Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

Learning dynamics from large biological data sets: Machine learning meets systems biology

William Gilpin[1], Yitong Huang[2] and Daniel B. Forger[1,3,4,5]

Abstract

In the past few decades, mathematical models based on dynamical systems theory have provided new insight into diverse biological systems. In this review, we ask whether the recent success of machine learning techniques for large-scale biological data analysis provides a complementary or competing approach to more traditional modeling approaches. Recent applications of machine learning to the problem of learning biological dynamics in diverse systems range from neuroscience to animal behavior. We compare the underlying mechanisms and limitations of traditional dynamical models with those of machine learning models. We highlight the unique role that traditional modeling has played in providing predictive insights into biological systems, and we propose several avenues for bridging traditional dynamical systems theory with large-scale analysis enabled by machine learning.

Addresses

[1] Quantitative Biology Initiative, Harvard University, USA
[2] Department of Mathematics, Dartmouth College, USA
[3] Department of Mathematics, University of Michigan, USA
[4] Department of Computational Medicine and Bioinformatics, University of Michigan, USA
[5] Michigan Institute for Data Science, University of Michigan, USA

Corresponding author: Forger, Daniel B (forger@umich.edu)

Current Opinion in Systems Biology 2020, **22**:1–7

This review comes from a themed issue on **Mathematical Modelling (2020)**

Edited by **Daniel Forger** and **Olivia Walch**

For complete overview of the section, please refer the article collection - Mathematical Modelling (2020)

Available online 30 July 2020

https://doi.org/10.1016/j.coisb.2020.07.009

Keywords

Machine learning, Systems biology, Neural networks, Unsupervised learning.

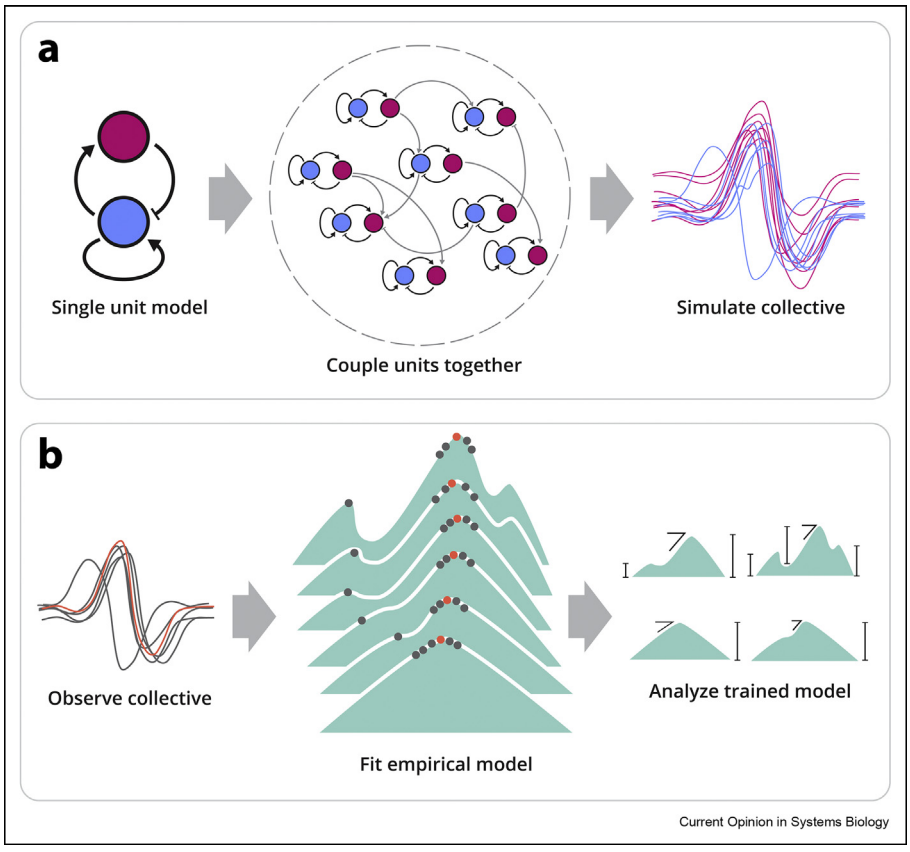
Introduction: the limitations of large mathematical models

The most informative biological models unify observations across scales. For example, the celebrated Hodgkin–Huxley model of neural excitation comprises a set of differential equations describing the coupled dynamics of current, potential, and polarization within a single neuron's lipid bilayer [38]. Hodgkin and Huxley derived these equations from a first-principles analysis that combined the theory of electrical circuits with detailed anatomical study and measurements of squid giant axons. In their model, detailed experiments determined the form and parameters describing a single neuronal unit, and when many such units are combined together, their collective dynamics display the same emergent properties as large-scale neural ensembles (Figure 1). But while the Hodgkin–Huxley model was developed and validated in this specific experimental system, its legacy stems from its relevance to diverse neurobiological problems — ranging from the beating of cardiac myocytes, to alpha wave propagation in the human thalamus [15,21]. Similar 'bottom-up' models recur throughout systems biology — for example, the Lotka–Volterra model of predator–prey competition may be scaled up to simulate entire multi-trophic food webs [17,33]; and the Goodwin model of negative feedback in an enzymatic network may be extended to describe large gene regulatory networks [19]. In these cases, the structure of the model (equations) was determined directly from experimental observations.

However, the bottom-up approach has a significant limitation: while the individual terms in each dynamical equation are usually well-justified and mechanistic, precise measurements are often unavailable for every parameter — such as rate constants and saturation values describing interactions. Modelers instead resort to imposing strong assumptions on the parameters defining the model, such as by choosing parameters that cause the model to exhibit the correct large-scale collective behavior as observed in data. In rare circumstances, mathematical analysis can be used to show that the structure of the model yields properties that can be tested experimentally. In general, the accuracy of these parameter choices often remains untested.

Recent advances in machine learning challenge the classical 'bottom-up' approach to biological modeling. In the language of modern statistical learning, traditional modeling consists of model selection (defining an equation via mechanistic studies), training (fitting rate constants and other parameters), and validation (prediction and confirmation of biologically realistic dynamics). In machine learning, however, large-scale observations often serve as a starting point for analysis — all of the complexities and peculiarities of the observed data are

Figure 1



Bottom-up versus top-down modeling. **(a)** In bottom-up models, experimental data are used to derive a model of single units (e.g. chemicals, neurons, and species). Many units are coupled together, and the resulting collective dynamics are analyzed. **(b)** In top-down models, experimental measurements of collective dynamics are measured directly and then fit to an empirical model (here, a time-evolving probability distribution). The parameters and properties of the best-fit model (here, properties of the distribution) may then be analyzed and compared across timepoints or replicates.

visible to the model during training. Recently, 'supervised' learning has generated significant interest among biologists due to remarkable advances in automating laboratory tasks requiring manual annotation of data — such as segmentation of cell images or classification of mutations. From the perspective of understanding biological dynamics, however, we are particularly interested in *unsupervised* learning, which seeks to identify structure and patterns indirectly using unlabeled raw experimental data (e.g. motif discovery).

Although many systems biology models blend aspects of bottom-up and top-down modeling, we believe that this dichotomy illustrates the potential power of modern statistical and machine learning to generate fundamentally new insights into systems biology. We argue that these techniques have the potential to work in tandem with traditional mathematical modeling techniques, revealing emergent simplicity in seemingly complex biological data sets.

Notions of time

Most classical models in systems biology are typically expressed in terms of differential equations of time. Given the initial state and input parameter values of a system, these equations tell us how the state of the system evolves over time. Machine learning models do not have an inherent concept of time; therefore, an auxiliary mapping between the model and time is required for proper comparison between these approaches. There are several ways to account for this:

1) One can run classical models until they reach an equilibrium state or end point and then compare their input (initial conditions) and output (final

state) to those achieved by machine learning approaches.
2) Some machine learning techniques take the full-time history of a system as an input for learning. These can be used to identify and classify different dynamical regimes within the time series, revealing different persistent behaviors (such as attractors).
3) Recurrent neural networks are a type of artificial neural network that explicitly include memory via an internal state, making them well-suited to modeling sequential data such as time series.
4) The individual layers of a deep neural network can be considered as timepoints, in which case training the network via backpropagation of gradients across layers becomes equivalent to the classical adjoint method for optimizing the parameters of an ordinary differential equation — making it possible to define networks with a continuous layer index [16]. Furthering this analogy, Pontryagin's maximum principle (a concept from control theory) can be used to improve training of strongly constrained models [3,5].

Many linear mathematical models, or nonlinear mathematical models with small magnitude inputs, obey the superposition property: the effect of an input at one time is independent of the effect of an input at another time. Such dynamics can readily be incorporated into a machine learning model by introducing an appropriate regularization, which is a constraint that guides learning and deters overfitting. However, most mathematical models of biological systems do not have this property, and this distinction plays an important role in information processing in biological systems. For example, consider the processing of inputs by a single biological neuron. For inputs that balance excitation and inhibition, the most effective signal to generate firing is often a growing sinusoid with a period matched to the resonant period of the neuron [19]. However, for primarily inhibitory inputs, the most effective stimulus is a postinhibitory rebound, or a sharp rise from inhibition. Moreover, a single neuron can also show different stable steady states (bistability) depending on its prior inputs. For example, it can exhibit quiescence or repetitive firing, and these states can persist until some other signal is strong enough to perturb the system state. Incorporating these properties into machine learning models requires explicit constraints to be imposed during training, whereas differential equation models naturally and concisely capture these properties.

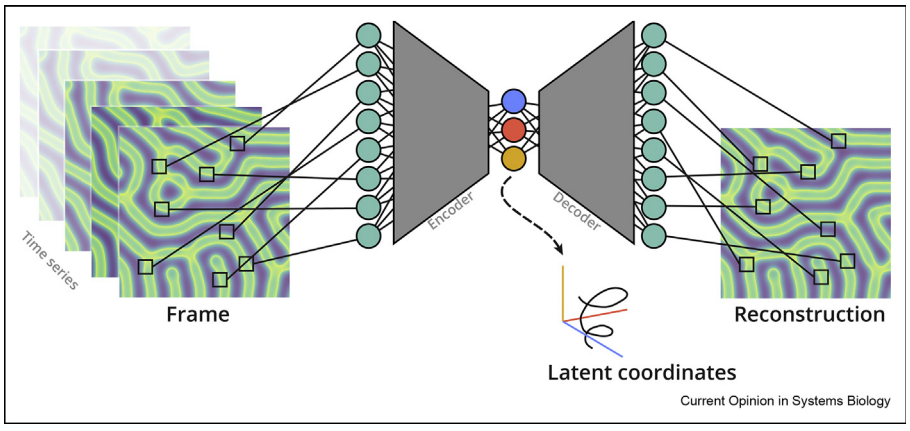
Reconstructing governing coordinates from experimental data

Various techniques based on unsupervised dimensionality reduction and manifold learning offer a promising application of machine learning to the problem of learning dynamics from experimental data. For example, researchers frequently record the simultaneous activity of thousands of neurons, under the implicit hypothesis that the dynamical variables necessary to model cognition may be reconstructed from these measurements alone [46,35]. These methods extract independent, low-dimensional governing coordinates from high-dimensional measurements using heuristics such as principal component analysis, which minimizes the covariance among extracted coordinates, or more sophisticated variants based on isolating timescales or recurring dynamical motifs [9]. Unsupervised learning of dynamics has broad application to not only neural recordings [35] but also to microbial ecology (Tikhonov 2017 [51]), organismal locomotion [23], and tracking animal behavior [23, 27, 30]. When large amounts of data are available, autoencoders (a type of neural network) are a particularly flexible technique for unsupervised learning (Figure 2). These networks seek to learn an identity function that maps the data onto itself, subject to constraints on how the data is stored and represented within the network. Autoencoders typically compress high-dimensional data into a low-dimensional 'latent' space, and additional constraints (e.g. orthogonality, sparsity, continuity) may be introduced in order to influence the representation of the data within the latent space.

Similar techniques exist for classical biological models. Even when models contain hundreds or thousands of variables, a system may often be accurately reduced to a low-dimensional system with variables that are nonlinear combinations of the original system variables. An example of this arises in mathematical models of biological clocks, which contain many variables, but which can often be reduced to a subsystem (manifold) consisting of two variables [2]. However, such low-dimensional characterizations of the original system have variables and parameters that can be very complicated combinations of the original system variables. If one seeks biologically testable hypotheses about these models (e.g. if protein X increases in concentration, will protein Y increase as well?), the easiest path may be to simulate the full original system, rather than to try to determine this underlying smaller system.

Unsupervised learning can also be applied to the inverse problem of dimensionality augmentation, for cases in which the number of independent measurement channels is lower than that of a system's underlying dynamics. For example, in an ecosystem comprising an interacting predator and prey, it may be possible to trap and measure only the predator's dynamics — introducing the question of whether the dynamics of the prey can be inferred from measurements of the predator alone [44]. Similar 'hidden variables' occur in molecular biology, in which time-resolved measurements of a handful of fluorescent reporters are used to infer the

Figure 2



A schematic of an autoencoder, a type of neural network with two halves: an 'encoder,' which maps high-dimensional input data to a low-dimensional 'latent' space, and a 'decoder,' which maps the low-dimensional latent representation back to a high-dimensional reconstruction of the original input. The two halves are connected and trained together, producing a network that effectively maps the input data onto itself, subject to constraints on the dimensionality of the data's representation within the latent space. After training, the decoder portion of the network may be set aside, and the encoder portion may be used to map additional input data onto lower-dimensional latent variables.

properties of a many-component gene regulatory network. In medical studies involving wearable sensors, low-dimensional time series measurements of skin conduction or accelerometer recordings are used to infer a subject's high-dimensional behavioral or physiological state [48,49]. Classical techniques for learning additional measurements — collectively known as 'empirical dynamic modeling" or 'state space reconstruction' — seek to use time series analysis to reconstruct the underlying state space or attractor of a system, allowing insights into laboratory or ecological systems for which observations are sparse [34,44]. Recently, deep neural networks have been applied to this problem [32], allowing the time series' full history to be mined in order to produce robust and consistent proxies for unobserved dynamical variables.

Merging machine learning with dynamical modeling

Recent studies have sought to combine large-scale analysis facilitated by unsupervised learning with the interpretability of low-dimensional dynamical systems models. For example, in a recent study of freely behaving mice, an autoencoder was trained on 128 x 128 pixel videos of a mouse undergoing a range of free behaviors [7]. This autoencoder extracted ~10 latent variables that were sufficient to describe the majority of the mouse's observed behavioral repertoire, and these latent variables were then used as inputs for a model of the mouse's behavior as a stochastic dynamical system. Another recent study builds upon previously developed mathematical models of human circadian timekeeping and combines this analysis with machine learning to infer sleep stages based on recordings from wearable devices [48].

Symbolic regression seeks to fit differential equations directly to observed data by drawing upon a dictionary of candidate basis functions (such as polynomials or trigonometric functions) [40,39,6,1]. This technique has been used to analyze problems ranging from interactions in biochemical networks [42] to the outbreak dynamics of infectious diseases [37] and the heat stimulus response of *Caenorhabditis elegans* [43]. The choice and size of this function dictionary acts as a prior for the problem, with most methods seeking a balance between accuracy and parsimony. However, models found by symbolic regression are most revealing when the functions comprising the dictionary library have a mechanistic interpretation for the system under study. For example, in systems governed primarily by mechanical and geometric constraints (such as biomechanical systems), a best-fit differential equation model expressed in terms of trigonometric functions can reveal subtle mechanical symmetries and constraints [40]. Likewise, using a dictionary of Hill functions in a symbolic regression model of a large gene regulatory network reveals the presence of specific governing interaction pathways and their rate laws [39]. However, for data sets such as neural recordings, a readily interpretable dictionary of functions describing the problem may not necessarily exist — limiting the applicability of symbolic regression.

For most biological systems, explicit but parsimonious enumeration of the dynamics in terms of known basis functions is likely impossible or computationally infeasible. These systems offer a promising platform for recent efforts to construct numerical approximations of time evolution operators, which computationally map an observation at one timepoint to its values at later timepoints [4,13,24,35]. These approaches have proven particularly applicable to structural biology problems such as protein folding, where they allow informative coarse-grained models to be algorithmically extracted from detailed molecular dynamics simulations [31,12]. These techniques are motivated by the general theoretical properties of nonlinear dynamical systems, which are often difficult to analyze using traditional mathematical tools.

After fitting, numerical time evolution operators represent an empirical analogue of a differential equation conditioned on the training data. For example, a recent study used dynamic mode decomposition (a method originally developed for fluid dynamics) to identify transient resting states in a large data set comprising fMRI readings of the human brain [41]. Unlike principal component analysis or other dimensionality reduction techniques, these methods associate dynamical motifs with scales in both space and time, providing more insight into spatiotemporal patterns than traditional decomposition techniques such as Fourier transforms [22]. Other recent applications of this technique have been for behavior and gait quantification [28] and for optimization of reporter placements for cell-state measurements [47]. While operator methods have not yet become widespread within the synthetic and systems biology communities, we anticipate that their strong theoretical underpinnings and efficient numerical implementations will facilitate their broad adoption.

Extrapolating predictions from limited data

A lingering question when applying unsupervised learning to biological data is whether the appearance of low-dimensional effective coordinates (and a dynamical model defined on those coordinates) reveals any new information about the system itself. Traditional first-principle models posit a hypothesis in their formulation: the choice of terms in the governing equation represents a falsifiable theory of a system's underlying mechanism. But when low-dimensional coordinates are discovered by unsupervised learning, the interpolative nature of the resulting model represents a re-packaging of the existing data, consistent with the constraints on the particular learning algorithm.

A key question in machine learning is whether the training data are sufficient to construct an accurate representation of the system for inputs that strongly differ from those encountered during training. In control theory, this is called the identifiability of a mathematical model. Only in select cases can this question be fully addressed. In a recent development, certain types of mathematical models can be shown to exhibit 'completely determinable dynamics,' meaning that one recording of the system suffices to determine the behavior of a system for all other cases over a range of system states [18,19]. Thus, the behavior of the system over a finite time (a one-dimensional observation) determines the behavior over a range of system states (having the same dimension as the original system). However, not all systems have this property — for example, chaotic systems exhibit very different behavior for small changes in their state.

In dynamical systems theory, differential equation models are often analyzed by determining their equilibrium points and then performing stability analysis of the model at these fixed points. In machine learning, a similar technique is saliency mapping, which calculates the approximate derivative of a trained model's outputs with respect to its inputs [9]. This process identifies the input features that most strongly determine the model's predictions. For example, in a recent study, a neural network was trained to predict the crawling trajectories of developing muscle cells (Kimmel et al., 2019 [50]). Saliency maps were calculated for the trained model with respect to the input time series, and the maps detected unusual bouts of locomotion that distinguished different cell types — thereby identifying qualitative traits that microscopists could search for when observing cells. Another common technique for analyzing differential equation models is sensitivity analysis, which seeks to determine how a model's outputs change as the parameters describing individual terms in the model change. For a machine learning model, a comparable technique is ablation, in which single components are removed from a trained model (such as architectural features of a neural network) in order to identify which terms in the model most strongly influence its accuracy [29].

Conclusion and outlook

We have shown diverse examples demonstrating the ability of machine learning models to distill insights into large-scale biological dynamics. Does the ability of unsupervised learning to identify low-dimensional coordinates in diverse biological systems suggest deeper biological principles?

A counterintuitive result of the ever-larger-scale deployment of artificial neural networks is the observation that many aspects of large models are surprisingly consistent across different implementations and applications [45]. Despite having thousands, or even millions, of trainable parameters, artificial neural networks often converge to similar solutions across replicates, to the degree that their training can sometimes be approximated by low-order ordinary differential equations [10,11,36]. Such results

mirror the phenomenology of large-scale biological regulatory networks, which, despite comprising many interacting genes and proteins, often have a small set of effective governing parameters — such as the kinetic constants of a few rate-limiting reactions — that jointly determine the system's large-scale dynamics [8]. Similar results are observed in large ecological communities, for which isolated study of individual species — and thus validated models of individual interactions — is difficult to achieve [20]. In neuroscience, many theoretical and experimental studies demonstrate cases in which thousands of interacting neurons collectively produce low-dimensional dynamics [15,19,14]. For example, by applying an unsupervised learning algorithm based on topological filtration, the spiking activity of thousands of neurons in a mouse thalamus can be robustly mapped onto a low-dimensional 'attractor manifold' encoding the angular bearing of the animal's head (Chaudhuri et al., 2019 [52]). Intriguingly, direct comparison of neural activity in the visual system to activation patterns in trained convolutional neural networks (the current state-of-the-art technique for image analysis) suggest that artificial and biological neural networks may even share common hierarchical representation and organization of image information [26,25].

Taken together, these results introduce the question of whether universal mathematical constraints determine certain aspects of large biological systems — that large, overparameterized systems of weakly interacting units spontaneously collapse onto a low-dimensional manifold. Has evolution driven complex biological networks toward these emergent motifs, and do they confer adaptive benefits from the perspective of control or stability — such as stabilizing an ecosystem against an invasive predator, or suppressing unwanted fluctuations in a large genetic circuit? We hope that further development of models at the intersection of machine learning and dynamical systems theory will provide unified insight into this question.

Conflict of interest statement

DBF is the CSO of Arcascope, which did not sponsor this work.

Acknowledgements

The authors thank the NSF-Simons Center for Mathematical and Statistical Analysis of Biology at Harvard University, NSF Grant No. DMS-1764269, NSF Grant No. DMS-1714094 and the Harvard FAS Quantitative Biology Initiative.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Udrescu Silviu-Marian, Tegmark Max: **AI Feynman: a physics-inspired method for symbolic regression**. *Science Advances* 2020, **6**. **16**:eaay2631.

2. Forger Daniel B, Kronauer Richard E: **Reconciling mathematical models of biological clocks by averaging on approximate manifolds**. *SIAM J Appl Math* 2002, **62**:1281–1296.

3. Zhang Dinghuai, *et al.*: **You only propagate once: accelerating adversarial training via maximal principle**. *Adv Neural Inf Process Syst* 2019:227–238.

4. Lusch Bethany, Kutz JNathan, Brunton Steven L: **Deep learning for universal linear embeddings of nonlinear dynamics**. *Nat Commun* 2018, **9**:1–10.

5. Weinan E: **A proposal on machine learning via dynamical systems**. *Communications in Mathematics and Statistics* 2017, **5**: 1–11.

6. ** Champion Kathleen, *et al.*: **Data-driven discovery of coordinates and governing equations**. *Proc Natl Acad Sci Unit States Am* 2019, **116**:22445–22451.
Demonstrates a systematic machine learning technique for extracting governing coordinates from spatiotemporal datasets, and then fitting the resulting coordinates to differential equations using sparse symbolic regression.

7. Batty Eleanor, *et al.*: **BehaveNet: nonlinear embedding and Bayesian neural decoding of behavioral videos**. *Adv Neural Inf Process Syst* 2019.

8. Gutenkunst Ryan N, *et al.*: **Universally sloppy parameter sensitivities in systems biology models**. *PLoS Comput Biol* 2007, **3**:10.

9. Hastie Trevor, Tibshirani Robert, Friedman Jerome: *The elements of statistical learning: data mining, inference, and prediction.* Springer Science & Business Media; 2009.

10. Jacot Arthur, Gabriel Franck, Hongler Clément: **Neural tangent kernel: convergence and generalization in neural networks**. *Adv Neural Inf Process Syst* 2018.

11. Bordelon Blake, Canatar Abdulkadir, Pehlevan Cengiz: **Spectrum dependent learning curves in kernel regression and wide neural networks**. In *International conference on machine learning*; 2020.

12. Ramsundar Bharath, Eastman Peter, Walters Patrick, Pande Vijay: *Deep learning for the life sciences: applying deep learning to genomics, microscopy, drug discovery, and more.* O'Reilly Media, Inc.; 2019.

13. Gilpin William: **Cellular automata as convolutional neural networks**. *Phys Rev* 2019, **100**, 032402.

14. Bick Christian, *et al.*: **Understanding the dynamics of biological and neural oscillator networks through mean-field reductions: a review**. *J Math Neurosci* 2020.

15. Lynn Christopher W, Bassett Danielle S: **The physics of brain network structure, function and control**. *Nature Reviews Physics* 2019, **1**:318.

16. * Chen Tian Qi, *et al.*: **Neural ordinary differential equations**. *Adv Neural Inf Process Syst* 2018.
Identifies an equivalence between backpropagation of errors across layers in deep neural networks, and the adjoint method used for optimization of parameters in ODEs and PDEs.

17. Lafferty Kevin D, *et al.*: **A general consumer-resource population model**. *Science* 2015, **349**:854–857.

18. Kim Jae K, Forger Daniel B: **On the existence and uniqueness of biological clock models matching experimental data**. *SIAM J Appl Math* 2012, **72**:1842–1855.

19. * Forger Daniel B: *Biological clocks, rhythms, and oscillations: the theory of biological timekeeping.* MIT Press; 2017.
An in-depth study of modeling and analysis of biological rhythms.

20. Gross Thilo, *et al.*: **Generalized models reveal stabilizing factors in food webs**. *Science* 2009, **325**:747–750.

21. Breakspear Michael: **Dynamic models of large-scale brain activity**. *Nat Neurosci* 2017, **20**:340.

22. Mezić Igor: **Analysis of fluid flows via spectral properties of the Koopman operator**. *Annu Rev Fluid Mech* 2013, **45**: 357–378.

23. Brown André EX, De Bivort Benjamin: **Ethology as a physical science**. *Nat Phys* 2018, **14**:653–657.

24. Pathak Jaideep, *et al.*: **Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir computing approach**. *Phys Rev Lett* 2018, **120**, 024102.

25. ** Bashivan Pouya, Kar Kohitij, DiCarlo James J: **Neural population control via deep image synthesis**. *Science* 2019, **364**: 6439. eaav9436.
The authors design an artificial neural network to mimic the visual system, and then use it to generate images that selectively activate different neuron populations in the macaque visual cortex. Using various metrics, the authors found that the artificial neural network reproduced several key aspects of the animals' neural activity patterns.

26. Jaegle Andrew, *et al.*: **Population response magnitude variation in inferotemporal cortex predicts image memorability**. *eLife* 2019, **8**.

27. Mathis Alexander, *et al.*: **DeepLabCut: markerless pose estimation of user-defined body parts with deep learning**. *Nat Neurosci* 2018, **21**:1281.

28. Fujii Keisuke, *et al.*: **Data-driven spectral analysis for coordinative structures in periodic human locomotion**. *Sci Rep* 2019, **9**:1–14.

29. Goodfellow Ian, Bengio Yoshua, Courville Aaron: *Deep learning*. MIT press; 2016.

30. Berman Gordon J: **Measuring behavior across scales**. *BMC Biol* 2018, **16**:23.

31. Mardt Andreas, *et al.*: **VAMPnets for deep learning of molecular kinetics**. *Nat Commun* 2018, **9**:1–11.

32. Gilpin William: *Deep reconstruction of strange attractors from time series*. arXiv preprint arXiv:2002.05909. 2020.

33. Grilli Jacopo, *et al.*: **Higher-order interactions stabilize dynamics in competitive network models**. *Nature* 2017, **548**: 210–213.

34. Clark TJ, Luis Angela D: **Nonlinear population dynamics are ubiquitous in animals**. *Nature Ecology & Evolution* 2020, **4**:75–81.

35. Pandarinath Chethan, *et al.*: **Inferring single-trial neural population dynamics using sequential auto-encoders**. *Nat Methods* 2018, **15**:805–815.

36. Lee Jaehoon, *et al.*: **Wide neural networks of any depth evolve as linear models under gradient descent**. *Adv Neural Inf Process Syst* 2019.

37. Horrocks Jonathan, Bauch Chris T: **Algorithmic discovery of dynamic models from infectious disease data**. *Sci Rep* 2020, **10**:1–18.

38. Hodgkin Alan L, Huxley Andrew F: **A quantitative description of membrane current and its application to conduction and excitation in nerve**. *J Physiol* 1952, **117**:500–544.

39. Daniels Bryan C, Ilya Nemenman: **Automated adaptive inference of phenomenological dynamical models**. *Nat Commun* 2015, **6**:1–8.

40. Schmidt Michael, Lipson Hod: **Distilling free-form natural laws from experimental data**. *Science* 2009, **324**:81–85.

41. Kunert-Graf James Michael, *et al.*: **Extracting reproducible time-resolved resting state networks using dynamic mode decomposition**. *Front Comput Neurosci* 2019, **13**:75.

42. Casadiego Jose, *et al.*: **Model-free inference of direct network interactions from nonlinear collective dynamics**. *Nat Commun* 2017, **8**:1–10.

43. Daniels Bryan C, Ryu William S, Nemenman Ilya: **Automated, predictive, and interpretable inference of Caenorhabditis elegans escape dynamics**. *Proc Natl Acad Sci Unit States Am* 2019, **116**:7226–7231.

44. Sugihara George, *et al.*: **Detecting causality in complex ecosystems**. *Science* 2012, **338**:496–500.

45. Bahri Yasaman, *et al.*: **Statistical mechanics of deep learning**. *Annual Review Condensed Matter Physics* 2020.

46. Marques João C, *et al.*: **Internal state dynamics shape brainwide activity and foraging behaviour**. *Nature* 2020, **577**: 239–243.

47. Hasnain Aqib, Boddupalli Nibodh, Yeung Enoch: *Optimal reporter placement in sparsely measured genetic networks using the Koopman operator*. arXiv preprint arXiv:1906.00944. 2019.

48. Walch Olivia J, Huang Yitong, Forger Daniel B, Goldstein Cathy: **Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device**. *Sleep* 2019, **42**:zsz180.

49. Huang Yitong, *et al.*: *Predicting circadian phase across populations: a comparison of mathematical models and wearable devices*. 2020. Under review.

50. Kimmel Jacob, Andrew Brack, Wallace Marshall: **Deep convolutional and recurrent neural networks for cell motility discrimination and prediction**. In *IEEE/ACM transactions on computational biology and bioinformatics*; 2019.

51. Tikhonov Mikhail: **Theoretical microbial ecology without species**. *Phys Rev E* 2017, **96.3**:032410.

52. Chaudhuri Rishidev, *et al.*: **The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep**. *Nat Neurosci* 2019, **22.9**:1512–1520.