Supplementary Information

Self-supervised Reservoir Computing with Spatial-temporal Encoding for Identifying Critical Transitions

Contents

Section S1. Supplementary Figures

Supplementary Figure S1. The sliding window scheme

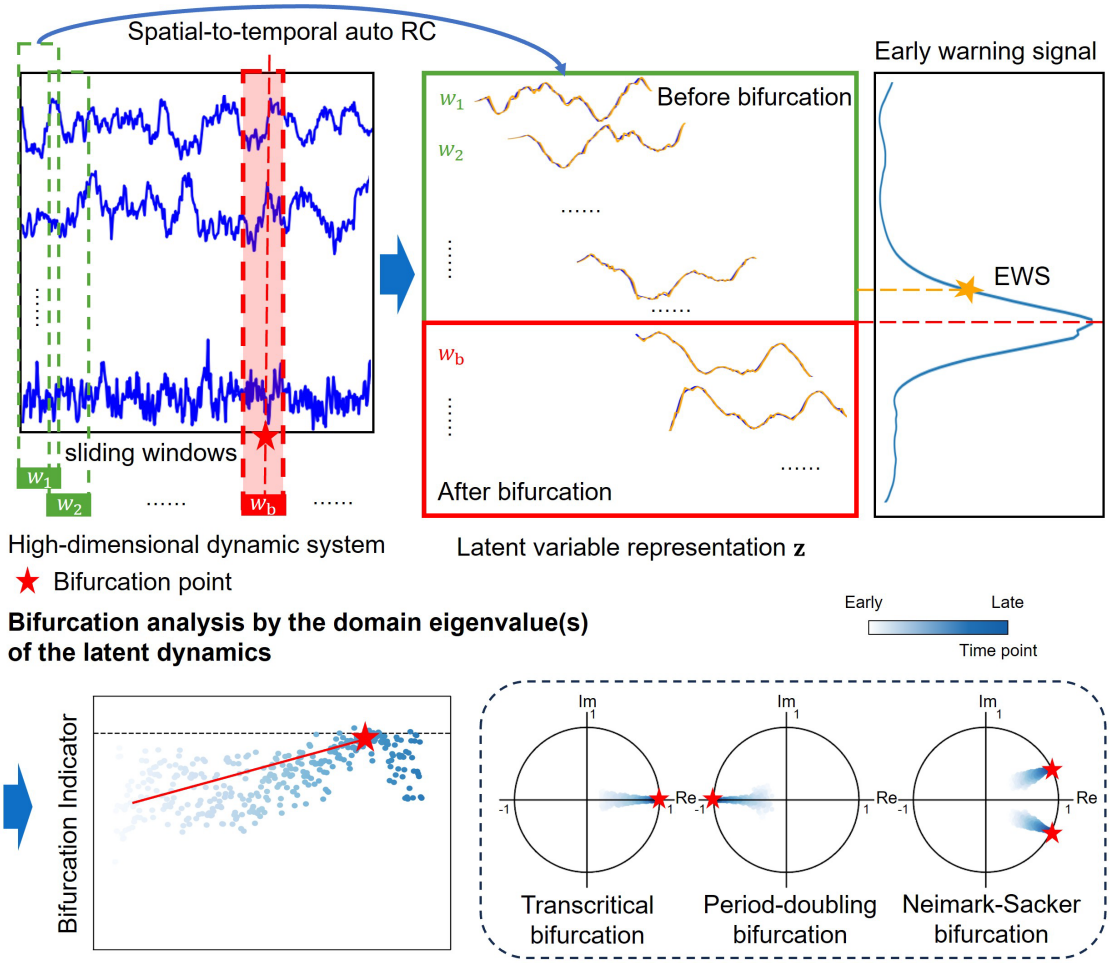


Figure S1. Illustration of the sliding window scheme. High-dimensional "spatial" information $\mathbf{X}^t$ is transformed by stARC into the "temporal" information $\mathbf{Z}^t$ of a one-dimensional latent variable z. Early warning signals (EWSs) are then identified from changes in the fluctuation of z, and bifurcation analysis is performed by estimating the dominant eigenvalue of the latent dynamics $\mathbf{z}$.

Supplementary Figure S2. The self-supervised framework of stARC

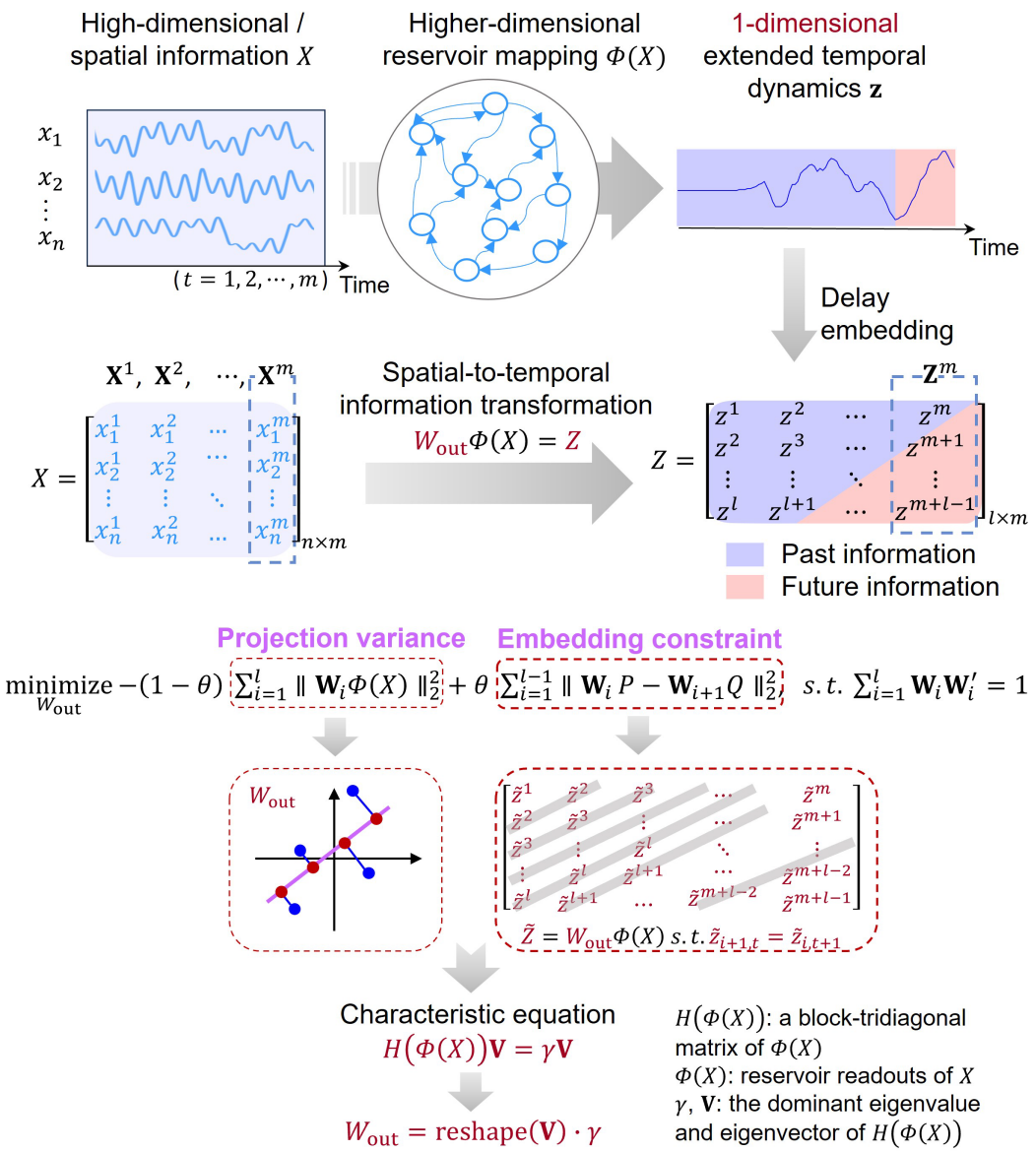


Figure S2. The self-supervised framework of stARC. Input high-dimensional spatial information $\mathbf{X}^t$ is first mapped into a higher-dimensional reservoir state space and then transformed into the temporal information $\mathbf{Z}^t$ of a one-dimensional latent variable z through the spatial-to-temporal information (STI) transformation equation, which is analytically derived from the delay-embedding theorem. For a high-dimensional time series X with m time points, stARC analytically derives the reservoir readout matrix W_{out} and the projected Hankel matrix Z. The optimization objective contains two terms: the first maximizes the variance of the latent variable z, and the second enforces the delay-embedding condition on the projected Hankel matrix Z. The optimization problem admits an analytical solution by solving the characteristic equation $H(\Phi(X))V = \gamma V$, where $H(\Phi(X))$ is a block-tridiagonal matrix.

Supplementary Figure S3. The spatial network of the system space

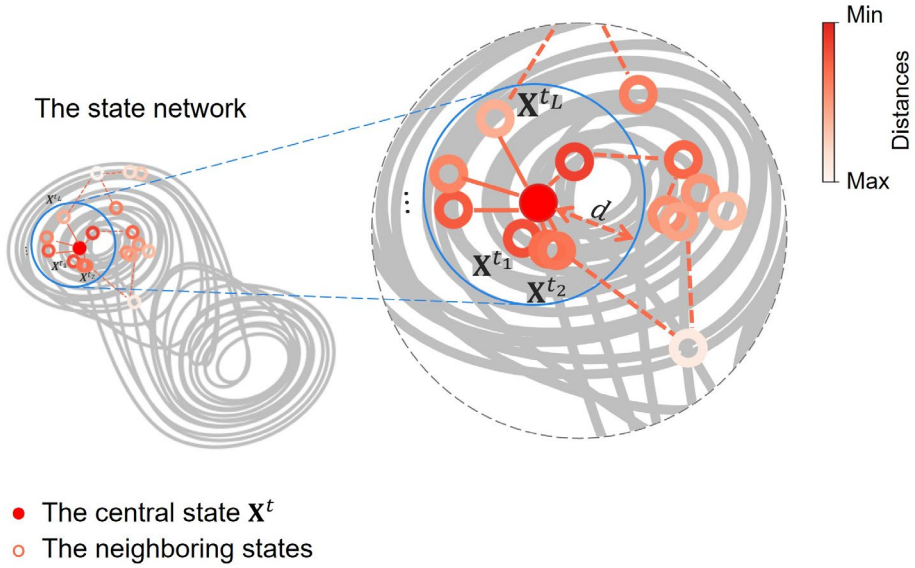
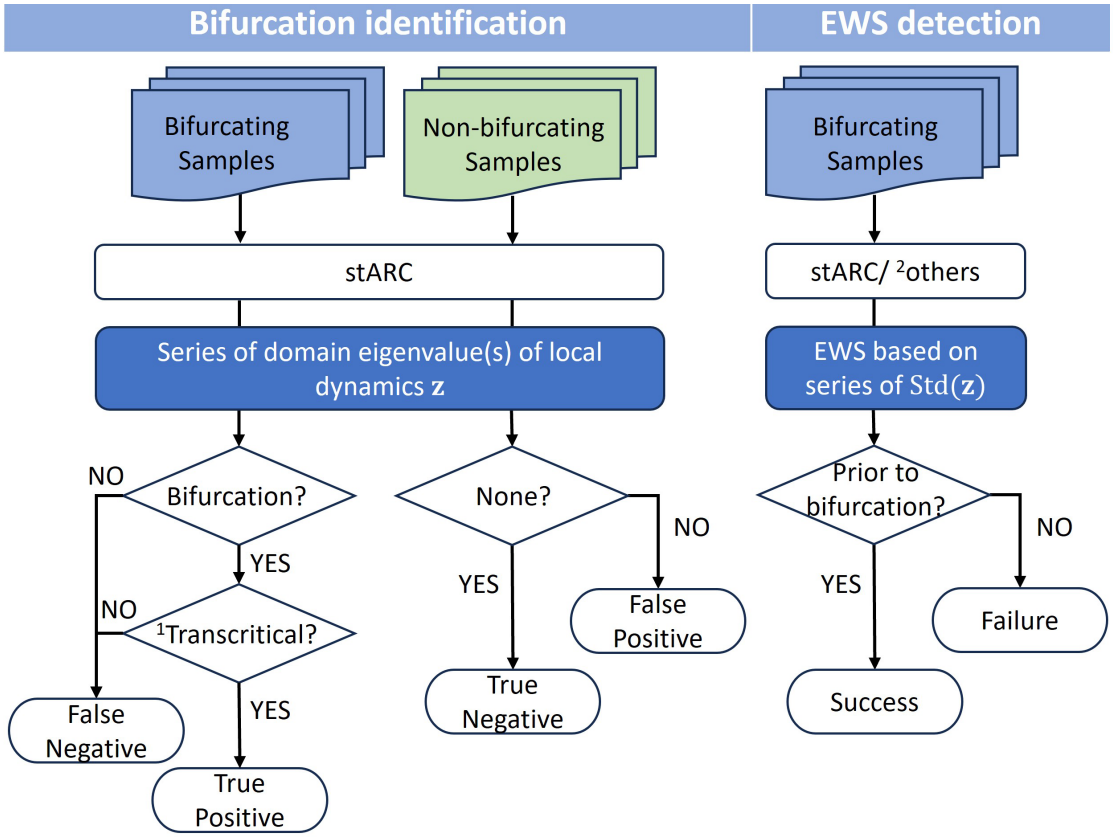


Figure S3. The spatial network of the system space. For any given high-dimensional spatial state $\mathbf{X}^t$, a local spatial network is established with $\mathbf{X}^t$ as the central state and its L nearest neighbors. The largest distance between $\mathbf{X}^t$ and these L nearest neighbors is denoted by d. To enhance robustness and stability, stARC constructs this local spatial network to capture the structural information of the local phase space.

Supplementary Figure S4. Evaluation workflow of bifurcation and EWS detection



[1]It is the responding bifurcation type according to the simulated model.

[2]Other methods include multiDEV, OCSVM, LSTM RNN, RDA, Ridge-DMD and Neural Koopman.

Figure S4. Construction of performance matrices. For bifurcation identification, 200 bifurcating and 200 non-bifurcating samples were generated from each simulation system under Gaussian noise, including the coupled Van der Pol-Duffing (ADVP, Neimark-Sacker bifurcation), Lorenz (transcritical bifurcation), and Hénon (period-doubling bifurcation) systems. Experiments were conducted under noise strengths $\sigma = 0.1, 0.2$, and 0.5 (with the window length fixed at $m = 100$) and under window lengths of $m = 100$ with variations of $\pm 10\%$ and $\pm 20\%$ (with the noise strength fixed at $\sigma = 0.1$). For early warning signals (EWS) detection, a detection was considered successful if the signal was identified before the bifurcation point. Each method was evaluated on 400 randomly generated samples per system under the same experimental settings.

Supplementary Figure S5. Input time series generated by synthetic models

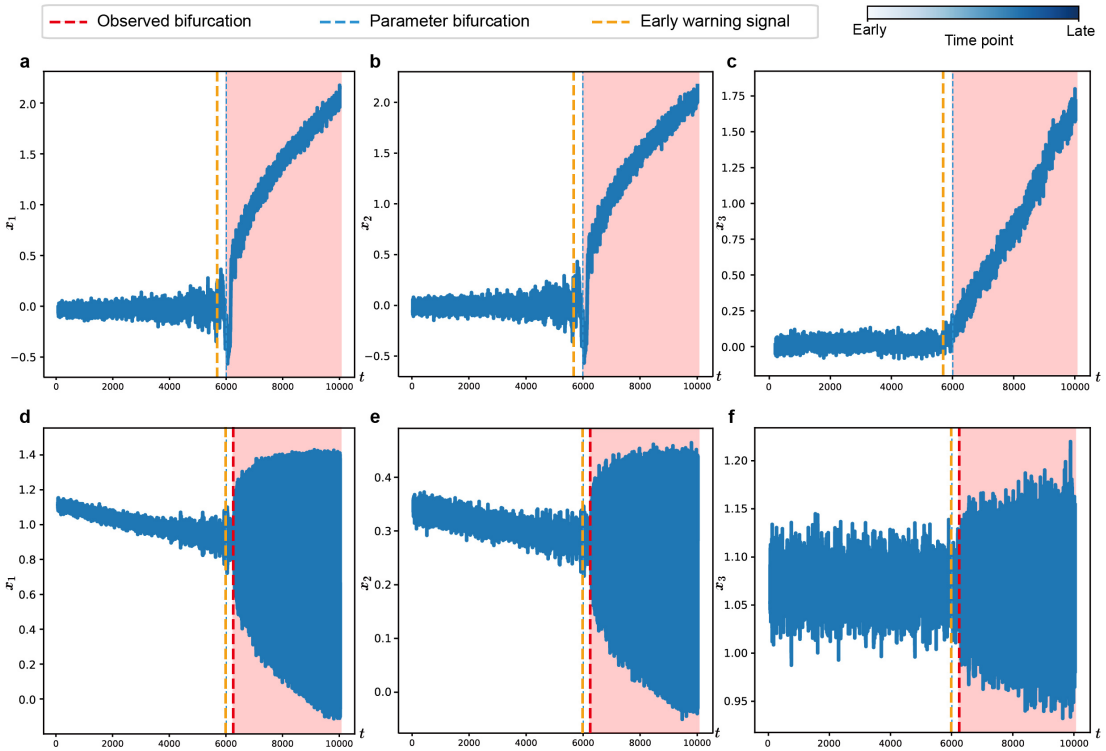


Figure S5. Time series of the coupled Lorenz system and Hénon map. (**a-c**) Time series of the three state variables x_1, x_2, and x_3 in the coupled Lorenz system, where a transcritical bifurcation is triggered at $t = 6000$ when the bifurcation parameter reaches its critical value (blue dashed line). Using stARC, the EWS is detected at $t = 5650$ (orange lines), preceding the parameter bifurcation point. (**d-f**) Time series of the three state variables x_1, x_2, and x_3 in the coupled Hénon map, where a period-doubling bifurcation is triggered at $t = 6000$ (blue dashed line) and becomes apparent around $t = 6250$ (red dashed line). Using stARC, the EWS is detected at $t = 5980$ (orange lines), preceding both the parameter and observed bifurcation points. Red shaded regions indicate post-bifurcation dynamics. Detailed results are presented in the Results section of the main text.

Supplementary Figure S6. Robustness analysis of EWS detection

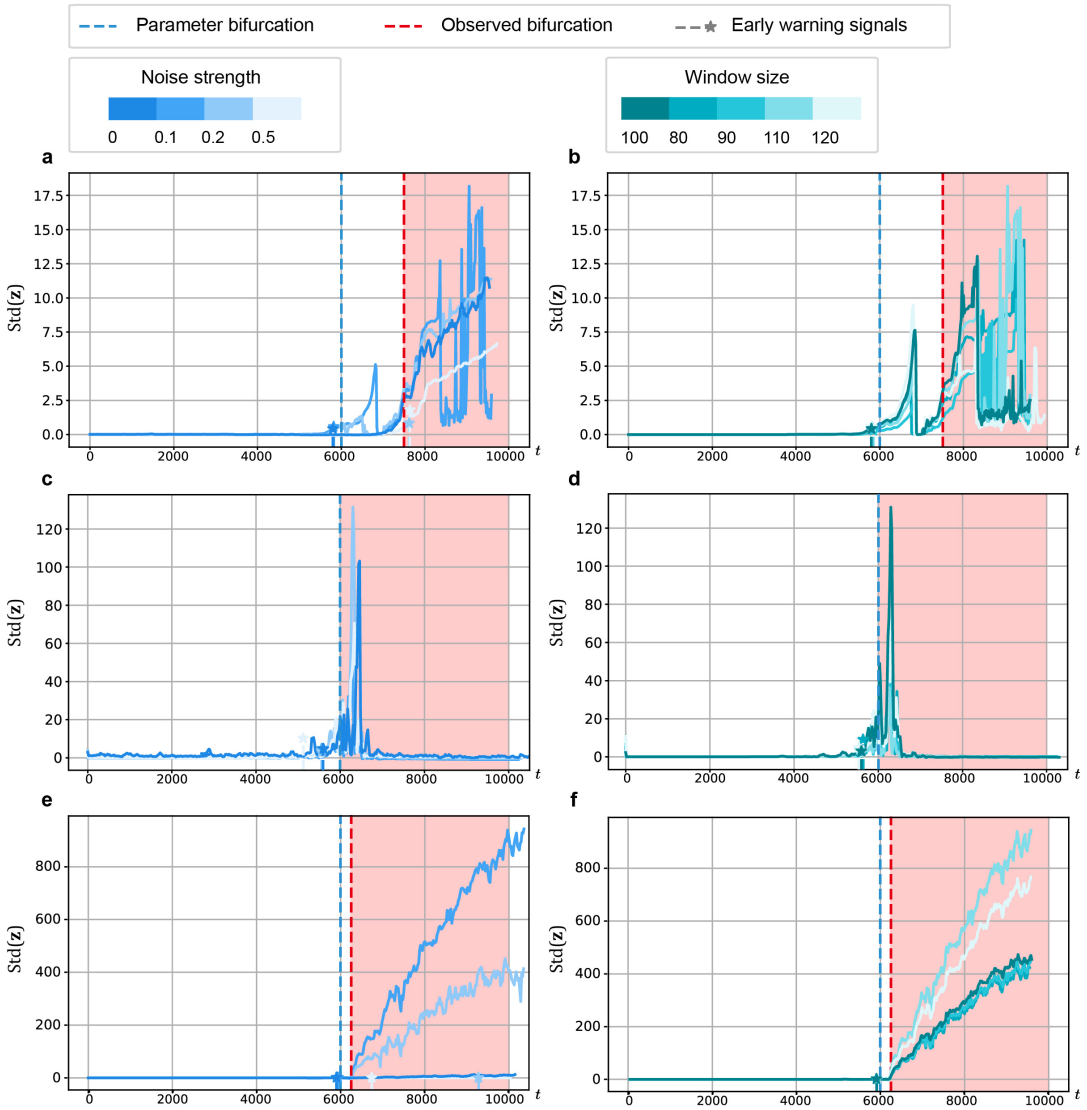


Figure S6. EWS detection performance of stARC on simulated systems under different conditions. (**a**, **c**, **e**) EWS detection under noise strengths $\sigma = 0, 0.1, 0.2$, and 0.5 (with the window length fixed at $m = 100$) for the coupled ADVP, Lorenz, and Hénon systems, respectively. (**b**, **d**, **f**) EWS detection under window lengths of $m = 100$ with variations of $\pm 10\%$ and $\pm 20\%$ (with the noise strength fixed at $\sigma = 0.1$) for the coupled ADVP, Lorenz, and Hénon systems, respectively. In all cases, the parameter bifurcation point occurs at $t = 6000$, where the bifurcation parameter reaches its critical value (blue dashed line). Star markers denote detected EWSs, and their colors match the corresponding system settings. Red dashed lines indicate observed bifurcation points where applicable, and red shaded regions denote post-bifurcation dynamics.

Supplementary Figure S7. Mineralogical and growth data from the Holocene stalagmite record

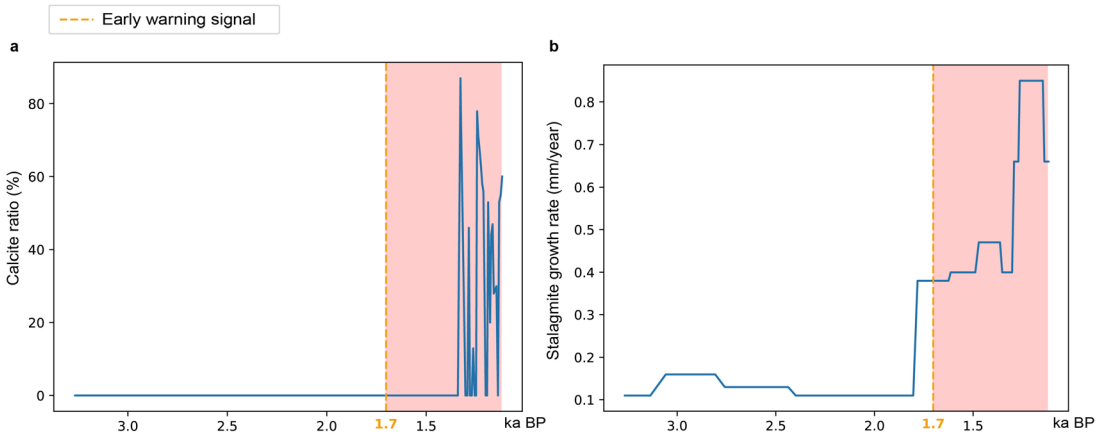


Figure S7. Mineralogical and growth data from the Holocene stalagmite record. (**a**) Calcite ratio (%) and (**b**) stalagmite growth rate (mm/year) of stalagmite DP1, a 1.6-meter-long speleothem from Dante Cave in northeastern Namibia, covering 3.3 to 1.1 ka BP. A calcite ratio of zero indicates that the stalagmite consists solely of aragonite. Orange dashed lines represent early warning signals, and red shaded regions highlight post-bifurcation dynamics.

Section S2. Supplementary Tables

Supplementary Table S1. Summary of parameters and variables in the stARC framework.

Table S1. Summary of parameters and variables in the stARC framework.

Symbols	Dimensions	Explanation
Ψ	$n \to l+1$	STI transformation mapping
R	$n \to N$	Function, a given reservoir network whose weights are randomly given
Φ	$n \to N$	Function, a structural readout regarding each window
$\mathbf{X}^t$	n	Vector, observed states of high-dimensional systems at time point t
$\mathbf{Z}^t$	l	Representative vector, delay embedding of the original system state at time point t
$\mathbf{z}$	$m+l-1$	Representative vector, extended dynamics of the original system within each window
$\mathbf{z}^t$	E	Vector, delayed dynamics of z for a local linear approximation, and $E < l$
z	1	Variable, the representative variable obtained by stARC
Z	$l \times m$	Matrix, low-dimensional Hankel matrix
$\mathbf{w}$	$L+1$	Vector, structural weights of state-neighbors
$\mathbf{r}^t$	N	Vector, reservoir state at time point t
$\Phi(\mathbf{X}^t)$	N	Vector, reservoir readout of stARC at time point t
$\mathbf{b}$	N	Vector, node bias vector
W_{in}	$N \times n$	Input matrix, defining the weights from the input layer to the reservoir
W_{r}	$N \times N$	Reservoir adjacency matrix, representing the reservoir's internal connectivity
W_{out}	$l \times N$	Readout matrix, the to-be-solved weights of $\Phi(\boldsymbol{X}^t)$
X	$n \times m$	Matrix, the observed high-dimensional time series with m time points
$\mathbb{R}^n$	n	n-dimensional Euclidean space
m	1	Scalar, the window size
l	1	Scalar, delay embedding dimension
d	1	Scalar, box-counting dimension
t	1	Scalar, time point
n	1	Scalar, the dimension of the observed variables in time-series

N	1	Scalar, the dimension of the reservoir states (i.e., the nodes of reservoir network)
L	1	Scalar, the number of neighbors
θ	1	Scalar, regularization parameter in stARC loss function
x_k^t	1	Scalar, the value of the k-th variable at time point t
y^t	1	Scalar, the value of the target variable y at time point t
$\circ$		Function composition operation
$'$		Transpose of a vector
Std		Standard deviation

Supplementary Table S2. Conceptual comparison between stARC and gradient-based learning approaches

Table S2. Conceptual comparison between stARC and gradient-based learning approaches

	stARC	Gradient-based Learning Approaches
Interpretability	Bifurcation analysis with theoretical foundation	Task-optimized black-box
Identification Principle	Intrinsic dynamical properties (Eigenvalue solution)	Loss-based classification
Data Requirement	Short unlabeled time series	Typically large/long training datasets
Training Mechanism	Analytical solution (self-supervised)	Gradient-based optimization (iterative backpropagation)
Robustness under limited data	Stable under short/nonstationary data	Sensitive to noise
Capacity	Prediction of critical transition; identification of bifurcation types	Prediction of critical transition; identification of bifurcation types
Applicability	Transcritical, period-doubling, and Neimark-Sacker bifurcations	Dependent on training datasets

Supplementary Table S3. Performance comparison between stARC and two representation-based methods.

Table S3. Performance comparison between stARC and two representation-based methods. Performance metrics include EWS detection accuracy, bifurcation identification accuracy (ACC), and true positive rate (TPR). An EWS detection is counted as correct if a signal is detected before the bifurcation point in bifurcating samples and absent in non-bifurcating samples. Each method was tested on 400 randomly generated samples per system (200 bifurcating and 200 non-bifurcating) under identical experimental settings.

Dataset	Metrics		stARC	Ridge DMD	Neural Koopman
Coupled Lorenz	Accuracy of EWS detection		96.54%	79.25%	87.00%
	Bifurcation Type	ACC	97.75%	79.25%	87.00%
		TPR	96.00%	92.50%	95.75%
Coupled Hénon	Accuracy of EWS detection		94.75%	61.00%	52.50%
	Bifurcation Type	ACC	97.75%	61.25%	96.50%
		TPR	99.50%	22.50%	98.50%
Coupled ADVP	Accuracy of EWS detection		96.15%	66.25%	71.75%
	Bifurcation Type	ACC	99.25%	66.25%	71.75%
		TPR	98.50%	99.00%	98.00%

Supplementary Table S4. Bifurcation identification performances of stARC under different neighbor settings.

Table S4. Bifurcation identification performances of stARC under different neighbor settings. Bifurcation identification performance is evaluated by accuracy (ACC) and true positive rate (TPR). For each simulated system, 200 bifurcating and 200 non-bifurcating samples were randomly generated. Results are reported for noise strengths $\sigma = 0.1,\ 0.2$, and 0.5, and for different neighborhood settings, including no neighbors (i.e., $L = 0$) and $L = 5,\ 10,\ 15$.

System	Noise Strength σ	Metrics	Exclude Neighbors ($L = 0$)	Neighbor Number L		
				$L = 5$	$L = 10$	$L = 15$
Coupled Lorenz	$\sigma = 0.1$	ACC	67.50%	97.25%	97.75%	98.25%
		TPR	37.00%	96.00%	96.00%	96.50%
	$\sigma = 0.2$	ACC	58.00%	98.50%	97.25%	98.00%
		TPR	22.00%	97.50%	95.92%	96.00%
	$\sigma = 0.5$	ACC	67.00%	93.00%	95.75%	96.25%
		TPR	36.00%	90.00%	91.50%	92.50%
Coupled Hénon	$\sigma = 0.1$	ACC	87.50%	97.75%	97.75%	98.50%
		TPR	75.00%	96.00%	99.50%	99.00%
	$\sigma = 0.2$	ACC	86.00%	96.75%	97.00%	97.00%
		TPR	72.50%	94.00%	94.00%	95.75%
	$\sigma = 0.5$	ACC	71.00%	95.00%	95.75%	99.50%
		TPR	42.75%	90.00%	91.50%	99.00%
Coupled ADVP	$\sigma = 0.1$	ACC	97.00%	97.50%	99.25%	98.50%
		TPR	94.50%	95.00%	98.50%	97.00%
	$\sigma = 0.2$	ACC	97.00%	99.00%	99.50%	99.50%
		TPR	94.00%	98.00%	99.00%	99.50%
	$\sigma = 0.5$	ACC	81.00%	97.25%	99.50%	98.75%
		TPR	62.00%	96.50%	99.50%	97.50%

Supplementary Table S5. Bifurcation identification performance of stARC with different regularization parameters.

Table S5. Bifurcation identification performance of stARC with different regularization parameters. Each method was evaluated on 400 randomly generated samples per system (200 bifurcating and 200 non-bifurcating) under consistent experimental settings. The performance is evaluated by accuracy (ACC) and true positive rate (TPR).

Dataset	**Metrics**	$\theta = 0.6$	$\theta = 0.8$	$\theta = 0.9$	$\theta = 0.99$
	ACC	82.50%	97.75%	96.00%	94.50%

Coupled Lorenz	TPR	66.00%	96.00%	92.00%	89.00%
Coupled Hénon	ACC	95.50%	97.75%	98.25%	60.25%
	TPR	97.00%	99.50%	99.00%	12.00%
Coupled ADVP	ACC	99.50%	99.25%	99.00%	50.05%
	TPR	99.00%	98.50%	98.00%	1.00%

Section S3. Supplementary Notes

Supplementary Note S1. Reservoir Computing

Reservoir computing (RC) is fundamentally built upon a fixed recurrent network of nonlinear neurons, allowing bidirectional signal propagation across its internal structure[1] and thereby enabling rich dynamics that capture complex temporal dependencies. A typical RC framework comprises an input layer, a reservoir layer with high-dimensional nonlinear dynamics, and an output layer. The input layer maps multivariate time series data into the high-dimensional state space of the reservoir, while the output layer, also known as the readout, linearly combines the evolving reservoir states to produce the final output, such as time series prediction or classification.

Mathematically, the evolution of the reservoir states can be expressed by the following equation:

$$\mathbf{r}^t = \alpha\mathbf{r}^{t-1} + (1-\alpha)f(W_{\text{in}}\mathbf{X}^t + W_{\text{r}}\mathbf{r}^{t-1} + \mathbf{b}), \tag{S1}$$

where t denotes the discrete time point, α $(0 \leq \alpha \leq 1)$ is the leakage factor, $\mathbf{r}^t \in \mathbb{R}^N$ is the reservoir state vector, also denoted by $R(\mathbf{X}^t)$, $\mathbf{X}^t \in \mathbb{R}^n$ $(n \ll N)$ is the n-dimensional input vector, $\mathbf{b}$ is a node bias vector, and f_i from $f = (f_1, f_2, \cdots, f_N)$ represents the activation function for the i-th reservoir node, which is typically a sigmoid- or ReLU-type activation function. Here, the input matrix W_{in} defines the weights from the input layer to the reservoir and the reservoir adjacency matrix W_{r} defines the reservoir's internal connectivity.

After the reservoir states are updated over time, the readout output at each time step is generated by a linear readout layer. Specifically, the output is often given by

$$\mathbf{Y}^{\mathrm{t}} = W_{\mathrm{out}}\Phi(\mathbf{X}^{t}) \tag{S2}$$

where $\mathbf{Y}^{t}$ is the RC output vector, $\Phi(\mathbf{X}^{t})$ is the reservoir readout, and W_{out} is the output weight matrix. In conventional RC, $\mathbf{r}^{t}$ is typically set as the reservoir readout and the matrix W_{out} is learned via supervised learning methods such as ridge regression, ordinary least squares, or other regularized techniques. In this paper, a self-supervised framework of reservoir computing, i.e., spatial-to-temporal auto reservoir computing (stARC), is proposed to anticipate critical transitions and identify bifurcation types (detailed in Methods section of the main manuscript) as follows:

$$W_{\mathrm{out}}\Phi(\mathbf{X}^{t}) = \mathbf{Z}^{t}. \tag{S3}$$

Our proposed stARC framework involves several types of parameters, which can be grouped into three categories according to their roles:(1) Hyperparameters of the reservoir network: This set of predetermined parameters defines the reservoir network, including the size N, average degree k, spectral radius ρ of reservoir adjacency matrix, scale parameter a and leakage factor α. There are standard methods such as Bayesian optimization, surrogate optimization (SGO), particle swarm optimization, and genetic algorithm, which can be used to determine an optimal set of hyperparameters. To ensure fair comparison across different dynamical systems, we employed the same set of hyperparameters for all experiments. This design choice eliminates biases that may arise from system-specific parameter tuning and facilitates reproducibility of the results. More importantly, the ability of stARC to maintain high performance without dataset-specific adjustments highlights its robustness and generalizability, which are essential for practical applications where exhaustive hyperparameter optimization is often infeasible. Specifically, the parameters are set to $N = 100$, $k = 20$, $\rho = 0.8$, $a = 1$, $\alpha = 0.5$. These values were chosen empirically and then fixed across all experiments.

(2) Elements of the input matrix W_{in}, bias vector $\mathbf{b}$ and the reservoir adjacency matrix W_{r}: These are chosen randomly and fixed for any given stARC realization. Their distributions are controlled by the hyperparameters. In this paper, each node in

the input layer is connected to all nodes in the reservoir network and the reservoir network is set to be an undirected (symmetric) random network with average degree k. Specifically, all elements of W_{in} and $\mathbf{b}$ are generated from a uniform distribution in the interval $[-a, a]$. The nonzero elements of W_{r} are sampled from a zero-mean normal distribution and then uniformly rescaled so that the matrix has a spectral radius ρ. The spectral radius ρ controls the dynamical regime of the reservoir: setting $\rho < 1$ generally helps preserve the echo state property (ESP) and system stability, while values close to unity allow the reservoir to retain longer memory of past inputs, thereby balancing stability and representational capacity.

(3) Training parameters: window size m, embedding dimension l, and number of considered neighbors L. These are the set of parameters associated with the training process of stARC. Notably, $m > 2l$ is required to guarantee the validityof stARC. The effects of different values of m and L on the performance of stARC are discussed in the section "Robustness Analysis of stARC" of the main text and in Supplementary Note S7, respectively.

Supplementary Note S2. Analytical solution of stARC

For each observed high-dimensional time series $\mathbf{X}^{\text{t}} = (x_1^t, x_2^t, \dots, x_n^t)'$ with n variables and $t = 1,2, \dots, m$, we construct a corresponding delayed vector $\mathbf{Z}^t = (z^t, z^{t+1}, \dots, z^{t+l-1})'$ by a delay embedding strategy with $l > 1$ as the embedding dimension ($l > 2d > 0$, where d is the box-counting dimension of the attractor of the original high-dimensional system), and the " $'$ " denotes the transpose of a vector. The stARC establishes the following equation

$$Z = W_{\text{out}}\Phi(X) \tag{S4}$$

where Φ is applied columnwise to X as

$$\Phi(X) = [\Phi(\mathbf{X}^1) \quad \Phi(\mathbf{X}^2) \quad \cdots \quad \Phi(\mathbf{X}^m)]_{N\times m}$$

and matrices W_{out} and Z are of form

$$W_{\text{out}} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1N} \\ w_{21} & w_{22} & & w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{l1} & w_{l2} & \cdots & w_{lN} \end{bmatrix}_{l\times N}, \; Z = \begin{bmatrix} z^1 & z^2 & \cdots & z^m \\ z^2 & z^3 & \cdots & z^{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ z^l & z^{l+1} & \cdots & z^{m+l-1} \end{bmatrix}_{l\times m}.$$

To simplify the notation, we henceforth denote $\Phi(X)$ by G. Based on the delay embedding requirement for Z, the latter $m-1$ elements in the i-th row are identical to the former $m-1$ elements in the $(i+1)$-th row of the matrix Z. In addition, we have vectors

$$\begin{aligned} \mathbf{W}_i P &= \left(z^{i+1}, z^{i+2}, \cdots, z^{i+m-1}\right) \\ \mathbf{W}_{i+1} Q &= \left(z^{i+1}, z^{i+2}, \cdots, z^{i+m-1}\right) \end{aligned} \tag{S5}$$

where row vector $\mathbf{W}_i = (w_{i1}, w_{i2}, \cdots, w_{iN})$, P and Q are two submatrices of $\Phi(\mathbf{X})$, i.e., $P = [\Phi(\mathbf{X}^2)\ \Phi(\mathbf{X}^3)\ \cdots\ \Phi(\mathbf{X}^m)]$, $Q = [\Phi(\mathbf{X}^1)\ \Phi(\mathbf{X}^2)\ \cdots\ \Phi(\mathbf{X}^{m-1})]$. Clearly, it holds that

$$\mathbf{W}_i P = \mathbf{W}_{i+1} Q, \qquad i = 1, 2, \cdots, l-1 \tag{S6}$$

1. Considering the reconstruction loss and the embedding constraints, the loss function is defined as

$$\begin{aligned} &\underset{W_{\text{out}}}{\text{minimize}} -(1-\theta)\sum_{i=1}^{l} \| \mathbf{W}_i G \|_2^2 + \theta \sum_{i=1}^{l-1} \| \mathbf{W}_i P - \mathbf{W}_{i+1} Q \|_2^2 \\ &s.t. \sum_{i=1}^{l} \mathbf{W}_i \mathbf{W}_i{}' = 1, \end{aligned} \tag{S7}$$

where $\theta \in [0, 1]$ is a predetermined regularization parameter.

2. On the basis of the loss function, the Lagrange multiplier is defined as follows:

$$\begin{aligned} L(W_{\text{out}}, \lambda) &= -(1-\theta)\sum_{i=1}^{l} \| \mathbf{W}_i G \|_2^2 + \theta \sum_{i=1}^{l-1} \| \mathbf{W}_i P - \mathbf{W}_{i+1} Q \|_2^2 + \lambda\left(\sum_{i=1}^{l} \mathbf{W}_i \mathbf{W}_i' - 1\right) \\ &= -(1-\theta)\cdot \sum_{i=1}^{l} tr(\mathbf{W}_i G G' \mathbf{W}_i') + \theta \sum_{i=1}^{l-1} tr((\mathbf{W}_i P - \mathbf{W}_{i+1} Q)'(\mathbf{W}_i P - \mathbf{W}_{i+1} Q)) \\ &+\lambda\left(\sum_{i=1}^{l} \mathbf{W}_i \mathbf{W}_i' - 1\right) \\ &= -(1-\theta)\cdot \sum_{i=1}^{l} tr(\mathbf{W}_i G G' \mathbf{W}_i') + \lambda(\mathbf{W}_i \mathbf{W}_i' - 1) \\ &+\theta \sum_{i=1}^{l-1} tr(\mathbf{W}_i P P' \mathbf{W}_i' + \mathbf{W}_{i+1} Q Q' \mathbf{W}_{i+1}' - \mathbf{W}_i P Q' \mathbf{W}_{i+1}' - \mathbf{W}_{i+1} Q P' \mathbf{W}_i') \end{aligned} \tag{S8}$$

where $\lambda \in \mathbb{R}$ is a Lagrange coefficient.

3. Taking the partial derivative of $L(W,\lambda)$ with respect to W_i and λ yields

$$\begin{aligned}\frac{\partial L(W_{\text{out}},\lambda)}{\partial \mathbf{W}_1} &= -2(1-\theta)\cdot GG'\mathbf{W}_1' + \theta(2PP'\mathbf{W}_1' - PQ'\mathbf{W}_2' - (\mathbf{W}_2QP')') + 2\lambda\mathbf{W}_1' \\ &= -2(1-\theta)\cdot GG'\mathbf{W}_1' + 2\theta\cdot(PP'\mathbf{W}_1' - PQ'\mathbf{W}_2') + 2\lambda\mathbf{W}_1' \\ &= -2\big(\big((1-\theta)\cdot GG' - \theta PP'\big)\mathbf{W}_1' + \theta PQ'\mathbf{W}_2'\big) + 2\lambda\mathbf{W}_1'\end{aligned} \tag{S9}$$

$$\begin{aligned}\frac{\partial L(W_{\text{out}},\lambda)}{\partial \mathbf{W}_i} &= -2(1-\theta)\cdot GG'\mathbf{W}_i' \\ &\quad +\theta((2QQ'\mathbf{W}_i' - (\mathbf{W}_{i-1}PQ')' - QP'\mathbf{W}_{i-1}') \\ &\qquad +(2P'\mathbf{W}_i' - PQ'\mathbf{W}_{i+1}' - (\mathbf{W}_{i+1}QP')')) \\ &\qquad +2\lambda\mathbf{W}_i' \\ &\qquad = -2(1-\theta)GG'\mathbf{W}_i' + 2\lambda\mathbf{W}_i' \\ &\qquad +2\theta(PP'\mathbf{W}_i' + QQ'\mathbf{W}_i' - QP'\mathbf{W}_{i-1}' - PQ'\mathbf{W}_{i+1}') \\ &\qquad = -2(\theta QP'\mathbf{W}_{i-1}' + \big((1-\theta)\cdot GG' - \theta PP' - \theta QQ'\big)\mathbf{W}_i' \\ &\qquad +\theta PQ'\mathbf{W}_{i+1}') + 2\lambda\mathbf{W}_i'\end{aligned} \tag{S10}$$

where $2 \le i \le l-1$.

$$\begin{aligned}\frac{\partial L(W_{\text{out}},\lambda)}{\partial \mathbf{W}_l} &= -2(1-\theta)\cdot GG'\mathbf{W}_l' + \theta(2QQ'\mathbf{W}_l' - (\mathbf{W}_{l-1}PQ')' - QP'\mathbf{W}_{l-1}')) \\ &\qquad +2\lambda\mathbf{W}_l' \\ &\qquad = -2(1-\theta)\cdot GG'\mathbf{W}_l' + 2\theta(QQ'\mathbf{W} - QP'\mathbf{W}_{l-1}') + 2\lambda\mathbf{W}_l' \\ &\qquad = -2\theta\big(QP'\mathbf{W}_{l-1}' + \big((1-\theta)\cdot GG' - \theta QQ'\big)\mathbf{W}_l'\big) + 2\lambda\mathbf{W}_l'\end{aligned} \tag{S11}$$

4. Letting $\frac{\partial L(W_{\text{out}},\lambda)}{\partial \mathbf{W}_i} = 0$, we obtain the following characteristic equation from Eq. (S9)-(S11):

$$\begin{bmatrix} (1-\theta)GG' - \theta PP' & \theta PQ' & 0 & \cdots & 0 & 0 & 0 \\ \theta QP' & (1-\theta)GG' - \theta PP' - \theta QQ' & \theta PQ' & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \theta QP' & (1-\theta)GG' - \theta PP' - \theta QQ' & \theta PQ' \\ 0 & 0 & 0 & \cdots & 0 & \theta QP' & (1-\theta)GG' - \theta QQ' \end{bmatrix}_{(Nl)\times(Nl)}$$

$$\cdot \begin{bmatrix} \mathbf{W}_1' \\ \mathbf{W}_2' \\ \vdots \\ \mathbf{W}_{l-1}' \\ \mathbf{W}_l' \end{bmatrix}_{(Nl)\times 1} = \lambda \cdot \begin{bmatrix} \mathbf{W}_1' \\ \mathbf{W}_2' \\ \vdots \\ \mathbf{W}_{l-1}' \\ \mathbf{W}_l' \end{bmatrix}_{(Nl)\times 1} \tag{S12}$$

$$H(G)\mathbf{V} = \gamma\mathbf{V} \tag{S13}$$

where H is a function that converts X into a block-tridiagonal matrix, i.e.,

$$H(G) = \begin{bmatrix} (1-\theta)GG' - \theta PP' & \theta PQ' & 0 & \cdots & 0 & 0 & 0 \\ \theta QP' & (1-\theta)GG' - \theta PP' - \theta QQ' & \theta PQ' & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \theta QP' & (1-\theta)GG' - \theta PP' - \theta QQ' & \theta PQ' \\ 0 & 0 & 0 & \cdots & 0 & \theta QP' & (1-\theta)GG' - \theta QQ' \end{bmatrix},$$

γ and $\mathbf{V}$ denote a top eigenvalue and eigenvector of $H(G)$.

5. W_{out} is obtained by reshaping the vector $\mathbf{V}$ by $W_{\text{out}} = \text{reshape}(\mathbf{V})\cdot\gamma$.

6. We obtain

$$\tilde{Z} = W_{\text{out}}\Phi(X). \tag{S14}$$

7. From Eq. (S14), let $z^t = \frac{\sum \hat{z}_{i,j}}{k}$, where $i + j = t + 1$, and $i = 1, \cdots, l;\ j = 1, \cdots, m;\ t = 1, \cdots, m + l - 1,$ and k is the number of $\hat{z}_{i,j}$. The representative vector $\mathbf{z} = (z^1, z^2, \cdots, z^m, z^{m+1}, \cdots, z^{m+l-1})'$ and a Hankel matrix Z can then be easily obtained from the one-dimensional variable z.

Supplementary Note S3. Detecting early warning signals (EWSs) of critical transitions

We consider a discrete-time dynamical system in generic form

$$\mathbf{X}^{t+1} = T(\mathbf{X}^t; \mathbf{P}) \tag{S15}$$

where $\mathbf{X}^t$ is an n-dimensional state vector at time instant t, $\mathbf{P} = (p_1, \ldots, p_s)$ is a parameter vector representing slowly changing factors and $T: \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^n$ is a nonlinear map. In addition, the following conditions are assumed to hold for Eq. (S15): (i) $\overline{\mathbf{X}}$ is a fixed point of the system such that $\overline{\mathbf{X}} = T(\overline{\mathbf{X}}; \mathbf{P})$. (ii) There is a value $\mathbf{P}_\mathrm{c}$ such that one or a pair of complex conjugate eigenvalues of the Jacobian matrix $\frac{\partial T(\mathbf{X};\mathbf{P}_\mathrm{c})}{\partial \mathbf{X}}|_{\mathbf{X}=\overline{\mathbf{X}}}$ has modulus 1. (iii) When $\mathbf{P} \neq \mathbf{P}_\mathrm{c}$, the eigenvalues of the system are not always 1 in modulus. Based on bifurcation theory, these three conditions with other transversal conditions imply that the system undergoes a codimension-one bifurcation near $\overline{\mathbf{X}}$ as $\mathbf{P}$ approaches the threshold $\mathbf{P}_\mathrm{c}$, thereby inducing a critical transition. Thus, the system in Eq. (S15) will asymptotically converge to a stable fixed point $\overline{\mathbf{X}}$ when the control parameter $\mathbf{P}$ has not reached $\mathbf{P}_\mathrm{c}$, in which case the moduli of all eigenvalues lie within [0, 1). The critical parameter value $\mathbf{P}_\mathrm{c}$ is referred to as the bifurcation thresholdor the critical transition threshold. Once $\mathbf{P}$ passes through $\mathbf{P}_\mathrm{c}$, at least one eigenvalue crosses the unit circle, leading to the loss of stability of the fixed point and the onset of a bifurcation, which gives rise to a qualitatively new dynamical regime.

Now, by considering the local linear approximation of Eq. (S15) and introducing new variables $\mathbf{Y}^t = (y_1^t, y_2^t, \ldots, y_n^t)'$ and a transformation matrix $A = \left[a_{ij}\right]_{n\times n}$, i.e., $\mathbf{Y}^t = A^{-1}(\mathbf{X}^t - \overline{\mathbf{X}})$, we have

$$\mathbf{Y}^t = \Lambda(P)\mathbf{Y}^{t-1} + \boldsymbol{\zeta}^{t-1}, \tag{S16}$$

where $\boldsymbol{\zeta}^t = (\zeta_1^t, \zeta_2^t, \cdots, \zeta_n^t)'$ denotes small Gaussian noise with zero mean and ζ_i^t has a small standard deviation σ_i for all i.

Without loss of generality, consider the diagonalized matrix $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, where the eigenvalues are ordered by decreasing modulus and satisfy $0 < |\lambda_i| < 1$. As the control parameter $\mathbf{P}$ tends to the critical value $\mathbf{P}_c$, the dominant eigenvalue λ_1 simultaneously approaches unity in modulus, which signals the occurrence of a local bifurcation in the neighborhood of the fixed point. Furthermore, the variance of y_i in $\mathbf{Y}$ associated with λ_i is demonstrated as follows:

$$\mathrm{Var}(y_i) = \frac{\sigma_i^2}{\left|1 - \lambda_i^2\right|}. \tag{S17}$$

Given that $x_i = a_{i1}y_1 + \cdots + a_{in}y_n + \overline{x_i}$ with $a_{i1} \neq 0$, the standard deviations of variables x_i increase sharply as the dominant eigenvalue $|\lambda_1| \to 1$. However, such x_i with $a_{i1} \neq 0$ are difficult to identify based on only time-series data (without a definite model).

The single latent variable z obtained by stARC can be interpreted as a representative variable of the dominant mode near the bifurcation point.Therefore, when only high-dimensional time series are observed, the standard deviation of $\mathbf{z}$, denoted by $\mathrm{Std}(\mathbf{z})$, can be utilized to evaluate the state whether a dynamical system approaches a bifurcation point from a stable steady state.

A sharp increase in a sequential data series can be identified by Bayesian online changepoint detection (BOCD), a method grounded in Bayesian inference[2]. Given the sequential data $s^{1:t}$, where $s^{1:t}$ denotes the sequence of $\mathrm{Std}(\mathbf{z})$ calculated from sliding windows ending at times $1, 2, \cdots, t$ (Supplementary Fig. S1), the objective is

to detect the onset of abrupt changes, i.e., changepoints. Such changepoints serve as early warning signals, as they indicate abrupt shifts in the system's underlying dynamics.

In Bayesian inference, the observation data are assumed to be generated from a distribution in the exponential family with conjugate priors. These priors are typically parameterized by compact hyperparameters, such as ν (representing prior strength) and χ (representing sufficient statistics), enabling efficient recursive posterior updates within each segment. Building on this assumption, Bayesian Online Changepoint Detection (BOCD) models the time since the most recent changepoint -- known as the run length -- based on the observed data up to the current time. Logically, the run length at time t, named b^t, can take one of binary values,

$$b^t = \begin{cases} 0, & \text{if changepoint at time } t, \\ b^{t-1} + 1, & \text{else.} \end{cases} \tag{S18}$$

and the transition probabilities are modeled by the changepoint prior $p(b^t|b^{t-1})$ under the assumption that b^t is conditionally independent of all other variables given b^{t-1}. The prior probability distribution over the interval between changepoints follows a discrete exponential (i.e., geometric) distribution $\mathrm{Geom}(\mu)$, characterized by a timescale parameter μ. This geometric distribution has a success probability of $p = 1 - \mathrm{e}^{-\mu}$. That is,

$$p(b^t|b^{t-1}) = \begin{cases} H(b^{t-1}+1) & \text{if } b^t = 0, \\ 1 - H(b^{t-1}+1) & \text{if } b^t = b^{t-1} + 1, \\ 0 & \text{else.} \end{cases} \tag{S19}$$

where the function $H(\cdot)$ is the hazard function:

$$H(x) = \frac{P(x)}{\sum_{t=x}^{\infty} P(t)} = 1 - e^{-\lambda}, \tag{S20}$$

and $P(t)$ denotes the probability distribution of the interval between changepoints. The constant hazard function $H(x)$ implies a memoryless process.

Coherently, a recursive message-passing algorithm for the joint distribution $p(b^t, s^{1:t})$ over the current run length and the data is generated as follows:

$$
\begin{aligned}
p(b^t, s^{1:t}) &= \sum_{b^{t-1}} p(b^t, b^{t-1}, s^{1:t}) \\
&= \sum_{b^{t-1}} p(b^t, s^t | b^{t-1}, s^{1:t-1}) p(b^{t-1}, s^{1:t-1}) \\
&= \sum_{b^{t-1}} p(b^t | b^{t-1}) p\left(s^t | b^{t-1}, s^t_{(r)}\right) p(b^{t-1}, s^{1:t-1})
\end{aligned} \tag{S21}
$$

where $s^t_{(r)}$ indicate the segment of observations associated with the run b^t, and then the following probabilities are computed in sequence, with corresponding updates to the hyperparameters ν and χ:

$$
\begin{aligned}
p(s^{1:t}) &= \sum_{b^t} p(b^t, s^{1:t}), \\
p(b^t | s^{1:t}) &= \frac{p(b^t, s^{1:t})}{p(s^{1:t})}, \\
p(s^{t+1} | s^{1:t}) &= \sum_{b^t} p\left(s^{t+1} | s^t_{(r)}, b^t\right) p(b^t | s^{1:t}).
\end{aligned} \tag{S22}
$$

Finally, we obtain the distribution matrix of the run length B in which each element $b_{ij} = p(b^t = j | s^{1:i})$.

To detect changepoints based on the run-length probability matrix B, we scan the matrix row by row starting from $i = 1$. For the i-th row, we consider first i columns $\{b_{i1}, \cdots, b_{ii}\}$, which represent the run-length probabilities up to time i, and identify the column where the maximum value occurs, denoted by $t_{\mathrm{col},i} = \arg \max_{1 \le j \le i} b_{ij}$If $t_{\mathrm{col},i} = i$, meaning the maximum probability lies on the diagonal (i.e., no changepoint is indicated till time i), we increment i and repeat the procedure until either a changepoint is detected or the end of the sequence is reached (i.e., $i = t$). If $t_{\mathrm{col},i} \neq i$, as evidence of a changepoint, the estimated changepoint position is then given by $i - t_{\mathrm{col},i}$, marking the most probable start of the current segment.

Supplementary Note S4. Sequential locally weighted global linear maps (S-map)

Sequential locally weighted global linear maps (S-map) is a locally weighted multivariate linear regression scheme that approximates the best local linear model,

involving sequentially calculated Jacobians, by assigning greater weights to points on the attractor that are closest to the current system state[3–6].

Considering the extended vector $\mathbf{z} = (z^1, z^2, \cdots, z^m, z^{m+1}, \cdots, z^{m+l-1})'$ obtained from stARC within a single window, let $\mathbf{z}^t = (z^t, z^{t-1}, \cdots z^{t-(E-1)})' \in \mathbb{R}^E$ denote the delay embedding vector of the one-dimensional variable z at time t where E is the optimal embedding dimension.The S-map method produces a local linear model with coefficient vector $\mathbf{C}^t$, which predicts the future value z^{t+1} from the reconstructed state-space vector $\mathbf{z}^t$. Then the forecast at $t+1$ is:

$$\widetilde{z^{t+1}} = \sum_{j=1}^{E} C_j^t z^{t-j+1} + C_0^t \tag{S23}$$

where C_0^t denotes intercept term, $\mathbf{C}^t = (C_1^t,\ C_2^t,\ \cdots, C_E^t)'$ is first estimated by singular value decomposition (SVD), after which C_0^t is determined accordingly. Specifically, $\mathbf{C}^t$ is obtained using historical embedded vectors from the library set, i.e., $\{\mathbf{z}^E, \mathbf{z}^{E+1}, \cdots, \mathbf{z}^{m+l-2}\}$ as follows:

$$\mathbf{B}_{m\times 1} = A_{m\times E} \cdot \mathbf{C}_{E\times 1}. \tag{S24}$$

In particular, $B_i = \omega(\|\mathbf{z}^i - \mathbf{z}^t\|)z^{i+1}$, $A_{ij} = \omega(\|\mathbf{z}^i - \mathbf{z}^t\|)z^{i-j+1}$, $\omega(d) = e^{-\delta d/\bar{d}}$ where $\delta \geq 0$ measuring the extent of nonlinearity, d is the distance between the $\mathbf{z}^t$ and the neighbor vector in the library set, and $\bar{d}$ is the average distance between neighbors. Note that if $\delta = 0$, the weighting is equal across all points on the attractor (and independent of distance), and hence S-maps reduces to simple vector autoregression (VAR) model. For $\delta > 0$, the coefficients of $\mathbf{C}$ can vary with location on the attractor and with increasing δ they can vary more strongly as the system changes state.

Eq. (S23) can be concisely rephrased as

$$\begin{pmatrix} z^{t+1} \\ z^t \\ \vdots \\ \vdots \\ z^{t-(E-2)} \end{pmatrix} = \begin{bmatrix} C_1^t & C_2^t & \cdots & C_{E-1}^t & C_E^t \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} z^t \\ z^{t-1} \\ \vdots \\ \vdots \\ z^{t-(E-1)} \end{pmatrix} + \begin{pmatrix} C_0^t \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix} \tag{S25}$$

to be consistent with the form $\mathbf{z}^{t+1} = J^t\mathbf{z}^t + \mathbf{v}^t$, which represents a local linear approximation around $\mathbf{z}^t$ where J^t denotes the local Jacobian matrix and $\mathbf{v}^t$ denotes the offset vector.

For high-dimensional systems $\mathbf{X}^t = (x_1^t, x_2^t, \cdots x_n^t)'$, a multivariate approach[3] is used instead of taking lags to reconstruct the attractor, that is,

$$\begin{pmatrix} x_1^{t+1} \\ x_2^{t+1} \\ \vdots \\ x_n^{t+1} \end{pmatrix} = \begin{bmatrix} j_{11} & \cdots & \cdots & j_{1n} \\ j_{21} & \cdots & \cdots & j_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ j_{n1} & \cdots & \cdots & j_{nn} \end{bmatrix} \begin{pmatrix} x_1^t \\ x_2^t \\ \vdots \\ x_n^t \end{pmatrix} + \begin{pmatrix} v_1^t \\ v_2^t \\ \vdots \\ v_n^t \end{pmatrix} \tag{S26}$$

where the i-th row of the Jacobian matrix consists of the coefficients from the multivariate S-map. Subsequently, the eigenvalues of the local Jacobian matrix can be computed to distinguish between different types of bifurcations.

Supplementary Note S5. Performance contrasts of stARC under different settings

Performance contrasts of stARC under different neighbor settings

The neighbor number L controls the extent of local spatial interaction incorporated into the stARC framework. To assess its influence, we conducted additional robustness analyses by varying L and noise strength σ over a range of values (see Table S4). In the absence of neighborhood information (i.e., $L = 0$), stARC exhibits a pronounced performance degradation in the coupled Lorenz and Hénon systems across all tested noise levels. In contrast, the ADVP system is less sensitive to the removal of neighborhood information, with noticeable deterioration observed only under noise strength $\sigma = 0.5$. Once neighborhood information is incorporated, stARC consistently achieves reliable performance across different neighborhood sizes ($L = 5,10,$ and 15). Specifically, both ACC and TPR remain above 90% across all tested noise levels. These results indicate that stARC performs stably over the tested range of neighborhood sizes, suggesting that the framework is robust to the specific choice of L.

Performance contrasts of stARC under different regularization parameters

As shown in Fig. S2, the regularization parameter θ $(0 \leq \theta \leq 1)$ is predetermined to balance the trade-off between maximizing the projection variance (the first term in the stARC loss) and enforcing delay-embedding consistency (the second term in the stARC loss). In the extreme case where $\theta = 1$, the optimization is governed solely by the delay-embedding constraint ("Embedding constraint" in Fig. S2), while the objective of maximizing the variance of $\tilde{Z}$ is no longer enforced. Although this setting may improve anti-diagonal consistency of $\tilde{Z}$, the resulting one-dimensional representation becomes less sensitive to dynamical instability and thus loses its effectiveness for critical transition detection. Conversely, when θ is too small, the delay-embedding constraint becomes insufficiently enforced, so that the elements along each anti-diagonal of $\tilde{Z}$ may differ significantly from one another. In this case, the optimization becomes dominated by the fluctuation term ("Projection variance" in Fig. S2) and tends to degenerate into a PCA-like low-dimensional approximation, which primarily captures variance structure rather than preserving the underlying dynamics without distortion in a one-dimensional representation. Therefore, θ should lie within an intermediate range in which delay-embedding consistency is preserved while instability-related fluctuations remain detectable.

As shown in Table S5, the bifurcation-identification performance of stARC depends on θ in a system-specific manner rather than following a universal trend. For the coupled Lorenz system, stARC remains robust for relatively large θ values ($\theta >$ 0.8), whereas performance degrades at $\theta = 0.6$. By contrast, for the coupled Hénon and ADVP systems, high ACC and TPR are achieved over an intermediate range ($\theta =$ 0.6-0.9), while performance drops sharply at $\theta = 0.99$. These results indicate that θ should be chosen within a system-dependent range that balances delay-embedding consistency and fluctuation sensitivity, which arise important directions for future work to explore how to adaptively select an appropriate θ in theory for a specific dataset.

Supplementary Note S6. Coupled simulation systems

Coupled ADVP systems

The Van der Pol–Duffing (ADVP) system is a classical nonlinear dynamical system that combines the features of the Van der Pol oscillator and the Duffing oscillator. In this study, we consider a coupled ADVP system with parallel resistance, modeled as a continuous-time dynamical system, to evaluate the performance of the proposed stARC. The coupled ADVP system is described by the following equations:

$$\begin{aligned} \dot{x}_i &= -\nu(x_i^3 - \mu x_i - y_i) + Cx_{i-1} + \omega\zeta_{x,i} \\ \dot{y}_i &= x_i - \alpha y_i - z_i + \omega\zeta_{y,i} \qquad\qquad i = 1,2,3, \\ \dot{z}_i &= \beta y_i + \omega\zeta_{z,i} \end{aligned} \tag{S27}$$

where $\zeta_{x,i}$, $\zeta_{y,i}$, and $\zeta_{z,i}$ denote independent Gaussian white noise terms. The coupling term Cx_{i-1} represents that the i-th subsystem is coupled with the $(i-1)$-th subsystem via the x component. When $i = 1$, we set $i - 1 = 3$ so that the system can be closed. Process noise was applied to the state variables in the form of Gaussian white noise, with ω denoting the noise strength. A Hopf bifurcation occurs at the equilibrium point $E_0 = (0,0,0)$ for given positive parameters β, ν with $\mu < 0$, in the neighborhood of $\alpha = (-b_0 + \sqrt{b_0^2 + 4\mu^2\nu^3})/(-2\mu\nu)$ where $b_0 = \beta - \nu + \mu^2\nu$[27]. In this study, we take $\beta = 200$, $\nu = 100$, $\mu = -0.1$, $\omega = 0.001$, $C = 0.1$, with the bifurcation parameter α decreasing from 4.6 to 3.8, thereby crossing the critical value $\alpha = 4.1421$ at which the Hopf bifurcation occurs.

Coupled Lorenz systems

The Lorenz system, originally derived from a simplified model of atmospheric convection, is a classical model in nonlinear dynamics known for its chaotic behavior. In this study, we consider a coupled Lorenz system governed by the following equations:

$$\begin{aligned} \dot{x}_i &= \sigma(y_i - x_i) + Cx_{i-1}^2 + \omega\zeta_{x,i} \\ \dot{y}_i &= \rho x_i - y_i - x_i z_i + \omega\zeta_{y,i} \qquad i = 1,2,3, \\ \dot{z}_i &= -bz_i + x_i y_i + \omega\zeta_{z,i} \end{aligned} \tag{S28}$$

where $\zeta_{x,i}$, $\zeta_{y,i}$, and $\zeta_{z,i}$ denote independent Gaussian white noise terms with intensity ω. The coupling term Cx_{i-1}^2 represents that the i-th subsystem is coupled with the $(i-1)$-th subsystem via the x component. When $i=1$, we set $i-1=3$ so that the system can be closed. In this study, we take $\sigma=4$, $b=8/3$, $\omega=0.1$, $C=0.1$, with the bifurcation parameter ρ varying within the range $[-3,2]$. Numerical simulations indicate that, after introducing the coupling term Cx_{i-1}^2, the coupled system undergoes a transcritical bifurcation near $\rho=1$.

Coupled Hénon map

The Hénon map was originally proposed by Michel Hénon to capture essential features of the Lorenz system while allowing an adjustable degree of dissipation[8]. Although the Hénon map is a two-dimensional discrete-time dynamical system, we construct a coupled Hénon map system to evaluate the efficacy of stARC in handling higher-dimensional dynamics:

$$\begin{aligned} x_i^{t+1} &= (1+l_i)x_i^t - 4(x_i^t)^2 - \frac{x_i^t y_i^t}{1+0.5x_i^t} + Cx_{i-1} + \omega\zeta_{x,i} \\ y_i^{t+1} &= -2y_i^t + \frac{6x_i^t y_i^t}{1+0.5x_i^t} + \omega\zeta_{y,i} \end{aligned} \qquad i=1,2,3, \qquad \text{(S29)}$$

where $\zeta_{x,i}$ and $\zeta_{y,i}$ denote independent Gaussian white noise terms. When $i=1$, we set $i-1=3$ so that the system can be closed. As l_i increases, the system's attractor transitions from a stable fixed point to a stable period-2 orbit. Subsequently, the system undergoes a period-doubling cascade toward chaos, interspersed with periodic windows[8]. In this study, we specifically consider the onset of period-doubling behavior, i.e., the first bifurcation from a fixed point to a period-2 orbit. To this end, we construct a coupled Hénon system, where the coupling strength and noise level are set to $\omega=0.02$, $C=0.02$, while the bifurcation parameter l_1 varies within the range $[0,0.5]$, which includes the critical value $l_1=0.3675$ at which the period-doubling bifurcation occurs. Bifurcation parameters l_2, l_3 are randomly assigned within the range of $[0,0.3]$, ensuring that the corresponding subsystems remain in a pre-bifurcating regime.

For each model, we first chose the fixed parameters (i.e., parameters other than the bifurcation parameter) such that the system converged to a steady state in the absence of noise when the bifurcation parameter was set to its initial value. The corresponding steady-state values were then used as the initial conditions for simulating the stochastic dynamics, thereby eliminating the need for a burn-in period. Finally, we generated the time series by linearly varying the bifurcation parameter over time. Each simulation yielded a time series containing 10,000 data points.

Supplementary Note S7. Real-world datasets

Holocene Stalagmite stable isotope and mineralogy data

Stalagmite DP1, a speleothem 1.6 m in length from Dante Cave in northeastern Namibia, provides a paleoclimate record of a gradual transition from wetter to drier conditions from 4.6 to 3.3 ka BP, a variable but pronounced dry period from 3.3 to 1.8 ka BP, and a wetter but variable period from 1.8 ka to the present[9]. This record is based on 30 U/Th radiometric dates and the resulting calculated growth rates, as well as C and O stable isotope data, the relative proportions of aragonite and calcite in the layers, measurements of stalagmite width along layers, and observations of petrographic surfaces suggestive of changes from drier to wetter conditions and vice versa. The stalagmite's initial deposition, which seemingly followed conditions too wet for stalagmite deposition, coincided with desiccation in the Sahara and the end of the African Humid Period there. Gradual drying continued and led to a sustained very dry period from 3322 ± 11 to 1786 ± 10 BP, a "2-3 ka BP Dry Period". That dry period began and ended abruptly. The abrupt transition from drier to wetter conditions at 1.8 ka coincides with the beginning of the Iron Age in southern Africa, suggesting that wetter conditions facilitated migrations and/or changes in food production that may have contributed to a transition in human technologies and lifestyles. This transition is coeval with transitions to colder conditions in ice core records from Greenland and Antarctica. The DP1 record suggests considerable change over the past 1800 years, with at least three wet/dry cycles. The wettest conditions may have occurred relatively

recently, between 230 and 100 BP (A.D. 1720 and 1850), so that early European explorers may have seen and/or heard reports of conditions among the wettest during the later Holocene in southern Africa.

Fish community structure data

Peter Lake, MI (89° 32' W, 46° 13' N), located at the University of Notre Dame Environmental Research Center, was dominated by a variety of small-bodied fish species (hereafter referred to as planktivores). It was dominated by a variety of small-bodied fish species including pumpkinseeds (*Lepomis gibbosus*), golden shiners (*Notemigonus crysoleucas*), fathead minnows (*Pimephales promelas*), dace (*primarily Phoxinos eos and Phoxinus neogaeus*), brook stickleback (*Culaea inconstans*), and central mudminnow (*Umbra limi*). These fishes consume a mix of plankton and benthos. This study focuses on their effect on chlorophyll through zooplanktivory, and refer to these fishes collectively as planktivores. On day 149 of 2008, 1200 golden shiners were added to Peter Lake to increase the dominance by planktivores. Approximately 39 largemouth bass were present in Peter Lake in early 2008. From 2008 to 2010, biomass of largemouth bass increased as a result of the manipulations and a major recruitment event in 2009. As numbers of 1+ and adult largemouth bass increased, populations of small-bodied planktivorous fishes decreased.

The resulting suppression of planktivorous fish led to cascading effects in the food web, culminating in cyclic oscillations of zooplankton and phytoplankton biomass[10]. Planktivorous fish density (animals caught per trap per hour) was estimated as catch per effort (CPE) in minnow traps deployed in the littoral zone and Phytoplankton biomass was measured by chlorophyll a.

Sediment cores data

The paleohumidity record from the Chew Bahir basin in the southern Ethiopian Rift documents the climate history of eastern Africa for the past ~620 kyrs. The climate in this part of Africa is determined by the interaction of several air streams and convergence zones, which, in combination with the varied topography, large lakes, and

the nearby Indian Ocean, result in a complex spatiotemporal distribution of precipitation. The dual passage of the tropical rain belt results in a bimodal distribution of rainfall in the Chew Bahir region today, whereby moisture reaching the Ethiopian highlands comes from the Mediterranean and Red Sea (55%), and from the Indian Ocean (31%). On annual to decadal timescales, the intensity of rainfall is related to the east–west adjustments in the Walker circulation associated with the Indian Ocean Dipole. During the African Humid Period (15-5 ka BP), the Chew Bahir basin was filled with a large freshwater lake up to the ~45 m overflow level to the Turkana basin.

In the years from 2009 to 2014, sediment cores up to ~278 m long were recovered[11] from the southern part of the Chew Bahir basin. High potassium (K) content in the sediment, determined by micro X-ray fluorescence (μXRF) scanning, was previously shown to be a reliable proxy for aridity in the Chew Bahir basin. The major controls on K concentrations are the hydrochemistry of the paleolake and porewater. These factors exert a direct control on the degree of authigenic mineral formation in the sediment, such as low-temperature illitization of smectites and analcime formation during episodes of higher alkalinity and salinity in the closed-basin lake resulting from a drier climate. The progressive authigenic K fixation in smectites with increasing evaporative conditions was found to be further enhanced by excess octahedral layer charge that is caused by Al-to-Mg substitutions in clay minerals with more alkaline and saline lake conditions.

Chick heart data

Experiments were carried out in accordance with the ethical and Health and Safety regulations at McGill University. Aggregates were prepared using the method proposed by DeHaan54. Ventricles of 7-day-old White Leghorn chicken embryo hearts were dissected and dissociated into single cells by trypsinization. The cells were then added to Erlenmeyer flasks containing a culture medium (818A) gassed with 5% CO_2, 10% O_2, 85% N_2 ($pH = 7.4$), and placed on a gyratory shaker for 24-48 hours at 37 °C. This generated aggregates with diameters of approximately 100-200 μm that displayed a beating pattern of period with a period of approximately 1-2 seconds. Experiments were

conducted 2-6 hours after the aggregates were plated in dishes maintained at 37 °C. The aggregates were treated with 0.5-2.5 μmol of E4031, a drug that blocks the human Ether-à-go-go-Related Gene (hERG) potassium channel55. The beating of the aggregates was recorded using phase-contrast imaging sampled at 40 Hz with a CCD camera (NeuroCD-SM; RedShirtImaging, LLC) at an 80 × 80 pixel spatial resolution, focusing on light-intensity variations at the edge of the aggregate. There were periodic interruptions in the recording (for 2-3 minutes) for data storage purposes. The light signal for each aggregate was processed through a band-pass filter (cutoff frequencies: 0.1-6.5 Hz). The timing of each beat was then determined as the moment when the signal passes a threshold (the mean of the record plus 0.7 times the standard deviation) with a positive slope. The interbeat intervals were computed as the time between consecutive beats, and were used in the analysis of this study.

Supplementary Note S8. Comparison methods

Multi-variable dynamic eigenvalue method

Multi-variable dynamic eigenvalue method (MultiDEV)[12] is an analytical method for estimating the local Jacobian matrix of high-dimensional dynamical systems without requiring time-delayed embedding for attractor reconstruction. The method is parameterized by the following set of equations:

$$\begin{pmatrix} x_1^{t+1} \\ x_2^{t+1} \\ \vdots \\ x_n^{t+1} \end{pmatrix} = \begin{bmatrix} j_{11} & j_{12} & \cdots & j_{1n} \\ j_{21} & j_{22} & \cdots & j_{21} \\ \vdots & \vdots & \ddots & \vdots \\ j_{n1} & j_{n2} & \cdots & j_{nn} \end{bmatrix} \begin{pmatrix} x_1^t \\ x_2^t \\ \vdots \\ x_n^t \end{pmatrix} + \begin{pmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_n \end{pmatrix} \tag{S30}$$

Where where x_i^t denotes the state of the i-th variable at time t, n is the number of variables, and ν_i denotes the intercept term in the local linear approximation. The i-th column of the Jacobian matrix is composed of the coefficients obtained from the multivariate S-map fitted to the local state patch x_i. The dominant eigenvalue is subsequently computed from the eigenvalues of the Jacobian matrix and is used to anticipate a critical transition when its modulus exceeds one, indicating a loss of local stability.

One-class support vector machine

One-Class Support Vector Machine[13] (OCSVM) algorithm is a widely used unsupervised learning method for outlier detection or critical transition detection. OCSVM learns a decision boundary that encloses the majority of the training data, which are assumed to represent the "normal" class, by solving a quadratic optimization problem in a high-dimensional feature space. During inference, samples falling outside the learned boundary are identified as anomalies.

In this study, we employed an OCSVM with a radial basis function (RBF) kernel. The model was trained using data sampled from the non-bifurcating period (specifically, the first 3,000 of the 10,000 time points) to learn the characteristic patterns of system stability. Once trained, the model was applied to the entire time series to identify potential deviations that might signal upcoming critical transitions.

Long-short term memory recurrent neural network

Long short-term memory recurrent neural networks[14] (LSTM RNNs) have been proposed as effective models for anomaly detection due to their ability to capture temporal dependencies in time series data. In this method, an LSTM RNN is first applied to predict the current value based on historical observations. To detect collective anomalies, a circular array is introduced to store prediction errors over recent time steps. If the prediction errors exceed a predefined threshold and persist over a specified duration, a collective anomaly is identified.

DeepCluster

Recently, deep clustering, which can learn clustering-friendly representations using deep neural networks, has been broadly applied in a wide range of clustering tasks. In particular, DeepCluster was introduced as an unsupervised method that jointly optimizes the neural network parameters and the cluster assignments of the learned feature representations. The method follows an iterative optimization procedure: in each iteration, a convolutional neural network is used to extract features from the input data, and then, a standard clustering algorithm such as *k*-means is applied to these

features to assign cluster labels. These cluster assignments are subsequently treated as pseudo-labels to supervise network training using a classification loss (e.g., cross-entropy). This process of alternately performing clustering and representation learning continues until convergence, typically determined by the stabilization of cluster assignments or the saturation of model performance.

Robust deep autoencoder

Deep autoencoders have proven effective in learning nonlinear feature representations across a wide range of domains. To enhance the robustness to noisy data, the Robust Deep Autoencoder[15] (RDA) was proposed, aiming to preserve the expressive power of deep autoencoders while simultaneously isolating outliers and noise, without requiring access to clean training data.

Let $\mathbf{X}^t = (x_1^t, x_2^t, \cdots, x_n^t)'$ denote the n-dimensional input vector at time t, X denote the input matrix formed by sequential time points. Inspired by Robust Principal Component Analysis (RPCA), RDA assumes that the input data X can be decomposed as $X = L_D + S$, where L_D denotes the component that can be well reconstructed by a deep autoencoder, and S captures the sparse noise and outliers present in the original data X.

The optimization objective is defined as:

$$\min_{\theta,S} \|X - f_\theta(X - S)\|_2 + \lambda \parallel S \parallel_1 \tag{S31}$$

where $f_\theta(\cdot)$ denotes the deep autoencoder parameterized by θ, $\|\cdot\|_2$ is the L_2 norm measuring reconstruction error, $\parallel S \parallel_1$ is the L_1 norm encouraging sparsity in the outlier component, and λ is a regularization parameter balancing the trade-off between reconstruction accuracy and robustness.

After optimization, the sparse component S serves as an anomaly indicator matrix. Specifically, if an element S_{jt} is significantly non-zero, it suggests that the corresponding input value x_j^t cannot be explained by the learned data manifold, and is thus likely to be an anomaly. Conversely, small or zero values in S indicate that the corresponding input conforms well to the learned structure. Accordingly, the anomaly

score for each input $\mathbf{X}^t$ is defined as $\| \mathbf{S}_t \|_1$ where $\mathbf{S}_t$ denotes the corresponding sparse error vector, i.e., $(S_{1t}, S_{2t}, \cdots, S_{nt})'$, and $\mathbf{X}^t$ is flagged as anomalous if this score exceeds a predefined threshold.

Ridge-regularized Dynamic Mode Decomposition

Ridge-regularized Dynamic Mode Decomposition (Ridge-DMD)[16] is a data-driven approach for approximating the underlying linear dynamics of high-dimensional systems from time-series data, with ridge regularization introduced to stabilize operator estimation and mitigate noise effects. Let $\mathbf{X}^t = (x_1^t, x_2^t, \cdots, x_n^t)'$ denote the n-dimensional input vector at time t, X denote the input matrix formed by sequential time points $t = 1,2, \cdots, m$. Ridge-DMD construct matrices

$$X_1 = [\mathbf{X}^1, \mathbf{X}^2, \cdots, \mathbf{X}^{m-1}], X_2 = [\mathbf{X}^2, \mathbf{X}^3, \cdots, \mathbf{X}^m],$$

and, to improve numerical stability and reduce noise effects, project the data onto a rank-k subspace using singular value decomposition (SVD):

$$X_1 \approx U_k \Sigma_k V_k', \tag{S32}$$

where $U_k \in \mathbb{R}^{n \times k}$ contains the leading left singular vectors. The projected data matrices are then given by $Y_1 = U_k' X_1$ and $Y_2 = U_k' X_2$. The reduced linear operator is estimated using Tikhonov (ridge) regularization:

$$\tilde{A} = Y_2 Y_1' (Y_1 Y_1' + \alpha I)^{-1}, \tag{S33}$$

where $\alpha > 0$ is the regularization parameter controlling the trade-off between data fidelity and numerical stability. Finally, the eigenvalues of $\tilde{A}$ are computed to characterize the dynamical behavior of the system:

$$\tilde{A} W = W \Lambda, \tag{S34}$$

where Λ contains the DMD eigenvalues. These eigenvalues are subsequently used to analyze the stability properties of the reconstructed dynamics for critical-transition detection and bifurcation identification.

Neural Koopman

Neural Koopman approaches[17] use deep neural networks to learn data-driven representations of Koopman eigenfunctions, thereby embedding nonlinear dynamics into a low-dimensional latent space[17]. Within this space, the evolution of the original system can be approximated by a linear operator, typically learned within an autoencoder framework.

Given a sequence of snapshot vectors $\{\mathbf{X}^1, \mathbf{X}^2, \cdots, \mathbf{X}^M\}$, the encoder maps each input into a latent variable:

$$\mathbf{Z}^t = \mathcal{E}(\mathbf{X}^t), t = 1, 2, \cdots, M, \tag{S35}$$

while the decoder reconstructs the original observations as $\widetilde{\mathbf{X}^t} = \mathcal{D}(\mathbf{Z}^t)$, $\mathbf{Z}^t \in \mathbb{R}^d$ denotes the d-dimensional latent representation of $\mathbf{X}^t$. Within the latent space, we assume that the dynamics evolve approximately linearly, i.e.,

$$\mathbf{Z}^{t+1} \approx K\mathbf{Z}^t, \tag{S36}$$

where K denotes the Koopman operator in the learned coordinates. For each sliding window of length m $(m < M)$, K is estimated using a Tikhonov-regularized least-squares formulation:

$$K = Z_2 Z_1' (Z_1 Z_1' + \alpha I)^{-1}, \tag{S37}$$

Where $Z_1 = [\mathbf{Z}^1, \mathbf{Z}^2, \cdots, \mathbf{Z}^{m-1}], Z_2 = [\mathbf{Z}^2, \mathbf{Z}^3, \cdots, \mathbf{Z}^m]$, and $\alpha > 0$ is a regularization parameter that improves numerical stability under noise.

The model is trained by jointly optimizing three objectives: (i) a reconstruction loss to preserve information in the latent representation, (ii) a one-step linear consistency loss enforcing $\mathbf{Z}^{t+1} \approx K\mathbf{Z}^t$, and (iii) a multi-step consistency loss that constrains the long-term evolution $\mathbf{Z}^{t+k} \approx K^k\mathbf{Z}^t$. The overall loss function is given by

$$\mathcal{L} = \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + \lambda_{\text{lin}}\mathcal{L}_{\text{lin}} + \lambda_{\text{multi}}\mathcal{L}_{\text{multi}}, \tag{S38}$$

where λ_{rec}, λ_{lin} and λ_{multi} are weighting coefficients.

After training, local Koopman operators are re-estimated within each sliding window in the latent space, and their eigenvalues are computed to characterize the stability of the reconstructed dynamics. In particular, the spectral radius $\rho = \max_i |\lambda_i|$ and the corresponding eigenvalue (i.e., the dominant eigenvalue) are used as indicators of dynamical stability to detect critical transitions and identify bifurcation types.

Supplementary Note S9. Performance matrices

To evaluate the performance of stARC on bifurcation identification task, the following evaluation metrics are employed.

1. True Positives (TP): The number of bifurcating samples correctly identified as bifurcations.
2. True Negatives (TN): The number of non-bifurcating samples correctly identified as non-bifurcations.
3. False Positives (FP): The number of non-bifurcating samples incorrectly identified as bifurcations. In addition, bifurcating samples that are misclassified into the wrong bifurcation type are also treated as false positives.
4. False Negatives (FN): The number of bifurcating samples incorrectly identified as non-bifurcations.

Based on the indicators above, some other assessment metrics are derived. The TPR (true positive rate), i.e., sensitivity or recall, refers to the ratio of true positive cases to all positive cases. Accuracy represents the ratio between the true positive and negative cases and all positive and negative cases. These metrics can be mathematically represented as:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{S39}$$

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}. \tag{S40}$$

Besides, to evaluate the performance on EWS detection task of a specific method, including stARC, the multi-variable dynamic eigenvalue method[12] (multiDEV), one-class support vector machine[13] (OCSVM), Long-Short Term Memory Recurrent Neural Network[14] (LSTM RNN), DeepCluster, Robust Deep Autoencoder[15] (RDA), Ridge-regularized Dynamic Mode Decomposition (Ridge-DMD)[16] and Neural Koopman[17], success rate (SR) and accuracy are employed under consistent experimental settings.

1. Successful detection (Success): The number of bifurcating samples in which an EWS is detected prior to the bifurcation point.

2. Failed detection (Failure): The number of bifurcating samples in which an EWS is detected after the bifurcation point.

3. Accurate detection: The number of bifurcating samples in which an EWS is detected prior to the bifurcation point and non-bifurcating samples in which an EWS is not detected.

The success rate can be mathematically represented as:

$$\mathrm{SR} = \frac{\mathrm{Success}}{\mathrm{Success} + \mathrm{Failure}}. \tag{S41}$$

Supplementary Reference

1. Jaeger, H. The “echo state” approach to analysing and training recurrent neural networks – with an Erratum note. **148**, 13 (2001).
2. Adams, R. P. & MacKay, D. J. C. Bayesian Online Changepoint Detection. Preprint at http://arxiv.org/abs/0710.3742 (2007).
3. Deyle, E. R., May, R. M., Munch, S. B. & Sugihara, G. Tracking and forecasting ecosystem interactions in real time. *Proc. R. Soc. B.* **283**, 20152258 (2016).
4. Nonlinear forecasting for the classification of natural time series. *Phil. Trans. R. Soc. Lond. A* **348**, 477–495 (1994).
5. Cenci, S. Non-parametric estimation of the structural stability of non-equilibrium community dynamics. *Nat. Ecol. Evol.* **3**, 912–918.
6. Dixon, P. A., Milicich, M. J. & Sugihara, G. Episodic Fluctuations in Larval Supply. *Science* **283**, 1528–1530 (1999).
7. Matouk, A. E. & Agiza, H. N. Bifurcations, chaos and synchronization in ADVP circuit with parallel resistor. *J. Math. Anal. Appl.* **341**, 259–269 (2008).
8. Strogatz, S. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. (Perseus Books, Reading, Mass., 1999).
9. Sletten, H. R. *et al.* A petrographic and geochemical record of climate change over the last 4600 years from a northern Namibia stalagmite, with evidence of abruptly wetter climate at the beginning of southern Africa’s Iron Age. *Palaeogeogr. Palaeocl. Palaeoecol.* **376**, 149–162 (2013).
10. Carpenter, S. R. *et al.* Early Warnings of Regime Shifts: A Whole-Ecosystem Experiment. *Science* **332**, 1079–1082 (2011).
11. Trauth, M. H. *et al.* Early warning signals of the termination of the African Humid Period(s). *Nat. Commun.* **15**, 3697 (2024).
12. Grziwotz, F. *et al.* Anticipating the occurrence and type of critical transitions. *Sci. Adv.* **9**, eabq4558 (2023).
13. Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J. & Williamson, R. C. Estimating the Support of a High-Dimensional Distribution. *Neural Comput.* **13**, 1443–1471 (2001).
14. Bontemps, L. Collective Anomaly Detection based on Long Short Term Memory Recurrent Neural Network. in *Future Data and Security Engineering (FDSE 2016)* (eds Dang, T. et al.) vol. 10018 (Springer, Cham, 2016).
15. Zhou, C. & Paffenroth, R. C. Anomaly Detection with Robust Deep Autoencoders. in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 665–674 (ACM, Halifax NS Canada, 2017). doi:10.1145/3097983.3098052.
16. Tu, J. H., Rowley, C. W., Luchtenburg, D. M., Brunton, S. L. & Kutz, J. N. On dynamic mode decomposition: Theory and applications. *J. Comput. Dyn.* **1**, 391–421 (2014).
17. Lusch, B., Kutz, J. N. & Brunton, S. L. Deep learning for universal linear embeddings of nonlinear dynamics. *Nat. Commun.* **9**, 4950 (2018).