

# Quantifying Limits to Detection of Early Warning for Critical Transitions

Carl Boettiger<sup>a,\*</sup>, Alan M Hastings<sup>a</sup>

<sup>a</sup>*Center for Population Biology, University of California, Davis, United States*

---

## Appendix A. The data sets

All data sets and analyses performed are included in the accompanying R package. The simulated data sets in Figure 1a of the text are produced with 100 data points sampled evenly sampled over a time interval (0,100) using the linearized models directly, see Appendix D. The deteriorating environment was produced under the LSN model (D.3), with  $\theta = 500$ ,  $\sigma = 5$ ,  $R(t) = 5 - 0.04999t$ . The constant data is produced using the simulation method for the constant environment model (D.11), with parameters estimated from the deteriorating data-set to correspond as nearly as possible to a similar system. The Glaciation data comes from Petit et al. (1999), accessible from NOAA: <http://www.ncdc.noaa.gov/paleo/metadata/noaa-icecore-2453.html>. The data is preprocessed by interpolation and detrending to be as consistent as possible with the analysis presented in Dakos et al. (2008), analyzing the third glaciation event. The processed data consists of 121 sample points. The match is not exact since Dakos et al. (2008) estimates the de-trending window size manually, but the estimated correlations in the first-order auto-regression coefficients are in close agreement with that analysis. The code for processing the data from the original Petit et al. (1999) file is included in the accompanying R package. The Algae example comes from the chemostat “H6” in the experiments in *Daphnia magna* populations of Drake and Griffen (2010). The analysis in that paper focuses on data averaged over the ensembles, though the supplement includes analysis over individual replicates. This individual replicates was chosen as an example of a single replicate that showed a statistically significant correlation in variance over window of time where critical slowing down was

---

\*Corresponding author.

Email address: [cboettig@ucdavis.edu](mailto:cboettig@ucdavis.edu) (Carl Boettiger)

25 expected. Further empirical examples that have been previously investigated for early warning  
 26 signs are also analyzed under our approach in [Appendix K](#).

## 27 **Appendix B. Averaging over ensembles vs averaging over time**

The approach of detecting early warning signals by summary statistics is motivated by the ensemble dynamics of a stochastic system. If we had an infinite number of identical, replicate systems, we could average across replicates to obtain asymptotic estimates of the (for instance) the variance and autocorrelation of the stochastic process  $X_t$  at each time  $t$ ,

$$V(X_t) = \int X_t^2 P(X_t) dX_t - \left( \int X_t P(X_t) dX_t \right)^2 \quad (\text{B.1})$$

$$\text{Autocorr}(X_\tau) = \int (X_t - X_{t-\tau})^2 P(X_t) dX_t \quad (\text{B.2})$$

28 and so forth for higher moments. All of the integrals are over replicates,  $X_t$ . To apply the theory  
 29 to a single replicate and instead compute the statistics over time windows one must assert that  
 30 the system is ergodic –roughly, independent of the time over which it is measured ([Gardiner,](#)  
 31 [2009](#)). This cannot be true if the system is experiencing a gradual loss of stability leading to a  
 32 critical transition – but may hold approximately if the change is gradual relative to the window,  
 33 and the window is long relative to the relaxation time of the system. Then if the moments are  
 34 calculated across the ensemble at sufficiently wide sampling intervals, one may argue that they  
 35 are approximately independent. Only under these assumptions can we apply the approach of  
 36 summary statistics to a single replicate. Even then, we would need techniques from time-series  
 37 analysis test, since we have not satisfied the assumptions of the standard correlations tests.  
 38 Much of the same data are used to compute sequential values of the moments, hence sequential  
 39 values are temporally correlated.

40 This lack of independence violates the assumptions under which the  $p$ -values associated with  
 41 typical correlation tests are calculated. For instance the rank correlations such as Kendall’s  $\tau$   
 42 test a null hypothesis that increases in the  $V(X_t)$ , etc. are independent of increases in time,  
 43 which is not satisfied even when the environment is constant and  $X_t$  is stationary. Calculating  
 44  $p$ -values for Pearson’s test needs parametric assumptions such as normal distributed variables as  
 45 well as independence to define the null distribution. As both the stationary and the deteriorating  
 46 environment violate the assumptions of the null, with sufficient data both appear significant.

Consequently the  $p$ -values seen in Figure 1a and common in the current approaches cannot be interpreted as the probability a warning signal would be observed by chance in a stable system.

It is possible to define a  $p$ -value for these correlation statistics by estimating a null distribution consistent with the null hypothesis, as we do through the model-based Monte Carlo estimation of these distributions in Figure 1b of the text, (Type I error).

## Appendix C. Error Types

The  $p$ -values given in Figure 1a of the text do not correspond to the probabilities that the measured value of the statistical test (Kendall's  $\tau$ ) or greater would be observed under a stable system (previous section, [Appendix B](#)). The approach illustrated in Figure 1b estimates the appropriate null distribution of  $\tau$  for each warning statistic (variance, autocorrelation, etc.) under the null hypothesis that the system is stable, following the assumption of the stable model defined in the text.<sup>1</sup>

The appropriate  $p$ -value for the test statistic can be determined directly from this distribution by integrating to the right of the observed value. This determines the probability of a false alarm, Type I error. We are fortunate to be able to estimate the distribution of the test statistic under the alternative hypothesis that the system is stable (red distributions), and hence can also estimate the power of the test, defined at the accepted false-positive rate. Having accepted a false-positive rate determines a position beyond which only less than that fraction of the null distribution lies. We can ask what fraction of the test distribution lies beyond this fraction as well. If the false-positive rate is set very low, we lose power. The fraction of the test distribution lying to the other side of this accepted rate (1 minus the power) is the probability of a missed event, Type II error. There is an obvious trade-off between the threshold set for type I error and the resulting Type II error.

Note that while the Type I error is dependent directly on the value observed in the data (black triangles of Figure 1b & 1c), the Type II error depends only on the distributions themselves

---

<sup>1</sup> While other stable models are possible, this model corresponds to the underlying bifurcation theory behind the design of early warning signals. Many systems may have dynamics that fall entirely outside the scope of early warning signals [Hastings and Wysham \(2010\)](#). Handling such systems as may have dynamics that correspond poorly to any of the models treated here can be addressed by tests of model adequacy, see [Appendix I](#).

72 (though it should be remembered that these are also estimated from the data by means of the  
73 model fitting).

74 While it is customary in the scientific literature to accept 5% false-alarm rate, we emphasize  
75 here that this is properly a choice set by management, under consideration of the relative costs  
76 of false alarms and missed events. In particular, having an estimate of power is very useful in  
77 selecting this rate – if the probability of a missed event is quite high at a 5% false alarm rate, it  
78 may make sense to use a less conservative estimate of when to act. We further note that the best  
79 management decisions will be based on the probabilities of both outcomes, rather than forcing  
80 a binary decision (Brozović and Schlenker, 2011). For this reason we emphasize reporting the  
81 entire distributions in Figure 1b & 1c, rather than only summary information about the error  
82 rates.

## 83 Appendix D. Model Derivations

84 The sudden collapses we wish to detect are driven by bifurcations when an eigenvalue changes  
85 sign in the corresponding deterministic model, resulting in the sudden loss of a stable state and  
86 the onset of a critical transition. There are many possible models that can contain such a  
87 transition. These may be highly nonlinear models with multiple attractors. The diversity of  
88 possible models can be classified by the type of bifurcation, such as the saddle-node bifurcation in  
89 which a stable point (node) and unstable point (saddle) collide and annihilate, or a trans-critical  
90 bifurcation, in which a stable point becomes unstable. (Any dynamical systems textbook, such  
91 as (Guckenheimer and Holmes, 1983) can give a good overview of such models and bifurcations).  
92 In general we will be unable to estimate such models accurately if the system has only been  
93 observed near the favorable attractor (i.e. before the critical transition occurs), even if we knew  
94 the functional form of the model. Fortunately the detection of early warning is based on a *local*  
95 theory, predicting changes in the dynamics near the current attractor. Consequently, it is as  
96 unnecessary as it is impossible to estimate a fully nonlinear model, but only the linearized model  
97 near the bifurcation.

98 Near the region of the bifurcation, we can use the simple normal form equation for the  
99 bifurcation rather than the full model, but these are still non-linear and difficult to estimate  
100 accurately from the data. Critical slowing down is based on taking this one step further –

101 looking at the eigenvalue around the stable point. This is equivalent to linearizing the normal  
 102 forms of the bifurcations around the stable point. By so doing we obtain a simple but general  
 103 models we can estimate from data. We are left with explicit, stochastic models that reflect the  
 104 behavior of the more complicated systems around the stable point, and we can see how they  
 105 change as they approach the bifurcation.

#### 106 *Appendix D.1. Saddle-node bifurcation*

107 The most frequently considered bifurcation model in early warning signals is the saddle-node  
 108 bifurcation. In this bifurcation, a gradual decline is interrupted by a transition when a stable  
 109 node and saddle node collide and annihilate, leaving the system to return to another attractor,  
 110 if one exists. Because this can occur while the system is far from the other attractor, this  
 111 transition appears particularly sudden and has been a focal concern (Scheffer et al., 2001, 2009).

Many models with alternative stable states considered in ecological systems can exhibit this kind of bifurcation. For instance, a common model is one with a saturating birth rate (such as Holling Type III) and linear death rate,

$$dX_t = \left( \frac{eKx^2}{X^2 + h_t^2} - eX_t - a_t \right) dt + \sigma \sqrt{\frac{eKx^2}{X^2 + h_t^2} + eX_t + a_t} dB_t \quad (\text{D.1})$$

This model can experience a saddle-node bifurcation through increased mortality by increasing either  $e$  or  $a$ . This parameterization is only an example, of course many others are possible (Scheffer, 2009; Scheffer et al., 2001; Strogatz, 2001; Guckenheimer and Holmes, 1983). In this case we have specified the noise dependence on  $X_t$  explicitly for the influence of demographic (intrinsic) noise. The saddle-node bifurcation has normal form

$$\frac{dx}{dt} = r_t - x^2. \quad (\text{D.2})$$

Linearizing this model differs from the transcritical bifurcation, since the location of stable point moves as the bifurcation parameter changes. Transforming the canonical form to allow for an arbitrary mean  $\theta$ , the bifurcation looks like  $dx/dt = r_t - (\theta - x)^2$ , with fixed point  $\hat{x} = \sqrt{r_t} + \theta =: \phi$ , which gives our second model, LSN:

$$dX = 2\sqrt{r_t}(\phi - X_t)dt + \sigma\sqrt{\phi}dB_t. \quad (\text{D.3})$$

112 *Appendix D.2. Transcritical Bifurcation*

113 [Drake and Griffen \(2010\)](#) induce a transcritical bifurcation in laboratory populations of  
 114 *Daphnia magna* for purpose of testing early warning signals ([Drake and Griffen, 2010](#)). The  
 115 transcritical bifurcation

For instance, consider the stochastic Levin's model ([Levins, 1969](#)) (a logistic model), where  $X_t$  is the number of occupied patches of some total number  $K$ ,  $c_t$  is a (time-dependent) colonization rate and  $e_t$  an extinction rate (both scaled by the number of patches  $K$ ).

$$dX_t = (c_t X_t (1 - X_t/K) - e_t X_t) dt + \sigma \sqrt{\frac{e_t}{c_t}} dB_t, \quad (\text{D.4})$$

such a model can be derived from an individual-based description with or without a stochastic environment, i.e. ([Kampen, 2007](#); [Nisbet and Gurney, 2004](#)). The model contains a transcritical bifurcation when  $c_t < e_t$ . The normal form of the bifurcation is

$$\frac{dx}{dt} = r_t x - x^2 \quad (\text{D.5})$$

We can rewrite the deterministic part of the Levins model, Eq [D.4](#) in normal form without loss of generality by taking  $r_t = 1 - e_t/c_t$ . We can linearize around the stable equilibrium value  $\hat{x}(t) = 1 - e_t/c_t = r_t$ ,

$$\dot{x} = f(x) \approx f'(x)|_{\hat{x}}(x - \hat{x}) + \dots, \quad (\text{D.6})$$

$$= (r_t - 2x_t|_{x=\hat{x}})(x - \hat{x}), \quad (\text{D.7})$$

$$= r_t(r_t - x). \quad (\text{D.8})$$

116 Note the equilibrium time dependence arises not from the internal dynamics but from the  
 117 changing parameter values. We can express the linearized stochastic dynamics then by the  
 118 time dependent mean-reverting process. We refer to this model as the linearized transcritical  
 119 bifurcation.

$$dX_t = r_t(r_t - X_t/K)dt + \sigma\sqrt{1 + r_t}dB_t \quad (\text{D.9})$$

120 The pattern of the saddle-node differs from that of the transcritical in the square-root de-  
 121 pendence on the bifurcation parameter.

### 122 *Appendix D.3. Modeling limitations*

Many models may differ substantially from (D.1) and (D.4) and yet correspond to the linearizations illustrated here. The linearizations have thus been constructed with enough free parameters to capture the gross features of the system, (scaling the equilibrium position, variance, and timescale) while corresponding to the approximations of normal forms of the bifurcations. Despite this, there are severe limitations in formulating even a linear approximation without ample knowledge of the system. Note that specifying this model such that likelihoods can be calculated still requires a specification of the rate of change of  $r_t$ . We assume a linear model,

$$r_t = r_0 - mt \tag{D.10}$$

123 though the approach could easily be extended to an arbitrary model. A gradual change can be  
 124 approximately linear over the time interval of interest, and is appropriate to distinguish from  
 125 a stable model. Nonlinear rates of change in the environmental conditions will generally be  
 126 harder to detect (either by summary statistics or the likelihood approach), thus this assumption  
 127 is consistent with testing the best case scenario for these methods, illustrating the intrinsic limits  
 128 of these approaches.

129 We also draw attention to the fact that noise scales with the mean dynamics. For intrinsically  
 130 stochastic processes, such as demographic stochasticity, this will scale as the square root, while  
 131 for external perturbations to the system dynamics (environmental stochasticity) this would scale  
 132 linearly. We have assumed an intrinsic noise process in both models (D.3) and (D.9), though  
 133 an alternative model with linear scaling would also be plausible. Note that in these models,  
 134 the variance observed is determined by the ratio of the Brownian terms (in front of  $dB_t$ ) to the  
 135 stabilizing force (in front of  $dt$ ), so that in general larger  $r_t$  will correspond to smaller system  
 136 variances regardless of this smaller correction due to the noise scaling with the mean dynamics.  
 137 Either should perform better than the stable model if a system is slowly losing stability, though  
 138 using the better-matching model will increase the power of the approach.

139 The models here are neither more nor less general than the seemingly model-free approach of  
 140 summary statistics. We have proposed these models to match the assumptions of the summary  
 141 statistics – that a system is approaching a bifurcation in which a node loses stability as an  
 142 eigenvalue passes through zero.

#### 143 *Appendix D.4. Stable Model*

When the system is stable,  $r_t$  is constant and both models (D.3) and (D.9) reduce to a simple Ornstein-Uhlenbeck process,

$$dX_t = r(\theta - X_t)dt + \sigma dB_t \quad (\text{D.11})$$

### 144 **Appendix E. Likelihood calculations**

145 The linearization above is not only justified by the nature of and the properties we wish  
 146 to estimate from the data, but also convenient for the calculations. In particular, the models  
 147 are specified by linear stochastic differential equations, hence their solutions are Gaussian. The  
 148 fundamental challenge is that in modeling a gradual loss of stability, the solutions are also ex-  
 149 plicitly time-dependent. Model fitting and comparison requires we can write down the likelihood  
 150 expression for each model given the time-series data.

The probability  $P(M|X)$  of the data  $X$  given the model  $M$  is the product of the probability of observing each point in the time series given the previous point and the length of the interval,

$$\log P(M|X) = \sum_i \log P(x_i|x_{i-1}, t_i) \quad (\text{E.1})$$

151 As the processes are Gaussian, the probability density  $P$  is normally distributed and we need  
 152 only calculate its first and second moments. For the OU process (OU process) this is trivial:

$$E(x_i) = X_{i-1}e^{-rt_i}\theta (1 - e^{-rt_i}) \quad (\text{E.2})$$

$$V(x_i) = \frac{\sigma^2}{2r} (1 - e^{-2rt_i}) \quad (\text{E.3})$$

153 For the time dependent models, we have analytic forms only for the dynamical equations of  
 154 these moments, which we must integrate numerically over each time interval. This is responsible  
 155 for most of the computational effort required of this approach. For the linearized transcritical  
 156 bifurcation:

$$\frac{d}{dt}E(x_i) = r(t)(r(t) - x_i) \quad (\text{E.4})$$

$$\frac{d}{dt}V(x_i) = -2r(t)V(x_i) + (1 + r(t))\sigma^2 \quad (\text{E.5})$$



For the linearized saddle-node bifurcation:

$$\frac{d}{dt}E(x_i) = 2\sqrt{r(t)}(\sqrt{r(t)} + \theta - x_i) \quad (\text{E.6})$$

$$\frac{d}{dt}V(x_i) = -2\sqrt{r(t)}V(x_i) + \sigma^2(\sqrt{r(t)} + \theta) \quad (\text{E.7})$$

These are numerically integrated using `lsoda` routine for the likelihood calculation.

## Appendix F. The likelihood statistic

Likelihood methods form the basis of much of modern statistics, in both Frequentist and Bayesian paradigms. The ability evaluate likelihoods directly by computation has made it possible to treat cases that do not conform to traditional assumptions more directly. The basis of likelihood comparisons has its roots in the Neyman-Pearson Lemma, which essentially asserts that comparing likelihoods is the most powerful test of a choice between two hypotheses (Neyman and Pearson, 1933), which motivates tests from the simple likelihood ratio test up through modern model adequacy methods.

The hypotheses considered here are more challenging then the original lemma, as they are composite in nature: they specify two model forms (stable and changing stability) but with model parameters that must be first estimated from the data. Comparing models whose parameters have been estimated by maximum likelihood is first treated by Cox (1961, 1962), and has since been developed in this simulation estimation of the null distribution (McLachlan, 1987), by parametric bootstrap estimate (Efron, 1987). Cox's  $\delta$  statistic is simply the difference between the log likelihoods of these maximum likelihood estimates, defined as follows.

Let  $L_0$  be the likelihood function for model 0, let  $\theta_0 = \arg \max \theta_0 \in \Omega_0$ , ( $L_0(\theta_0|X)$ ) be the maximum likelihood estimator for  $\theta_0$  given  $X$ , and let  $L_0 = L_0(\theta_0|X)$ ; and define  $L_1$ ,  $\theta_1$ ,  $L_1$  similarly for model 1. The statistic we will use is  $\delta$ , defined to be twice the difference in log likelihood of observing the data under the two MLE models,  $\delta = -2(\log L_0 - \log L_1)$

This approach has since been applied to the problem of model adequacy (Goldman, 1993) and model choice (Huelsenbeck and Bull, 1996). We have extended the approach by generating the test distribution as well as a null distribution, allowing estimate of power and Type II error.

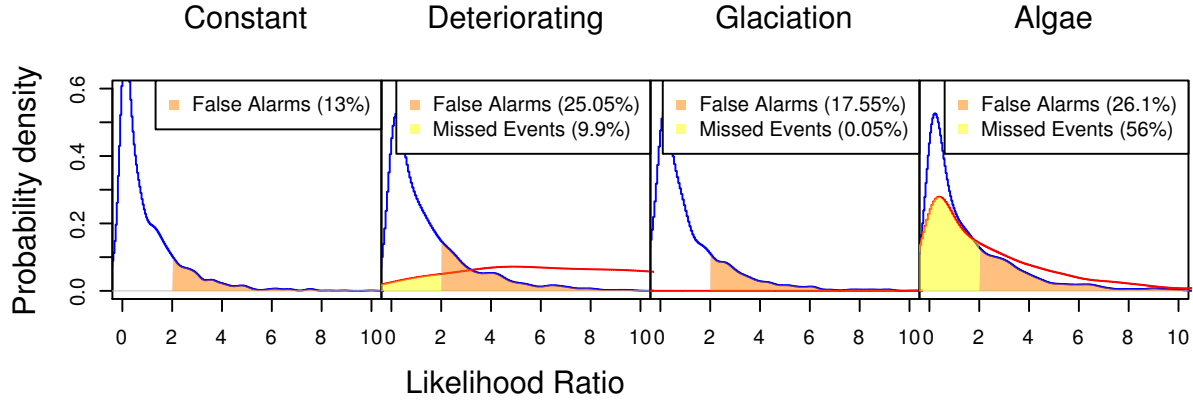


Figure G.1: False alarm and missed detection rate that result from an AIC decision criterion.

## 180 Appendix G. A note on information criteria

181 While it would be possible to compare these models directly by a simple test such as the  
 182 Akaike Information criterion to compare such models, this would be to attempt to answer the  
 183 question statistically that should be a management decision – whether to treat the data as  
 184 indicative of an impending critical transition – while leaving unanswered the statistical question  
 185 of the relative risks of missed events and false alarms. Using the distributions we compute  
 186 in Figure 1c, it is easy to assess the performance an AIC-based decision criteria would have,  
 187 Figure Appendix G. Without the approach described here, these error rates would be unknown.  
 188 Given the distributions, it is clear that there would be no need to accept the AIC criterion.  
 189 For instance, in the empirical data, third panel of Figure Appendix G, it would make sense to  
 190 use a much more stringent threshold, thereby decreasing the rate of false alarms without any  
 191 substantial increase in the rate of missed events. When power of detection is very low such, as  
 192 in the middle panel, error rates can be particularly high. Low power could result from only a  
 193 very gradual loss of power or from inadequate sampling, which can be estimated by comparing  
 194 parameter distributions, see section Appendix H.

## 195 Appendix H. Parameter Distributions

196 The approach taken in Figure 1c provides a powerful way to indicate when a warning signal  
 197 can be detected, but more care must be taken with the case in which a signal is not detected.

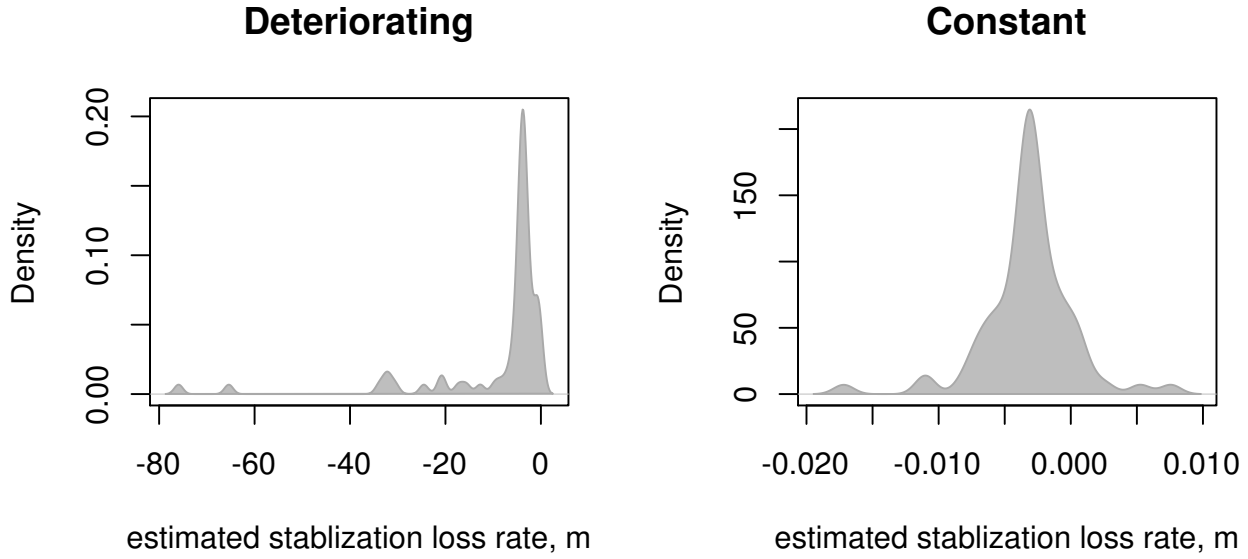


Figure H.2: .

198 It will always be easier to say a system is deteriorating (rejecting the null hypothesis that a  
 199 system is stable) then to establish that a system is stable, as we cannot rule out that we have  
 200 missed the detection for lack of power. Thus the left and right panels of Figure 1c make for clear  
 201 cases of destabilizing systems, while more care must be taken in drawing a conclusion about the  
 202 middle panel.

203 Can we argue that the middle system is truly stable, or could it be experiencing a very  
 204 gradual loss of stability? The lack of power could be the result of inadequate amount of data or  
 205 the result of a very gradual change that is difficult to distinguish from a stable system. In such  
 206 cases, the distribution of the estimated stability loss parameter can help establish a lack of power  
 207 (but potentially rapid rate of stability-loss) from a gradual loss. Figure H.2 shows an example  
 208 of this for the case of the simulated loss of stability and the simulated constant environment.  
 209 When estimating a changing-stability model on the constant environment data, the distribution  
 210 for the rate of change is tightly and symmetrically distributed around zero, suggesting that the  
 211 lack of a signal from this data in the likelihood test is not simply due to a lack of power to  
 212 detect a very gradual change. As this data is simulated under a constant environment, this is  
 213 exactly what we would expect to find given adequate sampling.

214 Distributions of all parameter estimates are obtained without extra computational cost by  
 215 the likelihood approach, and are reported by our package. Note that this parameter  $m$  represents  
 216 linearized or average rate of loss, and does not preclude the possibility of the rate of deterioration  
 217 accelerating.

## 218 Appendix I. Model choice and model adequacy

219 Any approach is only as good as its underlying assumptions. In this paper we have lever-  
 220 aged the machinery of likelihood-based approaches to both evaluate the performance of existing  
 221 methods and provide a statistical test that more closely matches the assumptions of the basic  
 222 theory. For this approach to be useful, the data must still be well-approximated by the models  
 223 used. We have provided two different simple bifurcation models that could each represent the  
 224 hypothesis of gradual decline in stability: linearized transcritical bifurcation and the linearized  
 225 saddle-node bifurcation; Equations (D.9), (D.3). As they both represent a gradual loss of stabil-  
 226 ity either can be used, but the test will be most powerful by using the one that best corresponds  
 227 to the data. While knowledge of the system alone can suggest this (such as in the case of [Drake](#)  
 228 [and Griffen \(2010\)](#)), it is straight forward to compare the models against each other using the  
 229 same likelihood framework, as illustrated in Figure I.3 for several empirical examples.

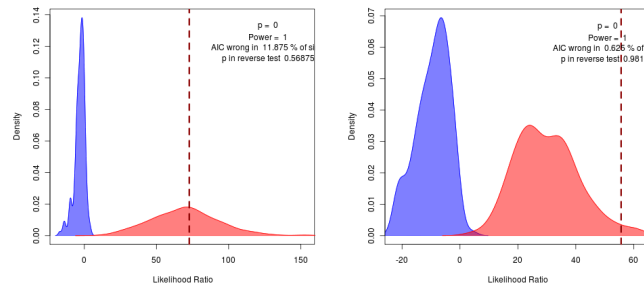


Figure I.3: Model

230 As discussed in the text, many systems may not correspond well to any of the simple dynamics  
 231 upon which the detection of early warning signals is based ([Hastings and Wysham, 2010](#)). Such  
 232 dynamics can be particularly misleading for current methods based on summary statistics. The  
 233 Monte Carlo bootstrapping of the likelihood approach can be applied as a test of model adequacy,

as illustrated in [Goldman \(1993\)](#) and discussed in [Sullivan and Joyce \(2005\)](#) in a phylogenetics context. If the observed likelihood ratio falls far from the distributions estimated under either model, the model is likely an inadequate description and subsequent inference may be misleading.

## Appendix J. R package tutorial

We provide an R package with simple implementation of the methods described here. The package is available from: <https://github.com/cboettig/warningsignals/archives/master>, where it will be actively maintained and developed. The package takes an R time-series object (or, for unevenly spaced data – a matrix or data-frame with sample times in the first column and observations in the second) as input and performs the likelihood analysis illustrated in Figure 1c:

```
> require(warningsignals)
>
> # Load a sample dataset
> data(glaciationIII)
>
> # Fit constant (OU) model and a stability-loss model (LSN);
> models <- fit_models(glaciationIII, 'LSN')
>
> results <- montecarlotest(models$const, models$timedep, n=2000, cpu=16)
> plot(results)
```

The package can generate traditional warning signals (as in Figure 1a) and bootstraps of their uncertainty (as in Figure 1b) for a variety of correlation tests (Pearson’s, Kendall’s, Spearman’s) and summary statistics (variance, autocorrelation, skew) as well. Copies of data used in the paper are provided as examples. The Monte Carlo methods can take advantage of multiple core processors or clusters by specifying the number of CPUs in the function argument. Full documentation of functions and data are provided in the package.

## Appendix K. Examples

In this section we illustrate the main results of the paper in three further empirical datasets from the climate record.

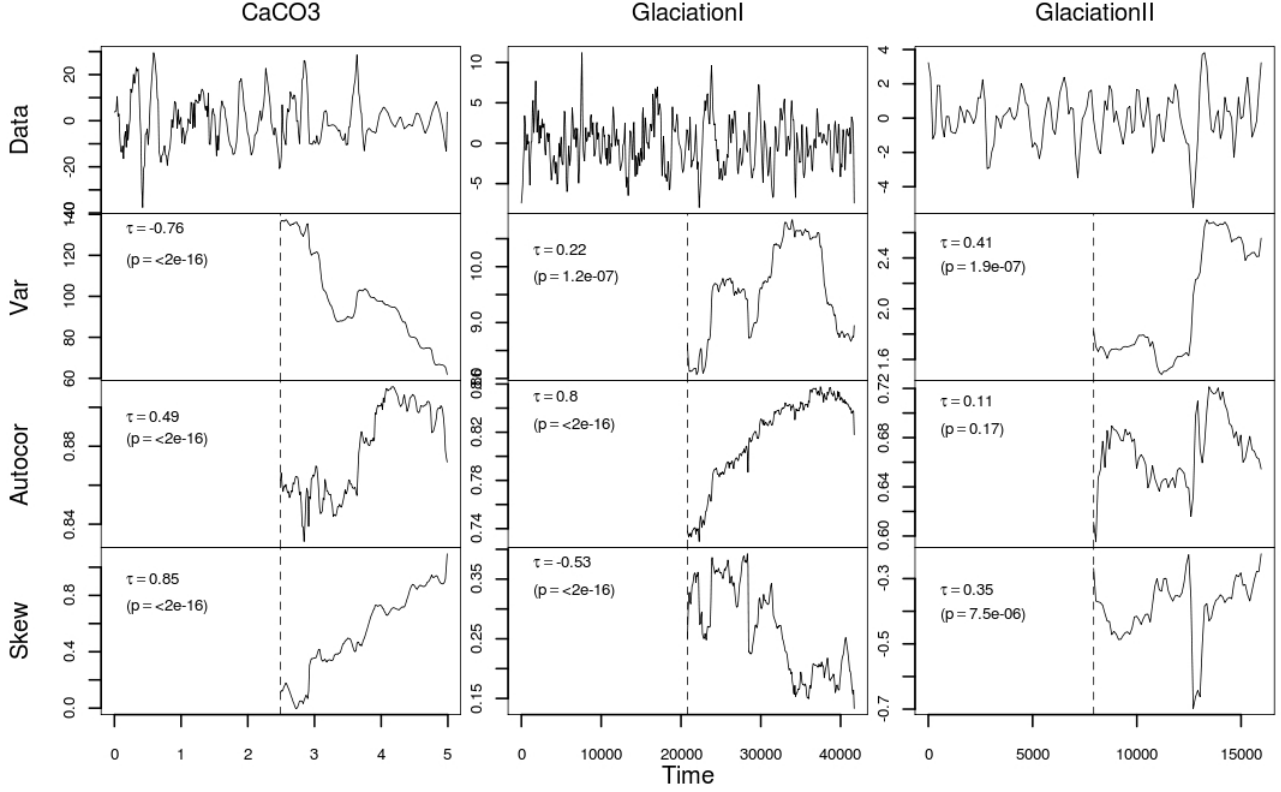


Figure K.4: Examples for three more empirical datasets from climate records, analysed for warning signals in Dakos et al. (2008). Calcium carbonate data originally from Tripathi et al. (2005), Deuterium concentrations in the periods before major ice ages from Petit et al. (1999).

## Appendix L. Comparisons for additional summary statistics

Higher moments have also been proposed for the detection of warning signals, such as the skewness (Guttal and Jayaprakash, 2008), illustrated for the datasets presented in the text in Figure L.7. The skewness appears to have even less power than the other indicators discussed in the text.

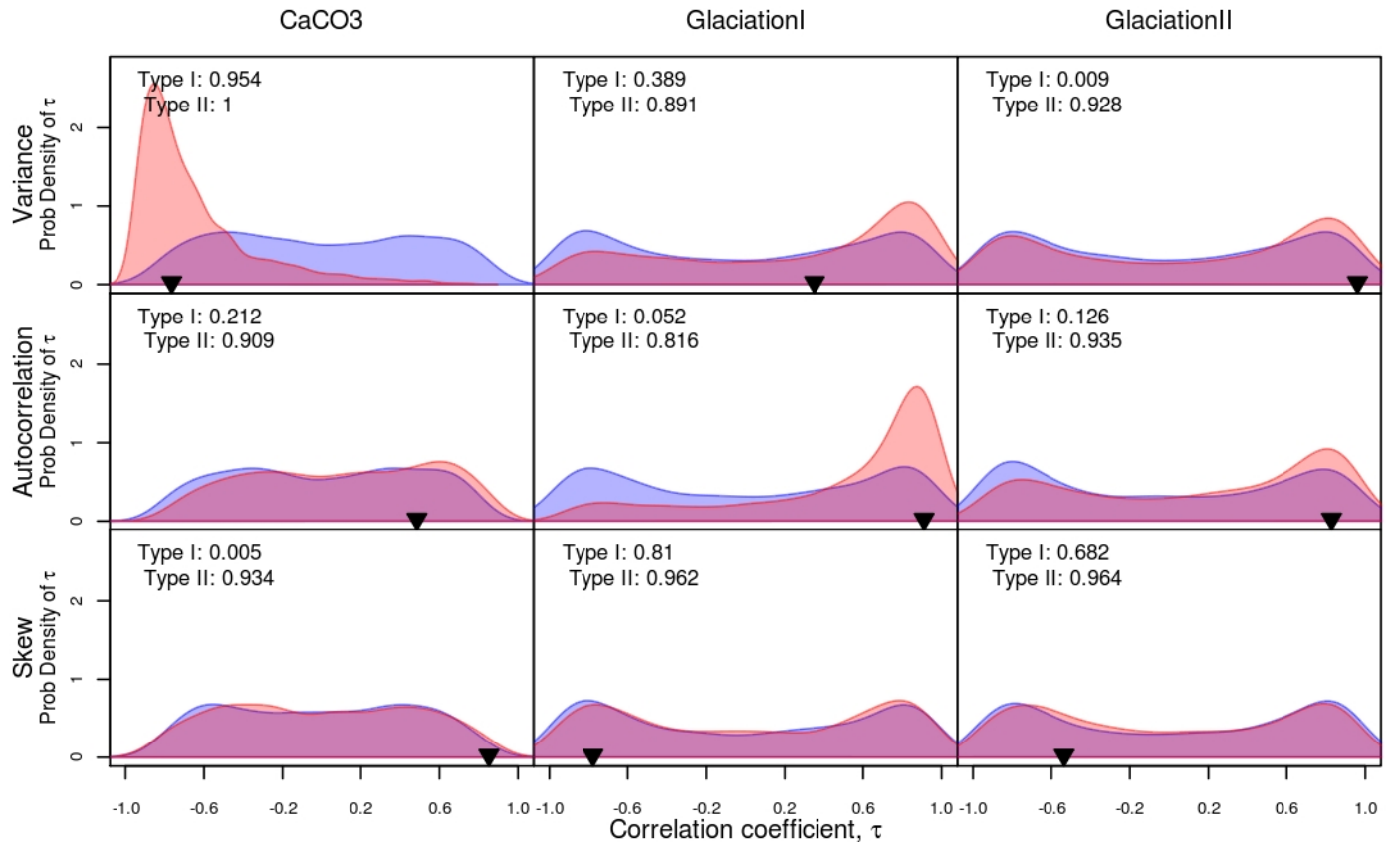


Figure K.5: Bootstraps of the correlations observed in Figure K.4, presented as in Figure 1b in the text. The distribution of the test statistic under the null (blue) and test (red) hypotheses are difficult to distinguish reliably, across all indicators for each data set.

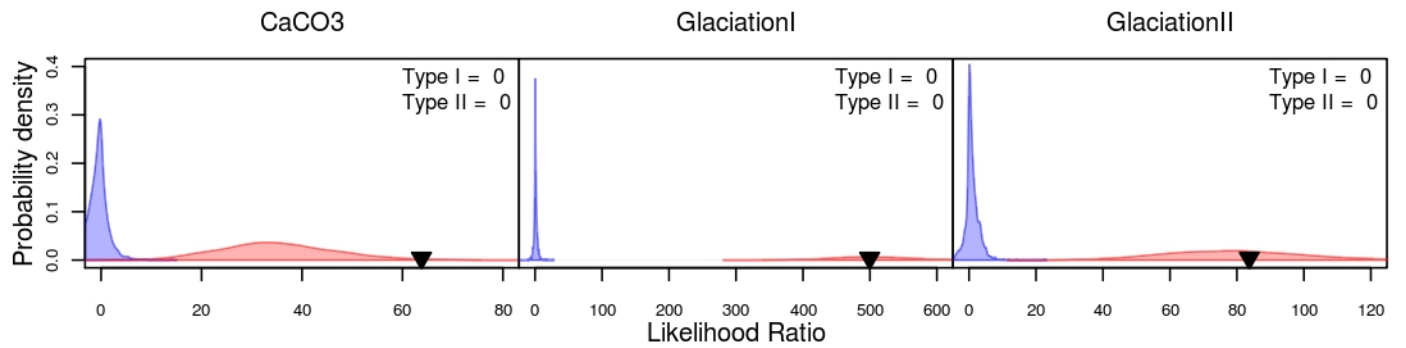


Figure K.6: Bootstraps of likelihood ratios for the data observed in Figure K.4, presented as in Figure 1c in the text. The distribution of the test statistic under the null (blue) and test (red) hypotheses are more clearly distinguished in this approach, supporting the evidence for a warning signal.

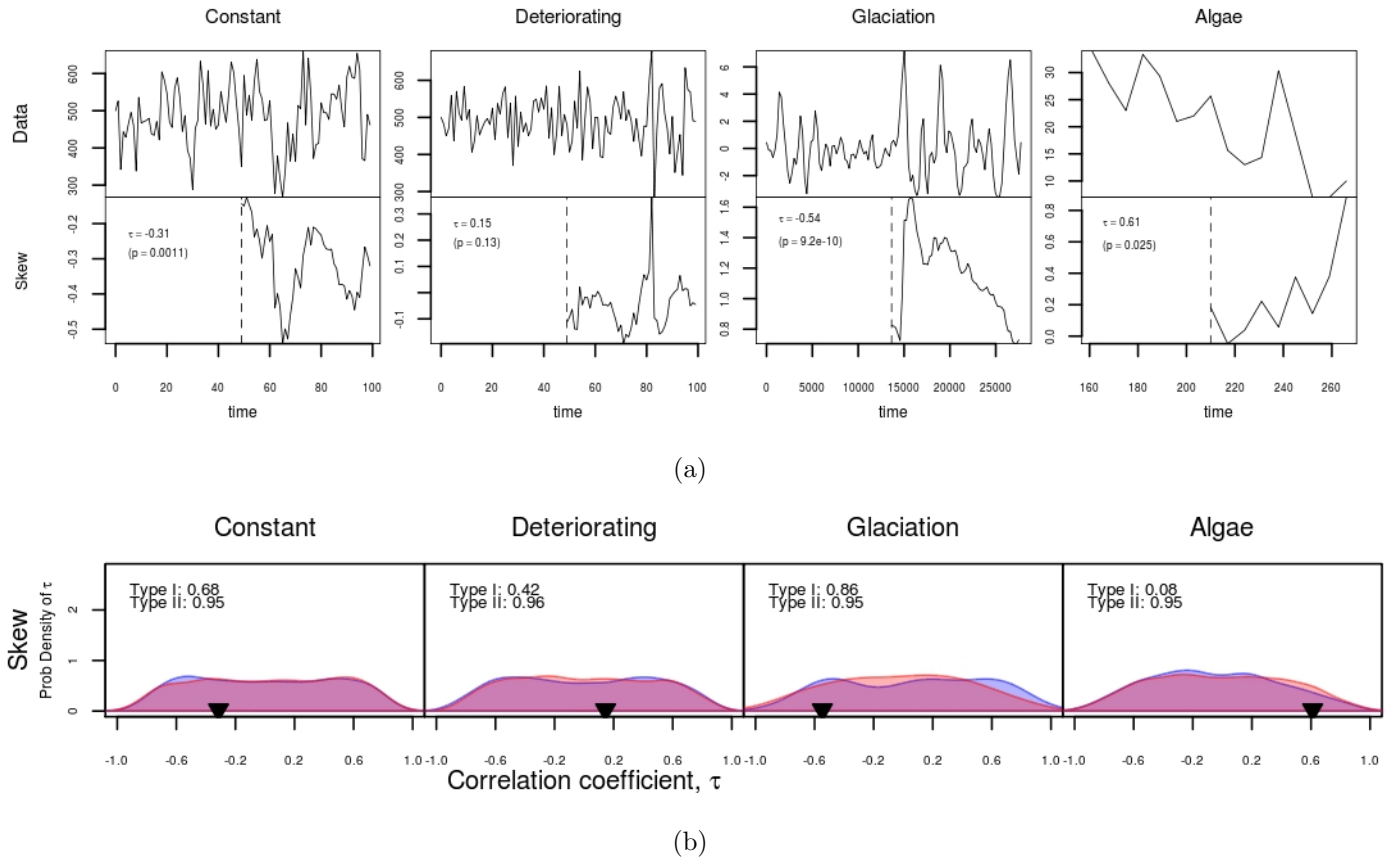


Figure L.7: Increasing skewness is no more likely under the stable hypothesis than the stability-loss hypothesis for each of these data.

268

269 Brozović, N., Schlenker, W., Jan. 2011. Optimal management of an ecosystem with an unknown  
 270 threshold. *Ecological Economics*, 1–14.

271 URL <http://linkinghub.elsevier.com/retrieve/pii/S0921800910004167>

272 Cox, D. R., 1961. Tests of Separate Families of Hypotheses. In: *Proceedings of the 4th Berkeley*  
 273 *Symposium*, University of California Press. No. 2. pp. 105 – 123.

274 Cox, D. R., 1962. Further results on tests of separate families of hypotheses. *Journal of the*  
 275 *Royal Statistical Society* 24 (2), 406–424.

276 URL <http://www.jstor.org/stable/2984232>

277 Dakos, V., Scheffer, M., van Nes, E. H., Brovkin, V., Petoukhov, V., Held, H., Sep. 2008. Slowing



278 down as an early warning signal for abrupt climate change. Proceedings of the National  
 279 Academy of Sciences of the United States of America 105 (38), 14308–12.  
 280 URL <http://www.ncbi.nlm.nih.gov/pubmed/18787119>

281 Drake, J. M., Griffen, B. D., Sep. 2010. Early warning signals of extinction in deteriorating  
 282 environments. Nature 467 (7314), 456–459.  
 283 URL <http://www.nature.com/doifinder/10.1038/nature09389>

284 Efron, B., 1987. Better bootstrap confidence intervals. Journal of the American Statistical As-  
 285 sociation 82 (397), 171– 185.  
 286 URL <http://www.jstor.org/stable/2289144>

287 Gardiner, C., 2009. Stochastic Methods: A Handbook for the Natural and Social Sciences  
 288 (Springer Series in Synergetics). Springer.  
 289 URL [http://www.amazon.com/Stochastic-Methods-Handbook-Sciences-Synergetics/](http://www.amazon.com/Stochastic-Methods-Handbook-Sciences-Synergetics/dp/3540707123)  
 290 [dp/3540707123](http://www.amazon.com/Stochastic-Methods-Handbook-Sciences-Synergetics/dp/3540707123)

291 Goldman, N., Feb. 1993. Statistical tests of models of DNA substitution. Journal of Molecular  
 292 Evolution 36 (2), 182–198.  
 293 URL <http://www.springerlink.com/index/10.1007/BF00166252>

294 Guckenheimer, J., Holmes, P., 1983. Nonlinear Oscillations, Dynamical Systems, and Bifurca-  
 295 tions of Vector Fields (Applied Mathematical Sciences Vol. 42). Springer.  
 296 URL [http://www.amazon.com/Nonlinear-Oscillations-Dynamical-Bifurcations-Mathematical/](http://www.amazon.com/Nonlinear-Oscillations-Dynamical-Bifurcations-Mathematical/dp/0387908196)  
 297 [dp/0387908196](http://www.amazon.com/Nonlinear-Oscillations-Dynamical-Bifurcations-Mathematical/dp/0387908196)

298 Guttal, V., Jayaprakash, C., May 2008. Changing skewness: an early warning signal of regime  
 299 shifts in ecosystems. Ecology letters 11 (5), 450–60.  
 300 URL <http://www.ncbi.nlm.nih.gov/pubmed/18279354>

301 Hastings, A., Wysham, D. B., 2010. Regime shifts in ecological systems can occur with no  
 302 warning. Ecology letters.  
 303 URL <http://www.ncbi.nlm.nih.gov/pubmed/20148928>

304 Huelsenbeck, J. P., Bull, J. J., Mar. 1996. A Likelihood Ratio Test to Detect Conflicting Phy-  
 305 logenetic Signal. *Systematic Biology* 45 (1), 92–98.  
 306 URL <http://sysbio.oxfordjournals.org/cgi/content/abstract/45/1/92>

307 Kampen, N. V., 2007. *Stochastic Processes in Physics and Chemistry*, Third Edition (North-  
 308 Holland Personal Library). North Holland.  
 309 URL [http://www.amazon.com/Stochastic-Processes-Chemistry-North-Holland-Personal/](http://www.amazon.com/Stochastic-Processes-Chemistry-North-Holland-Personal/dp/0444529659)  
 310 [dp/0444529659](http://www.amazon.com/Stochastic-Processes-Chemistry-North-Holland-Personal/dp/0444529659)

311 Levins, R., 1969. Some demographic and genetic consequences of environmental heterogeneity  
 312 for biological control. *Bulletin of the Entomological Society of America* 15 (3), 237–240.  
 313 URL [http://www.ingentaconnect.com/content/esa/besa/1969/00000015/00000003/](http://www.ingentaconnect.com/content/esa/besa/1969/00000015/00000003/art00009)  
 314 [art00009](http://www.ingentaconnect.com/content/esa/besa/1969/00000015/00000003/art00009)

315 McLachlan, G. J., 1987. On Bootstrapping the Likelihood Ratio Test Statistic for the Number  
 316 of Components in a Normal Mixture. *Applied Statistics* 36 (3), 318.  
 317 URL [http://links.jstor.org/sici?sici=0035-9254%281987%2936%3A3%3C318%](http://links.jstor.org/sici?sici=0035-9254%281987%2936%3A3%3C318%3A0BTLRT%3E2.0.CO%3B2-I&origin=crossref)  
 318 [3A0BTLRT%3E2.0.CO%3B2-I&origin=crossref](http://links.jstor.org/sici?sici=0035-9254%281987%2936%3A3%3C318%3A0BTLRT%3E2.0.CO%3B2-I&origin=crossref)

319 Neyman, J., Pearson, E., 1933. On the problem of the most efficient tests of statistical hypothe-  
 320 ses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers*  
 321 *of a Mathematical or Physical Character* 231 (694-706), 289.  
 322 URL <http://rsta.royalsocietypublishing.org/content/231/694-706/289.full.pdf>

323 Nisbet, R. M., Gurney, W., 2004. *Modelling Fluctuating Populations*. The Blackburn Press.  
 324 URL [http://www.amazon.com/Modelling-Fluctuating-Populations-R-Nisbet/dp/](http://www.amazon.com/Modelling-Fluctuating-Populations-R-Nisbet/dp/1930665903)  
 325 [1930665903](http://www.amazon.com/Modelling-Fluctuating-Populations-R-Nisbet/dp/1930665903)

326 Petit, J. R., Jouzel, J., Raynaud, D., Barkov, N. I., Barnola, J., Basile, I., Bender, M., Chap-  
 327 pellaz, J., Davis, M., Delaygue, G., Delmotte, M., Kotlyakov, V. M., Legrand, M., Lipenkov,  
 328 V. Y., Lorius, C., Pepin, L., Ritz, C., Saltzman, E., Stievenard, M., 1999. Climate and at-  
 329 mospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature*  
 330 399 (6735), 429–436.  
 331 URL <http://www.daycreek.com/dc/images/1999.pdf>

332 Scheffer, M., 2009. Critical Transitions in Nature and Society: (Princeton Studies in Complex-  
 333 ity). Princeton University Press.  
 334 URL [http://www.amazon.com/Critical-Transitions-Nature-Society-Complexity/dp/](http://www.amazon.com/Critical-Transitions-Nature-Society-Complexity/dp/0691122040)  
 335 [0691122040](http://www.amazon.com/Critical-Transitions-Nature-Society-Complexity/dp/0691122040)  
 336 Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H.,  
 337 van Nes, E. H., Rietkerk, M., Sugihara, G., 2009. Early-warning signals for critical transitions.  
 338 Nature 461 (7260), 53–9.  
 339 URL <http://www.ncbi.nlm.nih.gov/pubmed/19727193>  
 340 Scheffer, M., Carpenter, S., Foley, J. A., Folke, C., Walker, B., Oct. 2001. Catastrophic shifts in  
 341 ecosystems. Nature 413 (6856), 591–6.  
 342 URL <http://www.ncbi.nlm.nih.gov/pubmed/11595939>  
 343 Strogatz, S. H., 2001. Nonlinear Dynamics And Chaos: With Applications To Physics, Biology,  
 344 Chemistry, And Engineering (Studies in Nonlinearity). Westview Press.  
 345 URL [http://www.amazon.com/Nonlinear-Dynamics-Chaos-Applications-Nonlinearity/](http://www.amazon.com/Nonlinear-Dynamics-Chaos-Applications-Nonlinearity/dp/0738204536)  
 346 [dp/0738204536](http://www.amazon.com/Nonlinear-Dynamics-Chaos-Applications-Nonlinearity/dp/0738204536)  
 347 Sullivan, J., Joyce, P., Dec. 2005. Model Selection in Phylogenetics. Annual Review of Ecology,  
 348 Evolution, and Systematics 36 (1), 445–466.  
 349 URL [http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.ecolsys.36.](http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.ecolsys.36.102003.152633)  
 350 [102003.152633](http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.ecolsys.36.102003.152633)  
 351 Tripathi, A., Backman, J., Elderfield, H., Ferretti, P., Jul. 2005. Eocene bipolar glaciation asso-  
 352 ciated with global carbon cycle changes. Nature 436 (7049), 341–6.  
 353 URL <http://www.ncbi.nlm.nih.gov/pubmed/16034408>