Learning to learn ecosystems from limited data

Zheng-Meng Zhai[a], **Bryan Glaz**[b], **Mulugeta Haile**[c], **Alan Hastings**[d,e], **and Ying-Cheng Lai**[a,f,1]

This manuscript was compiled on June 13, 2025

A fundamental challenge in developing data-driven approaches to ecological systems for tasks such as state estimation and prediction is the paucity of the observational or measurement data. For example, modern machine-learning techniques such as deep learning or reservoir computing typically require a large quantity of data. Leveraging synthetic data from paradigmatic nonlinear but non-ecological dynamical systems, we develop a meta-learning framework with time-delayed feedforward neural networks to predict the long-term behaviors of ecological systems as characterized by their attractors. We show that the framework is capable of accurately reconstructing the "dynamical climate" of the ecological system with limited data. Three benchmark population models in ecology, namely the Hastings-Powell model, a three-species food chain, and the Lotka-Volterra system, are used to demonstrate the performance of the meta-learning based prediction framework. In all cases, enhanced accuracy and robustness have been achieved using five to seven times less training data as compared with the corresponding machine-learning method trained solely from the ecosystem data. In addition, two real-world ecological benchmark datasets: the microbial time series dataset and global population dynamics database, are tested to demonstrate the applicability of the meta-learning framework to the real world. A number of issues affecting the prediction performance are addressed.

meta learning | ecosystems forecasting | machine learning | nonlinear dynamics

Significance Statement

In recent years, machine learning has been successfully applied to complex and nonlinear dynamical systems for improved prediction of the future state, but ecological systems represent a great challenge because of the scarcity of the observational data. This work develops a meta-learning framework with time-delayed feed-forward neural networks to predict the long-term behaviors of ecological systems by leveraging synthetic data from paradigmatic nonlinear and non-ecological dynamical systems for effective machine-learning training. The capability of accurately reconstructing the "dynamical climate" of the system with limited data is demonstrated using three benchmark population models and two real-world ecological datasets. The meta-learning framework can be generalized to other fields where forecasting the dynamics is the goal but the available empirical data is limited.

Recent years have witnessed a growing interest in applying machine learning to complex and nonlinear dynamical systems for tasks such as prediction (1–23), control (24), signal detection (25), and estimation (26). For example, a seminal work (5) exploited reservoir computing (27, 28) to accurately predict the state evolution of a spatiotemporal chaotic system for about half dozen Lyapunov times (one Lyapunov time is the time needed for an infinitesimal error to grow by the factor of e) - a remarkable achievement considering the sensitivity of a chaotic system to uncertainties in the initial conditions). Subsequently, long-term prediction of chaotic systems with infrequent state updates was achieved (12), and a parameter-adaptive reservoir computing was developed to predict critical transitions in chaotic systems based on historical data (15, 21).

The demonstrated power of modern machine learning in solving challenging problems in nonlinear dynamics and complex systems naturally suggest applications to ecological systems that are vital to the well being of the humanity. Ecosystems in the modern era are nonautonomous in general due to the human-influences-caused climate change, and it is of paramount interest to be able to predict the future state of the ecosystems. However, to enable applications of machine learning to ecosystems, a fundamental obstacle must be overcome. Specifically, a condition under which the existing machine-learning methods can be applied to complex dynamical systems is the availability of large quantities of data for training. For physical systems accessible to continuous observation and measurements, this data requirement may not pose a significant challenge. However, for ecosystems, the available empirical datasets are often small and large datasets are generally notoriously difficult to obtain (29). A compounding factor is that ecosystems are subject to constant disturbances (30), rendering noisy the available datasets. In recent years, machine learning has been applied to ecosystems (31). For example, support vector machines and random forests were widely used in ecological science for tasks such as classifying invasive plant species, identifying the disease, forecasting the effects of anthropogenic (32, 33), estimating the hidden differential equations (34), and reconstructing the "climate" of the entire system (35). More recently, deep learning was applied in species recognition from video and audio analysis (36, 37). We note that the existing methods of finding equations require sparsity condition, a condition that many ecological dynamical systems do not satisfy, making such methods unsuitable for ecological systems (34, 38).

Author affiliations: [a]School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA; [b]Vehicle Technology Directorate, CCDC Army Research Laboratory, 2800 Powder Mill Road, Adelphi, MD 20783-1138, USA; [c]Vehicle Technology Directorate, CCDC Army Research Laboratory, 6340 Rodman Road, Aberdeen Proving Ground, MD 21005-5069, USA; [d]Department of Environmental Science and Policy, University of California, Davis, CA 95616; [e]Santa Fe Institute, Santa Fe, NM 87501; [f]Department of Physics, Arizona State University, Tempe, Arizona 85287, USA

Z.-M.Z., B.G., M.H., A.H. and Y.-C.L. designed the research project, the models, and methods. Z.-M.Z. performed the computations.Z.-M.Z. and Y.-C.L. analyzed the data. Z.-M.Z. and Y.-C.L. wrote the paper. Y.-C.L. edited the manuscript.

Competing interests statement:The authors declare no competing interest.

[1]To whom correspondence should be addressed. Email: Ying-Cheng.Lai@asu.edu

To overcome the data-shortage difficulty, we exploit meta-learning (39, 40) to predict the long-term dynamics or the attractors of ecosystems. Meta-learning is a learning-to-learn paradigm that enhances the learning algorithm through experience accumulation across multiple episodes. Differing from the conventional machine learning approaches, a well-trained meta-learning framework can adapt to new tasks more swiftly and efficiently by leveraging its prior experience, thus reducing the necessity for extensive retraining and data collection. Owing to its unique features, meta-learning has found broad applications in fields such as computer vision (41), time series forecasting (42, 43), reinforcement learning (44), and identification of special quantum states (45). Our goal is to use meta-learning to reconstruct the "climate" of the target ecosystems, addressing the challenge of data scarcity. Specifically, we take advantage of a number of classical chaotic systems for training a conventional machine-learning architecture to gain "experience" with complex dynamics anticipated to occur in ecosystems, and then update or fine-tune the machine-learning algorithm using the small amount of available data from the actual ecosystem. The outcome is a well-adapted machine-learning framework capable of predicting the complex dynamical behavior of ecosystems with only limited data.

What machine learning architecture is appropriate to combine the meta-learning algorithm for predictive modeling of ecosystems? Recurrent neural networks (24, 46, 47) such as reservoir computing can be a candidate since the prediction requires historical information. For computational efficiency and broad applicability, we choose a foundational architecture, time-delayed feedforward neural networks (FNNs, see SI Appendix, Note 1) (19, 20, 25, 48), a variant of reservoir computing, where the present and historical information of the time series is input into the neural network through time-delayed embedding. With time-delayed FNNs, the meta-learning framework becomes adept at handling the intricate and often nonlinear temporal dependencies typical of ecological data, thereby enabling it to adapt and learn rapidly from new, limited, and noisy data.

In this work, we demonstrate the capability of the meta-learning framework in predicting the long-term behavior of ecological systems with limited data on (1) three prototypical models: the chaotic Hastings-Powell system, a chaotic food chain, and the chaotic Lotka-Volterra system, and (2) two real-world ecological datasets: the microbial time series dataset and global population dynamics database. In all cases, the meta-learning based framework yields more accurate and robust predictions than the model without meta-learning (vanilla model).

Results

The proposed meta-learning framework consists of two distinct phases: adaptation and deployment. In the adaptation phase, a meta-learning neural-network architecture is exposed to a diverse array of synthetic datasets from a number of chaotic systems, allowing it to acquire a broad range of "experiences," as illustrated in Fig. 1(a). This phase is crucial as it equips the neural networks with a versatile learning strategy, nurturing its ability to tackle new and unseen tasks from ecosystems. The variables sampled depend on the dimension of the time series used in the deployment phase. For example, if the target ecological system is three-dimensional, the number of the sampled variables is three. For meta-learning of the empirical datasets during the adaptation phase, we choose the variable dimension to match that of the datasets. In particular, we choose the first dimension of the synthetic data. For meta-learning, the Reptile algorithm, a gradient-based method is implemented, as shown in Fig. 1(b). Figure 1(c) illustrates the deployment phase, in which the well-trained meta-learning scheme is applied to a specific ecosystem of interest. With only limited time series data from the target ecosystem, the scheme adeptly generates accurate long-term predictions of the "climate" of dynamical systems, as well as reliable short-term forecasts for real-world targets. An issue is, as compared with a vanilla machine-learning scheme, defined as one with the same neural network structure but without the adaptation phase with synthetic chaotic data, how much data reduction can be achieved with our meta-learning approach. This issue can be addressed by performing numerical experiments to determine the training duration required to achieve similar performance by meta-learning based framework and the conventional FNN model in reconstructing an ecosystem. Figure 1(d) presents a representative result from the Hastings-Powell ecosystem, where the meta-learning algorithm is able to reduce the length of the training data approximately five times.

The core of meta-learning is the gradient-based Reptile algorithm, as shown in Fig. 1(b). Articulated in (49), it has become a widely used method due to its simplicity and efficiency. In particular, differing from more complex meta-learning algorithms, Reptile requires less memory and computational resources, making it particularly suitable for ecosystem prediction from limited data. The algorithm begins by initializing the parameters. It then iteratively samples tasks, performs gradient descent, and updates the parameters. Let ϕ denote the parameter vector of the machine-learning architecture, s denote a task, and $\mathrm{SGD}(L, \phi, k)$ be the function performing k gradient steps on loss L starting with ϕ and returning the final parameter vector, where SGD stands for stochastic gradient descent. The Reptile algorithm can be described as:

- Initialize ϕ
- For iteration $= 1, 2, \ldots,$ sample tasks $s_1, s_2, \ldots, s_n$
- For $i = 1, 2, \ldots, n$, compute $W_i = \mathrm{SGD}(L_{s_i}, \phi, k)$
- Update $\phi \leftarrow \phi + \epsilon \frac{1}{k} \sum_{i=1}^{n} (W_i - \phi)$
- Continue

A more detailed analysis on Reptile, and comparative analysis of Reptile with other meta-learning methods such as MAML (Model Agnostic Meta-Learning) is presented in SI Appendix, Note 2, and the differences between meta-learning and traditional transfer learning is discussed in SI Appendix, Note 3.

We present forecasting results for three ecosystems models: the three-dimensional chaotic Hastings-Powell system (50), a three-species food chain system (51), and the Lotka-Volterra system (52) with three species. The hypothesis is that the observational data from each system is quite limited (to be quantified below), so it is necessary to invoke meta-learning by first training the neural network using

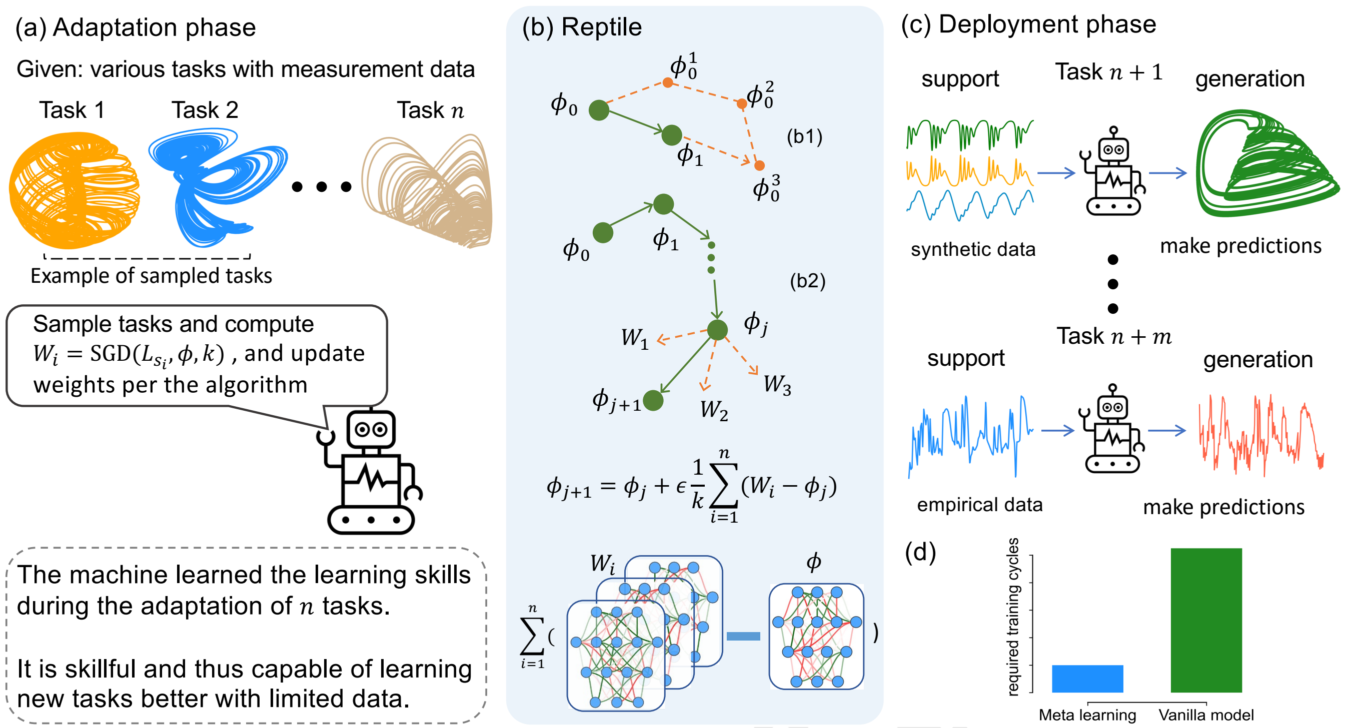

Fig. 1. Illustration of the proposed meta-learning framework for reconstructing ecosystems from limited data. (a) Adaptation phase, where the neural-network architecture is trained on various datasets from synthetic nonlinear chaotic systems so it learns the skill of learning and therefore can better learn the target ecosystem. (b) Illustration of the Reptile algorithm, a gradient-based meta-learning method - see text for details. The dashed arrows denote the intermediate states while the solid arrows indicate the final updating direction. (c) Deployment phase in which the trained meta-learning framework is applied to the target ecosystem, accomplishing the objective of predicting its long-term dynamics or attractor from limited time-series data. (d) An illustration of the comparison of the data requirements for achieving similar performance by the proposed meta-learning framework and standard machine-learning (vanilla model) in reconstructing the Hastings-Powell system.

synthetic time series from the computational models of a number of prototypical chaotic systems. Since the target ecosystems are three-dimensional, the chosen chaotic systems should have the same dimension. We prepare 27 such chaotic systems, as described in SI Appendix, Note 4. More specifically, during the adaptation phase, the neural-network architecture is trained and the values of the hyperparameters are determined with time-series data from the 27 synthetic systems. In the deployment phase, further training with appropriate adjustments to the hyperparameter values is done with the limited data from the target ecosystem, followed by prediction of its long-term dynamics. It is worth noting that the continuous adjustments and fine-tuning of the parameters is the key feature that distinguishes meta-learning from transfer leaning, as further explained in SI Appendix, Note 3. To demonstrate the superiority of meta-learning to conventional machine learning, we train the *same* neural-network architecture but using time series from the ecosystems only without any pre-training - the so-called vanilla or benchmark machine-learning scheme. For the vanilla scheme, typically much larger datasets are required to achieve comparable prediction performance by meta-learning. To validate the efficiency of the proposed framework, we tested it on real ecological systems, many of which exhibit chaos (53). In particular, we first adapt the meta-learning on synthetic chaotic systems and then deploy it on two real-world ecological benchmarks: the microbial time-series dataset and global population dynamics database, taking four time series from each. Since the datasets are one-dimensional, only the first dimension of the chaotic systems is used for adaptation.

To make the presentation succinct, in the main text we focus on the results from the chaotic Hastings-Powell system with a brief mentioning of the summarizing results for the three-species food chain and the chaotic Lotka-Volterra systems and the gut microbiome data. The detailed results from the two synthetic systems are presented in SI Appendix, Note 5. In addition, the detailed results from population database are presented in SI Appendix, Note 7. It is worth noting that the chaotic Hastings-Powell system is a seminal model in population dynamics. It describes the feeding relationships in a food chain from prey to predators (50) and has inspired numerous variations and studies. The three-species food chain system (51) is in fact one variant of the Hastings-Powell system, exhibiting a wide range of behaviors due to the incorporation of additional factors and bioenergetically derived parameters. There were also substantial works based on the original chaotic Hastings-Powell model (54–57), making it a benchmark and prototypical model in theoretical ecology.

Forecasting the chaotic Hastings-Powell system. The chaotic Hastings-Powell system (50) has three dynamical variables, corresponding to the resource, consumer, and predator abundances, respectively. The system is described by the

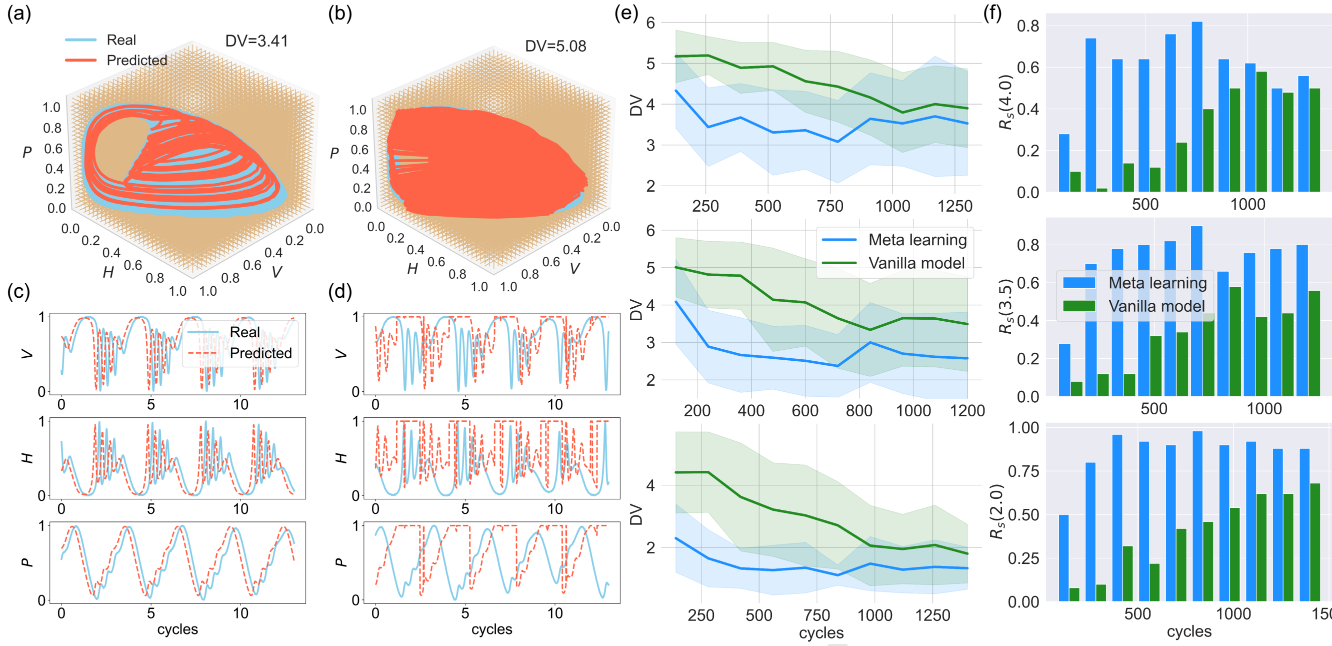

Fig. 2. Long-term ecosystems prediction by the meta-learning and vanilla frameworks. (a,b) Attractor reconstruction by the two frameworks. (c,d) Intercepted snippets of the three time series of the ground truth and prediction by the two frameworks. (e) DV versus the training length for the meta-learning and vanilla frameworks. (f) Stability indicator of prediction ($R_s(\mathrm{DV}_c)$) versus the training length for the meta-learning and vanilla frameworks. The upper, middle, and lower panels in (e) and (f) are from the chaotic Hastings-Powell, food chain, and Lotka-Volterra systems, respectively. To reduce the statistical fluctuations, the DVs, their shaded variabilities and the $R_s(\mathrm{DV}_c)$ values are calculated from an ensemble of 50 independently trained neural machines.

following set of differential equations (58):

$$\begin{aligned} \frac{dV}{dt} &= V(1-V) - \frac{a_1 V H}{b_1 V + 1}, \\ \frac{dH}{dt} &= \frac{a_1 V H}{b_1 V + 1} - \frac{a_2 H P}{b_2 H + 1} - d_1 H, \\ \frac{dP}{dt} &= \frac{a_2 H P}{b_2 H + 1} - d_2 P, \end{aligned} \qquad [1]$$

where V, H, and P are the biomass of the vegetation (or resource), herbivore (or consumer), and predator species, respectively. Alternatively, they can also represent vegetation, host, and parasitoid. The parameters a_1, a_2, b_1, b_2, d_1, and d_2 are chosen to be biologically reasonable (50) as $5, 0.1, 3, 2, 0.4$ and 0.01, respectively, which contain the information about the growth rate, the carrying capacity of the vegetation, etc.

To characterize the performance of long-term prediction of the attractor, we use two measures: deviation value (DV) and prediction stability, where the former describes the distance between the ground truth and predicted attractors and the latter (denoted as $R_s(\mathrm{DV}_c)$) is the probability that meta-learning generates stable dynamical evolution of the target ecosystem in a fixed time window. The definition of the three-dimensional DV here is extended from its two-dimensional version (22). To calculate the DV value, we place a uniform lattice in the three-dimensional phase space with the cell size $\Delta = 0.04$ and count the number of trajectory points in each cell for both the true and predicted attractors in a fixed time interval. The DV is given by

$$\mathrm{DV} \equiv \sum_{i=1}^{m_x} \sum_{j=1}^{m_y} \sum_{k=1}^{m_z} \sqrt{(f_{i,j,k} - \hat{f}_{i,j,k})^2}, \qquad [2]$$

where m_x, m_y, and m_z are the total numbers of cells in the x, y, and z directions, respectively, $f_{i,j,k}$ and $\hat{f}_{i,j,k}$ are the frequencies of visit to the cell (i,j,k) by the true and predicted trajectories, respectively. When the predicted trajectory leaves the square, we count them as if they belonged to the cells at the boundary where the true trajectory never visits. To obtain the prediction stability, we perform the experiment n times and calculate the probability that the DV is below a predefined stable threshold, which is given by

$$R_s(\mathrm{DV}_c) = \frac{1}{n} \sum_{i=1}^{n} [\mathrm{DV} < \mathrm{DV}_c], \qquad [3]$$

where DV_c is the DV threshold, n is the number of iterations and $[\cdot] = 1$ if the statement inside is true and zero otherwise.

Figure 2 presents the comparative forecasting results, where Figs. 2(a) and 2(b) show the ground truth and the predicted attractors in the three-dimensional space by the meta-learning and vanilla frameworks, respectively. It can be seen that the attractor predicted by meta-learning has a lower DV, indicating that the predicted attractor is closer to the ground truth. Figures 2(c) and 2(d) display some representative time-series segments of the predicted and true attractors from the meta-learning and vanilla frameworks, respectively, using the same training data from the ecosystem. Apparently, the vanilla framework fails to predict the attractor correctly. Figures 2(e) and 2(f) show, respectively, the DV and prediction stability values versus the length of the training data from the three ecosystems. The meta-learning framework yields a lower testing DV and higher prediction stability compared to those from the vanilla framework, indicating that meta-learning not only predicts more accurately the long-term dynamics on the attractor but the results are also more stable and reliable. These advantages are particularly pronounced with shorter

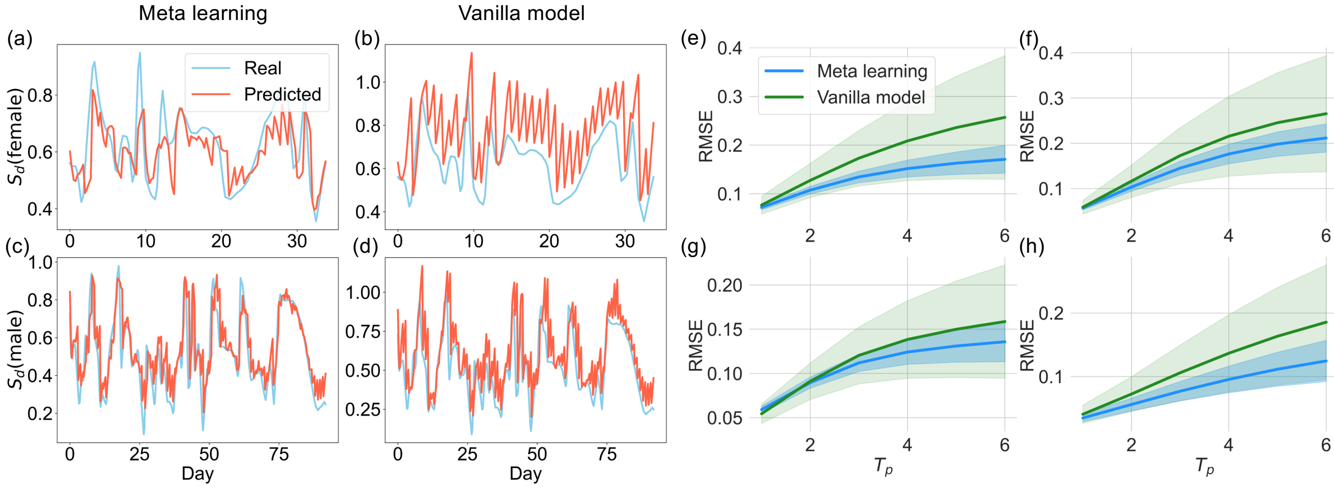

Fig. 3. Short-term gut microbiome prediction by meta-learning and the vanilla framework. Shannon diversity index (S_d) of female and male predictions by meta-learning (a, c) and vanilla model (b, d), respectively. (e-h) Testing RMSEs for the two frameworks, for female, male, donor A, and donor B, respectively. To reduce the statistical fluctuations, the RMSEs and their shaded variabilities are calculated from an ensemble of 50 independently trained machine-learning realizations.

training lengths. In terms of the DV indicator, the vanilla framework requires approximately 5 to 7 times the amount of training data in meta-learning to achieve a similar level of performance. In terms of the prediction stability, the vanilla framework fails to match the performance of meta-learning, regardless of the training length. In addition, by presenting the performance varying with the number of cycles, we can connect the simulated data with real-world experiments. For example, with different training cycles in the real data, the meta-learning framework can make predictions with varying levels of performance and outperforms the vanilla model.

In meta-learning, determining the optimal values of the hyperparameters is key to achieving reliable and accurate prediction results, which is done through standard Bayesian optimization. The procedure and the role of the optimal hyperparameter values are discussed in SI Appendix, Note 8.

Prediction of microbial time series data. To demonstrate the applicability of our proposed meta-learning framework in the real world, we use two empirical datasets: the microbial time series dataset [61] and the global population dynamics database [70]. For conciseness of presentation, we describe the first dataset and the machine-learning prediction results here in the main text, while providing the results from the second dataset in SI Appendix.

The temporal dynamics of the gut microbiome for both individual bacterial species and clusters are essential for understanding human health and disease. We utilize the preprocessed gut microbiome dataset (59), originally derived from two publicly accessible 16S rRNA gene sequencing datasets (60, 61). The dataset contains gut microbiome profiles from four healthy adult participants without any reported diseases. We preserve the name from the original source, i.e., the first dataset uses gender-based labels (female and male subjects), while the second dataset employs alphanumeric names (donor A and donor B). The Shannon diversity (S_d) index is used to characterize the collective dynamics. The numbers of points in the datasets are 185, 443, 365, and 252 for female, male, donor A, and donor B, respectively. Further details about data analysis can be found in the original paper (59).

We focus on short-term predictions for the following reasons. First, the dynamics generating the empirical data can be significantly more complicated than those described by a set of differential equations. For example, for the microbial time series dataset, the underlying dynamical system can be extremely high-dimensional due to external factors such as antibiotic treatments or travel, while the dataset is one-dimensional. Second, real-world ecological datasets often do not contain sufficient points for validating long-term predictions. Low resolution or low sampling density of the empirical datasets are also an issue, as the available data points are too few to describe the underlying dynamics. For example, to faithfully represent a cycle of oscillations, three or four points are not sufficient and can lead to misleading results. This issue of low resolution can be partly addressed by preprocessing the data using linear interpolation, where certain number l_i of points are added between any two original data points. We use $l_i = 3$ for the microbial time series dataset. Let T_p be the prediction horizon, i.e., the forward prediction time step, at each time step. The output of the machine learning model is fed back to the input, forming a closed-loop dynamical system generating T_p-step predictions. After making these predictions, we supply the historical ground truth data and make predictions T_p steps forward again. Iterating this process allows us to evaluate the machine-learning testing performance. For $T_p = 1$, the prediction is one-step, termed as nowcasting (62).

Figure 3 presents the comparative short-term prediction results, where Figs. 3(a,b) and 3(c,d) show representative prediction examples of $T_p = 4$ for the female and male dataset, respectively, from the meta-learning and vanilla models. The examples show that the short-term predictions from meta-learning are more accurate than those from the vanilla model. To quantify the performance of the short-term predictions, we use the root-mean-square error (RMSE) defined as (62)

$$\mathrm{RMSE}(\mathrm{y},\hat{\mathrm{y}}) = \sqrt{\frac{1}{T_p}\sum_{t=i}^{T_p}[y(t)-\hat{y}(t)]^2}, \quad [4]$$

where $y(t)$ and $\hat{y}(t)$ are the true and predicted time series, respectively. The RMSEs can be calculated by taking the average over the whole testing length. Figure 3 (e-h) displays the RMSEs versus the prediction step T_p from the four empirical gut datasets: female, male, donor A, donor B. The meta-learning framework yields lower testing RMSEs compared to those from the vanilla model, indicating that meta-learning is able to generate more accurate short-term prediction results. In addition, the variabilities (represented by the shaded region) by meta-learning are also smaller than those from the vanilla model, suggesting more robust and stable predictions. These advantages are particularly pronounced with longer prediction steps.

Optimal selection of synthetic systems for meta-learning. As described previously, the superiority of the meta-learning framework lies in gathering experience during the adaptation phase. However, indiscriminately utilizing different alternative chaotic systems as adaptation tasks in general does not lead to desired performances, and even worse, can destroy the training performance owing to the diversity of such systems. This raises the question of how to choose the proper adaptation systems for meta learning. To address this question, we employ the greedy algorithm to choose the optimal synthetic chaotic systems and use the chaotic Hastings-Powell system as a testing case for this algorithm. The test is performed, as follows. At each iteration, we perform a testing loop that involves adding a candidate system to the existing pool of chosen systems and monitoring the corresponding decrease in the resulting DV. Afterward, we remove this system and test another candidate system. After looping over all the candidate systems, we select one or several systems that lead to the maximal reduction of the average DV value calculated from 50 independent runs. Once a system has been selected, it will become a member of the chosen system pool in the iterative process that follows.

Figure 4(a) depicts this selection process, where the initial sampled system pool is $n_s = 3$. Looping at Stage one informs us that a new system should be added to the sampled system pool, so at Stage two the pool size becomes $n_s = 4$. Repeating this process iteratively, we collect the ensemble-averaged DV and the corresponding sampled system pool n_s, as shown in Fig. 4(b). We observe that the average DV decreases rapidly as the number of systems increases from one to twenty but begins to increase again when more systems are included. Consequently, guided by the greedy algorithm, we select the five most effective systems: Aizawa, Bouali, Chua, Sprott third, and Sprott fourteen (See SI Appendix, Note 4 for a detailed description of these systems). While certain synthetic systems are selected for the chaotic Hastings-Powell system, we apply them in the adaptation phase for the other two target model systems, and two target real datasets, which yields satisfactory performance as well. It is worth noting that we do not expect the greedy algorithm to produce globally optimal solutions in the space of all possible chaotic systems. It might miss useful systems, each alone would not reduce the DV but their combination would. Considering that this feature selection process is NP-hard, finding some locally optimal solutions is reasonable. This also implies that, while the performance of meta-learning is remarkable, there is ample room for improvement.

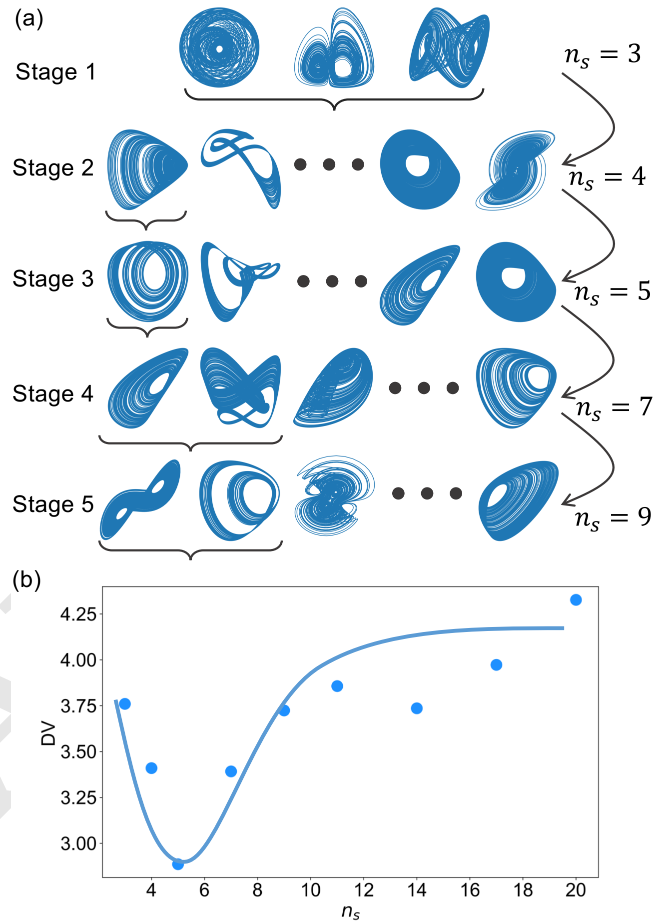

Fig. 4. Selecting the synthetic chaotic systems for the adaptation phase of meta-learning. (a) Illustration of greedy algorithm. Stated with the three systems in the sampled systems pool, one or several systems is (are) selected which lead to the best improvement in performance. (b) Ensemble averaged DV (with 50 independent realizations) versus the number of sampled systems pool n_s. As n_s increases, the average DV decreases rapidly but later increases again.

Discussion

Exploiting machine learning to predict the behaviors of dynamical systems has attracted extensive research in recent years, and it has been demonstrated that modern machine learning can solve challenging problems in complex and nonlinear dynamics that were previously deemed unsolvable. However, machine-learning algorithms often require extensive data for training, and this presents a significant challenge for ecosystems. Indeed, the observational datasets for ecosystems, especially those described by the population dynamics, are often small, preventing a straightforward and direct application of machine learning to these systems.

This work develops an "indirect," meta-learning framework for forecasting the long-term dynamical behaviors of chaotic ecosystems through a faithful reconstruction of the attractor using only limited data. Given a chaotic ecosystem of interest, the idea is to use a large number of alternative chaotic systems of the same dimension, which can be simulated to generate massive training data for a suitable machine-learning scheme such as the time-delayed feedforward neural-network architecture. The neural networks are trained using the synthetic data first, and are then "fine-tuned" with the data from the actual target ecosystem. As a result of the pre-training or first-stage training for adaptation, the neural machine is sufficiently exposed to the climate of the dynamical evolution of characteristically similar systems, which can then be readily adapted to the ecosystem. Specifically, we employed Reptile as the meta-learning algorithm. During the adaptation phase, the algorithm begins by gaining "experience" from learning a synthetic chaotic system. This process continues with data from different non-ecological chaotic systems until the machine is well-trained, experienced, and able to learn new tasks with limited data. In the deployment phase, the neural machine is further trained using the limited data from the target ecosystem - the second-stage training. we emphasize that the first-stage training uses massive data from a large number of model chaotic systems, and the second-stage training is done with limited data from the target ecosystem. After the second-stage training, the neural machine is capable of generating the correct attractor of the ecosystem, realizing accurate and reliable forecasting of its long-term dynamics.

For real ecological systems, due to the limited data, accurate predictions can be achieved but only for a limited number of time steps. While our results from the synthetic datasets demonstrate that meta-learning outperforms the vanilla model on longer predictions, long-term predictions based on the available empirical data are generally not reliable for both empirical datasets. In fact, with the limited data amount, the training of any machine-learning model can be done with at most a few hundred data points. As a result, the neural networks will not be able to fully learn the "dynamical climate" of the target ecological system that is likely to be vastly complex, rendering infeasible any long-term prediction.

One feature of our meta-learning framework is the integration of time-delayed feedforward neural networks for processing sequential data. It incorporates the concept of time delays into the conventional FNN architecture so as to take the advantage of the present and historical information in the time series. It is important to note that, while reservoir computing has demonstrated its capability in chaotic time series prediction and attractor reconstruction (63–66), time-delayed FNN is chosen for our meta-learning algorithms as they are effectively gradient descent-based networks. To our knowledge, so far reservoir computing has not been incorporated into meta-learning. We have tested the meta-learning framework on three benchmark ecosystems. More accurate and robust prediction was achieved by the meta-learning based framework, whereas the vanilla model requires $5-7$ times the training data to achieve a similar performance. Issues such as the effect of noise and the number of synthetic systems used in the adaptation phase of the training were addressed. Since the aim of this work is to facilitate the prediction of real ecological time series with limited data, and there are a variety of open-source datasets available online (59, 67–69), we selected two benchmarks - the microbiome dataset and the global population dynamics database - for validation. For these empirical datasets, meta-learning gave more accurate and stable predictions compared to the vanilla model.

In general, meta-learning is a powerful machine-learning tool for solving prediction and classification problems in situations where the available data amount is small. A recent example is detecting quantum scars in systems with chaotic classical dynamics. In particular, in a closed quantum system in the semiclassical regime where the particle wavelength is much smaller than the system size, a vastly large number of eigenstates are permitted, among which are those whose wavefunctions are not uniformly distributed in the physical space but instead concentrate on some classical periodic orbits of low periods. The emergence of such scarring eigenstates is counterintuitive, as the classical trajectories are uniform due to ergodicity (70, 71). In the field of quantum chaos, traditionally identifying quantum scarring states was done in a "manual" way through a visual check of a large number (e.g., 10^4) of eigenstates (72). This was challenging as the percentage of scarring states is typically small - less than 10% of all the eigenstates. A recent work demonstrated that meta-learning can be powerful for accurately detecting quantum scars in a fully automated and efficient way (45), where a standard large dataset called Omniglot from the field of image classification was used for training in the adaptation phase.

Our meta-learning framework incorporating time-delayed FNNs possesses a high level of sensitivity to the temporal variations in the data, making it potentially feasible for extension beyond ecosystems to challenging problems such as epidemic spread prediction and traffic forecasting, where effective data collection is often a hurdle. Moreover, our framework can potentially be used to improve the prediction performance of spatiotemporal chaotic systems or nodal dynamics of large networks (73). There is also room for enhancing the performance of the framework. For example, in the present work, we employed the greedy algorithm for selecting synthetic chaotic systems for the adaptation phase of the training and implement meta-learning using the Reptile algorithm. Alternative algorithms can be exploited to achieve better performance.

There have been recent works on exploiting reservoir computing for predicting system collapse induced by crisis (15, 21) and tipping (74) as a bifurcation parameter passes through a critical point. A basic requirement is that the machine-learning model learn how the dynamical climate of the target system changes with the bifurcation parameter, which can be accomplished by conducting the training from extensive time series from a number of distinct parameter values. For real-world ecological systems, due to the lack of such datasets, at the present time it is difficult to predict population collapse, bifurcations, or tipping points. To develop meta-learning based model tailored to ecological systems is an interesting but extremely challenging task worth further development.

The meta-learning framework has broad application potential in real-world ecological systems. We have demonstrated that the framework, when applied to synthetic systems, can improve the predictions on real ecological benchmark systems. It may also be useful to meta-learn multiple examples of short time-series from similar ecological systems and then train the neural network to predict new time series from the same ecosystem. This could potentially reveal the underlying dynamics in an effective manner (75). For instance, training on data from certain plankton populations and then predicting unused plankton data could provide better predictions due to similar dynamics.

Materials and Methods

Given a target system, time series are generated numerically by integrating the synthetic system models with the time step $dt = 0.01$. The initial states of both the dynamical process and the neural network are randomly chosen from a uniform distribution. An initial segment of the time series is removed to ensure that the trajectory has reached the attractor. The training and testing data are obtained by sampling the time series at the interval Δ_s. Specifically, for the chaotic Hastings-Powell, food-chain, and Lotka-Volterra systems, we set $\Delta_s = 60dt = 0.6$, 0.5, and 0.2, corresponding to approximately 1/77, 1/83, and 1/71 cycles of oscillation, respectively. The term "cycles" is referred to as the oscillations of the fast-evolving variable. Specifically, we estimated the average number of "cycles" for each system by counting the local minima within a specific range of the fast-evolving variable. The time series data are preprocessed by using min-max normalization so that they lie in the unit interval [0,1]. Considering the omnipresence of noise, we add Gaussian noise of amplitude $\sigma = 0.003$ to the normalized data. The training and predicting lengths of systems are set as 20,000 and 50,000, respectively.

For the time-delayed FNN, the embedding dimension is 1,000, so the dimension of the input vector is 3,000 (three-dimensional systems). The neural network comprises three hidden layers with the respective sizes [1024,512,128], and its output layer has the size of three (for three-dimensional target systems). The batch size is set to be 128. In meta-learning, we specify 20 inner iterations (I_i) and 30 outer iterations (I_o), with the inner and outer learning rates of $\alpha = 10^{-3}$ and $\epsilon = 1$, respectively. We apply Bayesian optimization to systematically determine the optimal hyperparameters and and test the effects of the hyperparameters on the performance. When studying real-world datasets, the machine-learning parameter settings are similar to those used for the synthetic data, such as the amplitude of added Gaussian noise and the chaotic systems used for adaptation. However, due to the fact that the empirical data are one-dimensional and short in length, we adjust two hyperparameters to better suit these data: reduced embedding time from 1000 to 30 and the dimension of the input vector set to be the embedding. Consequently, the neural network size is reduced to [128,64,16] and the batch size is set to 16. In the meta-learning process, the number of outer iterations (I_o) is set to 20. For the real data, the first 70% is used for training and the remaining 30% is reserved for testing to evaluate the performance. The simulations can be run locally without requiring high-performance computational resources. GPU computers are recommended to accelerate the experiments. In our study, simulations are run using Python on two desktop computers, each with 32 CPU cores, 128 GB memory, and one RTX 4000 NVIDIA GPU.

Data, Materials, and Software Availability

The simulated data are available at Zenodo:
https://doi.org/10.5281/zenodo.14261464.
The empirical ecological datasets are from:

https://github.com/bioinf-mcb/dynamo and
https://knb.ecoinformatics.org/view/doi:10.5063/F1BZ63Z8.
The code for reproducing the results presented in this work is available on GitHub: https://github.com/Zheng-Meng/Meta-learning-Ecosystems.

ACKNOWLEDGMENTS. This work was supported by the Air Force Office of Scientific Research under Grant No. FA9550-21-1-0438 and by the Army Research Office under Grant No. W911NF-24-2-0228..

1. ND Haynes, MC Soriano, DP Rosin, I Fischer, DJ Gauthier, Reservoir computing with a single time-delay autonomous Boolean node. *Phys. Rev. E* **91**, 020801 (2015).
2. L Larger, et al., High-speed photonic reservoir computing using a time-delay-based architecture: Million words per second classification. *Phys. Rev. X* **7**, 011015 (2017).
3. J Pathak, Z Lu, B Hunt, M Girvan, E Ott, Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data. *Chaos* **27**, 121102 (2017).
4. Z Lu, et al., Reservoir observers: Model-free inference of unmeasured variables in chaotic systems. *Chaos* **27**, 041102 (2017).
5. J Pathak, B Hunt, M Girvan, Z Lu, E Ott, Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Phys. Rev. Lett.* **120**, 024102 (2018).
6. TL Carroll, Using reservoir computers to distinguish chaotic signals. *Phys. Rev. E* **98**, 052209 (2018).
7. K Nakai, Y Saiki, Machine-learning inference of fluid variables from data using reservoir computing. *Phys. Rev. E* **98**, 023111 (2018).
8. ZS Roland, U Parlitz, Observing spatio-temporal dynamics of excitable media using reservoir computing. *Chaos* **28**, 043118 (2018).
9. A Griffith, A Pomerance, DJ Gauthier, Forecasting chaotic systems with very low connectivity reservoir computers. *Chaos* **29**, 123108 (2019).
10. J Jiang, YC Lai, Model-free prediction of spatiotemporal dynamical systems with recurrent neural networks: Role of network spectral radius. *Phys. Rev. Res.* **1**, 033056 (2019).
11. G Tanaka, et al., Recent advances in physical reservoir computing: A review. *Neural Net.* **115**, 100–123 (2019).
12. H Fan, J Jiang, C Zhang, X Wang, YC Lai, Long-term prediction of chaotic systems with machine learning. *Phys. Rev. Res.* **2**, 012080 (2020).
13. C Klos, YFK Kossio, S Goedeke, A Gilra, RM Memmesheimer, Dynamical learning of dynamics. *Phys. Rev. Lett.* **125**, 088103 (2020).
14. P Chen, R Liu, K Aihara, L Chen, Autoreservoir computing for multistep ahead prediction based on the spatiotemporal information transformation. *Nat. Commun.* **11**, 4568 (2020).
15. LW Kong, HW Fan, C Grebogi, YC Lai, Machine learning prediction of critical transition and system collapse. *Phys. Rev. Res.* **3**, 013090 (2021).
16. D Patel, D Canaday, M Girvan, A Pomerance, E Ott, Using machine learning to predict statistical properties of non-stationary dynamical processes: System climate, regime transitions, and the effect of stochasticity. *Chaos* **31**, 033149 (2021).
17. JZ Kim, Z Lu, E Nozari, GJ Pappas, DS Bassett, Teaching recurrent neural networks to infer global temporal structure from local examples. *Nat. Mach. Intell.* **3**, 316–323 (2021).
18. H Fan, LW Kong, YC Lai, X Wang, Anticipating synchronization with machine learning. *Phys. Rev. Res.* **3**, 023237 (2021).
19. E Bollt, On explaining the surprising success of reservoir computing forecaster of chaos? The universal machine learning dynamical system with contrast to VAR and DMD. *Chaos* **31**, 013108 (2021).
20. DJ Gauthier, E Bollt, A Griffith, WA Barbosa, Next generation reservoir computing. *Nat. Commun.* **12**, 1–8 (2021).
21. LW Kong, Y Weng, B Glaz, M Haile, YC Lai, Reservoir computing as digital twins for nonlinear dynamical systems. *Chaos* **33**, 033111 (2023).
22. ZM Zhai, LW Kong, YC Lai, Emergence of a resonance in machine learning. *Phys. Rev. Res.* **5**, 033127 (2023).
23. M Yan, et al., Emerging opportunities and challenges for the future of reservoir computing. *Nat. Commun.* **15**, 2056 (2024).
24. ZM Zhai, et al., Model-free tracking control of complex dynamical trajectories with machine learning. *Nat. Commun.* **14**, 5698 (2023).
25. ZM Zhai, M Moradi, LW Kong, YC Lai, Detecting weak physical signal from noise: A machine-learning approach with applications to magnetic-anomaly-guided navigation. *Phys. Rev. Appl.* **19**, 034030 (2023).
26. ZM Zhai, M Moradi, B Glaz, M Haile, YC Lai, Machine-learning parameter tracking with partial state observation. *Phys. Rev. Res.* **6**, 013196 (2024).
27. H Jaeger, The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Ger. Ger. Natl. Res. Cent. for Inf. Technol. GMD Tech. Rep.* **148**, 13 (2001).

28. W Mass, T Nachtschlaeger, H Markram, Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neur. Comp.* **14**, 2531–2560 (2002).
29. AR Rissman, S Gillon, Where are ecology and biodiversity in social–ecological systems research? A review of research methods and applied recommendations. *Conserv. Lett.* **10**, 86–93 (2017).
30. SN Wood, Statistical inference for noisy nonlinear ecological dynamic systems. *Nature* **466**, 1102–1104 (2010).
31. S Christin, É Hervet, N Lecomte, Applications for deep learning in ecology. *Methods Ecol. Evol.* **10**, 1632–1644 (2019).
32. JM Drake, C Randin, A Guisan, Modelling ecological niches with support vector machines. *J. Appl. Ecol.* **43**, 424–432 (2006).
33. DR Cutler, et al., Random forests for classification in ecology. *Ecology* **88**, 2783–2792 (2007).
34. K Course, PB Nair, State estimation of a physical system with unknown governing equations. *Nature* **622**, 261–267 (2023).
35. A Wikner, et al., Stabilizing machine learning prediction of dynamics: Novel noise-inspired regularization tested with reservoir computing. *Neural Net.* **170**, 94–110 (2024).
36. D Tuia, et al., Perspectives in machine learning for wildlife conservation. *Nat. Commun.* **13**, 792 (2022).
37. M Pichler, F Hartig, Machine learning and deep learning—A review for ecologists. *Methods Ecol. Evol.* **14**, 994–1016 (2023).
38. YC Lai, Finding nonlinear system equations and complex network structures from data: A sparse optimization approach. *Chaos* **31**, 082101 (2021).
39. C Finn, P Abbeel, S Levine, Model-agnostic meta-learning for fast adaptation of deep networks in *International Conference on Machine Learning.* (PMLR), pp. 1126–1135 (2017).
40. T Hospedales, A Antoniou, P Micaelli, A Storkey, Meta-learning in neural networks: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 5149–5169 (2021).
41. YX Wang, D Ramanan, M Hebert, Meta-learning to detect rare objects in *Proceedings of the IEEE/CVF International Conference on Computer Vision.* pp. 9925–9934 (2019).
42. TS Talagala, RJ Hyndman, G Athanasopoulos, Meta-learning how to forecast time series. *J. Forecast.* **42**, 1476–1501 (2023).
43. BN Oreshkin, D Carpov, N Chapados, Y Bengio, Meta-learning framework with applications to zero-shot time-series forecasting in *Proceedings of the AAAI Conference on Artificial Intelligence.* Vol. 35 (10), pp. 9242–9250 (2021).
44. A Nagabandi, et al., Learning to adapt in dynamic, real-world environments through meta-reinforcement learning. *arXiv preprint arXiv:1803.11347* (2018).
45. CD Han, CZ Wang, YC Lai, Meta-machine-learning-based quantum scar detector. *Phys. Rev. Appl.* **19**, 064042 (2023).
46. M Hermans, B Schrauwen, Memory in linear recurrent neural networks in continuous time. *Neural Net.* **23**, 341–355 (2010).
47. JZ Kim, DS Bassett, A neural machine code and programming framework for the reservoir computer. *Nat. Mach. Intell.* **5**, 622–630 (2023).
48. Y Zhang, SP Cornelius, Catch-22s of reservoir computing. *Phys. Rev. Res.* **5**, 033213 (2023).
49. A Nichol, J Achiam, J Schulman, On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999* (2018).
50. A Hastings, T Powell, Chaos in a three-species food chain. *Ecology* **72**, 896–903 (1991).
51. K McCann, P Yodzis, Nonlinear dynamics and population disappearances. *Am. Nat.* **144**, 873–879 (1994).
52. J Vano, J Wildenberg, M Anderson, J Noel, J Sprott, Chaos in low-dimensional Lotka-Volterra models of competition. *Nonlinearity* **19**, 2391 (2006).
53. TL Rogers, BJ Johnson, SB Munch, Chaos is not rare in natural ecosystems. *Nat. Ecol. Evol.* **6**, 1105–1111 (2022).
54. A Hastings, et al., Transient phenomena in ecology. *Science* **361**, eaat6412 (2018).
55. SR Proulx, DE Promislow, PC Phillips, Network thinking in ecology and evolution. *Trends Ecol. Evol.* **20**, 345–353 (2005).
56. M Kot, *Elements of Mathematical Ecology.* (Cambridge University Press), (2001).
57. J Terborgh, JA Estes, *Trophic Cascades: Predators, Prey, and the Changing Dynamics of Nature.* (Island press), (2013).
58. G Vissio, Weather-like forcing in population mortality rates (2023).
59. Z Karwowska, P Szczerbiak, T Kosciolek, Microbiome time series data reveal predictable patterns of change. *Microbiol. Spectr.* **12**, e04109–23 (2024).
60. JG Caporaso, et al., Moving pictures of the human microbiome. *Genome Biol.* **12**, 1–8 (2011).
61. LA David, et al., Host lifestyle affects human microbiota on daily timescales. *Genome Biol.* **15**, 1–15 (2014).
62. ZM Zhai, M Moradi, S Panahi, ZH Wang, YC Lai, Machine-learning nowcasting of the Atlantic Meridional Overturning Circulation. *APL Mach. Learn.* **2** (2024).
63. A Banerjee, JD Hart, R Roy, E Ott, Machine learning link inference of noisy delay-coupled networks with optoelectronic experimental tests. *Phys. Rev. X.* **11**, 031014 (2021).
64. A Banerjee, S Chandra, E Ott, Network inference from short, noisy, low time-resolution, partial measurements: Application to C. elegans neuronal calcium dynamics. *Proc. Natl. Acad. Sci.* **120**, e2216030120 (2023).
65. J Pathak, Z Lu, BR Hunt, M Girvan, E Ott, Using machine learning to replicate chaotic attractors and calculate lyapunov exponents from data. *Chaos* **27** (2017).
66. J Pathak, B Hunt, M Girvan, Z Lu, E Ott, Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Phys. Rev. Lett.* **120**, 024102 (2018).
67. M Holland, A Louchart, L Artigas, A McQuatters-Gollop, Changes in phytoplankton and zooplankton communities. *OSPAR, 2023: The 2023 Qual. Status Rep. for Northeast. Atl.* (2023).
68. J Prendergast, et al., The global population dynamics database. *Know. Net. Biocomp.* **10** (2010).
69. JM Cocciardi, et al., The value of long-term ecological research for evolutionary insights. *Nat. Ecol. Evol.* **8**, 1584–1592 (2024).
70. SW McDonald, AN Kaufman, Spectrum and eigenfunctions for a Hamiltonian with stochastic trajectories. *Phys. Rev. Lett.* **42**, 1189 (1979).
71. EJ Heller, Bound-state eigenfunctions of classically chaotic Hamiltonian systems: scars of periodic orbits. *Phys. Rev. Lett.* **53**, 1515 (1984).
72. H Xu, L Huang, YC Lai, C Grebogi, Chiral scars in chaotic Dirac fermion systems. *Phys. Rev. Lett.* **110**, 064102 (2013).
73. K Srinivasan, et al., Parallel machine learning for forecasting the dynamics of complex networks. *Phys. Rev. Lett.* **128**, 164101 (2022).
74. S Panahi, et al., Machine learning prediction of tipping in complex dynamical systems. *Phys. Rev. Res.* **6**, 043194 (2024).
75. S Löwe, D Madras, R Zemel, M Welling, Amortized causal discovery: Learning to infer causal graphs from time-series data in *Conf. Causal Learn. Reason.* (PMLR), pp. 509–525 (2022).