: abrupt global climate shifts and eutrophication in shallow lakes, and the power and limitations of each are discussed.

tipping point | critical slowing | process-based modelling | early warning signals

Introduction

Many real-life complex, dynamical systems show regions of stability—specific system states which are resilient to external changes, where negative feedback loops correct natural or man-made perturbations.

A catastrophe in these systems is typically some sudden shift between different stable regions—from some beneficial state, such as habitable environmental conditions, to a much less desirable state. This can be catastrophic in numerous ways: the destination state may be inhospitable, the sudden change to reach this may be dangerous, or the destination state may be chaotic and unpredictable.

A lot of research into the stability of systems focusses on developing explicit models of the system processes involved, and real-world catastrophes are predicted by considering the outputs from simulations of these models, given a range of feasible input values. While this approach has had some successes, they can be poorly generalisable, display have a poor track record by mask catastrophes caused by poorly understood mechanisms[9] or predict with too much uncertainty to be useful for policy-making[28, 10]. The successes and failures of some contemporary models for specific case system studies are detailed in this paper.

A more recent branch of research has looked at other approaches—testing a growing collection of indicators[23, 16], which are statistical measures from time series data. The most commonly studied indicators are detailed in this report, their effectiveness and applicability is discussed.

This report begins with a concise primer of dynamic systems theory, although the interested reader will find more detail in any good dynamical systems textbook, such as [26]. The report then focusses on two environmental case studies with known catastrophic critical points—abrupt climate change and shallow lake eutrophication. However, many of the techniques are applicable to a much wider array of dynamical systems, and have been used in academic areas as diverse as ecology, finance and neuroscience.

Dynamical systems

A *dynamical system* mathematically formalises a system whose state changes over time. The state can be represented as some d-dimensional vector $\mathbf{x} \in \mathcal{D}$, where $\mathcal{D} \subseteq \mathbb{R}^d$—a unique configuration of the system from a space of all possible configurations. A dynamical system also considers some *evolution rule*, detailing how the state changes over time. For continuous-valued models (*flows*), this is represented as an ordinary differential equation $^{d\mathbf{x}}/_{dt} = f(\mathbf{x})$, abbreviated by convention to $\dot{\mathbf{x}}$. We also require the initial system state at $t = 0$: $\mathbf{x}_0$.

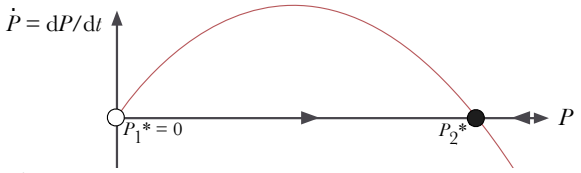


Fig. 1. Diagram of the simple logistic population growth dynamical system. The population P has positive growth in the range $0 \leq P \leq P_2^*$, and a negative growth rate for $P > P_2^*$—highlighted by the arrows on the P-axis. P_1^* is an unstable fixed point (open), while P_2^* is a stable fixed point (shaded).

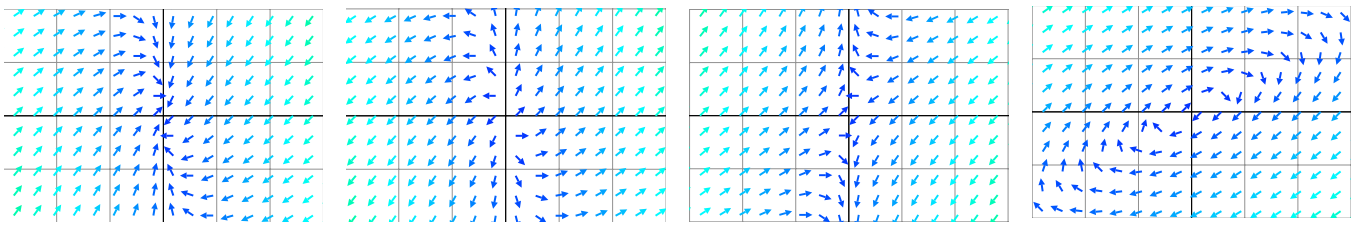
As a simple example, Figure 1 shows a graphical representation of the logistic growth model: a simple model for population dynamics. Here, we consider one variable—population, $P > 0$. We see compound positive growth ($\dot{P} > 0$) up to the critical point P_2, where the environment is saturated, and growth slows (i.e. $\dot{P} < 0$).

Firstly, we may notice the points P_1^* and P_2^*, where $^{dP}/_{dt} = 0$. If we are at either of these, we do not move elsewhere—these are *fixed points*. If we disturb P_1^* to $P_1^* + \epsilon$, then we see positive feedback propelling the population higher. However, if we disturb P_2^*, in either direction, then P is propelled back to P_2^*

We say that P_1^* is *stable* or *attracting*, while P_2^* is *unstable* or *repelling*.

Stability in Higher Dimensions. We can also determine stability for a constant-coefficient linear dynamical system, $\dot{\mathbf{x}} = A\mathbf{x}$, where $A \in \mathbb{R}^{n \times n}$, with a single fixed point at $\mathbf{x} = \mathbf{0}$. As the evolution rule is linear, we can assume a solution following a linear trajectory of the form $\mathbf{x} = \mathbf{v}e^{\lambda t}$ exists[26]. Therefore, $\dot{\mathbf{x}} = \lambda \mathbf{v} e^{\lambda t}$. We substitute this back, to achieve $\lambda \mathbf{v} e^{\lambda t} = A\mathbf{v}e^{\lambda t} \Rightarrow \lambda \mathbf{v} = A\mathbf{v}$, which is the eigenvector formulation. Hence, we can consider whether the system undergoes exponential growth or decay along vector $\mathbf{v}$ (representing $\mathbf{0}$

Reserved for Publication Footnotes



(a) $A = \begin{pmatrix} -2 & 0 \\ -2 & -1 \end{pmatrix}$ $\lambda = \{-2, -1\}$

(b) $A = \begin{pmatrix} 2 & 0 \\ 2 & 1 \end{pmatrix}$ $\lambda = \{2, 1\}$

(c) $A = \begin{pmatrix} -2 & 0 \\ -2 & 1 \end{pmatrix}$ $\lambda = \{-2, 1\}$

(d) $A = \begin{pmatrix} -1 & 2 \\ -1 & -1 \end{pmatrix}$ $\lambda = -1 \pm i\sqrt{2}$

Fig. 2. Phase portraits around the origin for different stability classes of linear dynamical system $\dot{\mathbf{x}} = A\mathbf{x}$. **(a)** Both $\lambda < 0$: stable fixed point. **(b)** Both $\lambda > 0$: unstable fixed point. **(c)** $\lambda_1 < 0 < \lambda_2$, stable along x, unstable along y: saddle point. **(d)** $\lambda \in \mathbb{C}$, $Re[\lambda] < 0$: stable spiral.

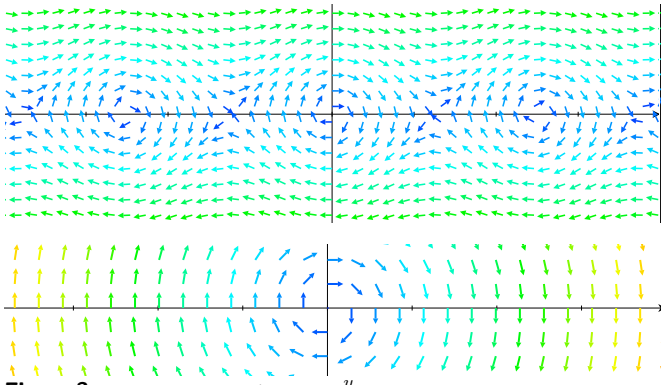


Fig. 3. Nonlinear system $\dot{\mathbf{x}} = \begin{matrix} y \\ -\sin x \end{matrix}$ above, with linear approximation $\dot{\mathbf{x}} = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)\mathbf{x}$ around origin below. The linear approximation accurately captures the stability details (clockwise spiral attracting fixed point) at the origin.

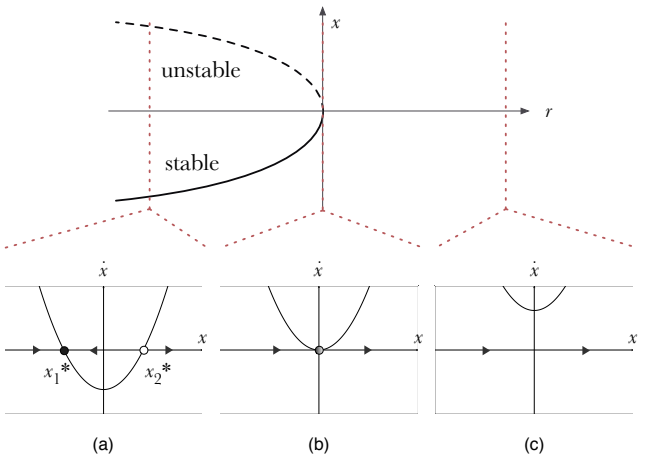


Fig. 4. Bifurcation diagram for the saddle-node bifurcation, $\dot{x} = r + x^2$, showing how the fixed points change in response to control parameter r. The solid line shows the position stable point; the dashed line, the unstable point. Below, the three distinct cases are demonstrated in more detail: $r < 0$ (a), $r = 0$ (b) and $r > 0$ (c) respectively.

being the unstable or stable fixed point, respectively) by considering the eigenvalues λ.

We often exclusively consider the eigenvalue with the largest absolute value λ_{max}, as this dominates as $t \to \infty$[26].

For real-valued eigenvalues, $\lambda_{max} > 0$ means the fixed point is unstable ($\mathbf{v}e^{\lambda t}$ will grow), and $\lambda_{max} < 0$ means the fixed point is stable ($\mathbf{v}e^{\lambda t}$ decays to $\mathbf{0}$). If some are positive and some are negative, this is a saddle point (stable in some directions, unstable in the rest). Complex eigenvalues lead to spiral points ($e^{i\theta} = \cos\theta + i\sin\theta$), and the stability is determined by the real component. Figure 2 shows phase portraits for the vector field $\dot{\mathbf{x}}$ for each of these cases, in two-dimensions.

We can use this approach to approximate a nonlinear system at a fixed point $\mathbf{x}^*$ and considering a first-order truncated Taylor expansion of the local gradient around the point, which requires the Jacobian matrix J, a matrix of local partial derivatives. The stability of the fixed point corresponds to the eigenvalues of J.

Figure 3 shows an example of linearisation for a nonlinear dynamical system.

Bifurcations. Often, we have dynamical systems with external parameters, and wish to consider how changes in these parameters qualitatively modify the phase portrait — such as the addition, removal or change in stability of fixed points. These discrete changes of phase space topology occur at *bifurcation points*.

A simple bifurcation family is the *saddle-node bifurcation*. In this case, the dynamical system is of the general form $\dot{x} = r + x^2$, with r as the control parameter. Figure 4 shows the effect of varying r: for $r > 0$ we have no fixed points, for $r = 0$ we have one semi-stable point (stable when approaching from the left) and for $r > 0$ we have an attractor at $-\sqrt{r}$ and a repeller at $\sqrt{r}$.

Of course real dynamical systems are not so straightforward, but simple algebraic idealisations of bifurcations, are referred to as *normal forms*, which strip unnecessary nonlinear terms from many common real bifurcation types, but still remain topologically equivalent.

The number of control parameter changes necessary to undergo a bifurcation is called the *codimension*. Many bifurcations which are not "truly" codimension-one can be considered as such, since their normal forms can be written as such.

It can often be helpful to picture the phase portrait as some energy function—with attracting fixed points as local minima. The *basin of attraction* of some fixed point $\mathbf{x}^*$ the set of surrounding points that converge to $\mathbf{x}^*$.

[1] Strictly speaking, we should talk about transitions between dynamic regimes, as we may transition into, for example, a stable orbit—however "state" is a commonly-used shorthand throughout the literature.

Earth as a dynamical system?

On an abstract level, environmental subsystems, or indeed the whole Earth can be considered as a complex dynamical systems—but the difficulty lies in accurately representing the real-world system accurately, without being overwhelmed by intractable detail. Thompson and Sieber (2011a) give the vital elements to consider in climate modelling as the atmosphere, ocean, land, ice and biosphere. Additionally, we consider some external forces which may have an impact, such as solar radiation. This is typically referred to as a *forcing* parameter.

The evolution rules are based on classical physics. In theory, for given starting conditions, a fully accurate model, with known deterministic forcing, would trace a single, unique trajectory through the phase space. However, complex processes are usually modelled stochastically. The most significant class of modelling error is in poor understanding of, or ignorance of subsystems—from faulty assumption or poor theory[9].

Even if we are not explicitly and exactly trying to model Earth processes, general theoretical findings from dynamical systems (particularly bifurcation theory) prove useful into gaining insight and information about such complex systems, as demonstrated with early warning signals.

However, not all catastrophes occur from a bifurcation controlled by some forcing parameter, and at least two other mechanisms have been proposed. Firstly, *noise-induced transitions*—where we experience some random, stochastic change in phosphorous concentration (P) large enough to drive the lake away from the stable oligotrophic point into the basin of attraction for the eutrophic state[28]. More recently, the method of *R-tipping* has been identified, where a stable point moves too quickly for the system to be decay to the point[1].

Tipping points and the environment

Many dynamical systems exhibit so-called *critical transitions*, where the system rapidly performs a discontinuous, nonlinear shift between states[1]. These are closely related to the concept of *tipping points*, which captures the idea of difficult-to-reverse changes—the tipping point is the cusp where the current stable state disappears, like a tightrope walker slowly leaning to one side until the centre of mass reaches a critical angle, causing them to fall.

In this report, we consider two case studies of environmental systems that undergo critical transition: abrupt climactic shifts, and sudden eutrophication in shallow lakes.

Abrupt climate change. Often the Earth rapidly shifts between climate states at a significantly faster rate than external forcing (slowly increasing solar activity)[28]. Abrupt climate change is of particular interest to researchers, as human-driven CO_2 exacerbates the greenhouse effect, and seems like it may drive the climate system beyond a tipping point in the future. Lenton (2008) captures this notion through *tipping elements*[17]—independent climate subsystems which are known to undergo abrupt regime changes. Additionally, to capture real-world political constraints retain a pragmatic focus on the most significant elements, they fulfil the following additional conditions:

1. There is empirical or theoretical evidence that there are external parameters which can be combined to a single control parameter α, and exceeding some $\alpha_{critical}$ causes a qualitative change in the phase portrait.
2. The control parameter α is at least partly controlled by human intervention, which can be changed in a reasonable *political time horizon*. i.e. avoid focussing on tipping points that seem unavoidable.
3. To realistically predict and respond to the change would take place within an *ethical time horizon*—catastrophe in the distant future is not likely influence contemporary policy.
4. There is significant interest in the effects of such changes, such as directly affecting human welfare or biodiversity.

A crucial tipping element identified by Lenton (2008) is the Atlantic Thermohaline Circulation (THC) system. As well as wind, tides and the Coriolis effect, ocean currents are also affected by a separate system of currents, driven by convection of differing densities of water masses. The density of water is primarily affected by two variables: heat (colder water is denser) and salinity (freshwater is lighter). The water heat is affected primarily by surface temperature (and to a lesser extent, underwater heat sources). Salinity can be increased by evaporation (also driven by surface temperature) and ice formation, and can be decreased by influxes of freshwater (such as from melted ice).

The circulation takes place underwater, and spans a large amount of the Earth's ocean system, and is largely driven by sinking of water masses at high latitudes. Much of the nonlinearity, and positive feedback come from the salinity—salinity in the deepwater formation regions encourages circulation, which in turn return more saline water to the region[22].

The wind-driven Gulf stream provides a relatively temperate climate to Western Europe, since the Atlantic THC "conveyor belt" carries warm water from equatorial regions—so disruptions to this (from, e.g. Greenland Ice-Sheet meltwater) could lead to regional cooling. Additionally, melting ice sheets may cause positive feedback, as the underlying land/water typically has lower albedo (reflectivity) than ice, leading to correspondingly higher rates of absorption of solar radiation and rising local air temperatures.

The THC system is well-understood from multiple models to have a specifically a fold bifurcation structure. Simulations and empirical records show bistability and hysteresis[28].

Lake eutrophication. As a second case study, we look at eutrophication in lakes. Typically, phosphates and other nutrients are the primary limiting growth factor for algae in a lake environment. Human agricultural interventions, such as phosphate-rich sewage or fertiliser runoff can spawn excessive algae growth. The decomposition of dead algae by bacteria uses oxygen, which is thus deprived from fish, shellfish and non-algal vegetation at the bottom of the lake[25].

The transition between being *oligotrophic* (low levels of biological activity) to *eutrophic* tends to be rapid, as the 'slow' variable of increasing phosphate concentration can combine

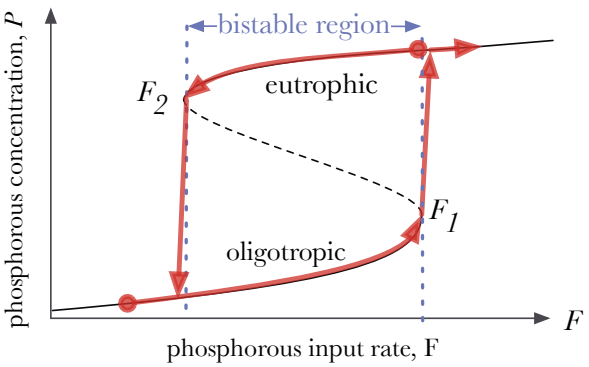


Fig. 5. Bifurcation diagram showing a caricature of the fold model. The axes are labelled with appropriate variables lake eutrophication. A path taken through a lake system as the forcing is varied is highlighted in red.

with 'fast' (sub-annual) variables such as water volume, and trigger the eutrophication process[30]. The clearest visual sign of eutrophication is the transparency of the lake, which tends to be affected little until the critical transition, when the lake suddenly becomes murkier.

Fold Bifurcation Model

Consider Figure 5, which is a simple bifurcation diagram for lake eutrophication. The system follows a steady oligotrophic state as we increase the rate of phosphorous entering the lake system, α, until the bifurcation point F_1, where this stable state disappears, and the system is rapidly driven towards the eutrophic state.

In the region $F_2 \leq \alpha \leq F_1$ the system could be in either stable state. This is known as a *bistable region*. This leads to the phenomenon of *hysteresis*, where the state of a system can depend on its previous path, leading to irreversibility: a small change in α can nudge the lake from oligotrophic to eutrophic, but a much larger reduction in α is then needed to reverse the lake back into an oligotrophic state.

Prediction through modelling

The majority of attempts to predict catastrophes in systems take place through explicit modelling of the processes and interactions, and tuning these models to empirical parameters. In this section, a high-level overview of some contemporary models and modelling techniques for global climate systems and lakes are offered.

Earth Climate Modelling. Modelling the global climate is a large area of research, especially due to the influence of the IPCC, and the crucial role that computational models play in policy-making. In this section, the ideas behind state-of-the-art modelling techniques for models featured in the latest IPCC report are broadly covered.

Climate process models are developed by firstly considering some mathematical idealisation of the system components' behaviour, from theoretical and empirical understanding. Following this, the system is discretised, typically as a 3-D grid, consisting of latitude, longitude and height. Some processes are too complex, or their behaviour is obscured by the coarseness of the grid, to be modelled in their entirety, so may be replaced with some intermediate conceptual model, referred to as a *parameterisation*, which are often used to model cloud cover or albedo. Simulating these models are complex, and the most significant bottleneck tends to be available computational power. Practically, this places upper bounds on the grid resolution—where higher resolution model is more accurate, although not necessarily more reliable. A typical model has a horizontal grid-size of 1° to 2° (latitude and longitude) and 30-40 vertical layers for the atmosphere. Additionally, there are also restrictions on which processes can be included. Often, the tradeoff is such that improving parameterisations, and adding more processes produces higher accuracy than improving resolution[10].

The old standard model class is the *Atmosphere-Ocean General Circulation Model* (AOGCM), which considers the dynamics of atmosphere, ocean, land and sea ice, with greenhouse gas and aerosol forcing. The current state of the art class is the *Earth System Model* (ESM), which extend AOGCMs by also considering biogeochemical cycles. Finally, some models fall under the class of *Earth system Models of Intermediate Complexity* (EMICs), which trade-off simulating all relevant components, at the expense of resolution—typically including processes which are absent from ESMs, such as ice sheets and ocean sediment[10].

The performance of climate models is determined by comparing simulation results with observed estimates, for dependent variables of interest, such as global temperature distributions. This typically involves *ensemble* results, which combine the results from multiple simulations. This could be from different research groups' models (*multi-model ensemble*), or on an individual model, with different parameters (*perturbed parameter ensemble*). While this reduces and quantifies the uncertainty of models, there are several limitations. Firstly, there is still bias in models, as researchers share many component resources, and the sample size of models is small. Additionally, there are difficulties in determining how to average the ensemble results. Techniques can range from a simple unweighted mean of individual results, to sophisticated Bayesian methods, and can have a significant effect on the end-result[10].

GENIE-2 is an example of a model which has been used to test predictions of tipping points—specifically the collapse of Atlantic THC. The existence of two stable regions (on/off), and hysteresis under reduction of freshwater forcing have been known from simple classical models, however the use of a full OAGCM can more thoroughly answer where the bifurcation point exists in the space of background model parameters, and determine the likelihood of collapse over an ensemble of forcing parameters. Under most scenarios, however, there tended to be disagreement between members of the ensemble. When testing an idealised CO_2 trajectory, some collapse reversibly, and some do not collapse at all. Whilst testing a mid-range estimate of 3° C warming, all members showed weakening of THC but the majority did not collapse. By incorporating extra sources of freshwater (such as Greenland ice sheets), a greater proportion collapsed, but was still far from definitive. The authors argue that better, higher resolution historic time series are needed, to constrain the variance of background parameters under consideration, which should lead to higher agreement amongst ensembles, and therefore more useful predictions[18].

Lake Modelling. The de facto lake models studying eutrophication are those developed by Carpenter[7, 6]. The first is a simple model[7] for the concentration of phosphorous, P:

$$\frac{dP}{dt} = l - sP + r\frac{P^q}{m^q + P^q} \qquad [1]$$

where l is phosphorous input from the lake's watershed; s is the rate of phosphorous loss to sediment/biomass sequestration or flowing into another water system; and the final term describes the recycling process—with maximum recycling rate r; m as a calibration term for the amount of concentration of phosphorous at the half-maximum recycling rate; and q as a steepness parameter, primarily determined by the depth of the lake. This model was later extended[6] to include the processes involving the internal forcing dynamics of the outflow of sequestered phosphorous from the lake's sediment, believed to be the significant factor for the hysteresis observed in eutrophication—however, the model is still in a simple form that allows symbolic analysis, unlike most contemporary climate models.

Carpenter (2005) estimates the tipping point for a specific example (Lake Mendota) is predicted, by using parameter estimates (from empirical data), and analytically determining the fixed points (roots) of the explicit dynamical system equations. This produces more precise answers than more complex models, like OAGCMs, but may masks much of the underlying uncertainty—and there is no evidence to suggest whether

the prediction of fixed points is accurate. There is reason to believe the model is too simple, as it fails to incorporate identified complexities in lake systems such as: complex population dynamics, dependence on short-term weather patterns and the topology of lake components (e.g. if shallow parts are connected to deeper parts)[25].

Generic early warning signals

By taking a dynamical systems perspective, we can look at the fold bifurcation as an idealisation of catastrophic bifurcations, and determine general properties about systems as they approach these bifurcation points.

Critical Slowing Down. As systems tend towards the bifurcation point, they often have a tendency to "slow down", meaning they take longer to return to their stable state when exposed to small perturbations. This can be viewed as the basin of attraction of the fixed point becoming shallower, as the system comes closer to the bifurcation point[29].

Ideally, we would determine how close a system is to the bifurcation point by subjecting the system to perturbations and measuring the response time to return to the stable state—however intervention is usually infeasible. Usually the driving forces tend to be more difficult to reverse (the problem in the first place), and we do not wish to risk perturbing too much, and tipping over. We can, however, observe indicators of critical slowing down from time series data, as the systems are typically already being subjected to natural stochastic perturbations.

Slowing down should correspond to a change in various computable metrics, which can then be used as indicator variables. These variables tend to be valid for any dynamical system whose tipping point sufficiently approximates the fold bifurcation, thus providing a class of early warning signals for a wide variety of dynamical systems.

Critical slowing down also offers some other helpful properties [29]. Firstly, the slowing down, when tested on models, tends to begin significantly before the threshold point is reached. Secondly, critical slowing down can indicate non-stable state tipping points, such as an oscillatory dynamic regime. Finally, many signals have shown a tendency to change at a linear rate when approaching the tipping point, which aids in prediction accuracy[17].

Formally, as the system tends closer to the bifurcation point, the dominant eigenvalue of the Jacobian tends to zero. This is demonstrated mathematically for the normal-form saddle-node bifurcation, as defined in Section 2:

The saddle-node bifurcation of the form $\frac{dx}{dt} = f(x)$ with $f(x) = x^2 - r$ has a stable point $x_1^* = -\sqrt{r}$ for $r \geq 0$, and an unstable point at $x_2^* = \sqrt{r}$ The size of the basin of attraction between the stable point and the unstable point is $B = 2\sqrt{r}$. We can consider the response to a perturbation of ε from the stable point by Taylor expansion around x_1^*:

$$\begin{aligned} \frac{d}{dt}\left[x_1^* + \varepsilon\right] &= f(x_1^* + \varepsilon) \simeq f(x_1^*) + \left.\frac{\partial f}{\partial x}\right|_{x=x_1^*} \varepsilon \\ &= f(x_1^*) - 2\varepsilon\sqrt{r} \\ &= f(x_1^*) + \lambda\varepsilon, \text{ such that } \lambda = -B. \end{aligned} \quad [2]$$

We define $\lambda \leq 0$ to encapsulate the *linear decay rate* (LDR). Notice that at the critical threshold, $r = 0$, that the recovery rate is 0, and scales linearly with the size of the basin, B. These results (including linearity) are common to many bifurcations[28].

Flickering. To detect critical slowing down, we need time series data of high enough resolution to capture the return from natural perturbations. If the sampling rate is lower than the average return time, then we cannot expect to notice the slowing down.

Another phenomenon that might be observed near to a tipping point, is 'flickering'—where the system begins to oscillate between multiple stable state basins. Flickering can also be a useful indicator when we expecting high levels of stochastic noise in our measurements[30] (which may otherwise obscure critical slowing down).

Early Warning Indicators

Some of the most commonly used metrics, on discrete univariate time series are detailed below. In these examples, we denote $Y_1, \ldots, Y_N$ as the measurements of the dependent variable at times $X_1, \cdots, X_N$. $\bar{Y}$ denotes the empirical mean over the time series.

Often, we consider a fixed-size sliding window: for a window length w, the measure at point $i \leq N$ is computed over the time series $Y_{i-w}, \cdots, Y_i$. This is typically done to avoid potential bias from a changing time series size, although then the window length w is left as a parameter to be selected—this is a tradeoff between accuracy (longer series produce more robust estimates) and having enough points to determine a trend. A common heuristic is to use $w = N/2$.

Autocorrelation. When slowing down, the state at each time step becomes progressively more dependent on the value at the previous step—i.e. the correlation between time-steps increases. Autocorrelation measures the correlation of a time series signal with itself, separated by a time lag k. Figure 6 shows some of the intuition behind this: the system state at time t is much more similar to the state at $t + 1$ when the potential surrounding the fixed point is shallower[13].

Calculating the autocorrelation is typically done by fitting an order-K autoregressive model, denoted $AR(K)$. This is a predictive linear model, which predicts Y_{t+1} from a linear combination of the previous K values $(Y_t, Y_{t-1}, \cdots, Y_{t-K})$.

Formally, an $AR(K)$ model looks like:

$$Y_{t+1} = \sum_{k=1}^{K} a_k Y_{t-k} + \eta_{t+1} \quad [3]$$

where $\mathbf{a} = (a_1 \cdots a_K)^T$ are the prediction coefficients to be fitted, and $\eta_{t+1} \sim \mathcal{N}(\mu, \sigma^2)$ is a Gaussian noise term.

This model differs from standard linear regression, as it does not assume that the values of Y are i.i.d. (identically and independently distributed), and the order of data points matters—there are dependencies between time-separated values. The AR model does, however, assume that the time-series Y is *weakly stationary*, which means it satisfies the following three properties across the whole series:

1. Mean $\mathbb{E}(Y_t) = 0$.
2. Variance is constant: $Var(Y_t) = Var(Y_{t+1})$.
3. Covariance/correlation between Y_t and Y_{t-k} is constant across the whole series.

We are primarily interested in the sample autocorrelation (ACF), which gives the correlations between Y_t and Y_{t-k}. For an $AR(1)$ model, of the form $Y_{t+1} = a_1 Y_t$, the lag-1 autocorrelation is simply a_1.

Thompson and Sieber (2011a) define a standard approach:

1. **Interpolation**. Most time series data are not spaced equally. The time series needs to be interpolated (usually

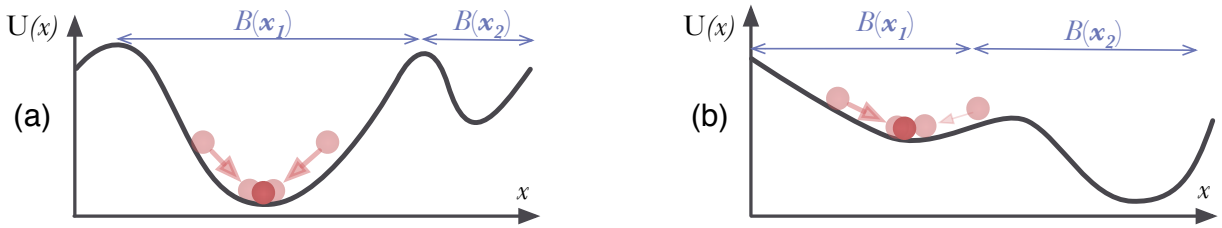


Fig. 6. This diagram shows the system as a potential landscape, with attracting (stable) fixed points x_1 (initial state) and x_2 (alternative state) with associated basins denoted B. The qualitative topology of the landscape changes as we approach a bifurcation point, which can be used to explain the phenomena of critical slowing down. In **(a)**, the system is far from transition, and the basin of the stable state has steep walls—so we return to the fixed point rapidly, following perturbations. In **(b)**, we are closer to the bifurcation point, and the basin of attraction becomes shallower in an asymmetric manner. Based on diagram from `www.early-warning-signals.com`.

linearly), and the interpolation is sampled at equidistant points.
2. **Detrending**. The time series is not necessarily stationary (as the fixed points may slowly drift through other forces). A moving average at each point $\bar{Y}_t$ can be computed, often using a Gaussian kernel approach. Subtracting this moving average gives a new time series $Z'_t = Y_t - \bar{Y}_t$ with a mean of zero..
3. **Moving Window**. The time series within this window should be fit to an $AR(1)$ model. In this case, a_1 is related to the LDR, α, as $\alpha = \exp(a_1 \Delta t)$.

Having fitted the AR model to the empirical data, we can use this model predict the tipping point, by extrapolating a_1 from each window, to find the t where $a_1 = 1$.

We can show that critical slowing down leads to increased lag-1 autocorrelation. Firstly, without loss of generality, we assume that the stochastic disturbance $\eta \sim \mathcal{N}(0, \sigma^2)$ from the stable point x^* occurs at at some regularly spaced intervals of Δt. The system returns to $x*$, between the perturbations with an LDR of λ. The discretised evolution rule (for $t = 0, \Delta t, 2\Delta t, \dots$) on the recovery of x can be approximated as:

$$x_{t+1} - x^* = e^{\lambda \Delta t}(x_n - x^*) + \eta_{t+1} \qquad [4]$$

i.e. the deviation from the stable point at time step $t+1$ is equal to the amount by which the system has recovered from the previous disturbance, and the next stochastic forcing. By centring around the stable point ($Y = x - x^*$), we can write this as an $AR(1)$ process:

$$Y_{t+1} = a_1 Y_t + \eta_{t+1}, \quad a_1 = e^{\lambda \Delta t} \qquad [5]$$

Critical slowing down shows that $\lambda \to 0$ as we approach the critical point, so the lag-1 ACF coefficient $a_1 \to 1$.

As a note, the disturbances modelled need not actually be identically and independently distributed (i.i.d.). The implicit assumptions made above involve the relative distinction of the time scales that certain dynamical processes operate on. The timescale of stochastic noise (τ_η) events needs to be faster than the timescale of the return rate (τ_λ), so that the compound effect of many individual disturbances can approximate a normal distribution (by central limit theorem). Additionally, we only observe non-zero autocorrelation when the system has not returned to x^* after the interval $\Delta t \sim \tau_\eta$.

Detrended Fluctuation Analysis (DFA). One significant limitation of autocorrelation, is the need for a relatively long time series to obtain a robust autocorrelation estimate. Additionally, autocorrelation struggles when analysing time series data produced by a process with *long-term memory*. DFA originated in the bioinformatics literature, where it was developed to detect long-range correlation in DNA sequences. Here the assumption is that the fractal has some form of fractal structure, with self-similarity preventing the existence of a characteristic scale [21, 20]. Formally, a fractal time-series $Y = Y_1 \dots Y_N$ satisfies

$$Y(t) \equiv a^\alpha Y(t/a) \qquad [6]$$

where $\equiv$ means that the first and second-order statistical properties are identical, and α is the self-similarity parameter. Under this relation, a self-similar process with $\alpha > 0$ would have to exponentially grow in amplitude, however we are primarily interested in finding self-similarity in time series data with physically bounded amplitude. The "trick" employed by DFA is to consider the integral of the time series, which is unbounded. The algorithm employs three steps:

1. **Integration**. This achieves an unbounded series $Y'(k) = \sum_{i=1}^{k} [Y_i - \bar{Y}]$, where $\bar{Y}$ is the empirical mean of Y, for all $k = 1..N$ points in Y.
2. **Detrending**. Firstly, Y' is split into equally-sized windows with n points. In each section, the local trend is estimated by fitting a line, with Y-coordinate of any point k denoted as $Y'_n(k)$.
3. **Determining characteristic scale**. The root-mean-square fluctuation from the trend-lines is calculated over all windows:

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} [Y'(k) - Y'_n(k)]}$$

This is repeated for different window sizes, and we can determine α by plotting n against $F(n)$ on a log-log graph—as $F(n) \propto n^\alpha$ (the fluctuations in small boxes should share the same statistical properties as fluctuations in large boxes).

A result of $0.5 < \alpha < 1.0$ means that long-range temporal correlations exist, so we would expect a process with long-term memory that is undergoing critical slowing to have a DFA exponent α that tends to 1.

Variance. Close to a bifurcation point, the system tends to move more widely, as the basin becomes shallower. Additionally, the response variable will move further under perturbations, so increased variance should correspond with approaching a bifurcation point[15, 30].

Formally, the variance over a time series is the expected value of the square of the deviation from the mean:

$$Var(Y) = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \bar{Y})^2 \qquad [7]$$

Additionally, increased variance can serve as an indicator of flickering, as the system oscillates between multiple stable

states[30]. It has been shown in eutrophication models with that variance can increase, despite noise heavy enough to show decreasing lag-1 autocorrelation and skewness.

The increase in variance can be shown formally, based on the $AR(1)$ model in Equation 3. By the assumption of stationarity, we can assume $Y_{t+1} = Y_t$.

$$\begin{aligned} Var(Y_{t+1}) &= Var(a_1 Y_t + \eta_{t+1}) \\ &= a_1^2 Var(Y_t) + Var(\eta_{t+1}) \\ &= a_1^2 Var(Y_{t+1}) + \sigma^2 \\ &= \frac{\sigma^2}{1 - a_1} \end{aligned} \quad [8]$$

As $\lambda \to \infty$, $a_1 \to 0$, so the variance increases towards infinity.

Skewness. This is a measure of asymmetry about the mean. For a discrete time series, the sample skew is:

$$Skew(Y) = \frac{\frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^3}{(\frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^2)^{\frac{3}{2}}} \quad [9]$$

For intuition, Figure 6a shows that far from the transition, the response to perturbations will tend to be symmetrical—if the attractor state has symmetrical potential gradients, and the stochastic noise is symmetric, and less than the size of the attractor basin. However, as we get closer to the saddle between seats, we see an asymmetry: disturbances towards the saddle will have a slower response (lower potential gradient) than disturbances away from the saddle—so we expect to see more time spent recovering from perturbations towards the saddle, assuming symmetric stochastic noise.

Empirical Results Summary

As a brief review of published results, there are numerous findings showing increased autocorrelation associated with THC tipping in models[13, 17] (`CLIMBER-2`, `GENIE-1`) and paleo-records (increasing ACF leading to 8/8 abrupt climate events)[8]. Additionally, DFA has been tested for THC collapse in models[20, 17] (`GENIE-1`, `GENIE-2`). These approaches all use the preparation conventions outlined earlier.

A less naïve analysis is performed by Lenton et al. 2012[19], across three ice-core records (`Vostok`, `GISP2` and `Cariaco`) and three models (`CLIMBER-1`, `GENIE-1` and `GENIE-2`)—with a distribution of early warning signals for differing window lengths and detrending Gaussian filter bandwidths. The distribution of Kendall's τ[2] for each is estimated, giving a broader picture of the resilience of each metric. Compared to ACF, DFA showed weaker signals (mean of τ closer to 0), but was less sensitive to the parameters than ACF (smaller s.d. of τ). Additionally, there was disagreement over variance between the empirical datasets, with 2/3 predicting *decreasing* variance.

Lake systems offers a partial reversal to expectations. The Carpenter model shows increased ACF and skew while approaching eutrophication[11] as expected, but simpler models and measurements from Lake Erhai show *decreasing* autocorrelation and skewness[30]. Variance, however, offers a more robust signal, increasing in all cases[11, 30].

Generic Indicator Limitations

An obvious limitation of early warning signals is their inability to detect what type of bifurcation is being approached, as well as the corresponding dynamical regime after the transition.

Assumptions. Firstly, Boettiger and Hastings (2012b) note two implicit assumptions. Firstly, that the system is approaching a saddle-node bifurcation. Other bifurcation types can exhibit critical slowing, without a tipping point, such as the transcritical pitchfork. Secondly, that the system is being brought to the critical point by a gradual, monotonic change in some forcing parameter. The authors also note other possibilities for regime changes: a large, discontinuous perturbation, stochastic noise-induced transitions or a bifurcation parameter that varies in a highly rapid/nonlinear way[3]. The saddle-node is likely to be a reasonable assumption in the case of Atlantic THC and lake eutrophication, as these have theoretically and empirically demonstrated behaviour approximating the fold bifurcation (well approximated by a saddle-node at the ends of the bistable region).

Additionally, Thompson and Sieber (2011a) note that there is no *a priori* reason to assume distinct timescales[2].

False Negatives. Both noise-induced transitions and R-tipping are two transitions that are not likely to show critical slowing, so indicators would fail to predict a critical transition.

If the noise level is relatively low, we would need to be close to the tipping point before undergoing a noisy transition, so these indicators retain a chance of still being useful. Thompson and Sieber (2011b) have attempted to derive a probabilistic model, to predict the likelihood of a noise-induced transition with respect to a bifurcation parameter drift.

Tests on THC collapse on the GENIE-2 AOGCM, which has a high level of stochastic noise (weather fluctuations) do show qualitatively increasing ACF and DFA values, in spite of the heavy noise[18].

False Positives. Boettiger and Hastings (2012b) identify that many published results fail to consider the "prosecutor's fallacy"[2]. Most analyses of early-warning signals (EWS) show an increase in the EWS ($\mathcal{E}$) in some time series selected because a known transition has occurred ($\mathcal{T}$), and is presented as evidence that the signal is powerful. To formalise further, this is usually argued to be evidence that $P(\mathcal{T}|\mathcal{E})$ is high—whereas it is actually only showing that $P(\mathcal{E}|\mathcal{T})$ is high. These two probabilities are not necessarily the same, particularly if there are alternative hypotheses for transition (e.g. noise-induced transition) or alternative causes for changes in EWS values (R-tipping seems likely to increase autocorrelation).

The authors demonstrate that a random sample of windows in replicates of a population dynamics system (not conditioned on tipping) showed a statistically significant skew towards increasing of autocorrelation and variance. This demonstrates that early warning signals are likely to exhibit a high rate of false positives.

Furthermore, Kéfi et al. (2012) demonstrate that signals based critical slowing down merely indicates an imminent regime change—but there is no reason to assume this will be catastrophic. Critical slowing may also precede, for example, a transcritical or Hopf bifurcation.

Hypothesis Modelling. Many of these issues can overcome through considering EWS detection as a probabilistic model selection problem, where the distinct model parameters are estimated from the data, and the assumptions are explicitly encoded in the model[3]. One model corresponds to the hypothesis that the time series is produced by a system approaching a

[2] A common rank correlation statistic, producing a value in the range $-1 < \tau < 1$. The coefficient is 1 if the time series monotonically increases (the series is already perfectly ordered in terms of rank).

saddle-node bifurcation, and the other (null) hypothesis that the data is produced by a system not approaching a bifurcation.

Firstly, if neither model matches well, this gives strong evidence that the system violates modelling assumptions. Secondly, we can go further than a simple maximum likelihood estimate of parameters, and train the models on multiple simulations/repeated measurements. This way, we can quantify the tradeoff between false positive/false negative through an ROC (Receiving Operator Curve) analysis, taking into account a range of threshold values between both models' posterior distributions. In some of the examined models (Daphnia population dynamics and glaciation), Boettiger and Hastings (2012b) demonstrate that the autocorrelation and variance overlap heavily between the models, implying they are not universally robust EWSs. Additionally, they show that the distribution of Kendall's τ on autocorrelation and variance series in multiple empirical datasets has a positive skew (towards 1), implying that autocorrelation and variance can have a strong tendency to increase during normal circumstances (no catastrophe approaching)[3].

Conclusions and Future Work

Whilst both process-based modelling and early warning signals are often used to predict catastrophes, there are crucial differences between their scope.

Models tend to give much richer information—such as full-time series projections and information about the bifurcation, including the potential impact of crossing a tipping point. An early-warning signal merely indicates that we may be approaching *some* regime change, which is significantly less useful. Additionally, model-based approaches allow the design of counterfactual experiments, and simulate the results of a wide range of policies.

Furthermore, process-based modelling has a stronger link with theory—we tend to have a much better understanding of where biases and inaccuracies lie. On the other hand, if an early-warning signal does not predict well, then the cause of failure is significantly less obvious, and it is not clear we can simply create better early warning signals to catch these missed phenomena. Similarly, our confidence in our predictions, with multi-model ensembles, whereas it is unclear how robust early warning signals are to being combined.

Early warning signals seem more appropriate as non-critical indicators for small-scale systems, like lakes, where the resources do not exist to build full models, as they are cheaper to compute, avoid much of the difficulty in building accurate models.

In general, many of the significant outstanding issues concern the much more complex social layer, which plays an integral role in determining the likelihood of events, considering multiple policies to prevent catastrophes, and an appropriate harm cost for the negative impacts of a catastrophe.

ACKNOWLEDGMENTS. Thankyou to my supervisor, Dr. James Dyke, for his guidance and feedback.

1. Peter Ashwin, Sebastian Wieczorek, Renato Vitolo, and Peter Cox. Tipping points in open systems: bifurcation, noise-induced and rate-dependent examples in the climate system. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 370(1962):1166–1184, 2012.
2. Carl Boettiger and Alan Hastings. Early warning signals and the prosecutor's fallacy. Proceedings of the Royal Society B: Biological Sciences, 279(1748):4734–4739, 2012.
3. Carl Boettiger and Alan Hastings. Quantifying limits to detection of early warning for critical transitions. Journal of The Royal Society Interface, 9(75):2527–2539, 2012.
4. SR Carpenter and WA Brock. Rising variance: a leading indicator of ecological transition. Ecology letters, 9(3):311–318, 2006.
5. SR Carpenter, WA Brock, JJ Cole, JF Kitchell, and ML Pace. Leading indicators of trophic cascades. Ecology Letters, 11(2):128–138, 2008.
6. Stephen R Carpenter. Eutrophication of aquatic ecosystems: bistability and soil phosphorus. Proceedings of the National Academy of Sciences of the United States of America, 102(29):10002–10005, 2005.
7. Stephen R Carpenter, Donald Ludwig, and William A Brock. Management of eutrophication for lakes subject to potentially irreversible change. Ecological applications, 9(3):751–771, 1999.
8. Vasilis Dakos, Marten Scheffer, Egbert H van Nes, Victor Brovkin, Vladimir Petoukhov, and Hermann Held. Slowing down as an early warning signal for abrupt climate change. Proceedings of the National Academy of Sciences, 105(38):14308–14312, 2008.
9. Peter D Ditlevsen and Sigfus J Johnsen. Tipping points: early warning and wishful thinking. Geophysical Research Letters, 37(19), 2010.
10. G Flato, J Marotzke, B Abiodun, P Braconnot, SC Chou, W Collins, P Cox, F Driouech, S Emori, V Eyring, et al. Evaluation of climate models. Contribution of Working Group I to the fifth assessment report of the intergovernmental panel on climate change, 2013.
11. Vishwesha Guttal, C Jayaprakash, and Omar P Tabbaa. Robustness of early warning signals of regime shifts in time-delayed ecological models. Theoretical Ecology, 6(3):271–283, 2013.
12. Alan Hastings and Derin B Wysham. Regime shifts in ecological systems can occur with no warning. Ecology letters, 13(4):464–472, 2010.
13. Hermann Held and Thomas Kleinen. Detection of climate system bifurcations by degenerate fingerprinting. Geophysical Research Letters, 31(23), 2004.
14. Sonia Kéfi, Vasilis Dakos, Marten Scheffer, Egbert H Van Nes, and Max Rietkerk. Early warning signals also precede non-catastrophic transitions. Oikos, 122(5):641–648, 2013.
15. Steven J Lade and Thilo Gross. Early warning signals for critical transitions: a generalized modeling approach. PLoS computational biology, 8(2):e1002360, 2012.
16. Timothy M Lenton. Early warning of climate tipping points. Nature Climate Change, 1(4):201–209, 2011.
17. Timothy M Lenton, Hermann Held, Elmar Kriegler, Jim W Hall, Wolfgang Lucht, Stefan Rahmstorf, and Hans Joachim Schellnhuber. Tipping elements in the earth's climate system. Proceedings of the National Academy of Sciences, 105(6):1786–1793, 2008.
18. Timothy M Lenton, Richard J Myerscough, Robert Marsh, Valerie N Livina, Andrew R Price, and Simon J Cox. Using genie to study a tipping point in the climate system. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 367(1890):871–884, 2009.
19. TM Lenton, VN Livina, V Dakos, EH Van Nes, and M Scheffer. Early warning of climate tipping points from critical slowing down: comparing methods to improve robustness. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 370(1962):1185–1204, 2012.
20. Valerie N Livina and Timothy M Lenton. A modified method for detecting incipient bifurcations in a dynamical system. Geophysical Research Letters, 34(3), 2007.
21. Margaret and H.A. Rey Institute for Nonlinear Dynamics in Medicine. Fractal analysis methods tutorial. http://reylab.bidmc.harvard.edu/tutorial/DFA/, 2014. Accessed 05-2014.
22. S Rahmstorf. Thermohaline ocean circulation. encyclopedia of quaternary sciences. Postdam Institute for Climate Impact Research, 2006.
23. Marten Scheffer, Jordi Bascompte, William A Brock, Victor Brovkin, Stephen R Carpenter, Vasilis Dakos, Hermann Held, Egbert H Van Nes, Max Rietkerk, and George Sugihara. Early-warning signals for critical transitions. Nature, 461(7260):53–59, 2009.
24. Marten Scheffer, Steve Carpenter, Jonathan A Foley, Carl Folke, and Brian Walker. Catastrophic shifts in ecosystems. Nature, 413(6856):591–596, 2001.
25. Marten Scheffer and Egbert H van Nes. Shallow lakes theory revisited: various alternative regimes driven by climate, nutrients, depth and lake size. Hydrobiologia, 584(1):455–466, 2007.
26. Steven H Strogatz. Nonlinear dynamics and chaos (with applications to physics, biology, chemistry a. Perseus Publishing, 2006.
27. J Michael T Thompson and Jan Sieber. Climate tipping as a noisy bifurcation: a predictive technique. IMA Journal of Applied Mathematics, 76(1):27–46, 2011.
28. J Michael T Thompson and Jan Sieber. Predicting climate tipping as a noisy bifurcation: a review. International Journal of Bifurcation and Chaos, 21(02):399–423, 2011.
29. Egbert H Van Nes and Marten Scheffer. Slow recovery from perturbations as a generic indicator of a nearby catastrophic shift. The American Naturalist, 169(6):738–747, 2007.
30. Rong Wang, John A Dearing, Peter G Langdon, Enlou Zhang, Xiangdong Yang, Vasilis Dakos, and Marten Scheffer. Flickering gives early warning signals of a critical transition to a eutrophic lake state. Nature, 492(7429):419–422, 2012.