Supplementary Information for

Digital twins of nonlinear dynamical systems

Ling-Wei Kong, Yang Weng, Bryan Glaz, Mulugeta Haile, and Ying-Cheng Lai

Corresponding author: Ying-Cheng Lai (Ying-Cheng.Lai@asu.edu)

CONTENTS

Supplementary Note 1. A DRIVEN CHAOTIC LASER SYSTEM

We consider the single-mode, class B, driven chaotic CO_2 laser system [1–4] described by

$$\frac{du}{dt} = -u[f(t) - z], \tag{S1.1}$$

$$\frac{dz}{dt} = \epsilon_1 z - u - \epsilon_2 zu + 1, \tag{S1.2}$$

where the dynamical variables u and z are proportional to the normalized intensity and the population inversion, $f(t) = A\cos(\Omega t + \phi)$ is the external sinusoidal driving signal of amplitude A and frequency Ω, ϵ_1 and ϵ_2 are two parameters. Chaos is common in this laser system [1, 2, 4]. For example, for $\epsilon_1 = 0.09$, $\epsilon_2 = 0.003$, and $A = 1.8$, there is a chaotic attractor for $\Omega < \Omega_c \approx 0.912$, as shown by a sustained chaotic time series in Fig. S1(a1). The chaotic attractor is destroyed by a boundary crisis [5] at Ω_c. For $\Omega > \Omega_c$, there is transient chaos, after which the system settles into periodic oscillations, as exemplified in Fig. S1(a2). Suppose chaotic motion is desired. The crisis bifurcation at Ω_c can then be regarded as a kind of system collapse.

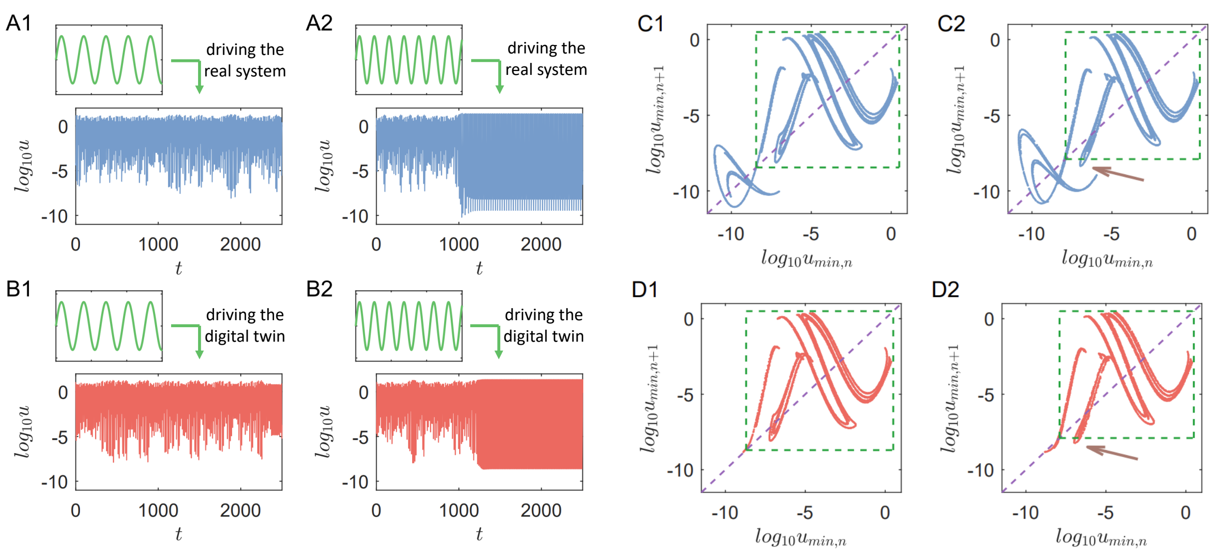

FIG. S1. Performance of the digital twin of the driven CO_2 laser system to extrapolate system dynamics under different driving frequencies. (A1, A2) True sustained and transient chaotic time series of $\log_{10} u(t)$ of the target system, for driving frequencies $\Omega = 0.905 < \Omega_c$ and $\Omega = 0.925 > \Omega_c$, respectively, where the sinusoidal driving signal $f(t)$ is schematically illustrated. In (A1), the system exhibits sustained chaos. In (A2), the system settles into a periodic state after transient chaos. (B1, B2) The corresponding time series generated by the digital twin. In both cases, the dynamical behaviors generated by the digital twin agree with the ground truth in (A1, A2): sustained chaos in (B1) and transient chaos to a periodic attractor in (B2). (C1, C2) The return maps constructed from the local minima of $u(t)$ from the true dynamics, where the green dashed square defines an interval that contains the chaotic attractor (C1) or a nonattracting chaotic set due to the escaping region (marked by the red arrow) leading to transient chaos (C2). (D1, D2) The return maps generated by the digital twin for the same values of Ω as in (C1, C2), respectively, which agree with the ground truth.

To build a digital twin for the chaotic laser system, we use the external driving signal as the natural control signal for the reservoir dynamical network. Different from the examples in the main text, here the driving frequency Ω, instead of the driving amplitude A, serves as the bifurcation parameter. Assuming observational data in the form of time series are available for several values of Ω in the regime of a chaotic attractor, we train the RC network using chaotic time series collected from four values of $\Omega < \Omega_c$: $\Omega = 0.81$, 0.84, 0.87, and 0.90. The training parameter setting is as follows. For each Ω value in the training set, the training and validation lengths are $t = 2,000$ and $t = 83$, respectively, where the latter corresponds to approximately five Lyapunov times. The "warming up" length is $t = 0.5$. The time step of the reservoir system is $\Delta t = 0.05$. The size of the random RC network is $D_r = 800$. The optimal hyperparameter values are determined to be $d = 151$, $\lambda = 0.0276$, $k_{\text{in}} = 1.18$, $k_c = 0.113$, $\alpha = 0.33$, and $\beta = 2 \times 10^{-4}$.

Figures S1(A1) and S1(A2) show two representative time series from the actual laser system (the ground truth) for $\Omega = 0.905 < \Omega_c$ and $\Omega = 0.925 > \Omega_c$, respectively: one is associated with sustained chaos (pre-critical) and the other is characteristic of transient chaos with a final periodic attractor (post-critical). The corresponding time series generated by the digital twin are shown in Figs. S1(B1) and S1(B2), respectively. It can be seen that the training aided by the control signal enables the digital twin to correctly capture the dynamical climate of the target system, e.g., sustained or transient chaos. The true return maps in the pre-critical and post-critical regimes are shown in Figs. S1(C1) and S1(C2), respectively, and the corresponding maps generated by the digital twin are shown in Figs. S1(D1) and S1(D2). In the pre-critical regime, an invariant region (the green dashed square) exists on the return map in which the trajectories are confined, leading to sustained chaotic motion, as shown in Figs. S1(C1) and S1(D1). Within the invariant region in which the chaotic attractor lives, the digital twin captures the essential dynamical features of the attractor. Because the training data are from the chaotic attractor of the target system, the digital twin fails to generate the portion of the real return map that lies outside the invariant region, which is expected because the digital twin has never been exposed to the dynamical behaviors that are not on the chaotic attractor. In the post-critical regime, a "leaky" region emerges, as indicated by the red arrows in Figs. S1(C2) and S1(D2), which destroys the invariant region and leads to transient chaos. The remarkable feature is that the digital twin correctly assesses the existence of the leaky region, even when no such information is fed into the twin during training. From the point of view of predicting system collapse, the digital twin is able to anticipate the occurrence of the crisis and transient chaos. A quantitative result of these predictions are demonstrated in Supplementary Note 6.

As indicated by the predicted return maps in Figs. S1(D1) and S1(D2), the digital twin is unable to give the final state after the transient, because such state must necessarily lie outside the invariant region from which the training data are originated. In particular, the digital twin is trained with time series data from the chaotic attractors prior to the crisis. With respect to Figs. S1(D1) and S1(D2), the digital twin can learn the dynamics within the dash green box in the plotted return maps, but is unable to predict the dynamics outside the box, as it has never been exposed to these dynamics.

A comparison of the real and predicted bifurcation diagram is demonstrated in Fig. S2. The strong resemblance between them indicate the power of the digital twin in extrapolating the correct global behavior of the target system. Moreover, this demonstrates that not only can this approach extrapolate with various driving amplitudes A (as demonstrated in the main text), but the approach can also work with varying driving frequencies Ω.

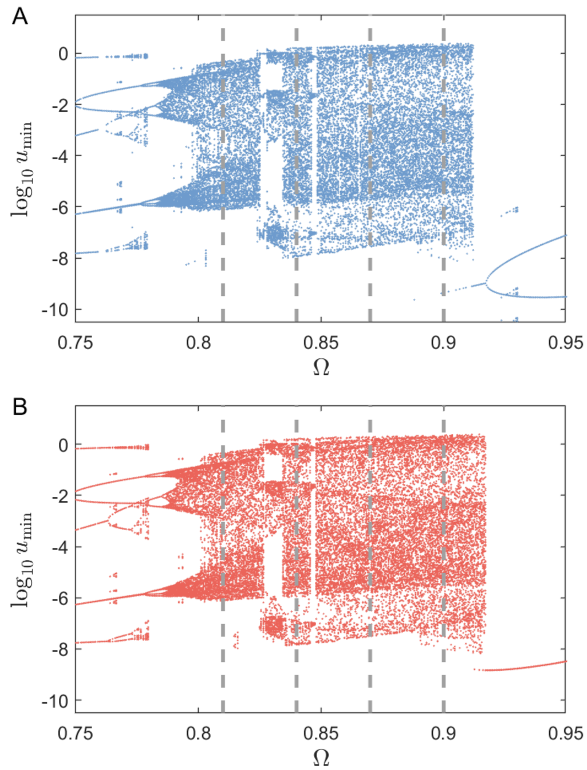

FIG. S2. Comparison of the real (A) and predicted (B) bifurcation diagrams of the driven laser system with varying driving frequencies. The four vertical grey dashed lines indicate the values of driving frequencies Ω used for training the RC neural network. The strong resemblance between the two bifurcation diagrams indicates the power of the digital twin in extrapolating the correct global behavior of the target system, and demonstrates that not only can this approach extrapolate system dynamics to various driving amplitudes A, but also to varying driving frequency Ω.

Supplementary Note 2. A DRIVEN CHAOTIC ECOLOGICAL SYSTEM

We study a chaotic driven ecological system that models the annual blooms of phytoplankton under seasonal driving [6]. Seasonality plays a crucial role in ecological systems and epidemic spreading of infectious diseases [7], which is usually modeled as a simple periodic driving force on the system. The dynamical equations of this model in the dimensionless form are [6]:

$$\frac{dN}{dt} = I - f(t)NP - qN, \tag{S2.3}$$

$$\frac{dP}{dt} = f(t)NP - P, \tag{S2.4}$$

where N represents the level of the nutrients, P is the biomass of the phytoplankton, the Lotka-Volterra term NP models the phytoplankton uptake of the nutrients, I represents a small and constant nutrient flow from external sources, q is the sinking rate of the nutrients to the lower level of the water unavailable to the phytoplankton, and $f(t)$ is the seasonality term: $f(t) = A\sin(\omega_{\text{eco}}t)$. The parameter values are [6]: $I = 0.02$, $q = 0.0012$, and $\omega_{\text{eco}} = 0.19$.

Climate change can dramatically alter the dynamics of this ecosystem [8]. We consider the task of forecasting how the system behaves if the climate change causes the seasonal fluctuation

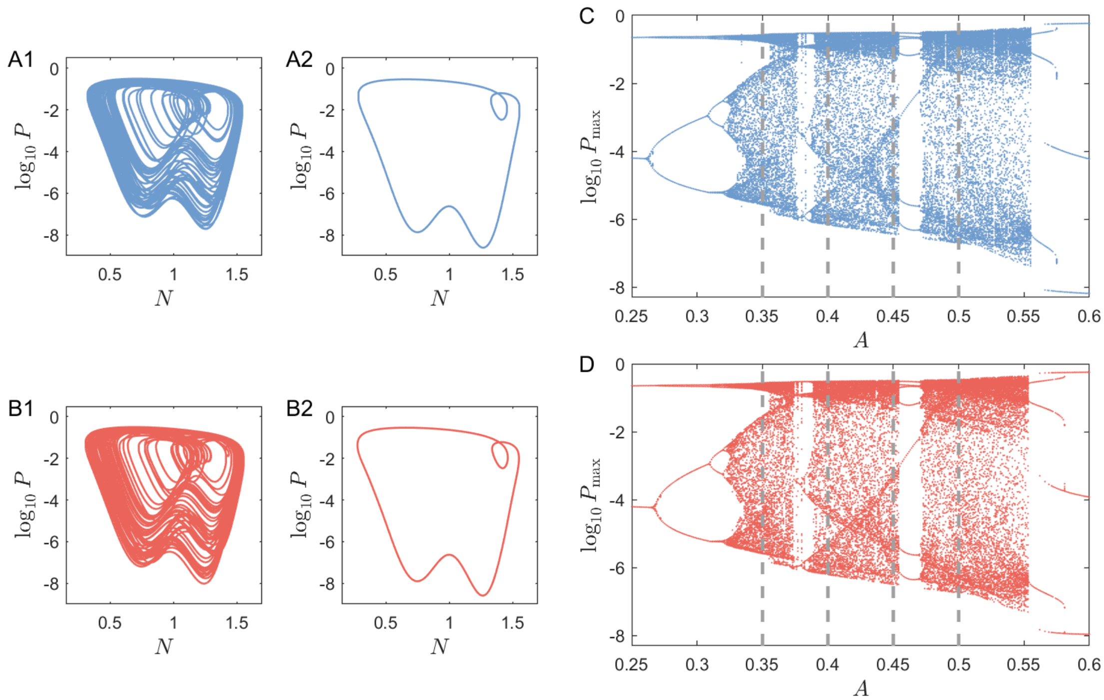

FIG. S3. Performance of the digital twin of an ecological system of the blooms of phytoplankton with seasonality. The effect of seasonality is modeled by a sinusoidal driving signal $f(t) = A\sin(\omega_{\text{eco}}t)$. (A1, A2) chaotic and periodical attractors of this system in the $(N, \log_{10} P)$ plane for $A = 0.45$ and $A = 0.56$, respectively. (B1, B2) The corresponding attractors generated by the digital twin under the same driving signals $f(t)$ as in (A1, A2). The digital twin has successfully extrapolated the periodical behavior outside the chaotic training region. (C) The ground-truth bifurcation diagram of the target system. (D) The digital-twin generated bifurcation diagram. In (C) and (D), the four vertical grey dashed lines indicate the values of driving amplitudes A used for training the RC network. The strong resemblance between the two bifurcation diagrams indicates the power of the digital twin in extrapolating the correct global behavior of the target system.

to be more extreme. In particular, suppose the training data are measured from the system when it behaves normally under a driving signal of relatively small amplitude, and we wish to predict the dynamical behaviors of the system in the future when the amplitude of the driving signal becomes larger (due to climate change). The training parameter setting is as follows. The size of the RC network is $D_r = 600$ with $D_{\text{in}} = D_{\text{out}} = 2$. The time step of the evolution of the network dynamics is $\Delta t = 0.1$. The training and validation lengths for each value of the driving amplitude A in the training are $t = 1,500$ and $t = 500$, respectively. The optimized hyperparameters of the RC are $d = 350$, $\lambda = 0.42$, $k_{\text{in}} = 0.39$, $k_c = 1.59$, $\alpha = 0.131$, and $\beta = 1 \times 10^{-7.5}$.

Figure S3 shows that training the digital twin using time series data from a few different values of the (small) driving amplitude enables it to learn the intrinsic dynamics of the system and generate the correct response of the system to a driving signal of larger amplitude. In particular, the training data are collected with the driving amplitude $A = 0.35, 0.4, 0.45$ and 0.5, all in the chaotic regions. Figures S3(A1) and S3(A2) show the true attractors of the system for $A = 0.45$ and 0.56, respectively, where the attractor is chaotic in the former case (within the training parameter regime) and periodic in the latter (outside the training regime). The corresponding attractors generated by the digital twin are shown in Figs. S3(B1) and S3(B2). It can be seen that the digital twin is able to not only replicate the chaotic behavior in the training data [Fig. S3(B1)] but also predict the transition to a periodic attractor under a driving signal of a larger amplitude (more extreme seasonality), as shown in Fig. S3(B2). In fact, the digital twin can faithfully produce the global dynamical behavior of the system, both inside and outside the training regime, as can be seen from the nice agreement between the ground-truth bifurcation diagram in Fig. S3(C) and the diagram generated by the digital twin in Fig. S3(D).

Supplementary Note 3. CONTINUAL FORECASTING UNDER NONSTATIONARY EXTERNAL DRIVING WITH SPARSE REAL-TIME DATA

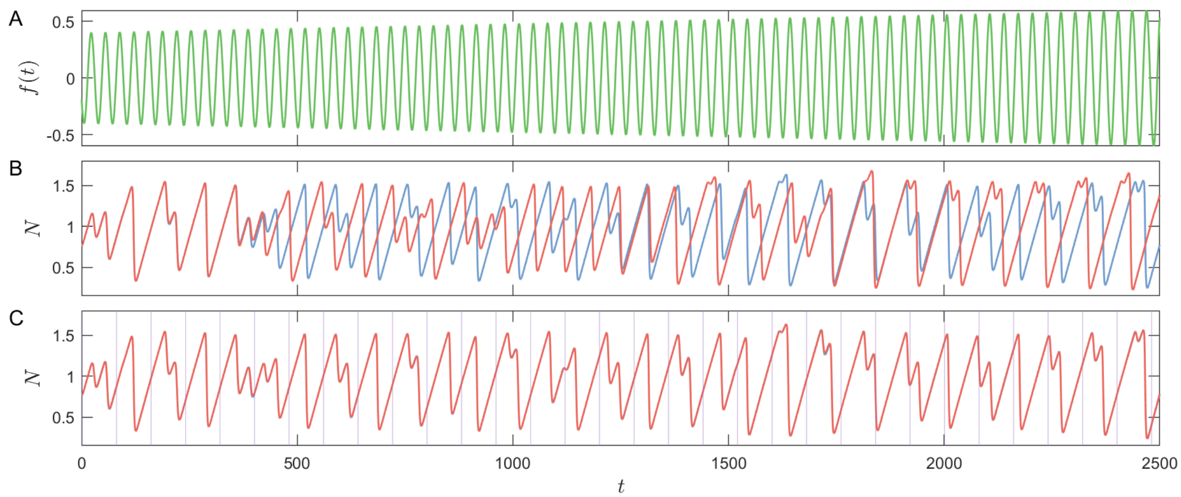

FIG. S4. Continual forecasting of the chaotic ecological system under nonstationary external driving with sparse updates. (A) A non-stationary sinusoidal driving signal $f(t)$ whose amplitude increases with time. The digital twin forecasts the response of the chaotic target system under this driving signal for a relatively long term. (B) The trajectory generated by the digital twin (red) in comparison with the true trajectory (blue). For $0 \leq t \lesssim 400$, the two trajectories match each other with small errors, but the digital-twin generated trajectory begins to deviate from the true trajectory at $t \sim 400$ (due to chaos). (C) With only sparse updates from real data at times indicated by the vertical lines (2.5% of the time steps in the given time interval), the digital twin can make relatively accurate predictions for a long term, demonstrating the ability to perform continual forecasting.

The three examples (Lorenz-96 climate network in the main text, the driven CO_2 laser and the ecological system) have demonstrated that our RC based digital twin is capable of extrapolating and generating the correct statistical features of the dynamical trajectories of the target system such as the attractor and bifurcation diagram. That is, the digital twin can be regarded as a "twin" of the target system only on a statistical sense. In particular, from random initial conditions the digital twin can generate an ensemble of trajectories, and the statistics calculated from the ensemble agree with those of the original system. At the level of individual trajectories, if a target system and its digital twin start from the same initial condition, the trajectory generated by the twin can stay close to the true trajectory only for a short period of time (due to chaos). However, with infrequent state updates, the trajectory generated by the twin can shadow the true trajectory (in principle) for an arbitrarily long period of time [9], realizing *continual forecasting* of the state evolution of the target system.

In data assimilation for numerical weather forecasting, the state of the model system needs to be updated from time to time [10–12]. This idea has recently been exploited to realize long-term prediction of the state evolution of chaotic systems using RC [9]. Here we demonstrate that, even when the driving signal is non-stationary, the digital twin is able to generate the correct state evolution of the target system. As a specific example, we use the chaotic ecosystem in Eqs. (S2.3-S2.4) with the same RC network trained in Sec. Supplementary Note 2. Figure S4(A) shows the non-stationary external driving $f(t) = A(t)\sin(\omega_{\rm eco}t)$ whose amplitude $A(t)$ increases linearly from $A(t=0) = 0.4$ to $A(t = 2500) = 0.6$ in the time interval $[0, 2500]$. Figure S4(B) shows the true (blue) and digital-twin generated (red) time evolution of the nutrient abundance. Due to chaos, without state updates, the two trajectories diverge from each other after a few cycles of oscillation. However, even with rare state updates, the two trajectories can stay close to each other for any arbitrarily long time, as shown in Fig. S4(C). In particular, there are 800 time steps involved in the time interval $[0, 2500]$ and the state of the digital twin is updated 20 times, i.e., 2.5% of the available time series data. We will discuss the results further discussion in the next section.

Supplementary Note 4. CONTINUAL FORECASTING WITH HIDDEN DYNAMICAL VARIABLES

In real world applications, it can occur that not all the variables of the system are accessible: often only a subset of the dynamical variables can be measured and the remaining variables are inaccessible or hidden from the outside world. Can a digital twin still make continual forecasting in the presence of hidden variables based on the time series data from the accessible variables? Also, Can the digital twin do this without knowing that there exists some hidden variables before training? In general, when there are hidden variables, the reservoir network needs to sense their existence, encode them in the hidden state of the recurrent layer, and constantly update them. As such, the recurrent structure of reservoir computing is necessary, because there must be a place for the machine to store and restore the implicit information that it has learned from the data. Compared with the case where complete information about the dynamical evolution of all the observables is available, when there are hidden variables, to predict the evolution of a target system driven by an nonstationary external signal using sparse observations of the accessible variables is significantly more challenging.

As an illustrative example, we again consider the ecosystem described by Eqs. (S2.3) and (S2.4). We assume that the dynamical variable N (the abundance of the nutrients) is hidden and $P(t)$, the biomass of the phytoplankton, can be externally accessible. Despite the accessibility to $P(t)$, we assume that it can be measured only occasionally. That is, only sparsely updated data of the variable $P(t)$ is available. It is necessary that the digital twin be able to learn some equivalent of $N(t)$ as the time evolution of $P(t)$ depends on the value $N(t)$ and to encode the equivalent in the reservoir network. In an actual application, when the digital twin is deployed, knowledge about the existence of such a hidden variable is not required.

Figure S5 present representative results, where Fig. S5(A) shows the non-stationary external driving signal $f(t)$ (the same as the one in Fig. S4(A) - included here as a time reference for the forecasting results). Figure S5(B) shows, when the observable $P(t)$ is not updated with the real data, the forecasted time series (red) $P(t)$ diverges from the true time series (blue) after about a

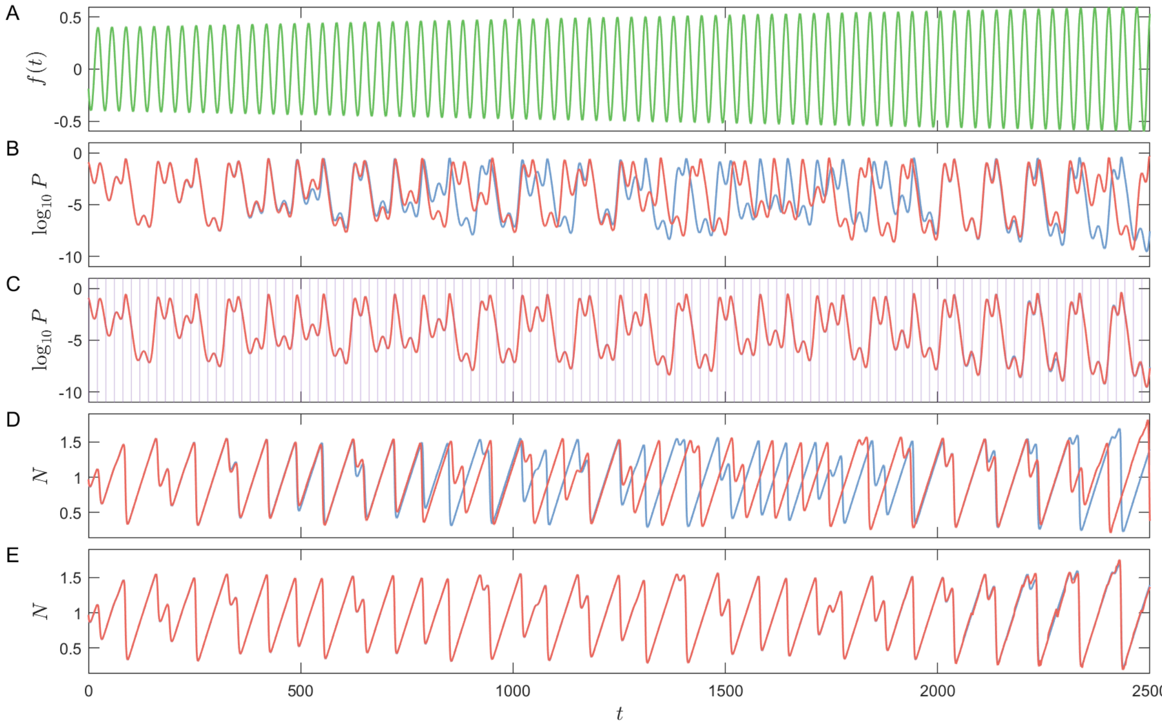

FIG. S5. Continual forecasting and monitoring of a hidden dynamical variable in the chaotic ecological system under non-stationary external driving with sparse updates from the observable. The system is described by Eqs. (S2.3) and (S2.4). The dynamical variable $N(t)$ is hidden, and the other variable $P(t)$ is externally accessible but only sparsely sampled measurement of it can be performed. (A) The nonstationary sinusoidal driving signal $f(t)$ with a time-varying amplitude. (B) Digital-twin generated time evolution of the accessible variable $P(t)$ (red) in comparison with the ground truth (blue) in the absence of any state update of $P(t)$. The predicted time evolution quickly diverges from the true behavior. (C) With sparse updates of $P(t)$ at the times indicated by the purple vertical lines (10% of the times steps), the digital twin is able to make an accurate forecast of $P(t)$. (D) Digital-twin generated time evolution of the hidden variable $N(t)$ (red) in comparison with the ground truth (blue) in the absence of any state update of $P(t)$. (E) Accurate forecasting of the hidden variable $N(t)$ with sparse updates of $P(t)$.

dozen oscillations. However, if $P(t)$ from the digital twin is updated with the true value at the times indicated by the purple vertical lines in Fig. S5(C), the forecasted time series $P(t)$ matches the ground truth for a much longer time. The results suggest that the existence of the hidden variable does not significantly impede the performance of continual forecasting.

The results in Fig. S5 motivate the following questions. First, has the reservoir network encoded information about the hidden variable? Second, suppose it is known that there is a hidden variable and the training dataset contains this variable, can its evolution be inferred with only rare updates of the observables during continual forecasting? Previous results [13–15] suggested that reservoir computing can be used to infer the hidden variables in a nonlinear dynamical system. Here we show that, with a segment of the time series of $N(t)$ used only for training an additional readout layer, our digital twin can forecast $N(t)$ with only occasional inputs of the observable time series $P(t)$. In particular, the additional readout layer for $N(t)$ is used only for extracting information about $N(t)$ from the reservoir network and its output is never injected back to the reservoir. Consequently, the states of the reservoir network during training or forecasting, the trained output layer for $P(t)$ and the forecasting results of $P(t)$ are not altered.

Figure S5(D) show that, when the observable $P(t)$ is not updated with the real data, the digital twin is able to infer the hidden variable $N(t)$ for several oscillations. If $P(t)$ is updated with the true value at the times indicated by the purple vertical lines in Fig. S5(C), the dynamical evolution of the hidden variable $N(t)$ can also be accurately predicted for a much longer period of time, as shown in Fig. S5(E). It is worth emphasizing that during the whole process of forecasting and monitoring, no information about the hidden variable $N(t)$ is required - only sparse data points of the observable $P(t)$ are used.

The training and testing settings of the digital twin for the task involving a hidden variable are as follows. The input dimension of the reservoir is $D_{\text{in}} = 1$ because there is a single observable $\log_{10} P(t)$. The output dimension is $D_{\text{out}} = 2$ with one dimension of the observable $\log_{10} P(t+\Delta t)$ plus one dimension of the hidden variable $N(t+\Delta t)$. Because of the higher memory requirement in dealing with a hidden variable, a somewhat larger reservoir network is needed, so we use $D_r = 1,000$. The times step of the dynamical evolution of the neural network is $\Delta t = 0.1$. The training and validating lengths for each value of the driving amplitude in the training are $t = 3,500$ and $t = 350$, respectively. Other optimized hyperparameters of the reservoir are $d = 450$, $\lambda = 1.15$, $k_{\text{in}} = 0.32$, $k_c = 3.1$, $\alpha = 0.077$, $\beta = 1 \times 10^{-8.3}$, and $\sigma_{\text{noise}} = 10^{-3.0}$.

It is also worth noting that Figs. S4 and S5 have demonstrated the ability of the digital twin to extrapolate beyond the parameter regime of the target system from which the training data are obtained. In particular, the digital twin was trained only with time series under stationary external driving of the amplitude $A = 0.35, 0.4, 0.45$, and 0.5. During the testing phase associated with both Figs. S4 and S5, the external driving is nonstationary with its amplitude linearly increasing from $A = 0.4$ to $A = 0.6$. The second half of the time series $P(t)$ and $N(t)$ in Figs. S4 and S5 are thus beyond the training parameter regime.

The results in Figs. S4 and S5 help legitimize the terminology "digital twin," as the reservoir computers subject to the external driving are dynamical twin systems that evolve "in parallel" to the corresponding real systems. Even when the target system is only partially observable, the digital twin contains both the observable and hidden variables whose dynamical evolution is encoded in the recurrent neural network in the hidden layer. The dynamical evolution of the output is constantly (albeit infrequently) corrected by sparse feedback from the real system, so the output trajectory of the digital twin shadows the true trajectory of the target system. Suppose one wishes to monitor a variable in the target system, it is only necessary to read it out from the digital twin instead of making more (possibly costly) measurements on the real system.

Supplementary Note 5. DIGITAL TWINS UNDER EXTERNAL DRIVING WITH VARYING WAVEFORM

In the main text, it is demonstrated that dynamical noise added to the driving signal during the training can be beneficial. Figure S6 presents a comparison between the noiseless training and the training with dynamical noise of the form $\delta_{\text{DB}} = 3 \times 10^{-3}$ (as in the main text). The ground-truth bifurcation diagram is shown in Fig. S6(A) and three examples with different reservoir neural networks for the noiseless (B1, B2, B3) and noisy (C1, C2, C3) training schemes are shown, where the settings are the same as that in Fig. 3 in the main text. It can be seen that adding dynamical noise into the training data can produce bifurcation diagrams that are closer to the ground truth than without noise.

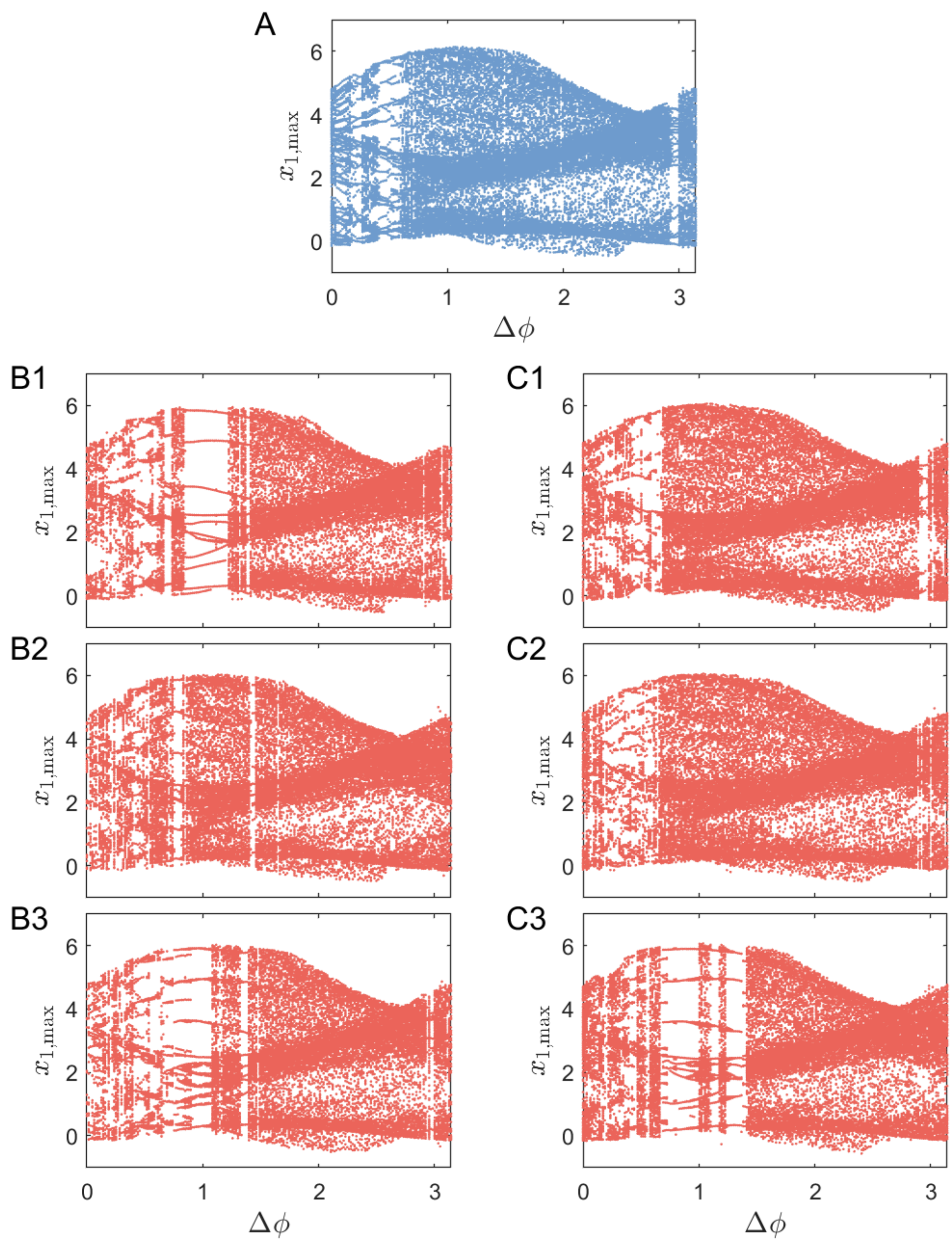

FIG. S6. Comparison of prediction performance between the noiseless (left) and noisy (right) cases on the task of predicting under external driving with varying waveforms. The target system is a six dimensional Lorenz-96 system. Panel (A) shows the true bifurcation diagram. (B1-B3) The prediction results without any dynamical noise in the training data with three realizations of the reservoir network. (C1-C3) Prediction results with dynamical noise of the form $\delta_{\text{DB}} = 3 \times 10^{-3}$ in the training data. The settings are the same as that in Fig. 3 in the main text.

To further demonstrate the beneficial role of noise, we test the additive training noise scheme using the ecological system. The training process and hyperparameter values of the digital twin are identical to these in Supplementary Note 2. A dynamical noise of amplitude $\delta_{\text{DB}} = 3 \times 10^{-4}$ is added to the driving signal $f(t)$ during training in the same way as in Fig. 3 in the main text.

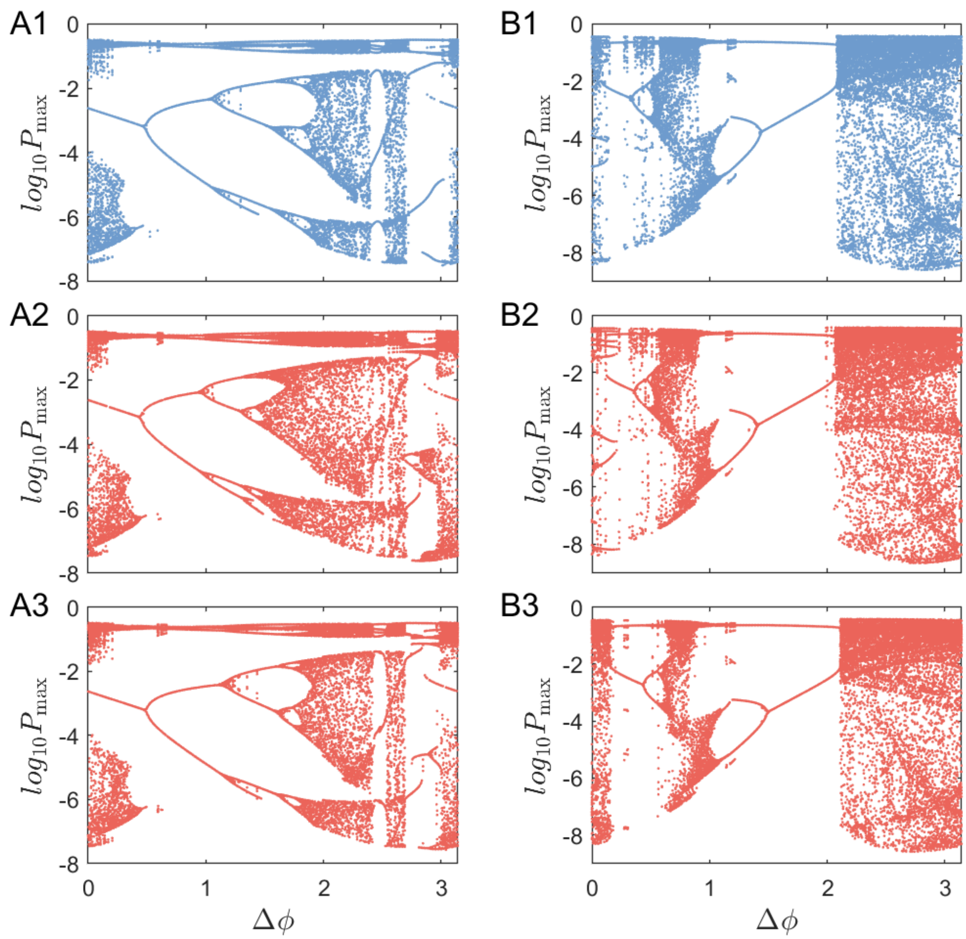

FIG. S7. Performance of the digital twin for the ecological model under driving signals with waveforms different from the training set. The testing driving signals are described by Eq. S5.5 while the training driving signals are sinusoidal waves with small dynamical noise. (A1) The real bifurcation diagram for $A_{\text{test}} = 0.3$. (A2, A3) Predicted bifurcation diagrams for $A_{\text{test}} = 0.3$ with two realizations of the reservoir network. (B1-B3) Same as (A1-A3) but for $A_{\text{test}} = 0.4$.

During testing, the driving signals is altered to

$$f_{\text{test}}(t) = A_{\text{test}} \sin(\omega_{\text{eco}} t) + \frac{A_{\text{test}}}{2} \sin(\frac{\omega_{\text{eco}}}{2} t + \Delta\phi) \tag{S5.5}$$

where $\omega_{\text{eco}} = 0.19$ as in Supplementary Note 2. Two sets of testing signals $f_{\text{test}}(t)$ are used, with $A_{\text{test}} = 0.3$ and 0.4, respectively. Figure S7 show the true and predicted bifurcation diagrams of $\log_{10} P_{max}$ versus $\Delta\phi$ for $A_{\text{test}} = 0.3$ (left column) and $A_{\text{test}} = 0.4$ (right column). It can be seen that the bifurcation diagrams generated by the digital twin with the aid of training noise are remarkably accurate. We also find that, for this ecological system, the amplitude δ_{DB} of the dynamical noise during training does not have a significant effect on the predicted bifurcation diagram. A plausible reason is that the driving signal $f(t)$ is a multiplicative term in the system equations.

Supplementary Note 6. QUANTITATIVE CHARACTERIZATION OF PERFORMANCE OF DIGITAL TWIN

In the main text, we demonstrate the performance of the digital twin qualitatively based on visually comparing the predicted bifurcation diagram with the ground truth. For a bifurcation diagram, the parameter values at which the various bifurcations occur are of great interest, as they define the critical points at which characteristic changes in the system can occur. We focus on the crisis point at which sustained chaotic motion on an attractor is destroyed and replaced by transient chaos. And, accordingly, we use two quantities to characterize the performance of the digital twin in extrapolating the dynamics of the target system: the errors in the predicted critical bifurcation point and average lifetime of the chaotic transient after the bifurcation.

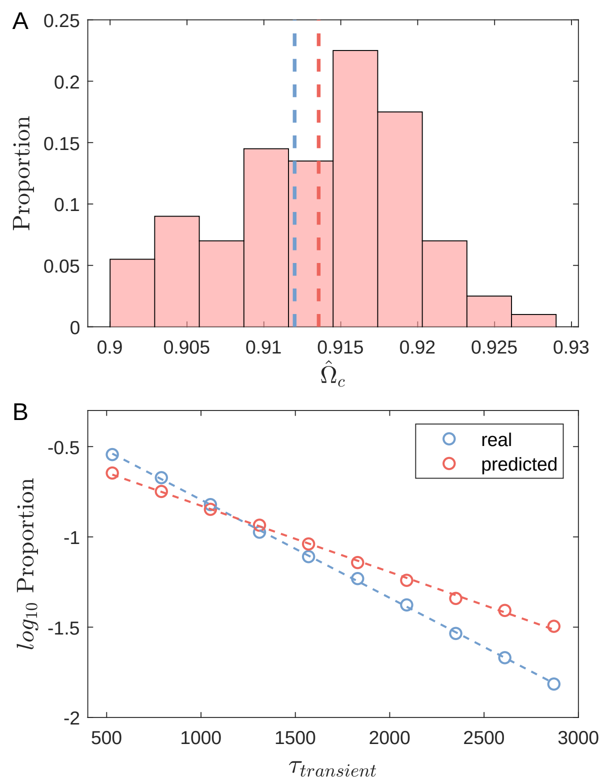

FIG. S8. Quantitative performance of the digital twin for a chaotic driven laser system. (A) Distribution of the predicted values of the crisis bifurcation point $\hat{\Omega}_c$, at which a chaotic attractor is destroyed and replaced by a non-attracting chaotic invariant set leading to transient chaos. The blue and red vertical dashed lines denote the true value $\Omega_c \approx 0.912$ and the average predicted value $\langle\hat{\Omega}_c\rangle$, respectively, where 200 random realizations of the reservoir neural network are used to generate this distribution. Despite the fluctuations in the predicted crisis point, the ensemble average value of the prediction is quite close to the ground truth. (B) Exponential distribution of the lifetime of transient chaos slightly beyond the crisis point: true (blue) and predicted (red) behaviors. The predicted distribution is generated using 100 random reservoir realizations, each with 200 random initial 'warming up" data.

As an illustrative example, we take the driven chaotic laser system in Sec. Supplementary Note 1, where a crisis bifurcation occurs at the critical driving frequency $\Omega_c \approx 0.912$ at which the chaotic attractor of the system is destroyed and replaced by a non-attracting chaotic invariant set

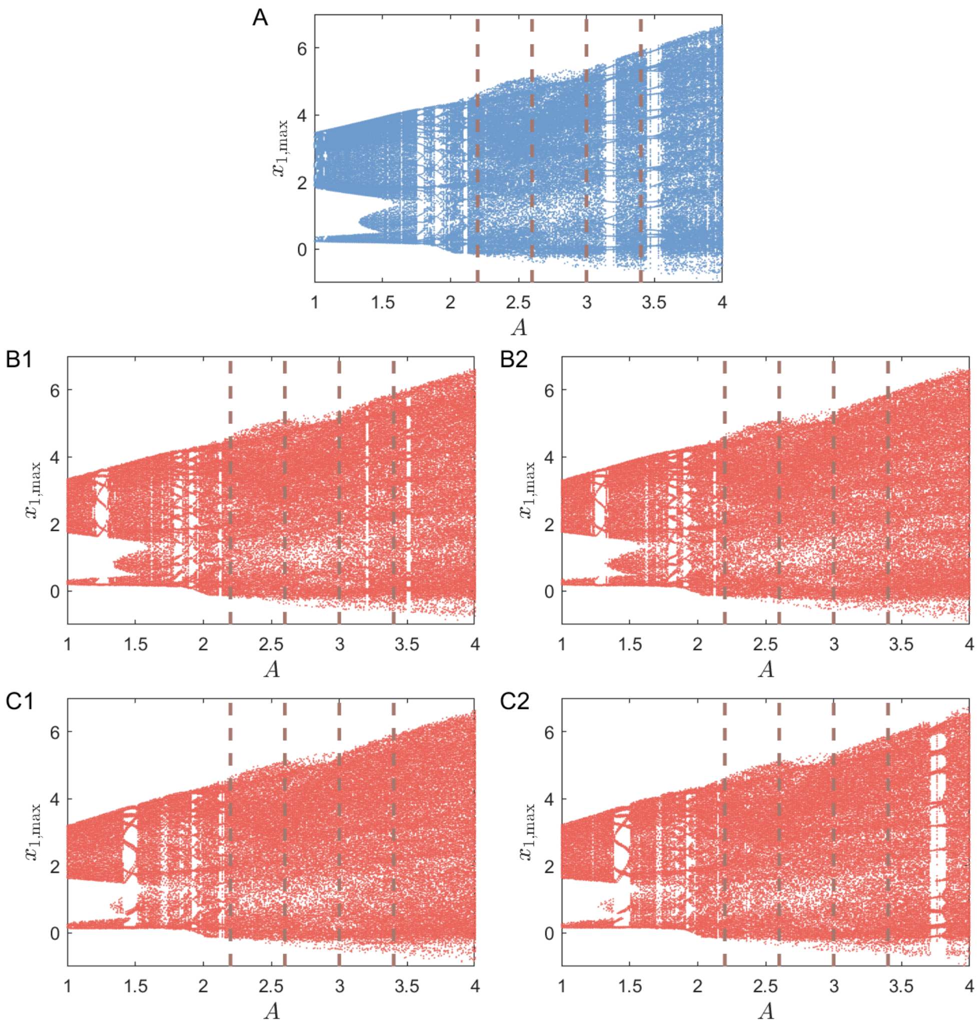

FIG. S9. Robustness of digital twin against combined dynamical and observational noises. The setting is the same as that in Fig. 1 in the main text, except with additional noises in the training data. (A) A true bifurcation diagram of the six-dimensional Lorenz-96 system. (B1, B2) Two examples of the bifurcation diagram predicted by the digital twin with training data under dynamical noise of amplitude $\sigma_{\text{dyn}} = 10^{-2}$ and observational noise of amplitude $\sigma_{\text{ob}} = 10^{-2}$. (C1, C2) Two examples of the predicted bifurcation diagrams under the two kinds of noise with $\sigma_{\text{dyn}} = 10^{-1}$ and $\sigma_{\text{ob}} = 10^{-1}$. Both the dynamical and observational noises are additive Gaussian processes. It can be seen that the additional noises do not have a significant effect on the predicted bifurcation diagrams. The settings of the training data and reservoir neural networks are the same as those in Sec. IIA in the main text. The dynamical noises are added to the equations of the state variables. There is no noise in the sinusoidal external driving.

leading to transient chaos. We test to determine if the digital twin can faithfully predict the crisis point based only on training data from the parameter regime of a chaotic attractor. Let $\hat{\Omega}_c$ be the digital-twin predicted critical point. Figure S8(A) shows the distribution of $\hat{\Omega}_c$ obtained from 200 random realizations of the reservoir neural network. Despite the fluctuations in $\hat{\Omega}_c$, its average value is $\langle \hat{\Omega}_c \rangle \approx 0.914$, which is close to the true value $\Omega_c = 0.912$. A relative error ε_Ω of $\hat{\Omega}_c$ can then be defined as

$$\varepsilon_\Omega = \frac{|\Omega_c - \hat{\Omega}_c|}{D(\Omega_c, \{\Omega_{\text{train}}\})}, \tag{S6.6}$$

where $D(\Omega_c, \{\Omega_{\text{train}}\})$ denotes the minimal distance from Ω_c to the set of training parameter points $\{\Omega_{\text{train}}\}$, i.e., the difference between Ω_c and the closest training point. For the driven laser system, we have $D(\Omega_c, \{\Omega_{\text{train}}\}) \approx 10\%$.

The second quantity is the lifetime $\tau_{\text{transient}}$ of transient chaos after the crisis bifurcation [16, 17], as shown in Fig. S8(B). The average transient lifetime is the inverse of the slope of the linear regression of predicted data points in Fig. S8(B), which is $\langle\tau\rangle \approx 0.8 \times 10^3$. Compared with the true value $\langle\tau\rangle \approx 1.2 \times 10^3$, we see that the digital twin is able to predict the average chaotic transient lifetime to within the same order of magnitude. Considering that key to the transient dynamics is the small escaping region in Fig. S1(D2), which is sensitive to the inevitable training errors, the performance can be deemed as satisfactory.

Supplementary Note 7. ROBUSTNESS OF DIGITAL TWIN AGAINST COMBINED DYNAMICAL/OBSERVATIONAL NOISES

Can our reservoir-computing based digital twins withstand the influences of different types of noises? To address this question, we introduce dynamical and observational noises in the training data, which are modeled as additive Gaussian noises. Take the six-dimensional Lorenz-96 system in Sec. IIA in the main text as an example. Figure S9(A) shows the true bifurcation diagram under different amplitudes of external driving, where the vertical dashed lines specify the training points. Figures S9(B1) and S9(B2) show two realizations of the bifurcation diagram generated by the digital twin under both dynamical and observational noises of amplitudes $\sigma_{\text{dyn}} = 10^{-2}$ and $\sigma_{\text{ob}} = 10^{-2}$. Two bifurcation diagrams for noise amplitudes of an order of magnitude larger: $\sigma_{\text{dyn}} = 10^{-1}$ and $\sigma_{\text{ob}} = 10^{-1}$, are shown in Figs. S9(C1) and S9(C2). It can be seen that the additional noises have little effect on the performance of the digital twin in generating the bifurcation diagram.

---

[1] Dangoisse, D., Glorieux, P. & Hennequin, D. Laser chaotic attractors in crisis. *Phys. Rev. Lett.* **57**, 2657 (1986).

[2] Dangoisse, D., Glorieux, P. & Hennequin, D. Chaos in a CO_2 laser with modulated parameters: experiments and numerical simulations. *Phys. Rev. A* **36**, 4775 (1987).

[3] Solari, H. G., Eschenazi, E., Gilmore, R. & Tredicce, J. R. Influence of coexisting attractors on the dynamics of a laser system. *Opt. Commun.* **64**, 49–53 (1987).

[4] Schwartz, I. B. Sequential horseshoe formation in the birth and death of chaotic attractors. *Phys. Rev. Lett.* **60**, 1359 (1988).

[5] Grebogi, C., Ott, E. & Yorke, J. A. Crises, sudden changes in chaotic attractors and chaotic transients. *Physica D* **7**, 181–200 (1983).

[6] Huppert, A., Blasius, B., Olinky, R. & Stone, L. A model for seasonal phytoplankton blooms. *J. Theo. Biol.* **236**, 276–290 (2005).

[7] Stone, L., Olinky, R. & Huppert, A. Seasonal dynamics of recurrent epidemics. *Nature* **446**, 533–536 (2007).

[8] Winder, M. & Sommer, U. Phytoplankton response to a changing climate. *Hydrobiologia* **698**, 5–16 (2012).

[9] Fan, H., Jiang, J., Zhang, C., Wang, X. & Lai, Y.-C. Long-term prediction of chaotic systems with machine learning. *Phys. Rev. Research* **2**, 012080 (2020).

[10] Kalnay, E. *Atmospheric Modeling, Data Assimilation and Predictability* (Cambridge university press, 2003).

[11] Asch, M., Bocquet, M. & Nodet, M. *Data Assimilation: Methods, Algorithms, and Applications* (SIAM, 2016).

[12] Wikner, A. *et al.* Using data assimilation to train a hybrid forecast system that combines machine-learning and knowledge-based components. *Chaos* **31**, 053114 (2021).

[13] Lu, Z. *et al.* Reservoir observers: Model-free inference of unmeasured variables in chaotic systems. *Chaos* **27**, 041102 (2017).

[14] Roland, Z. S. & Parlitz, U. Observing spatio-temporal dynamics of excitable media using reservoir computing. *Chaos* **28**, 043118 (2018).

[15] Weng, T., Yang, H., Gu, C., Zhang, J. & Small, M. Synchronization of chaotic systems and their machine-learning models. *Phys. Rev. E* **99**, 042203 (2019).

[16] Kong, L.-W., Fan, H.-W., Grebogi, C. & Lai, Y.-C. Machine learning prediction of critical transition and system collapse. *Phys. Rev. Research* **3**, 013090 (2021).

[17] Kong, L.-W., Fan, H., Grebogi, C. & Lai, Y.-C. Emergence of transient chaos and intermittency in machine learning. *J. Phys. Complexity* **2**, 035014 (2021).