

Regime shifts: early warnings

Florian Wagener^{*}

7th March 2012

A system that is at a steady state responds most of the time gradually to external changes. But in exceptional circumstances, it can exhibit a sudden ‘catastrophic’ shift to a different regime. It is of great practical interest to develop early warning indicators that signal the imminence of such a shift. A promising class of such indicators uses the universal fact that the average return time to a stable steady state after a small disturbance increases sharply close to a catastrophic shift. It is however important to realise that there are classes of dynamic regime shifts that cannot be predicted in this way. After reviewing the mathematical ideas behind these indicators, this article discusses their scope and their limitations.

Glossary

autocorrelation coefficient The first (second, third, ...) order autocorrelation coefficient is the correlation of the time series $\{x_t\}$ with the series where all elements are one (two, three, ...) step shifted $\{x_{t+1}\}$ ($\{x_{t+2}\}$, $\{x_{t+3}\}$, ...).

complex unit circle See eigenvalue.

eigenvalue A matrix can be fully described by its actions on characteristic directions. Barring exceptional cases, these are either invariant lines or invariant planes. In the first situation, the action of the matrix either contracts or expands the line by a fixed factor λ , the so-called eigenvalue of the characteristic direction. In the second situation, the action of the matrix is first a uniform contraction or expansion of all lines through the origin by a fixed factor r , and then a rotation around an angle ϑ of all these lines around the origin.

It is mathematically convenient to describe the second situation in terms of complex numbers, as the invariant plane then becomes an invariant complex line, with an associated complex eigenvalue $re^{i\vartheta}$. The statement that an eigenvalue is inside the complex unit circle is then equivalent to the fact that the action on the corresponding characteristic direction is a contraction, possibly, for complex eigenvalues that are not real, followed by a rotation. If an eigenvalue is outside

^{*}CeNDEF, Amsterdam School of Economics, Universiteit van Amsterdam, and Tinbergen Institute.
E-mail: F.O.O.Wagener@uva.nl

the unit circle indicates that the action is an expansion, possibly followed by a rotation.

Jacobian matrix For real-valued functions of one variable, the Jacobian matrix is just the derivative. For a vector-valued function $F = F(x)$, where the vector F has components F_i and where x has components x_j , the Jacobian matrix DF is the appropriate generalisation of the derivative. Its elements are of the form $(DF)_{ij} = \partial F_i / \partial x_j$, that is, the partial derivatives of the components of F .

smooth dependence The dependence $y = f(x)$ of the quantity y on the quantity x is smooth, if f is a differentiable function of x for all x .

stochastic process A stochastic process is a series of random variables, indexed by a time variable.

Taylor's theorem This is a mathematical statement about the approximation of a function around a given point. In the article only a very simple version is needed. If all partial derivatives of a map $F(x)$ exist at a point \bar{x} and are continuous functions there, then there is a function $N(x)$ such that

$$F(x) = F(\bar{x}) + DF(\bar{x})(x - \bar{x}) + N(x),$$

and which is such that $N(x)/\|x - \bar{x}\| \rightarrow 0$ as $x \rightarrow \bar{x}$; here $\|x - \bar{x}\|$ is the distance from x to \bar{x} . The mathematics express the fact that the difference N between the function $F(x)$ and its 'linear approximation' $F(\bar{x}) + DF(\bar{x})(x - \bar{x})$ is more and more negligible as x comes closer to \bar{x} .

1 Evolution equations

Systems that evolve in time are thought of and modelled as *dynamical systems*. The state of the system at a given point t in time is described by a *state vector* x_t , which is an ordered list of numbers that contains all information needed in order to describe the system at that particular instance in time. The length of the list is the *dimension* of the vector, and in practical applications the dimension of a state vector can be high. For instance, in a climate model, this would include the value of temperature, wind speed and pressure at many altitudes over many points on the earth surface.

Mathematically, a dynamical system consists of a set of state vectors, constituting the state space, and an evolution law, which determines how the state vector x_t evolves as time changes. In a model, the evolution law describes the effect of short term interactions of the components of the state vector. The general aim of dynamical system theory is to derive from these short time interactions the long run system behaviour.

There are two classes of evolution equations, describing the time evolution either as a continuous flow or as a succession of discrete time steps. As the distinction

has no significant implications to the following discussion, this article only considers discrete time systems. For these systems, the short time system interactions give rise to an evolution map F . The associated *evolution equation*

$$x_{t+1} = F(x_t) \quad (1)$$

governs the time path of the state. The map F is often called ‘the dynamical system’. The initial state x_0 and the evolution law F determine the evolution of a dynamical system of the form (1): such systems are called *deterministic*.

In analogy to the situation in mechanics, the quantity $f(x) = F(x) - x$ is called the force acting on the system. In terms of the force, the evolution equation takes the form

$$x_{t+1} = x_t + f(x_t). \quad (2)$$

If no force is acting on the system, then the state of the system will not change: that is, if $f(x_t) = 0$, then $x_{t+1} = x_t$ for all t . The system is said to be at a *steady state*.

Practical systems are subject to external disturbances. These could be internalised by considering a bigger system that also models the disturbance, but it is more common to model the disturbances as random shocks to the system. This leads to considering evolution equations of the form

$$x_{t+1} = F(x_t, \eta_t), \quad (3)$$

where $\{\eta_t\}$ is a sequence of stochastic variables that model random shocks. A system of the form (3) is called a *stochastic dynamical system*.

Typically, the evolution of a system confines itself for long periods of time to a small part of the whole state space of the system, a ‘regime’: states in a given regime have similar characteristics. Occasionally the system can switch to a different regime. For instance, the earth’s climate system has shifted between ‘glaciation’ and ‘greenhouse’ regimes in the past.

It is of interest to be able to predict these regime shifts before they actually occur, based on observations of the system. If the evolution map F is known, this is a question of identifying the present state x_t and to compute its future evolution. Often, however, the evolution map is known imperfectly or not at all and the only information available are a series of observations $\{x_t\}$ of the state, a so-called *time series*. General dynamical system theory can be used to develop early warning indicators for impending regime shifts even for this situation.

2 Regime shifts in deterministic systems

This section introduces some basic notions from general dynamical systems theory.

2.1 Steady states

As mentioned above, if at a given state \bar{x} no force acts on the system, then the system will not leave this state, and \bar{x} is a steady state. However, a steady state is only dy-

namically relevant if it is stable. This means the following: if all state vectors that are initially close to the steady state give rise to time evolutions of the system that tend to the steady state as time increases, the steady state is called stable, or in technical language ‘locally asymptotically stable’. Unstable steady states will almost never be observed, as in practice there are always small perturbations pushing a system out of the state and the system is not necessarily pushed back towards it.

To analyse the dynamics close to a steady state \bar{x} , it is convenient to write the state x_t of the system as a deviation y_t from the steady state, that is

$$x_t = \bar{x} + y_t. \quad (4)$$

In terms of deviations, the evolution equation (1) takes the form

$$y_{t+1} = F(\bar{x} + y_t) - \bar{x}. \quad (5)$$

This system has a steady state at $y_t = 0$.

Invoking Taylor’s theorem, and using the fact that $F(\bar{x}) = \bar{x}$ at steady state, this can be written as

$$y_{t+1} = Ay_t + N(y_t), \quad (6)$$

where $A = DF(\bar{x})$ is the Jacobian matrix of F at \bar{x} , and where the term $N(y_t)$ is typically much smaller than the linear term Ay_t when the deviation y_t is close to the steady state $\bar{y} = 0$. Put differently, often the dynamics of the system is already well described by the ‘linear approximation’

$$y_{t+1} = Ay_t. \quad (7)$$

A sufficient condition for the local stability of the steady state $\bar{y} = 0$ of the system (6) is that all eigenvalues of the matrix A are inside the complex unit circle. In the simplest situation, where the state space is one-dimensional, this implies that close to the steady state, the system is equivalent to the simple linear system

$$y_{t+1} = \lambda y_t \quad (8)$$

with $|\lambda| < 1$. Equivalence means here that the evolutions of the linear system close to the steady state $\bar{y} = 0$ are a faithful image of the evolutions close to the steady state \bar{x} of the original system; this implies that it is sufficient to consider the linear dynamics (8). Those dynamics have the structure of a negative feedback loop: the quantity y_t measures the displacement of the system from the steady state, and the new displacement y_t will be smaller in absolute value than y_t . As y_{t+1} feeds back into the right hand side of equation (8) when t is replaced by $t + 1$, it follows that the sequence of disturbances $y_t, y_{t+1}, y_{t+2}, \dots$ decays towards zero.

The characteristic time T of this decay is inversely related to the magnitude of $|\lambda|^{-1}$. More precisely, defining the characteristic time as the time needed for a disturbance to decay towards $e^{-1} \approx 0.37$ times the original level, then the characteristic decay time reads as

$$T = \frac{1}{\log |\lambda|^{-1}}. \quad (9)$$

For a stable steady state in a high-dimensional system, the characteristic time is also given by equation (9), but with λ replaced by the eigenvalue that is largest in absolute value.

If however only a single eigenvalue of the Jacobian matrix is outside the unit circle, then the steady state is unstable. Again equation (8), with $|\lambda| > 1$, can explain this: as now the absolute value of y_t is amplified at each time step by the factor $|\lambda|$, the system constitutes a positive feedback loop which will in time drive the system from the steady state.

2.2 Parameters

Dynamical systems often depend on parameters: think of these as state variables that do not change. Parameters are additional variables, separate from state variables, that determine the characteristics of the system; each value of the vector of parameters then corresponds to a different dynamical system. The totality of the systems obtained in this way is called a (parametrised) family F_μ of systems:

$$x_{t+1} = F(x_t, \mu) = F_\mu(x_t). \quad (10)$$

For instance, climate models operating on historical time scales assume the earth's axial tilt, which is the inclination of the rotation axis of the earth relative to its orbital plane, to be constant. In reality the axial tilt changes periodically over a time period of approximately 41 000 years. Other examples are the terrestrial albedo or the output of greenhouse gases in different locations around the globe. In general, as long as the value of a quantity changes sufficiently slowly it is admissible to treat it as a parameter. In a further analysis, it can and should be treated as a slowly varying state variable.

2.3 Generic properties

When discussing dynamical systems without any information given about the specific structure of the map F , it is helpful to restrict the discourse to *generic* properties of systems. Loosely speaking, these are the properties of 'typical' dynamical systems, or of typical families of systems.

Genericity of a property means the following two things: any system F , whether it possesses the property or not, can be arbitrarily well approximated by systems having the property: the property is said to be pervasive or *dense* in the space of all systems. And secondly, if a system F possesses the property, then it is not possible to make a modification to F that destroys the property, at least as long as the modification is sufficiently small: the property is said to be stable or *open* in the space of systems. Of course, the notions 'arbitrarily well approximated' and 'sufficiently small' have to be made more precise before this definition is operational, but this is the basic idea.

For instance, for the system given in equation (1), let \bar{x} be a locally asymptotically stable steady state of F . In the space of all systems that have a stable steady state,

the property that the eigenvalues of the Jacobian matrix $DF(\bar{x})$ are inside the complex unit circle is a generic property.

2.4 Loss of stability

Consider now a *family* F_μ of systems that depends on a real-valued parameter μ : each particular value of μ singles out a member of the family. Assume that F_{μ_0} has a stable steady state \bar{x}_0 such that all eigenvalues of $DF_{\mu_0}(\bar{x}_0)$ are inside the complex unit circle.

Then locally around the parameter value μ_0 , the steady state $x = \bar{x}(\mu)$ varies smoothly with the parameter μ . This rather abstract result has a concrete interpretation: generically, a stable steady state responds gradually to external changes in the surrounding conditions.

As the eigenvalues of the Jacobian matrix DF_μ evaluated at the steady state $\bar{x}(\mu)$ depend smoothly on μ , the steady state can only lose its stability when one of the eigenvalues crosses the complex unit circle. What happens in that situation depends on the local structure of the dynamical system at the steady state. As the structure of the resulting dynamics changes at a stability loss, the dynamical system is said to go through a qualitative change of dynamics or, more succinctly, a *bifurcation*.

There are two kinds of stability loss: soft and hard. After a soft loss of stability, the system settles down to a time evolution that is still in the vicinity of the steady state. After a hard loss of stability, the system may evolve towards an entirely different part of the state space and an entirely different dynamical behaviour, both unpredictable from the previous steady state dynamics. It is clear that only a hard loss of stability can induce a regime shift.

It is here that the concept of genericity turns out to be useful: whereas for arbitrary families of systems the dynamics might change in a multitude of different ways, for generic families that depend on a single scalar parameter there are precisely three different bifurcation scenarios through which a hard loss of stability can take place. These are, respectively, the saddle-node, the subcritical period-doubling and the subcritical Hopf bifurcation. Moreover, there are suitable descriptions of the state space such that these bifurcations take place on a low-dimensional subspace, a so-called centre manifold, and the dynamics of the whole system is entirely determined by its restriction to these centre manifolds: below, these dynamics will be called the ‘essential’ dynamics of the system.

There are three scenarios of hard stability loss that are universal to all dynamical systems: though the consequences are different for every system, the mathematical mechanisms are equal. Choosing variables appropriately reduces a system at a hard loss of stability to one of three normal forms that describe the bifurcation mechanisms.

Almost all of the literature on early warning signals considers only the saddle-node bifurcation, though often without mentioning it explicitly. The next subsection will treat this bifurcation in some detail. Afterwards, the distinctive features of the other two bifurcations will be touched upon briefly.

2.5 The saddle-node bifurcation

2.5.1 Mechanics of the bifurcation

A saddle-node bifurcation occurs if by varying a parameter μ , the biggest eigenvalue λ , in absolute value, of the Jacobian matrix $DF_\mu(\bar{x})$ of a steady state exits the complex unit circle at the point $\lambda = 1$. In this case, the essential part of the system dynamics is one-dimensional. Choosing variables in a certain suitable way, close to the steady state and for parameter values μ close to the bifurcation value μ_c , the system takes the *normal form*

$$y_{t+1} = y_t + \mu_c - \mu - y_t^2. \quad (11)$$

That is, locally around the steady state the essential part $G : \mathbb{R} \rightarrow \mathbb{R}$ of the evolution map has the form $G(y) = y + g(y)$, with restoring force $g(y) = \mu_c - \mu - y^2$. For values of μ smaller than μ_c , there are two steady states

$$\bar{y}_1 = \sqrt{\mu_c - \mu} \quad \text{and} \quad \bar{y}_2 = -\sqrt{\mu_c - \mu}, \quad (12)$$

while for $\mu > \mu_c$ there is no steady state. It is already apparent from the change in the number of steady states that the dynamics bifurcate at the critical parameter value $\mu = \mu_c$.

The Jacobian matrix of G reduces to the derivative G' , taking the values

$$\lambda_1 = G'(\bar{y}_1) = 1 - 2\sqrt{\mu_c - \mu} \quad \text{and} \quad \lambda_2 = G'(\bar{y}_2) = 1 + 2\sqrt{\mu_c - \mu}. \quad (13)$$

When μ is such that $\mu_c - \mu$ is positive but close to 0, the value of λ_1 is inside the complex unit circle, while λ_2 is outside. Moreover λ_1 , as well as λ_2 , equals 1 at bifurcation. Consequently, the steady state \bar{y}_1 is stable, while \bar{y}_2 is unstable, and as μ approaches μ_c from below, the ‘attractiveness’ of \bar{y}_1 decreases successively.

For $\mu > \mu_c$ there is no steady state left, and the normal form only indicates that the system will leave the neighbourhood of $\bar{y} = 0$ eventually, moving to a different part in state space. Figure 1 gives the corresponding *bifurcation diagram* of the saddle-node bifurcation. The set of equilibria forms a surface in the product of parameter and state space. At a saddle-node bifurcation this surface folds back unto itself, as in Figure 2 at $\mu = \mu_c$. On increasing the value of μ starting from low values, the stable and the unstable steady state merge and disappear in a saddle-node bifurcation.

Figure 1 also shows the normal form dynamics (11) for a range of parameter values. Initial states in the vicinity of the stable state lead to evolutions that move towards the steady state, whereas evolutions starting close to the unstable state move away from that point. In fact, it is apparent that the *basin of attraction* of the stable state \bar{y}_1 , that is the set of initial states which eventually will tend towards this state, is bounded by the unstable state \bar{y}_2 . This basin of attraction is a *regime* of the system, as all systems whose initial state are in the basin will eventually display the same dynamic behaviour.

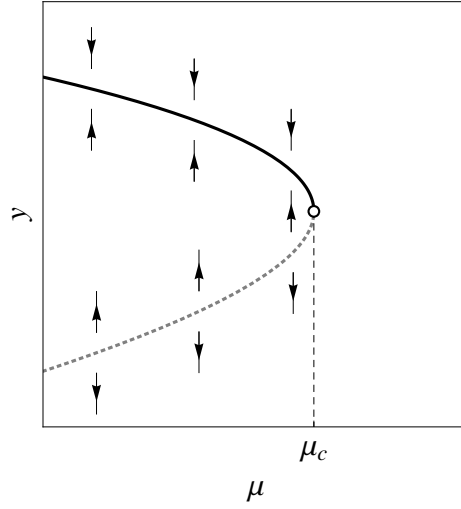


Figure 1: Bifurcation diagram of the saddle-node bifurcation. The solid line indicates stable steady states, the broken line indicates unstable steady states. The arrows show the direction of the dynamics.

2.5.2 Resilience of the steady state

Holling (1973) puts the difference between resilience and stability of a regime of a system as follows:

Resilience determines the persistence of relationships within a system and is a measure of the ability of these systems to absorb changes of state variables, driving variables, and parameters, and still persist. In this definition resilience is the property of the system and persistence or probability of extinction is the result. Stability, on the other hand, is the ability of a system to return to an equilibrium state after a temporary disturbance. The more rapidly it returns, and with the least fluctuation, the more stable it is. In this definition stability is the property of the system and the degree of fluctuation around specific states the result.

More precisely, *resilience* R of a stable steady state is defined as the minimal size of a perturbation that will shift the system into a different basin of attraction, whereas *stability* S is the speed by which a system returns to the steady state after a perturbation. The inverse of the characteristic time T , see (9), quantifies this: $S = 1/T$. It turns out that close to stability loss, both notions are closely related.

From the expressions for the stable and unstable equilibrium, it appears readily that

$$R = 2\sqrt{\mu_c - \mu}. \quad (14)$$

In Figure 1 this quantity is the distance of the unstable steady state, which bounds the basin of attraction of the stable steady state, to the stable steady state itself. Using expression (13) of the eigenvalue λ_1 of the stable steady state, also the characteristic

decay time can be expressed in terms of the difference between the actual and the critical value of the parameter:

$$T = -\frac{1}{\log(1 - 2\sqrt{\mu_c - \mu})}. \quad (15)$$

This leads to the following relation between resilience and stability at a saddle-node bifurcation

$$R = 1 - e^{-\frac{1}{T}} \approx \frac{1}{T} = S, \quad (16)$$

where the approximation is better as T takes larger values. We see that the resilience is inversely related to the characteristic decay time; put differently, close to a saddle-node bifurcation the measures for resilience and stability are approximately equal.

Mechanical engineers have known this relation between resilience and stability of steady states for a long time. For mechanical systems, characteristic frequencies replace characteristic times; in their 1978 book, *Catastrophe Theory*, Poston and Stewart describe this in the case of a strut: “Thus unloaded it will go ‘ting’, moderately loaded ‘bong’, and near buckling point ‘boinnngggg’”. (This is eminently familiar to the practical engineer, but should warn the general reader to prefer soprano structures to bass.)”

2.5.3 Slowly varying parameters

To summarise: knowing that a family of system F_μ is at bifurcation if the system parameter takes some critical value $\mu = \mu_c$ gives information about all the systems that are close to the bifurcating one. It is a statement about the structure of the family of systems F_μ , rather than about single members of this family.

A common fallacy in the interpretation of bifurcation theory is to think of the parameter μ as slowly changing in time, and not to realise that then there will be a dynamic interplay between the system dynamics and the parameter dynamics. This is the province of a different theory, the field of slow-fast systems. Though closely related to bifurcation theory, it studies different phenomena, like solution trajectories that remain close to unstable steady states for long stretches of time.

A slow-fast system is a family F_μ of systems where the parameter μ depends on a slow time parameter τ . The slow time is related to the fast time t by

$$\tau = \varepsilon t, \quad (17)$$

The requirement that $0 < \varepsilon \ll 1$ is a small positive number expresses that τ is much slower than t : one unit of t -time corresponds to ε units of τ -time. The parameter then evolves as

$$\mu = \mu(\tau), \quad (18)$$

A steady state $\bar{x}(\mu)$ of the system with constant parameters, where $\varepsilon = 0$, is usually not a steady state of the slow-fast system where $\varepsilon > 0$; it will be called a ‘quasi-static’ steady state in the following. The location of a quasi-static stable state evolves as $x =$

$\bar{x}(\mu(\tau))$. In fact, the substitution (18) replaces the multi-dimensional parameter μ by the single-dimensional parameter τ . Equivalently, we can think of the parameter μ being one-dimensional, and that it moves sufficiently slowly that the state has time to decay towards the quasi-static steady state, at least, if μ is far from bifurcation. Close to bifurcation this tracking of the quasi-static state by the actual state will break down. For it has already been seen that approaching the bifurcation, the characteristic decay time will tend to infinity, and will therefore at a certain point be slower than the change in the system parameter. See also Figure 2 below.

2.5.4 Catastrophic regime shifts

Consider now the system (11) with a slowly varying parameter $\mu(\tau) = \mu_c + \tau = \mu_c + \varepsilon t$: the parametrisation is chosen such that at $t = 0$ the parameter μ equals the critical value μ_c . The system then takes the form

$$y_{t+1} = \varepsilon t + y_t - y_t^2. \quad (19)$$

For $t < 0$, the state y_t remains close to the quasi-static steady state

$$\bar{y}_1(\mu) = \sqrt{\mu_c - \mu} = \sqrt{-\varepsilon t}. \quad (20)$$

At $t = 0$, the system is close to $y = 0$, and for $t > 0$, the system drifts away from $y = 0$ at a rate that is linearly increasing in t . Figure 2 illustrates this, indicating the regime shift for an infinitely slow changing parameter by a dashed line, and the shift for a more rapidly changing parameter by connected dots. In the second case, the system is still close to the region where the steady state used to be, while the steady state itself, and its basin of attraction, have already disappeared. This is as far as the local analysis

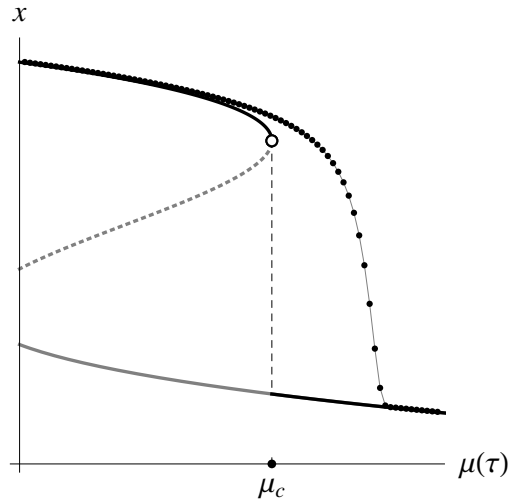


Figure 2: As μ increases, the upper steady state disappears in a saddle-node bifurcation at $\mu = \mu_c$. It then shifts catastrophically towards a new regime.

of the saddle-node bifurcation will take us: if $\mu > \mu_c$, the system shifts to a different

part in state space, possibly far away. This is commonly expressed by saying that the system readjusts to a new stable regime by going through a *catastrophic regime shift*.

Catastrophic system changes of this sort have been studied extensively from the 1970s onward. The mathematicians René Thom and Christopher Zeeman pointed out that since the underlying mechanisms are universal for all dynamical systems, they are expected to be relevant for a wide range of vastly different kinds of systems. Since then, many instances of such changes have been documented.

Actually, the saddle-node bifurcation described above is associated to the *fold catastrophe*, which is the simplest, and therefore the most prevalent, of Thom's classification of elementary catastrophes.

2.6 The subcritical bifurcations

There are two other scenarios for a hard loss of stability, the subcritical period-doubling and the subcritical Neimark-Sacker bifurcation. In both bifurcations the steady state loses stability through the onset of oscillations, and in both cases, a regime shift ensues. The corresponding supercritical bifurcations are soft and do not give rise to a shift.

2.6.1 Subcritical period-doubling

The normal form dynamics of the *subcritical period-doubling* bifurcation read as

$$y_{t+1} = G(y_t) = -(1 + \mu)y_t - y_t^3, \quad (21)$$

which is valid for values of y_t and μ both close to 0. There is only a single steady state $\bar{y} = 0$. For small negative values of μ this state is stable, while for positive values of μ it is unstable. There is consequently a qualitative change at $\mu_c = 0$. Foregoing a full analysis of the dynamics, Figure 3 displays the summarising bifurcation diagram. For $\mu < \mu_c$ there is a stable steady state at $\bar{y} = 0$. An unstable periodic trajectory of period two bounds its basin of attraction.

A period-two trajectory is a state which under the evolution returns to itself after two time steps, and, consequently, each second time step. Denoting the lower and upper bounds of the basin by y_ℓ and y_u , then

$$y_\ell = -\sqrt{\mu_c - \mu}, \quad y_u = \sqrt{\mu_c - \mu}, \quad (22)$$

and $G(y_\ell) = y_u$ and $G(y_u) = G(G(y_\ell)) = y_\ell$. If the initial state y_0 is equal to y_u , then y_1 is equal to y_ℓ , y_2 to y_u , y_3 to y_ℓ etc. The system is said to exhibit period-two cyclic behaviour. From the expressions of y_u and y_ℓ , it follows that the relation between the resilience R of the stable steady state and the parameter μ is again given by equation (14).

The derivative λ of G at the stable steady state reads as

$$\lambda = G'(\bar{y}) = -1 + (\mu_c - \mu). \quad (23)$$

The distinctive feature of the period-doubling bifurcation is that λ approaches the value -1 at bifurcation, and not 1 as in the saddle-node bifurcation.

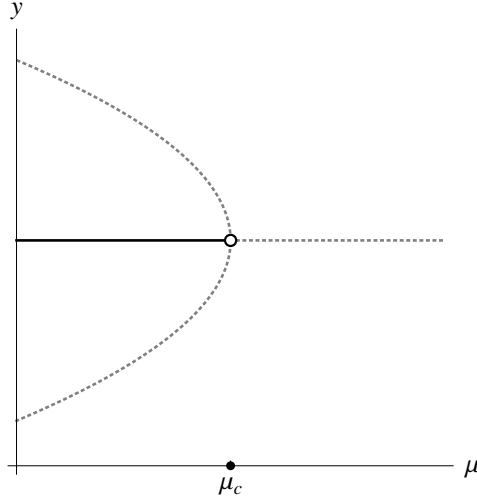


Figure 3: *Period-doubling bifurcation diagram. For $\mu < \mu_c$, there is a single stable steady state, surrounded by an unstable periodic orbit. At $\mu = \mu_c$, the steady state loses stability and becomes unstable for $\mu \geq \mu_c$.*

The relation between the resilience of the attracting steady state and the difference of the actual parameter value to the critical value is in this situation identical to that of the saddle-node bifurcation scenario, given in equation (14). As however λ has a different form, the relation between resilience and characteristic decay time takes here the form

$$R = \sqrt{1 - e^{-\frac{1}{T}}} \approx T^{-\frac{1}{2}} = \sqrt{S}; \quad (24)$$

again, the quality of the approximation improves as T takes larger values. This relation expresses that resilience decays much more slowly with the decay time than in the saddle-node situation. Put differently, the kind of relation between decay time and resilience depends on the kind of stability loss.

At $\mu = \mu_c$, the unstable period-two system trajectory merges with the stable steady state, which then turns into an unstable steady state. As in the case of the saddle-node bifurcation, the system will then shift towards a different regime.

Figure 4 shows this shift for the model system

$$x_{t+1} = F_\mu(x_t) = -(1 + \mu)x_t - x_t^3 + 10x_t^5. \quad (25)$$

This system has a stable steady state for $\mu < \mu_2 = 0$. Moreover, it goes through a saddle-node bifurcation of period-two orbits at $\mu = \mu_1 < \mu_2$. For $\mu_1 < \mu < \mu_2$, there are therefore two coexisting regimes. The regime associated to the stable steady state $\bar{x} = 0$ disappears in a subcritical period doubling bifurcation at $\mu = \mu_2$, after which the system settles on the stable period-two cycle that remains. Remark that the jump towards the new regime is not immediate: if $\mu(\tau) > \mu_2$, the system remains for some time in the vicinity of the, now unstable, steady state $\bar{x} = 0$. Clearly, even though the system is in the vicinity of a steady state for some time, it does not necessarily

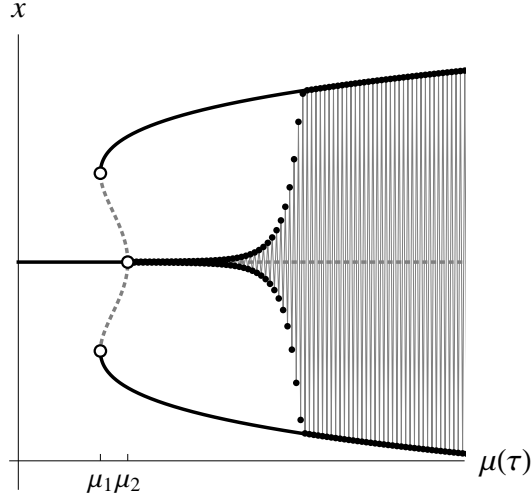


Figure 4: Dynamics for the system (25) in a (μ, x) -diagram, where $\mu(\tau) = \mu_c + \tau = \mu_c + \varepsilon t$. For $\mu = \mu_2 = 0$ the system goes through a period doubling bifurcation. The system keeps tracking the unstable equilibrium for some time, before shifting to a different regime, in this case a stable periodic orbit.

guarantee that the state is stable. This is another mechanism, again due to a slow-fast interaction between system dynamics and parameter dynamics, how the stability loss of a steady state is apparent only after some delay.

2.6.2 Subcritical Neimark-Sacker

The last bifurcation that leads to a hard loss of stability is the *subcritical Hopf* or *subcritical Neimark-Sacker* bifurcation. The normal form map is a map $G : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ on the plane, given as

$$z_{t+1} = (1 + \mu)U_{\alpha(\mu)}z_t + \|z_t\|^2 U_{c(\mu)}z_t + \dots, \quad (26)$$

where $z_t = (z_{1t}, z_{2t}) \in \mathbb{R}^2$, $\|z_t\|^2 = z_{1t}^2 + z_{2t}^2$, and where

$$U_{\vartheta} = \begin{pmatrix} \cos \vartheta & -\sin \vartheta \\ \sin \vartheta & \cos \vartheta \end{pmatrix} \quad (27)$$

is the matrix of a rotation through a positive angle ϑ . The bifurcation takes place at $\mu_c = 0$; the constants $\alpha(\mu_c)$ and $c(\mu_c)$ have to satisfy some technical conditions, which will not be specified, and the dots indicate terms of higher than third order in $|z_t|$ that have been omitted. Expressing the map in polar coordinates

$$z_{1t} = r_t \cos \varphi_t, \quad z_{2t} = r_t \sin \varphi_t, \quad (28)$$

and taking into account the technical conditions, the map takes the form

$$\begin{aligned} r_{t+1} &= r_t + r_t (\mu + r_t^2 + \dots), \\ \varphi_{t+1} &= \varphi_t + \alpha(\mu) + \beta(\mu)r_t^2 + \dots \end{aligned}$$

This time, the dots indicate terms of order four or more in r_t and of order two or more in μ , which will be disregarded in the following. In this approximation the evolution of r_t is independent of that of φ_t . In particular, there is a steady state $\bar{r}_1 = 0$ that is attracting for $\mu < \mu_c$ and repelling for $\mu > \mu_c$, and a second steady state $\bar{r}_2 = \sqrt{\mu_c - \mu}$ that is repelling for all $\mu < \mu_c$ and which bounds the basin of attraction of the stable steady state.

In the original two-dimensional dynamics, the unstable steady state $r = \bar{r}_2$ corresponds to an unstable invariant circle

$$C = \{z \in \mathbb{R}^2 : \|z\| = \sqrt{\mu_c - \mu}\}, \quad (29)$$

which bounds the basin of attraction of the stable steady state $\bar{z} = 0$.

As in the previous situation, the relation between the resilience of the stable steady state and the characteristic decay time is given by relation (24).

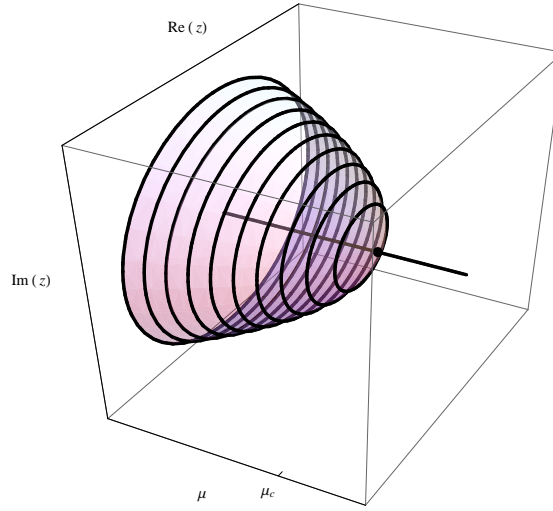


Figure 5: Neimark-Sacker bifurcation diagram. For $\mu < \mu_c$, there is a single stable steady state, surrounded by an unstable invariant circle orbit. At $\mu = \mu_c$, the steady state loses stability and becomes unstable for $\mu \geq \mu_c$.

3 Regime shifts in stochastic systems

For a practical system at this point a problem arises. Assuming that the system is in a stable steady state, we should like to have information about its resilience. Close to bifurcation, the resilience is inversely related to the characteristic decay time. But it is not feasible to subject a large system like, for instance, the earth's climate to a deliberate perturbation to determine the average return time of the climate system to steady state.

3.1 Linear stochastic systems

Fortunately, this is also not necessary as most systems are constantly subjected to small random perturbations. It is appropriate to approximate a nonlinear system

$$x_{t+1} = F_\mu(x_t) + \sigma\eta_t \quad (30)$$

at a steady state \bar{x} of F_μ by the linear stochastic system

$$y_{t+1} = A(\mu)y_t + \sigma\eta_t. \quad (31)$$

Here $A(\mu) = D_x F_\mu(\bar{x})$ and $y_t = x_t - \bar{x}$. The components of the vectors η_t of random variables are assumed to be identically and independently distributed, all of mean 0 and variance 1; the matrix σ models then the correlation structure of the noise.

The stochastic properties of the process $\{y_t\}$ should give information about the matrix A . Again for a one-dimensional system, this appears most readily. Consider

$$y_{t+1} = \lambda y_t + \sigma\eta_t.$$

This generates a linear stationary stochastic process. Its most important characteristic is the autocovariance function $\mathbb{Cov}(y_t, y_s)$, which, by stationarity, is only dependent on the difference $|t - s|$, and which reads as

$$\mathbb{Cov}(y_t, y_s) = \frac{\sigma^2}{1 - \lambda^2} \lambda^{|t-s|},$$

In particular, the variance of y_t is

$$\mathbb{Var}(y_t) = \mathbb{Cov}(y_t, y_t) = \frac{\sigma^2}{1 - \lambda^2}. \quad (32)$$

Also, the first order autocorrelation coefficient

$$\rho_1 = \frac{\mathbb{Cov}(y_t, y_{t+1})}{\mathbb{Var}(y_t)} \quad (33)$$

of the process is equal to λ .

In a saddle-node bifurcation scenario, as the parameter μ approaches the critical value, the value of λ approaches 1, and all autocorrelations will increase indefinitely. As these quantities can be readily estimated from time series, this opens an approach to determine the characteristic decay time and hence the resilience of the steady state in practical situations.

3.2 Model system: stochastic saddle node

To test whether this indeed leads to a method that can predict regime shifts, we simulate the nonlinear stochastic system

$$x_{t+1} = x_t - \mu_t - x_t^2 + \sigma\eta_t, \quad 0 \leq t \leq T - 1, \quad (34)$$

where x_t is the state variable, taking values in the real numbers, where μ_t is a slowly moving parameter given as

$$\mu_t = \mu_0 + \varepsilon t, \quad (35)$$

and where the η_t are normally distributed variables with mean 0 and variance 1. The critical parameter value is $\mu_c = 0$. Figure 6 gives the resulting time series in a (μ, x) -diagram. It tracks the location of the stable steady state closely before shifting towards a different regime.

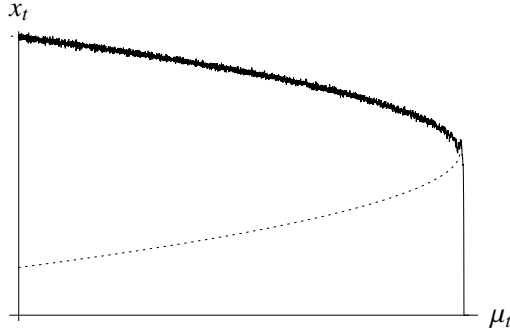


Figure 6: Stochastic fluctuations around a stable steady state in system (34), depicted in the (μ_t, x_t) -diagram. Parameters are $\varepsilon = 0.0001$ and $\sigma = 0.01$.

To obtain an estimate of the autocorrelation coefficient, first the deviation

$$y_t = x_t - \bar{x}(\mu_t) \quad (36)$$

of x_t from the steady state $\bar{x}(\mu) = \sqrt{\mu_c - \mu}$ is computed. As the dynamics of y_t are expected to be well described by the linear system (31), the linear model

$$y_{t+1} = \lambda y_t \quad (37)$$

is fitted to segments of length w of the series $\{y_t\}$. That is, for a segment

$$S(t) = \{y_{t-w+1}, y_{t-w+2}, \dots, y_t\}, \quad (38)$$

the ordinary least squares estimator of the first order autocorrelation coefficient λ in (37) is determined, based on the data in $S(t)$. This yields for each t an estimate $\hat{\lambda}_t$ of λ . As μ_t changes with time, the $\hat{\lambda}_t$ will change with time as well. Figure 7 gives the estimates $\hat{\lambda}_t$, together with their 95% confidence intervals.

The data indicates that $\hat{\lambda}_t$ is increasing with t . But the final value is still far away from 1. Moreover, compared to the quasi-static value

$$\lambda_t = 1 - 2\sqrt{\mu_c - \mu_t} \quad (39)$$

of λ , the estimated value is far off.

Of course, the problem is that the length of the segments used to estimate λ_t has been taken too long. On taking shorter segments, the error bounds increase, but on the other hand they contain the true value of λ_t in the great majority of cases: see Figure 8. There the estimated value of λ increases towards $\lambda = 1$.

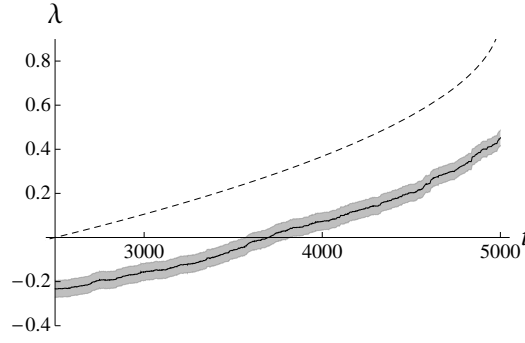


Figure 7: Estimates $\hat{\lambda}_t$ of the first order autocorrelation coefficient λ , with 95% confidence intervals. The dashed curve indicates $\lambda_t = 1 - 2\sqrt{-\mu_t}$. Parameters: $\varepsilon = 0.0001$, $\sigma = 0.01$ and $w = 2500$.

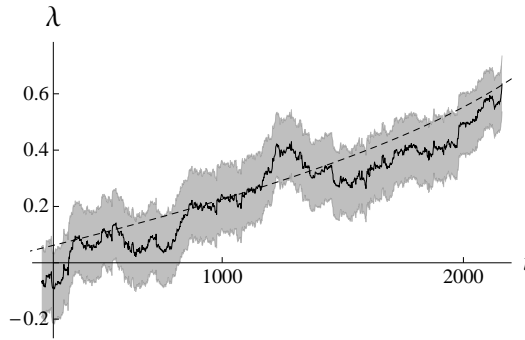


Figure 8: As in figure 7, but with $w = 250$.

3.3 The end of the last glaciation period

Using the same techniques, we examine a temperature time series obtained from analysing the Vostok ice core data. Figure 9 shows a segment of this time series. The sudden warming of the earth climate that started around 17,000 years ago is clearly visible.

A problem here is that the steady state $\bar{x}(\mu)$ is unknown. But this can be resolved by filtering out the high frequency oscillations to obtain a low-frequency component \bar{x}_t of the time series, and then subtracting out the low-frequency component to obtain

$$y_t = x_t - \bar{x}_t. \quad (40)$$

Again a linear model (37) is estimated over segments $S(t)$ of length w . Figure 10 depicts the resulting estimates $\hat{\lambda}_t$.

This seems to result in a sequence that is increasing, and a rank correlation test does amply confirm that. However, the subsequent estimates $\hat{\lambda}_t$ are highly dependent on each other. Moreover, the vertical scale in figure 10 is distorted. Figure 11(a) gives a version with larger scale and error bounds. There the increase in $\hat{\lambda}_t$ is far less pronounced. What is true, however, is that the value of $\hat{\lambda}_t$ is uniformly near the value 1. Together, this suggests that towards the end of the last glaciation period the earth cli-

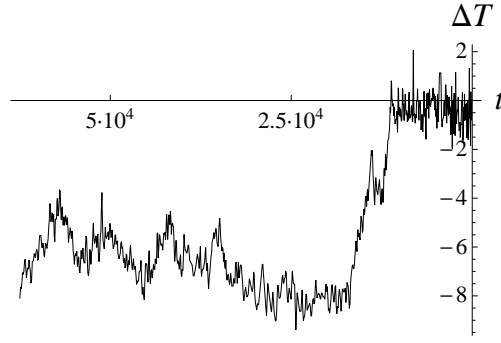


Figure 9: *Temperature dynamics at the end of the last glaciation period. Time is given in years to present, temperature as temperature difference to present day mean temperature. From Petit et al. (2001).*

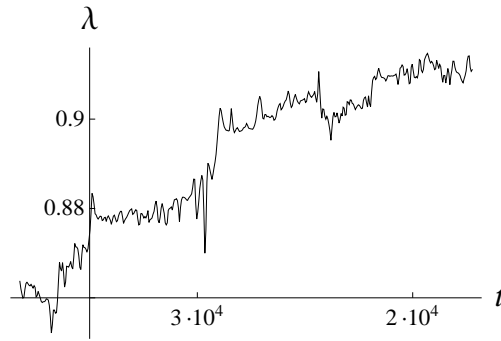


Figure 10: *Estimates $\hat{\lambda}_t$ of the first order autocorrelation coefficient λ . Parameters: $T = 911$ and $w = 455$.*

mate system was in a steady state whose resilience was small. However, the evidence of Figure 11(a) points against the hypothesis that the transition mechanism that led to the present state was a catastrophic shift.

Estimates performed with a smaller time window seem to confirm this, as the increase over time of $\hat{\lambda}_t$ is replaced by oscillatory behaviour, as in Figure 11(b).

3.4 Noise-induced regime shifts

Recall that the resilience of a steady state is the smallest distance of the steady state to a point on the boundary of its basin of attraction. If this quantity is of the same order of magnitude as σ , then there is a sizable probability that a single large shock, or a succession of medium-sized ones, may push the system outside the basin. This is a different mechanism how a system can change regimes; it is commonly called a *noise-induced regime shift*.

Figure 12(a) illustrates such a transition for the model system (34). Figure 12(b) gives the corresponding estimates of $\hat{\lambda}_t$ for times just prior to the shift. Though the

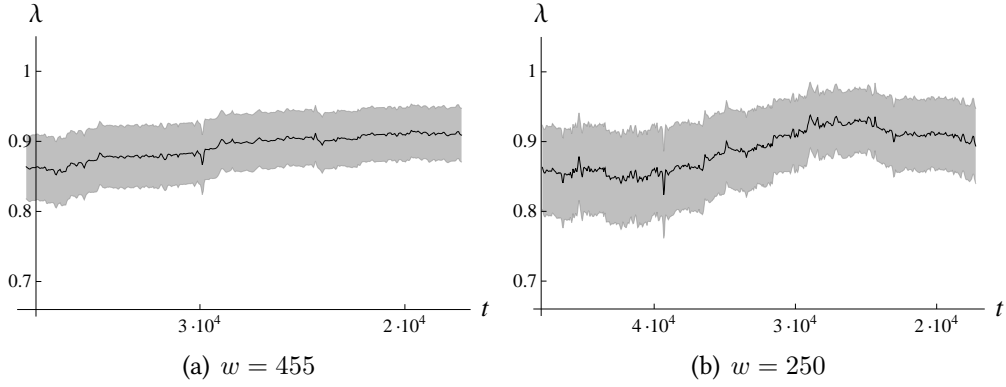


Figure 11: Estimates $\hat{\lambda}_t$ of the first order autocorrelation coefficient λ , together with 95% confidence intervals, for $w \leq t \leq 911$. Left: $w = 455$, right: $w = 250$.

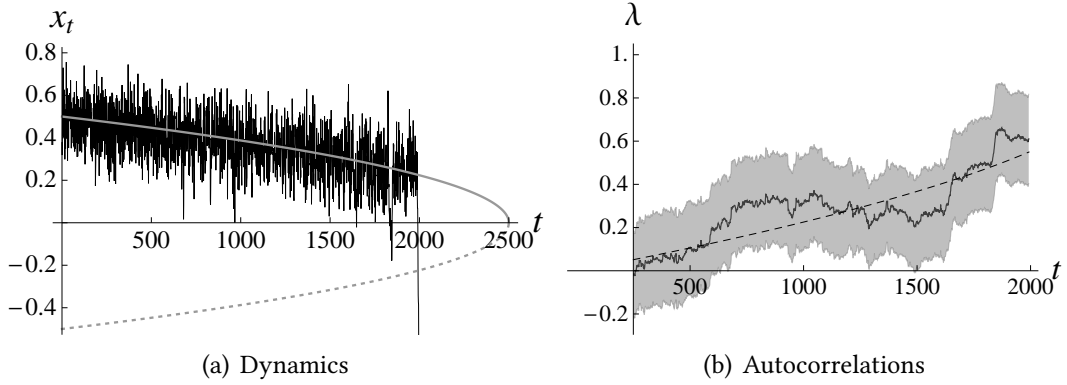


Figure 12: Left: stochastic fluctuations around a stable steady state in system (34), depicted in the (t, x_t) -diagram. A broken curve denotes the location of the unstable steady state that bounds the regime. At the noise-induced transition, the system leaves the regime. Parameters are $\varepsilon = 0.0001$ and $\sigma = 0.1$. Right: autocorrelations with 95% confidence intervals, computed on segments of length $w = 250$.

values of $\hat{\lambda}_t$ increase, fitting and extrapolating a functional relationship of the form

$$\lambda = C\sqrt{\mu_c - (\mu_0 + \varepsilon t)} \quad (41)$$

would result in a prediction for the catastrophic regime shift for values of t about 2500, whereas the noise-induced shift already occurred for $t \approx 2150$.

Comparing Figures 8 and 12(b) to Figures 11(a) and 11(b), it seems probable that the end of the last glaciation period was brought about by a noise-induced transition.

3.5 Global bifurcations

All of the above argumentation presupposes that the system is at or near a steady state. There are however other types of dynamic system behaviour. For instance, a system can oscillate regularly with one or more frequencies: the evolution of the system then

converges to a set in state space, an *attractor*, which is equivalent to a circle or a torus, that is, a Cartesian product of circles. Another possibility is that a system oscillates chaotically: the attracting set is then a *strange attractor*, an intricate fractal set carrying complex dynamics. In this context, *chaotic* refers to the phenomenon that trajectories which have very similar initial conditions almost always diverge from each other and eventually generate wildly different time series: chaotic systems are unpredictable in the long run.

The resilience of these attractors can be defined as before as the smallest perturbation that can push some point on the attractor out of the attractors basin of attraction. As in the case of attracting steady states, the resilience drops down to zero when the attractor touches the boundary of its basin of attraction: the system is then said to go through a *basin-boundary collision* or a *global bifurcation*.

The point relevant for the present discussion is the following: contrary to steady states, general dynamic attractors have non-trivial *internal* dynamics. A system at steady state is at rest, and only changes its state if it is perturbed from its steady state: the autocorrelations of the associated time series give information about the characteristic return time. A system that is on a dynamic attractor is perpetually in motion, even if not perturbed, and it is a hard filtering problem to decompose the time series of such a system in a component corresponding to the internal dynamics and a component that is generated by perturbations away from the attractor; only the latter will give information about the characteristic return time and the resilience of the system. In order to perform this decomposition, the strange attractor and its internal dynamics have to be estimated from the data. To obtain significant estimates in this manner, usually long time series of observations are needed.

A failure to take the possibility of global bifurcations into account can result in unexpected behaviour, treating for instance a time series coming from a strange attractor as if it were generated by a noisy system at a steady state. In the following example, a time series with vanishing first order autocorrelations is generated by the internal dynamics of an attractor in a deterministic, but chaotic, one-dimensional dynamical system. An observer assuming that a noisy system at a steady state generated the series and estimating the autocorrelations will conclude that the system is far from bifurcation. The dynamics is however critical: changing the parameter by a small amount, the chaotic dynamics loses the property of being attracting, and it escapes to a different region. In this case there is no warning signal given by increasing autocorrelations.

The example considers a deterministic evolution

$$x_{t+1} = F_\mu(x_t) \quad (42)$$

on a one-dimensional state space given as

$$F_\mu(x) = \begin{cases} \mu x(1 - x), & x \geq 0, \\ 2 \frac{e^{2\mu x}}{e^{2\mu x} + 1} - 1, & x < 0. \end{cases} \quad (43)$$

Figure 13 represents the evolution law graphically. Recall that intersections of the graph $x_{t+1} = F_\mu(x_t)$ and the line $x_{t+1} = x_t$ yield steady states of the system. The cru-

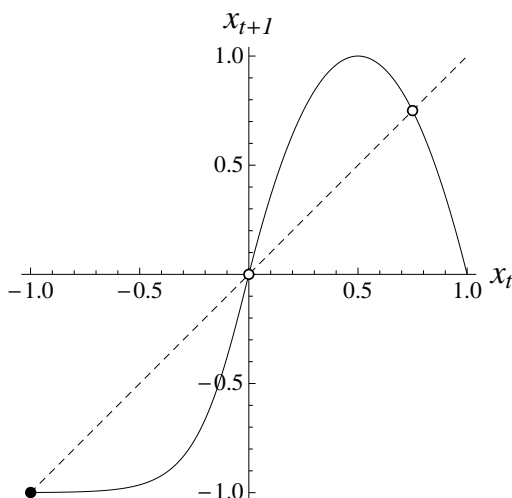


Figure 13: Evolution law (42) for $\mu = 4$. White dots indicate unstable steady states, black dots stable steady states. If $0 \leq x_0 \leq 1$, then $0 \leq x_t \leq 1$ for all $t \geq 0$.

cial property is that for $0 \leq \mu \leq 4$ the local maximum of F_μ at the critical point $c = 1/2$ is smaller than or equal to 1. This implies that if $0 \leq x_0 \leq 1$, then the inequalities $0 \leq x_t \leq 1$ will also hold for all future states x_t . If however $\mu > 4$, there is a small interval I around c such that if $x_{t_0} \in I$ for some t_0 , then $x_{t_0+1} > 1$, $x_{t_0+2} < 0$ and x_t tends to the stable steady state $\bar{x} = -1$.

This is another kind of catastrophic shift. Figure 14 illustrates the corresponding time series: as for $\mu = 4$ the dynamics cannot escape from the interval $[0, 1]$ the oscillations will go on indefinitely in that case. For $\mu > 4$, the dynamics escapes for almost all initial values from the interval $[0, 1]$, sooner or later, and ends up at the stable steady state that is close to $x = -1$.

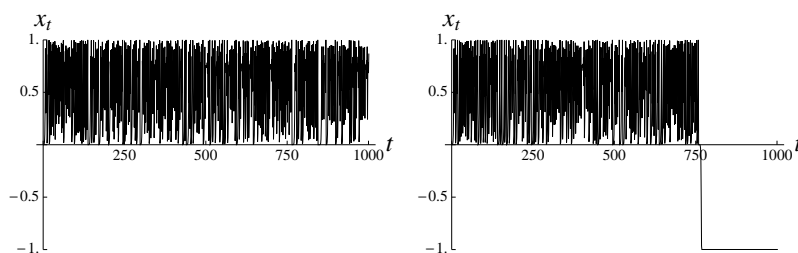


Figure 14: Time series of the deterministic system (42) for $\mu = 4$ (left) and $\mu = 4.00001$ (right). To the casual observer, these time series may seem to be stochastic.

However, estimates of the first order autocorrelation coefficient yield a value of almost zero in both cases, and the series that exhibits a catastrophic shift has no increase in the autocorrelations before the shift. See Figure 15.

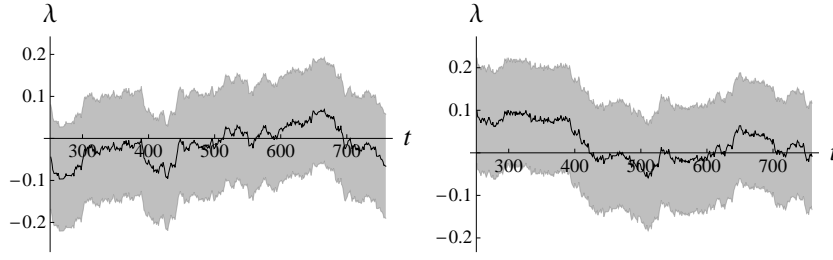


Figure 15: Autocorrelations estimated on the time series of the deterministic system (42) for $\mu = 4$ (left) and $\mu = 4.00001$ (right). Window length $w = 250$.

4 Conclusion

Imminent catastrophic regime shifts may be predicted from the rise of characteristic decay times, or, equivalently, from the increase of first order autocorrelation coefficients towards 1. This indicator by itself is crude: a rise gives a strong indication of an impending regime shift, but it may overestimate the time to the regime change if the parameter changes quickly (Figures 2 and 4), if the estimation windows are large (Figure 7), or if the stochastic perturbations are large (Figures 11 and 12). Moreover, if the indicator is far from the critical value, the system may still be close to a regime shift (Figure 14).

This should not give the impression that this method is impractical. On the contrary, the hallmark of a scientific theory is that it gives strong and testable implications. If estimated autocorrelations exhibit the square-root type of growth of equation (41), this strongly points to a catastrophic regime shift at

$$t_c = \frac{\mu_c - \mu_0}{\varepsilon}. \quad (44)$$

The counterexamples given emphasise that there are other types of regime shifts that are not picked up by the indicator. The glaciation example suggests that other mechanisms may be better explanations for a given shift. Moreover, the absence of a trend in the first order autocorrelation coefficient, or even their vanishing, does not necessarily imply a large resilience of the system.

Considering the possibility of a regime shift in, for instance, the climate system of the earth, it is perhaps worthwhile to point out that there we know that a system parameter – the amount of greenhouse gases in the earth’s atmosphere – is increasing, and that we have data on this increase. This gives more information to a statistical procedure, as the parameter ε , the speed of increase of the system parameters, can be estimated much more precisely. Moreover, the increase is presumably rapid compared to the natural climate dynamics, which makes the probability of a catastrophic regime shift, as opposed to a noise-induced shift, much larger. Based on the techniques presented, a statistical theory of early warning signals for such a shift can be built. However, for shifts that are associated to global bifurcations of more complex attractors than steady states, more sophisticated methods have to be developed.

5 Further reading

Carpenter, S.R. and Brock, W.A., Rising variance: a leading indicator of ecological transition, *Ecology Letters*, 9, 311-318.

Dakos, V., Scheffer, M., van Nes, E.H., Brovkin, V., Petoukhov, V. and Held, H., Slowing down as an early warning signal for abrupt climate change, *PNAS*, 105, 14308-14312, 2008.

Held, H. and Kleinen, T., Detection of climate system bifurcations by degenerate fingerprinting, *Geophysical Research Letters*, 31, L23207, 2004.

Holling, C.S., Resilience and stability of ecological systems, *Annual Review of Ecology and Systematics*, 4, 1-23, 1973.

Petit, J.R., et al., *Vostok Ice Core Data for 420,000 Years*, IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series #2001-076. NOAA/NGDC Paleoclimatology Program, Boulder CO, USA, 2001.

Poston, T. and Stewart, I., *Catastrophe theory and its applications*, London, Pitman, 1978.

Scheffer, M., *Critical transitions in nature and society*, Princeton University Press, Princeton, 2009.

Scheffer, M., Bascompte, J., Brock, W.A., Brovkin, V., Carpenter, S.R., Dakos, V., Held, H., van Nes, E.H., Rietkerk, M. and Sugihara, G., Early-warning signals for critical transitions, *Nature*, 461, 53-69, 2009.

Thompson, J.M.T. and Sieber, J., Climate tipping as a noisy bifurcation: a predictive technique, *IMA Journal of Applied Mathematics*, 76, 27-46, 2011.