

Internal noise interference to warnings of tipping points in generic multi-dimensional dynamical systems

Andreas Morr^{*†‡}, Niklas Boers^{†‡§}, and Peter Ashwin[§]

Abstract. A deterministic dynamical system that slowly passes through a generic fold-type (saddle-node) bifurcation can be reduced to one-dimensional dynamics close to the bifurcation because of the centre manifold theorem. It is often tacitly assumed that the same is true in the presence of stochasticity or noise so that, for example, critical slowing down (CSD) indicators can be applied as if the system were one-dimensional. In this work, we show that this is only true when given suitable system observables; specifically, we demonstrate that noise in other dimensions may interfere with indicators of CSD, also referred to as early warning signals (EWS). We point out a generic mechanism by which both variance and lag-1 autocorrelation [AC(1)], as well as other EWS, can fail to signal an approaching bifurcation. This can, in principle, occur whenever one noise source drives multiple system components simultaneously. Even under the favourable assumptions of uncoupled deterministic dynamics and stationary noise, some system observables can then exhibit false negative or false positive CSD indications. We isolate this phenomenon in an example that represents a generic two-dimensional fold-type bifurcation setting.

Key words. Abrupt transitions, Critical Slowing Down, Early Warning Signals, Bifurcation theory

MSC codes. 37G10, 37H05, 37H20

1. Introduction. Many physical systems exhibit abrupt transitions or are suspected to harbour the potential for them [29, 1, 23, 36, 6, 30]. There has been an increased interest in the analysis and detection of bifurcation-induced transitions [7, 5, 9, 10, 34, 14]. Such events occur if the system’s current equilibrium state destabilises in response to an external change to the underlying dynamics. The simplest model for a generic bifurcation-induced transition is that of a one-dimensional fold-type bifurcation [37, 39, 21]. This approach is reasonable if there is one essential component of the system experiencing negative feedback on its fast time-scale disturbances. Moreover, bifurcation theory tells us that this is one of only two possible generic one-parameter bifurcations of an attracting equilibrium, the other being the Hopf bifurcation. Close enough to a fold-type bifurcation, one can rigorously and generically reduce a multi-dimensional system to a one-dimensional model on a centre manifold. The negative feedback on this essential component weakens in advance of the critical transition brought about by the external change in dynamics. The resulting observational characteristic of slower and weaker responses to perturbations is called critical slowing down (CSD). If the model component undergoing such a bifurcation is a direct observable of the physical system, one can hope to employ the conventional techniques for detecting CSD, i.e. searching for positive trends in variance or lag-1 autocorrelation [AC(1)] in the time-series data. These

*Corresponding author: andreas.morr@tum.de

[†]Earth System Modelling, School of Engineering and Design, Technical University of Munich, 80333 Munich, Germany

[‡]Complexity Science, Potsdam Institute for Climate Impact Research, 14473 Potsdam, Germany

[§]Department of Mathematics and Statistics, University of Exeter, Exeter EX4 4QF, United Kingdom

trends are often referred to as early warning signals (EWS) of an impending tipping point in the literature [24, 25, 29, 8]. These methods are typically motivated by assuming proximity to a fold-type bifurcation of a one-dimensional system whose state is denoted by $x(t) \in \mathbb{R}$ and whose dynamics depend on a parameter α . In such a case, there will generically be a transformation of the dynamics to a local topological bifurcation normal form that, without loss of generality, is at $x = \alpha = 0$ and can be written as:

$$(1.1) \quad \dot{x} = -x^2 + \alpha.$$

For $\alpha > 0$, the system admits one stable equilibrium at $x^*(\alpha) = \sqrt{\alpha}$ and one unstable equilibrium, whereas for $\alpha < 0$, it admits no equilibria.

Complex systems that exhibit a clear separation of dynamical time scales can often be effectively represented by a stochastic differential equation (SDE). This may be motivated, for example, by the projection on macroscopic dynamics [41] or by leveraging the ergodic nature of rapidly varying chaotic system components [31]. Adding Gaussian white noise to model the omitted fast dynamics of the original high-dimensional physical system reduces the problem of determining the statistical characteristics (near a stable equilibrium) to the analysis of a one-dimensional Ornstein-Uhlenbeck process. This is because, in the case of small noise strength σ , the deterministic dynamics around the equilibrium can be represented by their first-order approximation:

$$dX_t = -2\sqrt{\alpha}(X_t - x^*(\alpha))dt + \sigma dW_t,$$

where W is a Wiener process supported on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, \mathbb{P})$. For variance and $\text{AC}(1)$ of the observable X sampled at time interval $\Delta t = 1$, one finds in the stationary limit

$$(1.2) \quad \text{Var}[X] = \frac{\sigma^2}{4\sqrt{\alpha}}, \quad \text{AC}_X(1) = \exp(-2\sqrt{\alpha}),$$

which both increase as $\alpha > 0$ decreases towards the bifurcation at $\alpha = 0$. We may conclude that given time-series observations of a one-dimensional system approaching a fold-type bifurcation, one can reliably detect CSD employing such EWS. However, the coarse-grained dynamics of real-world systems generally consist of multiple coupled dimensions. It is *a priori* not clear that a given system observable contains information regarding CSD and related EWS of an impending bifurcation.

In a multi-dimensional system, we should expect the noise disturbances within the system to drive multiple components simultaneously, meaning that for a “typical” observable being a function of the system state, the noise may interfere with itself and impede any warning of an approach to a tipping point. In particular, we investigate the implications of multiple independent noise disturbances on critical slowing down in observables near a fold-type bifurcation.

In the following [Section 2](#), we unpack some of the tacit assumptions underlying the conventional application of EWS methods. We point out that even though one can isolate the critical component or centre manifold in the multi-dimensional deterministic dynamics, a stochastic coupling of dimensions will persist in general. This will prove to be problematic when

considering non-trivial observables Ψ of the system. [Section 3](#) is dedicated to an analytical derivation of the statistical properties of a linear observable Ψ in the two-dimensional case and an assessment of the potential for “deceitful” warnings. Having identified possible pitfalls to the EWS in variance and AC(1), we check that those predicted behaviours indeed manifest in a numerically integrated non-linear system in [Section 4](#).

2. Centre manifold theory as the motivation for one-dimensional EWS analysis. The fact that a multi-dimensional system with a generic fold-type bifurcation can be transformed to the topological normal form [\(1.1\)](#) is often employed to motivate the use of a single system component in analysing EWS. To make this statement more concrete and to explore its limitations, we invoke the theory of centre manifolds at a bifurcation point (see [\[27, Theorem 5.2\]](#) and [\[20, Theorem 3.4.1\]](#) for a more detailed discussion). Consider a dynamical system whose state is denoted by $\mathbf{z}(t) \in \mathbb{R}^n$, $n \geq 1$, and whose dynamics follow

$$(2.1) \quad \dot{\mathbf{z}} = \mathbf{f}(\mathbf{z}, \alpha),$$

dependent on a parameter $\alpha \in \mathbb{R}$. Let us assume that the right-hand side $\mathbf{f} \in C^\infty(\mathbb{R}^n \times \mathbb{R}, \mathbb{R}^n)$ has a bifurcating equilibrium at (\mathbf{z}_c, α_c) that is close to stability. More precisely, suppose that

$$(2.2) \quad \mathbf{f}(\mathbf{z}_c, \alpha_c) = \mathbf{0}, \quad \sigma[\nabla_{\mathbf{z}} \mathbf{f}(\mathbf{z}_c, \alpha_c)] = \{0\} \cup S,$$

where σ denotes the spectrum of the Jacobian $\nabla_{\mathbf{z}} \mathbf{f}$. Equation [\(2.2\)](#) means we assume this spectrum consists of a single zero and of $n - 1$ eigenvalues $S \subset \mathbb{C}$ (counting multiplicity) with negative real part.

Generically, there will be a fold-type bifurcation at this point on varying α [\[27\]](#). Note that the centre manifold of the system at (\mathbf{z}_c, α_c) extended by a $\dot{\alpha} = 0$ is tangent to the two-dimensional plane spanned by the eigenvector associated with the zero eigenvalue of $\nabla_{\mathbf{z}} \mathbf{f}(\mathbf{z}_c, \alpha_c)$ and the α -direction. This centre manifold is invariant, exponentially attracting and, to linear order, uncoupled from the stable $(n - 1)$ -dimensional eigenspace. This implies that the system [\(2.1\)](#) can be reformulated by taking coordinates $\mathbf{z} = (x, \mathbf{y})$ relative to the eigenspaces for 0 and S and assuming without loss of generality that $(\mathbf{z}_c, \alpha_c) = (\mathbf{0}, 0)$:

$$\begin{aligned} \dot{x} &= ax^2 + b\alpha + \mathcal{O}(2), \\ \dot{\mathbf{y}} &= C\mathbf{y} + \mathcal{O}(2). \end{aligned}$$

Here, all eigenvalues of the $(n - 1) \times (n - 1)$ matrix C have negative real parts, and a and b are normal form coefficients that are generically non-zero. The $\mathcal{O}(2)$ denotes terms that are quadratic in x , \mathbf{y} and α . If a and b are non-zero (this corresponds to the genericity conditions (SN2) and (SN3) in [\[20, Theorem 3.4.1\]](#)) then this can be transformed to

$$(2.3) \quad \begin{aligned} \dot{x} &= ax^2 + b\alpha, \\ \dot{\mathbf{y}} &= C\mathbf{y}. \end{aligned}$$

With addition of noise, [\(2.3\)](#) becomes the Itô SDE

$$(2.4) \quad d\mathbf{Z}_t = \mathbf{f}(\mathbf{Z}_t, \alpha)dt + \Sigma d\mathbf{W}.$$

Because a Wiener distribution is determined purely by its covariance matrix $\Sigma^T \Sigma$, without loss of generality, and by Cholesky decomposition, we can choose

$$\Sigma = \begin{pmatrix} \sigma_x & 0 \\ \boldsymbol{\sigma}_y & \Sigma_y \end{pmatrix}$$

to be a lower triangular matrix. This gives a system

$$(2.5) \quad \begin{aligned} dX_t &= [aX_t^2 + b\alpha]dt + \sigma_x dW_t^{(1)}, \\ d\mathbf{Y}_t &= [C\mathbf{Y}_t]dt + \boldsymbol{\sigma}_y dW_t^{(1)} + \Sigma_y d\mathbf{W}_t^{(2)}, \end{aligned}$$

where σ_x is a scalar, $\boldsymbol{\sigma}_y$ is a vector, Σ_y is a lower triangular $(n-1) \times (n-1)$ matrix, and $W^{(1)}$ and $\mathbf{W}^{(2)}$ are standard Wiener processes that are independent in each component. Note that we choose here a lower triangular Σ so that dX is formulated with a single white-noise term. Later on, we will equivalently choose an upper triangular Σ for analysis. A more general noise model would include a state dependence of Σ . In this work, we restrict ourselves to the case of an approximately constant Σ . The application of the above normal-form theory demands low-amplitude noise.

An important implication of Eq. (2.5) is that, close to the bifurcation, the X coordinate functions as an optimal observable, i.e., it exhibits the expected behaviour with respect to CSD and its expression of EWS approximated in Eq. (1.2). This confirms it is reasonable to expect that there is a destabilising system component which exhibits EWS: [22, 12, 19, 40, 18]. The reduction of the deterministic dynamics seems to offer a valuable simplification: The critical dimension or centre manifold dynamics in X are uncoupled from all other coarse-grained equilibrium dynamics subsumed in \mathbf{Y} under such a transformation. However, this simplification is deceptive, as it does not consider possible couplings in the stochastic fast dynamics. We explore this generic circumstance in more detail for a two-dimensional system.

Consider the case $n = 2$ (so $y \in \mathbb{R}$) and assume $a = -1$, $b = 1$ and $C = -1$ so that:

$$\begin{aligned} \dot{x} &= -x^2 + \alpha, \\ \dot{y} &= -y. \end{aligned}$$

Adding a generic white-noise term yields the following SDE with a non-diagonal (anisotropic) noise coupling matrix:

$$(2.6) \quad \begin{pmatrix} dX_t \\ dY_t \end{pmatrix} = \begin{pmatrix} -X_t^2 + \alpha \\ -Y_t \end{pmatrix} dt + \Sigma \begin{pmatrix} dW_t^{(1)} \\ dW_t^{(2)} \end{pmatrix}.$$

One can arrive at Eq. (2.6) by transforming some original model to show a centre manifold which is isolated from the remaining deterministic dynamics, as we have discussed above. No such isolation can be simultaneously achieved on the side of the stochastic dynamics. One can hope to observe CSD in the X component of the system, as this sub-system is equivalent to the one-dimensional case discussed in Section 1. However, in applications, one may not be free to choose a system observable. An available observable will be of the form $\Psi_t = g(X_t, Y_t)$,

where g could be any non-linear function. We show that already the subset of simple linear observables

$$(2.7) \quad \Psi_t := \cos(\beta)X_t + \sin(\beta)Y_t$$

proves problematic in the context of anticipating critical transitions using EWS. Equation (2.7) covers all possible linear combinations of X_t and Y_t , and the phenomenon we discuss is impervious to an absolute scaling of the observable.

We investigate the linearised system corresponding to Eq (2.6) around $(x^*(\alpha), y^*(\alpha)) = (\sqrt{\alpha}, 0)$. If the noise amplitude is small, this will be an exceptionally good approximation of the full non-linear dynamics, as the system remains largely in the linear regime:

$$(2.8) \quad \begin{pmatrix} dX_t \\ dY_t \end{pmatrix} = \begin{pmatrix} -2\sqrt{\alpha} & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} X_t - x^*(\alpha) \\ Y_t \end{pmatrix} dt + \begin{pmatrix} \sigma_X & c \\ 0 & \sigma_Y \end{pmatrix} \begin{pmatrix} dW_t^{(1)} \\ dW_t^{(2)} \end{pmatrix}.$$

When defining the noise term through the matrix Σ , we are given three degrees of freedom in the form of the respective noise couplings σ_X , σ_Y and c . Note that any matrix with an additional coupling of the white-noise term $dW^{(1)}$ to dY can always be renormalised to an upper-triangular matrix. This is equivalent to the lower-triangular formulation of the noise term in Eq. (2.5). Similarly, any third white-noise term could always be represented in the above model equation containing two white-noise terms. We call c the cross-coupling of the noise, which, as discussed above, will be non-zero in the general case. We will see that introducing only this additional coupling to the model can have grave implications on the ability to detect the approach of a bifurcation through conventional CSD methods.

3. Statistical properties of the observable Ψ under linearised dynamics. To generalise the linear setting to applications beyond the particular parametrisation of the normal form given in the previous section, we now write the 2D-Ornstein-Uhlenbeck process (2.8) in question generally as

$$(3.1) \quad \begin{aligned} \begin{pmatrix} dX_t \\ dY_t \end{pmatrix} &= \begin{pmatrix} -\lambda_X & 0 \\ 0 & -\lambda_Y \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} dt + \begin{pmatrix} \sigma_X & c \\ 0 & \sigma_Y \end{pmatrix} \begin{pmatrix} dW_t^{(1)} \\ dW_t^{(2)} \end{pmatrix} \\ &=: A \begin{pmatrix} X_t \\ Y_t \end{pmatrix} dt + \Sigma \begin{pmatrix} dW_t^{(1)} \\ dW_t^{(2)} \end{pmatrix} \end{aligned}$$

The deterministic linear dynamics are still uncoupled, and we have carried along the cross-coupling c of the noise terms. Without loss of generality, we have also centred the dynamics of the X dimension around 0. In this more abstract formulation, we are interested in the statistical properties of the observable Ψ and their dependence on the value of λ_X . This is because, while approaching the fold-type bifurcation, λ_X will generically decrease towards zero and canonically incur CSD in the X dimension. Whether this CSD is also detectable in the observable Ψ is the main subject of this work. The two-dimensional SDE (3.1) together with the initial condition $(X_0, Y_0)^\top = (0, 0)^\top$ has the following closed-form solution [17]:

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \int_0^t \exp((t-s)A) \Sigma \begin{pmatrix} dW_s^{(1)} \\ dW_s^{(2)} \end{pmatrix}.$$

Consider the covariance matrix V of this zero-mean process, i.e.

$$\begin{aligned} V &:= \mathbb{E} \left[(X_t, Y_t)^\top (X_t, Y_t) \right] \\ &= \mathbb{E} \left[\int_0^t \int_0^t \exp((t-s_1)A) \Sigma \begin{pmatrix} dW_{s_1}^{(1)} \\ dW_{s_1}^{(2)} \end{pmatrix} \begin{pmatrix} dW_{s_2}^{(1)} & dW_{s_2}^{(2)} \end{pmatrix} \Sigma^\top \exp((t-s_2)A)^\top \right] \\ &= \int_0^t \exp((t-s)A) \Sigma \Sigma^\top \exp((t-s)A)^\top ds, \end{aligned}$$

where we have employed the Itô isometry in the last step. Taking the time derivative on both sides and investigating the stationary limit yields

$$\frac{d}{dt}V = AV + VA^\top + \Sigma \Sigma^\top = 0.$$

For the system in (3.1), this so-called Lyapunov matrix-equation [2] is solved by

$$V = \begin{pmatrix} \frac{\sigma_X^2 + c^2}{2\lambda_X} & \frac{\sigma_Y c}{\lambda_X + \lambda_Y} \\ \frac{\sigma_Y c}{\lambda_X + \lambda_Y} & \frac{\sigma_Y^2}{2\lambda_Y} \end{pmatrix}.$$

Adopting the initial condition $(X_0, Y_0)^\top \sim \mathcal{N}(0, V)$, the process is stationary, and the time-covariance matrix for all $\tau \geq 0$ can be similarly deduced as

$$R(\tau) := \mathbb{E} \left[(X_t, Y_t)^\top (X_{t+\tau}, Y_{t+\tau}) \right] = \exp(\tau A) V.$$

Note that $R(\tau)$ for $\tau \neq 0$ will in general not be a symmetric matrix, since $\mathbb{E}[X_t Y_{t+\tau}] = \mathbb{E}[X_{t+\tau} Y_t]$ if and only if $c = 0$ or $\lambda_X = \lambda_Y$:

$$R(\tau) = \begin{pmatrix} \frac{\sigma_X^2 + c^2}{2\lambda_X} \exp(-\lambda_X |\tau|) & \frac{\sigma_Y c}{\lambda_X + \lambda_Y} \exp(-\lambda_X |\tau|) \\ \frac{\sigma_Y c}{\lambda_X + \lambda_Y} \exp(-\lambda_Y |\tau|) & \frac{\sigma_Y^2}{2\lambda_Y} \exp(-\lambda_Y |\tau|) \end{pmatrix}.$$

The same holds true for the time-correlation matrix with entries $r(\tau)_{i,j} := R(\tau)_{i,j} / \sqrt{V_{i,i} V_{j,j}}$:

$$r(\tau) = \begin{pmatrix} \exp(-\lambda_X |\tau|) & \frac{2c}{\sqrt{\sigma_X^2 + c^2}} \frac{\sqrt{\lambda_X \lambda_Y}}{\lambda_X + \lambda_Y} \exp(-\lambda_X |\tau|) \\ \frac{2c}{\sqrt{\sigma_X^2 + c^2}} \frac{\sqrt{\lambda_X \lambda_Y}}{\lambda_X + \lambda_Y} \exp(-\lambda_Y |\tau|) & \exp(-\lambda_Y |\tau|) \end{pmatrix}.$$

For any linear observable $\Psi(\beta) = \cos(\beta)X + \sin(\beta)Y$ of the 2D-system, we may now compute the variance and the autocorrelation function in the stationary limit:

$$\text{Var}[\Psi] = \cos^2(\beta)V_{1,1} + \sin^2(\beta)V_{2,2} + 2\cos(\beta)\sin(\beta)V_{1,2},$$

where $V_{i,j}$ is the respective entry of the covariance matrix V . For the autocorrelation in time, we have:

$$\begin{aligned} \text{AC}_\Psi(1) &:= \mathbb{E}[\Psi_t \Psi_{t+\tau}] / \text{Var}[\Psi] \\ &= (\cos^2(\beta)R(\tau)_{1,1} + \sin^2(\beta)R(\tau)_{2,2} + \cos(\beta)\sin(\beta)(R(\tau)_{1,2} + R(\tau)_{2,1})) / \text{Var}[\Psi]. \end{aligned}$$

As expected from the uncoupled nature of the deterministic dynamics, if we choose $\beta = 0$, i.e. $\Psi = X$, we essentially retrieve the conventional quantities from the one-dimensional case. This observable should be considered as the optimal observable of the system. Also, not surprisingly, if we only inspect the second dimension $\Psi = Y$, the variance and $\text{AC}(1)$ will not depend on the destabilisation of X , i.e. $\lambda_X \rightarrow 0$, and can not function as EWS. See [Table 1](#) for explicit expressions.

Table 1: Edge cases of the linear observable setting. The conventional EWS information is perfectly conserved in the first case and entirely lost in the second.

β	Ψ	$\text{Var}[\Psi]$	$\text{AC}_\Psi(1)$
0	X	$(\sigma_X^2 + c^2)/2\lambda_X$	$\exp(-\lambda_X)$
$\pi/2$	Y	$\sigma_Y^2/2\lambda_Y$	$\exp(-\lambda_Y)$

In all other cases, the observable Ψ consists of some amplitude of X and Y , i.e. $\beta \neq k\pi/2$ for $k \in \mathbb{Z}$. Even here, the EWS seem to be reliable at first glance. We have

$$\text{Var}[\Psi] \xrightarrow{\lambda_X \rightarrow 0} \infty, \quad \text{AC}_\Psi(1) \xrightarrow{\lambda_X \rightarrow 0} 1.$$

These increases are monotonic in proximity to the bifurcation, where $\lambda_X = 0$. More precisely, for any parameter setting $\lambda_Y > 0$ and $\sigma_X, \sigma_Y, c \neq 0$, there exists a $\lambda_X^* > 0$ with the property that both variance and $\text{AC}(1)$ are monotonically increasing as λ_X decreases from λ_X^* to 0. In that sense, there still is a time span in the advent of the fold-type bifurcation, in which the linearised dynamics predict a clear increase in both variance and $\text{AC}(1)$, making them applicable as EWS in this case.

However, the practical caveat to this finding is the value of λ_X^* and the behaviour of the EWS before this threshold. We give the following example, which illustrates that using EWS has potential pitfalls and may lead to a misinterpretation of the actual destabilisation in the underlying dynamics. For this, choose $\lambda_Y = 1$, $\sigma_X = 0.1$, $\sigma_Y = 2$ and $c = 1$, meaning

$$(3.2) \quad \begin{pmatrix} dX_t \\ dY_t \end{pmatrix} = \begin{pmatrix} -\lambda_X & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} dt + \begin{pmatrix} 0.1 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} dW_t^{(1)} \\ dW_t^{(2)} \end{pmatrix},$$

and examine the observable $\Psi(\beta)$ associated with the mixing parameter $\beta = -\frac{\pi}{4}$, i.e. $\Psi_t = 2^{-1/2}(X_t - Y_t)$. [Figure 1](#) shows the variance and $\text{AC}(1)$ of Ψ as a function of the destabilising eigenvalue λ_X . Although the expected increasing behaviour can be observed when decreasing λ_X from $\lambda_X^* \approx 0.6$ towards 0, both CSD indicators, in fact, decrease as λ_X decreases toward λ_X^* . It is this veiling of the critical slowing down of the X dimension that we would like to point out in this work.

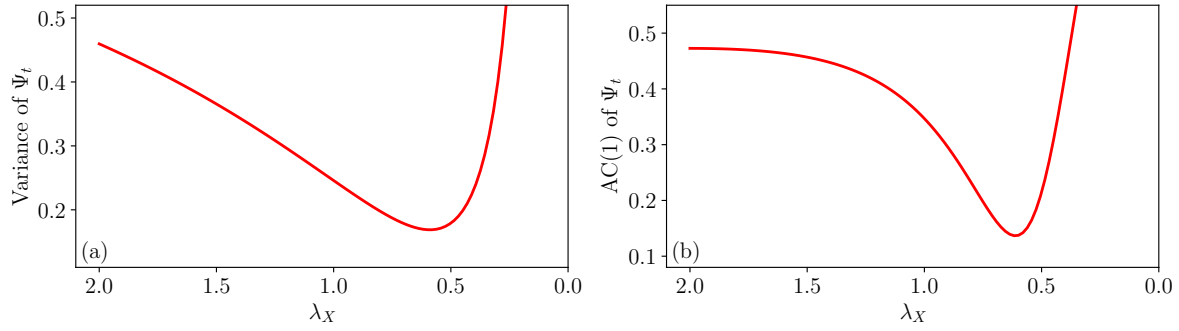


Figure 1: **(a)** Variance and **(b)** $AC(1)$ of the observable Ψ for $\beta = -\frac{\pi}{4}$ in the linearised setting of Eq. (3.2) described in the main text. While these two classical CSD indicators ultimately increase to their expected limits ∞ and 1 respectively as $\lambda_X \rightarrow 0$, the behaviour up until that increase can be deceitful with respect to critical slowing down.

The underlying prerequisite of the phenomenon explored above is the following configuration: The noise component $dW^{(2)}$ is positively coupled to both dimensions of the deterministic dynamics. At the same time, the observable Ψ is defined such that the two linear components have opposing signs. This leads to the disturbances in X and Y interfering with each other in the summary observable Ψ . The appendix gives a result stating that this effect is ubiquitous with respect to the specific parameter choices. Theorem A.1 implies that, as long as we have a cross-coupling $c \neq 0$, the phenomenon persists for all generic choices of σ_X , σ_Y , c and λ_Y , and it can occur in arbitrarily small proximity to the bifurcation point. On the other hand, the extent to which it quantitatively inhibits the use of EWS in an application setting will heavily depend on these values.

Giving a general guideline of which configurations exhibit a lengthy and pronounced decrease in variance and $AC(1)$, even though CSD would imply an increase, is not feasible given the number of degrees of freedom. We provide a Python script with which the prevalence of the effect can be assessed for different parametrisations of the discussed two-dimensional model. Central to this script is the plotting of the approach to the critical value $\lambda_X^{\text{crit}} = 0$ for different values of β and λ_X (see Figure 2 and its description).

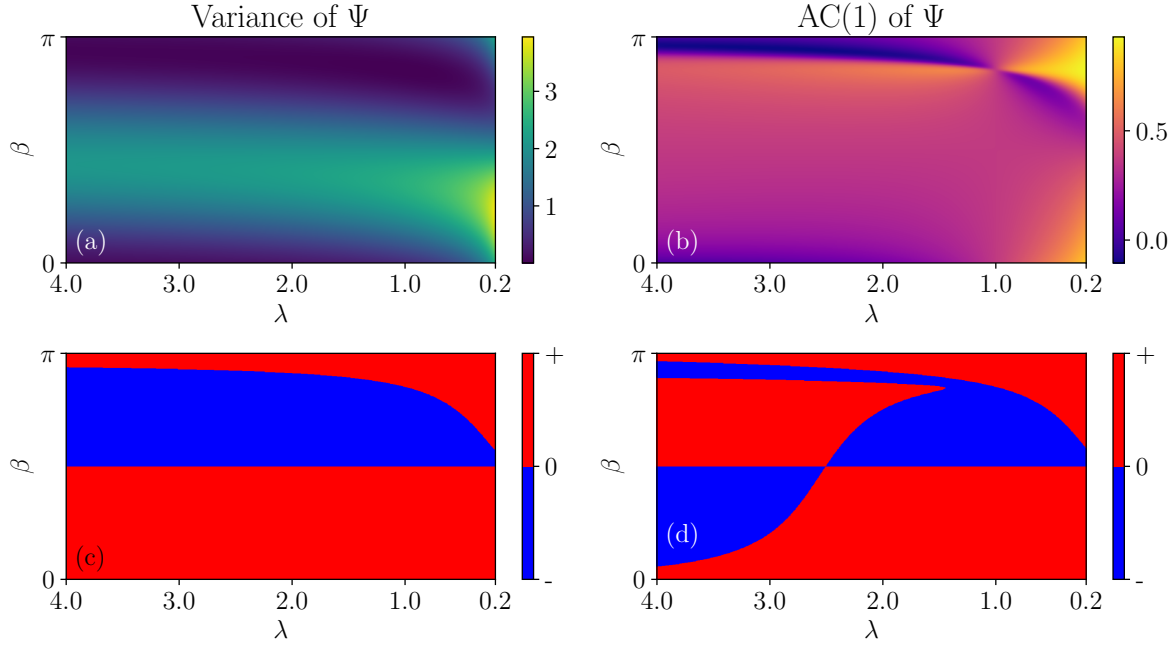


Figure 2: **(a)** variance and **(b)** AC(1) of Ψ for different values of β and λ_X in the 2D-Ornstein-Uhlenbeck setting of Eq. (3.2). For $\beta = \frac{3\pi}{4}$ one obtains Figure 1, since variance and AC(1) of Ψ are unchanged by a transformation $\beta + k\pi$, $k \in \mathbb{Z}$. In **(c)** and **(d)**, the colour signifies the trend of variance and AC(1), respectively, as one decreases λ_X from left to right during CSD. The plot is red, where this trend is positive, and blue, where it is negative. In this sense, all blue regions represent the values of β and λ_X where deceitful CSD indications are prevalent.

4. Observations in model data. After identifying potential pitfalls to employing variance and AC(1) on linear observables of the linearised system (2.8), we test these predictions on synthetic data of the true fold-type bifurcation in Eq. (2.6). To this end, we integrate the following system of SDEs using the Euler-Mayurama scheme:

$$(4.1) \quad \begin{pmatrix} dX_t \\ dY_t \end{pmatrix} = \begin{pmatrix} -X_t^2 + \alpha(t) \\ -Y_t \end{pmatrix} dt + \varepsilon \begin{pmatrix} 0.1 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} dW_t^{(1)} \\ dW_t^{(2)} \end{pmatrix},$$

$$\alpha(t) = 1 - \frac{11}{10} \frac{t}{T}.$$

Over the time span of $T = 10^4$, this integration is performed at a time-step of $\delta t = 1/30$ and subsequently sampled at every 30th value. In order to ensure that premature noise-induced tipping is restricted to a relatively short period before the bifurcation, we introduce the noise-scaling parameter $\varepsilon = 0.1$. For better illustration, we artificially introduce a second stable state for X past the bifurcation point. These dynamics are omitted in the SDE (4.1). The

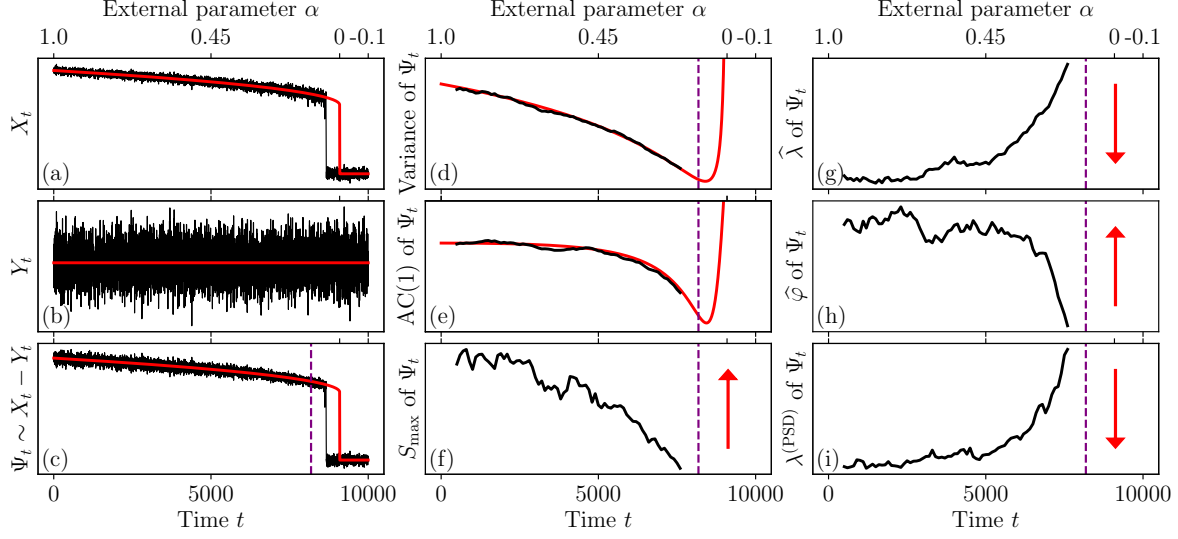


Figure 3: **(a) - (c)** Sample paths of X_t , Y_t and Ψ_t following the SDE (4.1) in the main text with $\beta = -\frac{\pi}{4}$. Up until the purple dashed line, after which noise-induced tipping may frequently come into play, the variance and AC(1) of Ψ are each estimated over windows of length $N = 1000$. The averaged results over 20 runs are plotted in **(d)** and **(e)**, respectively, along with the theoretically computed values from the linearised model. Their decrease would be misinterpreted as a stabilisation of the system rather than an approach to an abrupt transition. In **(f) - (i)**, alternative indicators for CSD are calculated. These are **(f)** the maximum of the observed power spectral density (PSD) [9], **(g)** the linear restoring rate estimated through a Langevin approach [33], **(h)** a generalised least squares estimator for an AR(1) model driven by correlated noise [5] and **(i)** the linearised restoring rate estimated via the PSD [32]. These quantities, too, exhibit the respective opposite trend of what would be expected of a destabilising system (indicated by the red arrows).

results for the linear observable

$$\Psi_t = \cos(\beta)X_t + \sin(\beta)Y_t = \frac{1}{\sqrt{2}}(X_t - Y_t)$$

with $\beta = -\frac{\pi}{4}$ are given in Figure 3. Even though the dynamics are now non-linear and non-autonomous, the contemporary linearised system still allows for a good assessment of the expected time-series statistics. The corresponding estimators of variance and AC(1) indeed give the expected results. This is due to the relatively small noise strength, which keeps the system close to the equilibrium point and thus in an approximately linear regime. The non-autonomous nature of the system is negligible, as the rate of change introduced in $\alpha(t)$ is slow in comparison with the internal variability induced by the driving noise and the time scales of the restoring dynamics. The investigated time series windows are short enough with respect to the change in α , such that the true variance and AC(1) are approximately constant during

that time interval. The quantities decrease with time, even though the destabilisation of the system in the dimension of the X component should conventionally incur an increase. This is caused by an interference of the shared driving noise in the observable Ψ . Note that similarly, a stabilisation of the system, i.e. an increase in λ_X , would cause the variance and $AC(1)$ to increase. This is a setting in which a false alarm with respect to CSD would be raised.

5. Discussion. We consider a generic fold-type bifurcation in a class of multi-dimensional dynamical systems, where we assume that noise drives multiple components of the system. Even though the centre manifold theorem yields one distinct, deterministically uncoupled dimension exhibiting pure CSD, this optimal observable will generally not be available for analysis. We demonstrate that such a setup can easily produce observables that, because of internal noise interference, exhibit deceiving trends in variance and lag-1 autocorrelation with respect to CSD. False or missed alarms in EWS have been shown to be an issue in systems driven by non-stationary or non-white noise [16, 26, 5, 8, 35, 32, 33] or systems exhibiting deterministic coupling [9]. It has also been shown that a misrepresentation of the natural system as a fold-type bifurcation with direct noise influence [4, 28, 13] or issues with the availability of time series data [15, 3, 38] may lead to similar problems. This work introduces a previously unknown complication with the detection of CSD via variance and $AC(1)$, as well as several other techniques introduced in Refs. [9, 33, 5, 32, 11] (see also Figure 3f-i). Even though the severity of the effect described here depends heavily on the relative values of the system’s restoring rates and noise couplings, its presence is ubiquitous (see Theorem A.1). We have also shown that in close proximity to the bifurcation point, EWS can still be expected to deliver information about the impending transition. However, since non-zero noise always incurs the possibility of noise-induced tipping well before the actual vanishing of a contemporary equilibrium, the discussed complication poses a serious challenge to the reliability of EWS [1]. Furthermore, if present, the effect can only be ruled out by choice of a different observable Ψ , which is often not achievable in the observations of natural systems.

Whenever possible, a sufficient physical understanding and previous model validation should confirm that the conventional assumption of a directly observed noisy dynamical system along its centre manifold is warranted. In this case, the presented results do not call into question the expressiveness of conventional EWS.

Data Availability. Visit the GitHub repository [LinearObservablesCSD](#) to access the code used to generate all figures in this manuscript.

Acknowledgments. This project has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No. 101137601 (ClimTip). This is ClimTip contribution #14. N.B. acknowledges further funding from the Volkswagen Stiftung and the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956170. Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. For the purpose of open access, the authors have applied a Creative

Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

REFERENCES

- [1] P. ASHWIN, S. WIECZOREK, R. VITOLO, AND P. COX, *Tipping points in open systems: bifurcation, noise-induced and rate-dependent examples in the climate system*, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 370 (2012), pp. 1166–1184, <https://doi.org/10.1098/rsta.2011.0306>.
- [2] S. BARNETT AND C. STOREY, *Some Applications of the Lyapunov Matrix Equation*, IMA Journal of Applied Mathematics, 4 (1968), pp. 33–42, <https://doi.org/10.1093/imamat/4.1.33>.
- [3] M. BEN-YAMI, V. SKIBA, S. BATHIANY, AND N. BOERS, *Uncertainties in critical slowing down indicators of observation-based fingerprints of the Atlantic Overturning Circulation*, Nature Communications, 14 (2023), p. 8344, <https://doi.org/10.1038/s41467-023-44046-9>.
- [4] M. C. BOERLIJST, T. OUDMAN, AND A. M. DE ROOS, *Catastrophic collapse can occur without early warning: Examples of silent catastrophes in structured ecological models*, PLOS ONE, 8 (2013), pp. 1–6, <https://doi.org/10.1371/journal.pone.0062033>.
- [5] N. BOERS, *Observation-based early-warning signals for a collapse of the Atlantic Meridional Overturning Circulation*, Nature Climate Change, 11 (2021), pp. 680–688, <https://doi.org/10.1038/s41558-021-01097-4>.
- [6] N. BOERS, M. GHIL, AND T. F. STOCKER, *Theoretical and paleoclimatic evidence for abrupt transitions in the Earth system*, Environmental Research Letters, 17 (2022), <https://doi.org/10.1088/1748-9326/ac8944>.
- [7] N. BOERS, N. MARWAN, H. M. J. BARBOSA, AND J. KURTHS, *A deforestation-induced tipping point for the South American monsoon system*, Scientific Reports, 7 (2017), <https://doi.org/10.1038/srep41489>.
- [8] C. BOETTNER AND N. BOERS, *Critical slowing down in dynamical systems driven by nonstationary correlated noise*, Physical Review Research, 4 (2022), <https://doi.org/10.1103/PhysRevResearch.4.013230>.
- [9] T. M. BURY, C. T. BAUCH, AND M. ANAND, *Detecting and distinguishing tipping points using spectral early warning signals*, Journal of The Royal Society Interface, 17 (2020), <https://doi.org/10.1098/rsif.2020.0482>.
- [10] T. M. BURY, R. I. SUJITH, I. PAVITHRAN, M. SCHEFFER, T. M. LENTON, M. ANAND, AND C. T. BAUCH, *Deep learning for early warning signals of tipping points*, Proceedings of the National Academy of Sciences, 118 (2021), <https://doi.org/10.1073/pnas.2106140118>.
- [11] J. J. CLARKE, C. HUNTINGFORD, P. D. L. RITCHIE, AND P. M. COX, *Seeking more robust early warning signals for climate tipping points: the ratio of spectra method (ROSA)*, Environmental Research Letters, 18 (2023), <https://doi.org/10.1088/1748-9326/acbc8d>.
- [12] V. DAKOS, *Identifying best-indicator species for abrupt transitions in multispecies communities*, Ecological Indicators, 94 (2018), pp. 494–502, <https://doi.org/10.1016/j.ecolind.2017.10.024>.
- [13] V. DAKOS, E. H. VAN NES, P. D’ODORICO, AND M. SCHEFFER, *Robustness of variance and autocorrelation as indicators of critical slowing down*, Ecology, 93 (2012), pp. 264–271, <https://doi.org/10.1890/11-0889.1>.
- [14] P. DITLEVSEN AND S. DITLEVSEN, *Warning of a forthcoming collapse of the Atlantic meridional overturning circulation*, Nature Communications, 14 (2023), p. 4254, <https://doi.org/10.1038/s41467-023-39810-w>.
- [15] P. D. DITLEVSEN AND S. J. JOHNSEN, *Tipping points: Early warning and wishful thinking*, Geophysical Research Letters, 37 (2010), <https://doi.org/10.1029/2010GL044486>.
- [16] P. S. DUTTA, Y. SHARMA, AND K. C. ABBOTT, *Robustness of early warning signals for catastrophic and non-catastrophic transitions*, Oikos (Copenhagen, Denmark), 127 (2018), pp. 1251–1263, <https://doi.org/10.1111/oik.05172>.
- [17] C. W. GARDINER, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, vol. 13 of Springer Series in Synergetics, Springer Berlin Heidelberg, 1985, <https://doi.org/10.1007/978-3-662-02452-2>.

- [18] S. V. GEORGE, S. KACHHARA, AND G. AMBIKA, *Early warning signals for critical transitions in complex systems*, Physica Scripta, 98 (2023), p. 072002, <https://doi.org/10.1088/1402-4896/acde20>.
- [19] A. GHADAMI, S. CHEN, AND B. I. EPUREANU, *Data-driven identification of reliable sensor species to predict regime shifts in ecological networks*, Royal Society Open Science, 7 (2020), <https://doi.org/10.1098/rsos.200896>.
- [20] J. GUCKENHEIMER AND P. HOLMES, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, vol. 42 of Applied Mathematical Sciences, Springer New York, 1983, <https://doi.org/10.1007/978-1-4612-1140-2>.
- [21] V. GUTTAL AND C. JAYAPRAKASH, *Impact of noise on bistable ecological systems*, Ecological Modelling, 201 (2007), pp. 420–428, <https://doi.org/10.1016/j.ecolmodel.2006.10.005>.
- [22] H. HELD AND T. KLEINEN, *Detection of climate system bifurcations by degenerate fingerprinting*, Geophysical Research Letters, 31 (2004), <https://doi.org/10.1029/2004GL020972>.
- [23] P. O. HOPCROFT AND P. J. VALDES, *Paleoclimate-conditioning reveals a North Africa land-atmosphere tipping point*, Proceedings of the National Academy of Sciences, 118 (2021), <https://doi.org/10.1073/pnas.2108783118>.
- [24] C. KUEHN, *A mathematical framework for critical transitions: Bifurcations, fast-slow systems and stochastic dynamics*, Physica D: Nonlinear Phenomena, 240 (2011), pp. 1020–1035, <https://doi.org/10.1016/j.physd.2011.02.012>.
- [25] C. KUEHN, *A mathematical framework for critical transitions: Normal forms, variance and applications*, Journal of Nonlinear Science, 23 (2013), pp. 457–510, <https://doi.org/10.1007/s00332-012-9158-x>.
- [26] C. KUEHN, K. LUX, AND A. NEAMȚU, *Warning signs for non-Markovian bifurcations: colour blindness and scaling laws*, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 478 (2022), <https://doi.org/10.1098/rspa.2021.0740>.
- [27] Y. A. KUZNETSOV, *Elements of Applied Bifurcation Theory*, vol. 112 of Applied Mathematical Sciences, Springer New York, 2004, <https://doi.org/10.1007/978-1-4757-3978-7>.
- [28] S. KÉFI, V. DAKOS, M. SCHEFFER, E. H. VAN NES, AND M. RIETKERK, *Early warning signals also precede non-catastrophic transitions*, Oikos (Copenhagen, Denmark), 122 (2013), pp. 641–648, <https://doi.org/10.1111/j.1600-0706.2012.20838.x>.
- [29] T. M. LENTON, V. N. LIVINA, V. DAKOS, E. H. VAN NES, AND M. SCHEFFER, *Early warning of climate tipping points from critical slowing down: comparing methods to improve robustness*, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 370 (2012), pp. 1185–1204, <https://doi.org/10.1098/rsta.2011.0304>.
- [30] C. MEISEL, A. KLAUS, C. KUEHN, AND D. PLENZ, *Critical slowing down governs the transition to neuron spiking*, PLOS Computational Biology, 11 (2015), pp. 1–20, <https://doi.org/10.1371/journal.pcbi.1004097>.
- [31] I. MELBOURNE AND A. M. STUART, *A note on diffusion limits of chaotic skew-product flows*, Nonlinearity, 24 (2011), pp. 1361–1367, <https://doi.org/10.1088/0951-7715/24/4/018>.
- [32] A. MORR AND N. BOERS, *Detection of Approaching Critical Transitions in Natural Systems Driven by Red Noise*, Physical Review X, 14 (2024), <https://doi.org/10.1103/PhysRevX.14.021037>.
- [33] A. MORR, K. RIECHERS, L. R. GORJÃO, AND N. BOERS, *Anticipating critical transitions in multidimensional systems driven by time- and state-dependent noise*, Physical Review Research, 6 (2024), p. 033251, <https://doi.org/10.1103/PhysRevResearch.6.033251>.
- [34] F. NAZARIMEHR, S. JAFARI, M. PERC, AND J. C. SPROTT, *Critical slowing down indicators*, Europhysics Letters, 132 (2020), p. 18001, <https://doi.org/10.1209/0295-5075/132/18001>.
- [35] D. PROVERBIO, A. SKUPIN, AND J. GONÇALVES, *Systematic analysis and optimization of early warning signals for critical transitions using distribution data*, iScience, 26 (2023), p. 107156, <https://doi.org/10.1016/j.isci.2023.107156>.
- [36] H. REN AND D. WATTS, *Early warning signals for critical transitions in power systems*, Electric Power Systems Research, 124 (2015), pp. 173–180, <https://doi.org/10.1016/j.epsr.2015.03.009>.
- [37] M. SCHEFFER, J. BASCOMPTE, W. A. BROCK, V. BROVKIN, S. R. CARPENTER, V. DAKOS, H. HELD, E. H. VAN NES, M. RIETKERK, AND G. SUGIHARA, *Early-warning signals for critical transitions*, Nature, 461 (2009), pp. 53–59, <https://doi.org/10.1038/nature08227>.
- [38] T. SMITH, R.-M. ZOTTA, C. A. BOULTON, T. M. LENTON, W. DORIGO, AND N. BOERS, *Reliability of resilience estimation based on multi-instrument time series*, Earth System Dynamics, 14 (2023),

- pp. 173–183, <https://doi.org/10.5194/esd-14-173-2023>.
- [39] J. M. T. THOMPSON AND J. SIEBER, *Climate tipping as a noisy bifurcation: a predictive technique*, IMA Journal of Applied Mathematics, 76 (2010), pp. 27–46, <https://doi.org/10.1093/imamat/hxq060>.
- [40] E. VAN NES AND M. SCHEFFER, *Slow Recovery from Perturbations as a Generic Indicator of a Nearby Catastrophic Shift*, The American Naturalist, 169 (2007), pp. 738–747, <https://doi.org/10.1086/516845>.
- [41] R. ZWANZIG, *Nonequilibrium statistical mechanics*, Oxford University Press, 2001, <https://doi.org/10.1093/oso/9780195140187.001.0001>.

Appendix A. Noise interference of early warning signals in a 2D linear SDE. For completeness, we state and prove the following result about the ubiquity of noise interference to early warning signs.

Theorem A.1. *Consider the linear SDE (3.1) with non-zero constants $\lambda_Y > 0$, σ_X , σ_Y and c . Choose any small $\delta > 0$ and consider the decrease of λ_X from λ_Y down to δ , representing a destabilisation of the system. There exists an open interval of mixing parameters $\beta \in (\beta^1, \beta^2)$ for which both variance and $AC(1)$ of Ψ defined through Eq. (2.7) decrease with decreasing λ_X .*

Proof. We can find such an interval of mixing parameters $\beta \in (\beta^1, \beta^2)$ close to the choice $\Psi = Y$, i.e. $\beta = \pi/2$. We have already derived analytical expressions of $\text{Var}[\Psi]$ and $AC_\Psi(1)$. They are smooth functions of the system’s parameters. The quantities decrease with a decreasing λ_X if their partial derivative with respect to λ_X is positive. We find at the trivial choice of $\Psi = Y$ that for all $\lambda_X \in \mathbb{R}$

$$(A.1) \quad \left. \partial_{\lambda_X} \text{Var}[\Psi] \right|_{\beta=\pi/2} = 0 \quad \text{and} \quad \left. \partial_{\lambda_X} AC_\Psi(1) \right|_{\beta=\pi/2} = 0.$$

We find that the derivative of these quantities with respect to the mixing parameter β is non-zero and equal in sign for both of the quantities:

$$\begin{aligned} \left. \partial_\beta \partial_{\lambda_X} \text{Var}[\Psi] \right|_{\beta=\pi/2} &= \frac{-2c\sigma_Y}{(\lambda_X + \lambda_Y)^2} \\ \left. \partial_\beta \partial_{\lambda_X} AC_\Psi(1) \right|_{\beta=\pi/2} &= \frac{-2c\lambda_Y (e^{-\lambda_X} (1 + \lambda_X + \lambda_Y) - e^{-\lambda_Y})}{(\lambda_X + \lambda_Y)^2 \sigma_Y} \end{aligned}$$

This sign is equal to $-1 \cdot \text{sgn}(c\sigma_Y)$, and it is uniform across all choices of $\lambda_X \leq \lambda_Y$. Consider without loss of generality a positive sign. Then for the compact interval $\lambda_X \in [\delta, \lambda_Y]$ we can find $\gamma_\delta > 0$ such that the quantities in Eq. (A.1) are positive on the interval $(\pi/2, \pi/2 + \gamma_\delta)$. The counterfactual to this statement implies the existence of a value $\hat{\lambda}_X$ with, e.g.,

$$\left. \partial_{\lambda_X} \text{Var}[\Psi] \right|_{\beta=\pi/2, \lambda_X=\hat{\lambda}_X} = 0 \quad \text{and} \quad \left. \partial_\beta \partial_{\lambda_X} \text{Var}[\Psi] \right|_{\beta=\pi/2, \lambda_X=\hat{\lambda}_X} > 0$$

while

$$\max_{\beta \in [\pi/2, \gamma]} \left\{ \left. \partial_{\lambda_X} \text{Var}[\Psi] \right|_{\beta=\pi/2, \lambda_X=\hat{\lambda}_X} \right\} = 0 \text{ for some } \gamma > 0,$$

a contradiction. For a negative sign, an analogous interval $(\pi/2 - \gamma_\delta, \pi/2)$ can be found.

This interval constitutes the desired range of mixing parameters. Note that the chosen interval is not the only region for which the phenomenon discussed in this work is prevalent (see e.g. [Figure 2c](#) and d). \square