REPORT

OPEN ACCESS

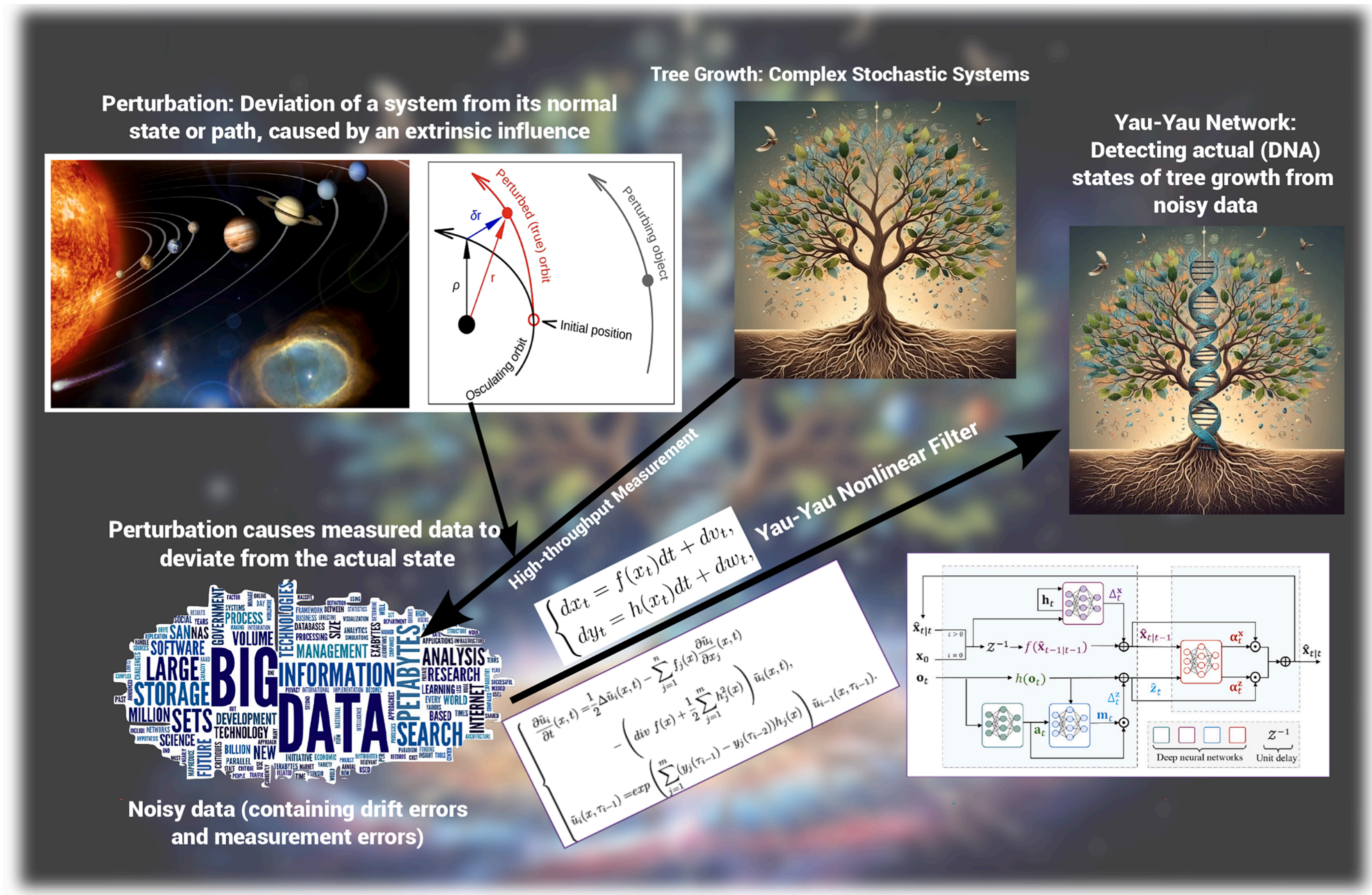
Statistical learning of stochastic complex systems via the Yau-Yau nonlinear filter

Shuyuan Xu,[1,2,3] Yu Wang,[1] Shuang Wu,[1] Yu Wang,[1] Ang Dong,[1] Stephen Shing-Toung Yau,[1] Shing-Tung Yau,[1,4] and Rongling Wu[1,4,5,*]

*Correspondence: ronglingwu@bimsa.cn

Received: March 22, 2025; Accepted: January 10, 2026; https://doi.org/10.1016/j.xinn.2026.101267

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- Complex stochastic systems are characterized by many interacting components that evolve randomly over time.
- We develop a unifying framework of Yau-Yau nonlinear filter and idopNetworks to infer Yau-Yau networks.
- Nodes and edges of Yau-Yau networks stochastically perturb around and across time.
- Yau-Yau networks reveal at which level of perturbation the system transmits from order to disorder.

Statistical learning of stochastic complex systems via the Yau-Yau nonlinear filter

Shuyuan Xu,[1,2,3] Yu Wang,[1] Shuang Wu,[1] Yu Wang,[1] Ang Dong,[1] Stephen Shing-Toung Yau,[1] Shing-Tung Yau,[1,4] and Rongling Wu[1,4,5,*]

[1]Beijing Key Laboratory of Topological Statistics and Applications for Complex Systems, Beijing Institute of Mathematical Sciences and Applications, Beijing 101408, China

[2]State Key Laboratory of Mathematical Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

[3]School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

[4]Yau Mathematical Sciences Center, Tsinghua University, Beijing 100084, China

[5]Shanghai Institute for Mathematics and Interdisciplinary Sciences, Shanghai 200433, China

*Correspondence: ronglingwu@bimsa.cn

Received: March 22, 2025; Accepted: January 10, 2026; https://doi.org/10.1016/j.xinn.2026.101267

Citation: Xu S., Wang Y., Wu S., et al., (2026). Statistical learning of stochastic complex systems via the Yau-Yau nonlinear filter. The Innovation **7(6)**, 101267.

Complex systems are characterized by nonlinearity, uncertainty, and noise interference, making it challenging to unravel their internal operation mechanisms. Here, we overcome this issue by implementing the Yau-Yau nonlinear filter theory to reconstruct stochastic networks from noisy data. Beyond ordinary differential equation-based deterministic networks, Yau-Yau stochastic networks can estimate the real states of actions and interactions among agents within dynamical systems and characterize how these states fluctuate due to stochastic perturbations. We perform Monte Carlo simulation to validate the statistical behavior of the Yau-Yau algorithm under different application scenarios. We employ the new approach to learn microbial stochastic interaction networks from a tri-cultural data of bacterial species, gaining new insight into how microbial interactions facilitate information processing and leverage robustness against perturbations. Yau-Yau networks open up a new avenue to reveal the emergence of order from disorder for complex systems.

INTRODUCTION

Many physical, natural, and social phenomena, ranging from the global climate change to cells, the human brain, organisms, microbial communities, ecosystems, social and economic organizations, and even the universe, may emerge as complex systems composed of many interacting parts.[1–6] Given their central relevance to the understanding of nature and human society, complex systems have increasingly become a focus of interest and research across a wide range of scientific areas.[6–12] Numerous mathematical, statistical, and computational approaches have been developed for disentangling complex systems, among which network modeling has been recognized as a mainstream means given its capacity to reveal hidden mechanisms.[13–21] Specifically, it can not only characterize intricate patterns of interactions but also reveals essential emergent attributes of complex systems to answer why the whole is greater than the sum of its individual parts.[22,23]

Complex networks model and describe the internal components and their relationship characteristics in complex systems by encoding the components as nodes and the relationships as edges between nodes into graphs. Existing network methods have advantages and disadvantages in describing the properties of interactions. Correlation-based networks can estimate the strength of interactions but fail to identify their direction.[16] Bayesian networks can infer causal relationships by calculating conditional probabilities, but their strength and sign remain unknown.[17] All missing information about interactions can be retrieved by a generalized statistical mechanics model derived from the integration of evolutionary game theory, community ecology theory, and graph theory.[21,24–26] This model decomposes values of an entity into its independent components (assumed to occur in an isolated condition) and dependent components (resulting from directed influences of other agents on this agent) and assembles these independent and dependent components into informative, dynamic, omnidirectional, and personalized networks (idopNetworks).[27,28] The integration of idopNetworks and the GLMY (A. Grigor'yan, Y. Lin, Y. Muranov, and S.-T. Yau) theory, a computational topological theory pioneered by Yau S.S.T.,[29] has been shown to be powerful for studying complex systems.[30,31]

Current network modeling is deterministic, assuming that complex systems do not fluctuate over time and context. However, complex systems may not only display nonlinearity but also stochasticity and noise interference, all of which are inevitable and result from interactions with the environment.[32] While nonlinear interactions are the main cause of emergence of the whole being more than the sum of parts,[6,11,12,33,34] stochasticity can help the systems alter internal structure for adapting changing environments,[7–10,35] improving information processing,[36,37] and increasing robustness.[38,39] To the end, complex systems can be better described and disentangled when three characteristics mediating their internal operation mechanisms[3]—nonlinearity, uncertainty, and noises—are all taken into account in network reconstruction.

In this study, we incorporate the Yau-Yau nonlinear filtering model to identify stochastic networks from noisy data. The core objective of filtering is to pursue accurate estimation or prediction for the actual state of stochastic complex systems based on a series of noisy data.[38,39] Filtering theory was dated back to Gauss' time two centuries ago, but it was well established by R. E. Kalman in the 1960s[40,41] The resulting Kalman theory has been instrumental for understanding stochastic systems, but its widespread application is limited by the linear assumption. Many nonlinear filter theories have been subsequently proposed, aiming to enhance practical applications for prediction.[42,43] A global approach for computing these theories is derived on the basis of the conditional probability of the state that satisfies the Duncan-Mortensen-Zakai (DMZ) equation.[44–46] However, it is difficult to directly solve the DMZ equation, since it is parabolic with coefficients containing observations. Yau and Yau[47,48] derived an elegant algorithm for solving the DMZ equation by reducing it to a forward Kolmogorov equation (FKE). This algorithm, now referred to as the Yau-Yau filtering algorithm, has been theoretically proven to converge to the true solution, provided that the growth rate of the observation is greater than that of the drift.[49] In both theory and practice, the Yau-Yau algorithm has been shown to be advantageous over more classic ensemble Kalman filters (EnKFs), unscented Kalman filters (UKFs), and particle filtering methods.[50–53] Therefore, it is regarded as groundbreaking in the field of control theory, displaying great potential to magnify filtering applications for various complex systems. We show that stochastic networks by the Yau-Yau filter can capture hidden patterns and relationships of complex systems due to randomness and uncertainties, providing unique insight into revealing the fluctuating mechanisms of natural and social phenomena.

MATERIALS AND METHODS

Deterministic interaction networks

Mathematical modeling of evolutionary game theory. Consider a dynamic system, such as the gut microbiota, a multicellular tumor, or the climate change, with m interacting agents (variables), which are measured at a series of T time points. Let $y_j(t)$ ($j = 1, \ldots, m$; $t = 1, \ldots, T$) denote the observed value of the jth variable of the system measured at time t. Sun et al.[25] integrated evolutionary game theory to characterize how interactions between two variables behave like a game and further derived a system of mixed ordinary differential equations (mODEs) to quantify this process. These mODEs are expressed as

$$\frac{dy_j(t)}{dt} = Q_j\left(y_j(t)\right) + \sum_{j'=1, j' \neq j}^{m} Q_{j \leftarrow j'}\left(y_{j'}(t)\right), \quad \text{(Equation 1)}$$

where the observed value of a variable can be decomposed into its independent component $Q_j(y_j(t))$ due to the intrinsic strategy of this variable and dependent component $Q_{j \leftarrow j'}\left(y_{j'}(t)\right)$ arising from the extrinsic influence of other variables through their strategies acting on this variable. The independent strategy component of variable j occurs when it is

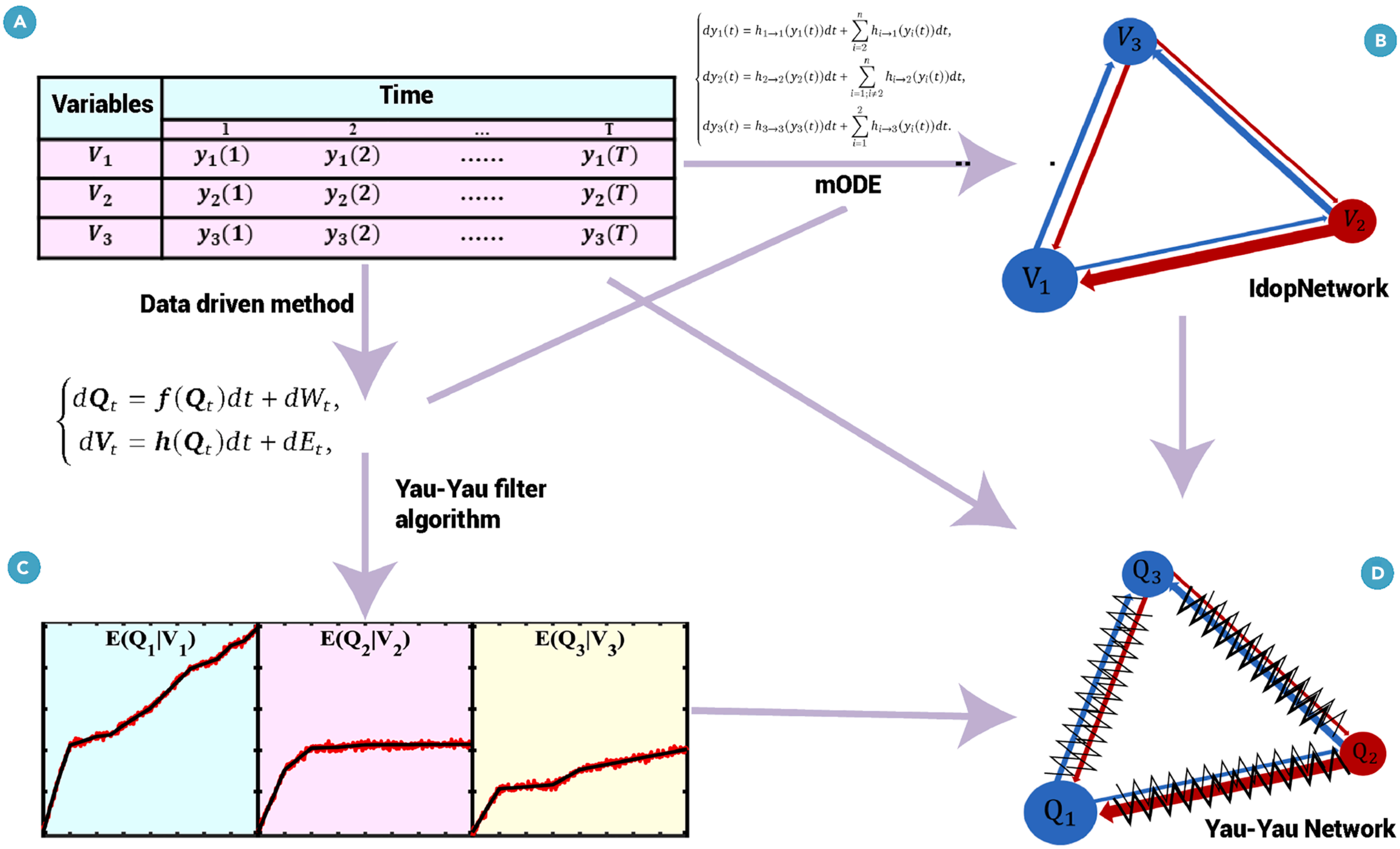


Figure 1. Diagram of reconstructing stochastic networks from noisy observations (A) Data structure for a complex system whose three components (variables) are measured at a series of T time points. (B) A deterministic interaction network among three components reconstructed by a system of mixed ordinary differential equations (mODEs). (C) The data are analyzed by the Yau-Yau filter to find the real states of three components. (D) The integration of mODE and Yau-Yau filter leads to the reconstruction of a stochastic interaction network (Yau-Yau network).

assumed to be in an isolated condition. For growth variables, the independent strategy component may obey the growth equation, expressed as

$$Q_j\left(y_j(t)\right) = \alpha_j y_j(t)\left(1-\left(\frac{y_j(t)}{K_j}\right)^{\beta_j}\right), \qquad \text{(Equation 2)}$$

where α_j is the relative change rate of variable j over time, K_j is the asymptotic value of variable j when time tends to be infinite, and β_j is the shape parameter of the growth curve. The dependent component reflects how and how much a variable j' affects the focal variable j, expressed as a function of the value of variable j'. In general, the forms of dependent components are unknown so that a nonparametric approach is implemented.

LOP fitting of dependent components. Because of its favorable mathematical properties, a Legendre Orthogonal Polynomial (LOP)-based nonparametric approach can be used to fit the time trajectories of dependent components.[25,26] Let $\mathbf{u}_{j\leftarrow j'} = (u_{j\leftarrow j'0}, u_{j\leftarrow j'1}, \ldots, u_{j\leftarrow j'R})$ denote a vector of basis values at degree R related to $Q_{j\leftarrow j'}\left(y_{j'}(t)\right)$, multiplied by $\mathbf{P}_{j\leftarrow j'}(t) = \left(P_{j\leftarrow j'0}(t), P_{j\leftarrow j'1}(t), \ldots, P_{j\leftarrow j'R}(t)\right)$ at degree R. Then, the dependent components in Equation 1 are formulated as

$$Q_{j\leftarrow j'}\left(y_{j'}(t)\right) = \mathbf{u}_{j\leftarrow j'}\mathbf{P}^{\mathrm{T}}_{j\leftarrow j'}(t)\hat{y}_{j'}(t), \qquad \text{(Equation 3)}$$

where $\hat{y}_{j'}(t)$ is the predicted value of $y_{j'}(t)$ using the growth Equation 1. For each variable, its optimal order of LOP can be determined via Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC).

Integrating graph theory. The mODEs of Equation 1 are solved under the fitting conditions given in Equations 2 and 3 by implementing the fourth-order Runge-Kutta algorithm. Ultimately, we obtain the estimates of two terms of Equation 1, independent components and dependent components. The independent components are only related to individual variables, whereas the dependent components describe how one variable affects, or is affected by, the others. For a specific pair of variables, there are two possible reciprocal dependent components, which represent the bidirectionality of variable-variable interactions. The estimated values of the independent components are associated with the strength of interactions, whereas their positive or negative estimation implies the relationship of promotion and repression between two given variables, respectively. Thus, according to graph theory, we encode the estimates of the independent components as nodes and the estimates of the dependent components as edges into informative networks that can fully capture the properties of bidirectional, signed, and weighted (bDSW) interactions. Since nodes and edges estimated by Equation 1 are a function of time, bDSW networks may dynamically change in structure and, thereby, its relationship with function. Equation 1, subject to variable selection by which a small set of the most significant variables linked to the focal variable j is chosen, can characterize all possible functionally existing links, forming sparse but omnidirectional networks. Network reconstruction by Equation 1 does not require data from multiple samples, meaning that bDSW networks can be reconstructed for each individual sample. Taken together, the model allows an idopNetwork to be reconstructed from a broad spectrum of data domains.[26,27] Given the nonlinear forms of $Q_j(y_j(t))$ and $Q_{j\leftarrow j'}\left(y_{j'}(t)\right)$, the idopNetwork provides a state-of-the-art means of disentangling hidden patterns of complex nonlinear systems.

Stochastic interaction networks

IdopNetworks are the networks that capture nonlinear relationships between different nodes. The behavior of many complex systems exhibits not only nonlinearity but also stochasticity.[32] One crucial step toward complexity research is to incorporate stochasticity into idopNetworks to better understand the internal workings of complex systems. It is not unreasonable to assume that the nodes and edges in the networks perturb stochastically over spatiotemporal scales to shape complex systems, whereas the perturbing pattern of the nodes and edges can be characterized from the noisy measurements of variables at multiple time points. This procedure can be formulated through the lens of the Yau-Yau nonlinear filter theory,[47,48] involving two coordinated state or signal and observation processes (Figure 1).

Yau-Yau nonlinear filter. A filtering system involves the actual state of a stochastic process that cannot be measured directly during the process but can be estimated from an observation process. Let $\mathbf{x}(t) = (x_1(t), \ldots, x_p(t)) \in \mathbb{R}^p$ denote the actual state vector of p agents comprising the system at time t, and let $\mathbf{y}(t) = (y_1(t), \ldots, y_m(t)) \in \mathbb{R}^m$ denote the observation vector of m variables describing the system measured at time t. The nonlinear filtering model[47,48] is built to determine approximate states from a given observation history, written as

$$\begin{cases} d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t))dt + d\mathbf{V}(t)\ \mathbf{x}(0) = \mathbf{x}_0 \\ d\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t))dt + d\mathbf{W}(t)\ \mathbf{y}(0) = \mathbf{y}_0 \end{cases}, \quad \text{(Equation 4)}$$

where $\mathbf{f}(\mathbf{x}(t))$ is the vector-valued state transition function in a nonlinear form, $\mathbf{h}(\mathbf{x}(t))$ is the vector-valued nonlinear function describing the hidden pattern of observation process, and $\mathbf{V}(t) \in \mathbb{R}^p$ and $\mathbf{W}(t) \in \mathbb{R}^m$ are mutually independent standard vector-valued Brownian processes with covariance matrices $E\left(d\mathbf{V}(t), d\mathbf{V}^{\mathrm{T}}(t)\right) = \mathbf{C}dt \in \mathbb{R}^p$ and $E\left(d\mathbf{W}(t), d\mathbf{W}^{\mathrm{T}}(t)\right) = \frac{1}{d}I_m dt \in \mathbb{R}^m$, respectively.

Equation 1 represents a generic model of the nonlinear filtering system, whose internal workings can be approached by natural law. For example, the observed system behavior is influenced by its actual state and measurement errors. How the system's actual state determines its observed behavior may be manifested through the operation mechanism of how the underlying agents act and interact with each other. In this study, we assume that p agents governing the actual state of the system correspond to m variables measured for the system. Taking into account these considerations, we rewrite the filtering Equation 4 as

$$\begin{cases} d\mathbf{x}(t) = \mathbf{f}(x_1(t), \ldots, x_m(t))dt + d\mathbf{V}(t) \\ dy_1(t) = \left(h_1(x_1(t)) + \sum_{j=2}^{m} h_{1\leftarrow j}(x_j(t))\right)dt + dW_1(t) \\ \vdots \\ dy_m(t) = \left(h_m(x_m(t)) + \sum_{j=1}^{m-1} h_{m\leftarrow j}(x_j(t))\right)dt + dW_m(t) \end{cases}, \quad \text{(Equation 5)}$$

where the second to last equations specify the observation processes of m variables, with $h_j(x_j(t))$ and $h_{j\leftarrow j'}(x_{j'}(t))$ ($j' \neq j$) representing the independent strategy component of variable j due to its intrinsic capacity and dependent strategy component of variable j resulting from the influence of variable j' on this variable, respectively. In Equation 4, $W_j(t) \in \mathbb{R}^m$ is a standard Brownian process of variable j with covariance matrix $E\left(dW_j(t), dW_j^{\mathrm{T}}(t)\right) = \mathbf{D}_j dt \in \mathbb{R}^m$. While the state equation describes how the state of variables fluctuates due to the recurrence of random diffusions, the observation equations specify how observed variables are affected by the state of variables through their self-regulation and interactions, subject to measurement errors. Equation 5 provides a unified framework of integrating idopNetworks into the Yau-Yau filtering model.

Fitting the state process. The Yau-Yau filter result shows that the state vector $\mathbf{x}(t)$ can be estimated from the observation vectors $\mathbf{y}(t)$ by solving the Kolmogorov equations. The state transition function $\mathbf{f}(x_1(t), \ldots, x_m(t))$ is locally Lipschitz continuous. In real data processing, $\mathbf{f}(x_1(t), \ldots, x_m(t))$ can be approximated by a data-driven nonparametric method, such as recurrent neural networks (RNNs) and sparse identification of nonlinear dynamics (SINDy) methods. Here, we consider SINDy methods. Let $\widehat{\mathbf{y}} = (\widehat{\mathbf{y}}(t_1), \cdots, \widehat{\mathbf{y}}(t_m))$ denote the estimated mean values of $\mathbf{y}(t)$ at a series of time points $(t_1, \ldots, t_T)$. Let

$$\mathbb{Y} = \begin{pmatrix} \dfrac{\widehat{y}(t_2) - \widehat{y}(t_1)}{\Delta t} \\ \dfrac{\widehat{y}(t_3) - \widehat{y}(t_2)}{\Delta t} \\ \vdots \\ \dfrac{\widehat{y}(t_m) - \widehat{y}(t_{m-1})}{\Delta t} \end{pmatrix}$$

$$\mathbb{Z} = \begin{pmatrix} \dfrac{(\widehat{y}(t_2) - \widehat{y}(t_1) - f(\widehat{y}(t_1))\Delta t)^2}{\Delta t} \\ \dfrac{(\widehat{y}(t_3) - y(t_2) - f(\widehat{y}(t_2))\Delta t)^2}{\Delta t} \\ \vdots \\ \dfrac{(\widehat{y}(t_m) - \widehat{y}(t_{m-1}) - f(\widehat{y}(t_{m-1}))\Delta t)^2}{\Delta t} \end{pmatrix},$$

which are estimates of drift terms and diffusion terms of the state model in the Yau-Yau filtering Equation 4, respectively. We construct a library $\mathbf{\Phi}$ consisting of nonlinear candidate functions of $\widehat{\mathbf{y}}$, which may be constant, polynomial, or more exotic functions, such as trigonometric and rational terms, and so on, which is expressed as

$$\mathbf{\Phi} = \left(1, \widehat{\mathbf{y}}, \widehat{\mathbf{y}}^2, \cdots, \sin(\widehat{\mathbf{y}}), \cos(\widehat{\mathbf{y}}), \sin(2\widehat{\mathbf{y}}), \cdots\right), \quad \text{(Equation 6)}$$

which is used to characterize the structure of stochastic differential equations (SDEs). We transform the identification problem of SDEs into a linear regression problem using the following models:

$$\mathbb{Y} = \mathbf{\Phi}\mathbf{c}_y + \boldsymbol{\epsilon}_y \quad \text{(Equation 7A)}$$

$$\mathbb{Z} = \mathbf{\Phi}\mathbf{c}_z + \boldsymbol{\epsilon}_z, \quad \text{(Equation 7B)}$$

where $\mathbf{c}_y$ and $\mathbf{c}_z$ are the regression coefficients that determine the form of the state transition function, and $\boldsymbol{\epsilon}_y$ and $\boldsymbol{\epsilon}_z$ are residual errors following a multivariate normal distribution. The least-squares method is implemented to estimate $\mathbf{c}_y$ and $\mathbf{c}_z$ as the inference of the state transition function. Because the library $\mathbf{\Phi}$ is constructed as a function space from which only an appropriate function is searched, $\mathbf{c}_y$ and $\mathbf{c}_z$ each require a certain level of sparsity, which can be determined by Least Absolute Shrinkage and Selection Operator (LASSO), sparse Bayesian learning, and some other methods.

Fitting the observation process. The observation process function includes an independent strategy component, $h_j(\cdot)$, and dependent strategy component, $h_{j\leftarrow j'}(x_{j'}(t))$, both of which can be model by a parametric or nonparametric approach. Here, we use an LOP-based nonparametric approach. Let $\mathbf{u}_j = (u_{j0}, u_{j1}, \ldots, u_{jR})$ denote a vector of basis values related to $h_j(x_j(t))$, multiplied by $\mathbf{P}_j(x_j(t)) = (P_{j0}(x_j(t)), P_{j1}(x_j(t)), \ldots, P_{jR}(x_j(t)))$ which denote LOP values at degree R, and let $\mathbf{u}_{j\leftarrow j'} = \left(u_{j\leftarrow j'0}, u_{j\leftarrow j'1}, \ldots, u_{j\leftarrow j'R}\right)$ denote a vector of basis values related to $h_{j\leftarrow j'}(x_{j'}(t))$, multiplied by $\mathbf{P}_{j\leftarrow j'}(x_{j'}(t)) = \left(P_{j\leftarrow j'0}(x_{j'}(t)), P_{j\leftarrow j'1}(x_{j'}(t)), \ldots, P_{j\leftarrow j'R}(x_{j'}(t))\right)$ which denote LOP values at degree R. Then, the independent and dependent components are formulated as

$$h_j(x_j(t)) = \mathbf{u}_j \boldsymbol{P}_j^{\mathrm{T}}(x_j(t)) \quad \text{(Equation 8A)}$$

$$h_{j\leftarrow j'}(x_{j'}(t)) = \mathbf{u}_{j\leftarrow j'} \boldsymbol{P}_{j\leftarrow j'}^{\mathrm{T}}(x_{j'}(t)). \quad \text{(Equation 8B)}$$

It is possible that $h_j(x_j(t))$ and $h_{j\leftarrow j'}(x_{j'}(t))$ can be better fitted by different LOP degrees. This difference can be tested and determined by AIC or BIC.

Yau-Yau network reconstruction. We incorporate the Yau-Yau filter algorithm[47,48] to solve the filtering system described by Equation 4. The estimated independent strategy components and dependent strategy components are perturbed. We code the independent components as perturbed nodes and the dependent components as perturbed edges into graphs, forming stochastically perturbed networks (coined Yau-Yau networks). Each node and edge are perturbed by implementing random (co)diffusion processes related to variable state values.

Application

A worked example. To demonstrate the reconstruction procedure of Yau-Yau networks, we analyze microbial data from a tri-cultural experiment. Three bacterial species, *Escherichia coli*, *Staphylococcus aureus*, and *Pseudomonas aeruginosa*, were jointly cultivated in the same medium to form a triad. There were a total of 100 such triads generated independently by 100 strains from each species. Each strain in the triads was measured for its abundance at a series of time points, which is used for network reconstruction.

In each triad as a small society, three co-existing bacterial species interact with each other in a nonlinear manner to shape microbial community behavior. We apply Equation 1 to reconstruct two deterministic interaction networks for triads #64 (Figure 2A) and #91 (Figure 2B). We find that species *E. coli* exerts an antagonistic relationship with both species *S. aureus* and *P. aeruginosa*, whereas there is little interaction between *S. aureus* and *P. aeruginosa* in both triads. We also find that interspecific interactions change over time, showing the dynamic property of microbial interaction networks (Video S1). In general, *E. coli* receives stronger but more cyclically changing inhibition from its co-existing counterparts than do the other two species. Microbial inhibition estimated by a deterministic model is thought to be stable at each time, which cannot reflect the reality of microbial perturbation to better adapt to stochastic errors.

Yau-Yau networks can capture how microbial interactions fluctuate around their real states. To reconstruct Yau-Yau networks, we implement SINDy equations. From the library $\mathbf{\Phi}$, given in SINDy Equation 6, we determine the overall core power of the dynamical microbial system. This power system is driven by the means of microbial growth over 100 tri-cultures. The form of the state equation is determined by selecting as few bases as possible without losing too much information, which is expressed as

$$d\mathbf{x}(t) = \left(A_0 + A_1\mathbf{x}(t) + A_2\mathbf{x}^2(t) + A_s \sin(\mathbf{x}(t))\right)dt + dV(t), \quad \text{(Equation 7)}$$

where A_1, A_2, A_s, and A_0 are matrices $\in \mathbb{R}^{3\times3}$ and $E\left(dV(t)dV^T(t)\right) = \frac{1}{10}I_3 dt$. Figure 3 illustrates the fitness of the above SINDy equation to the abundance data of three species from two randomly selected triads, #64 and #91. It can be seen that all species in the community fluctuate around their real time-varying state, but to different extents, depending on

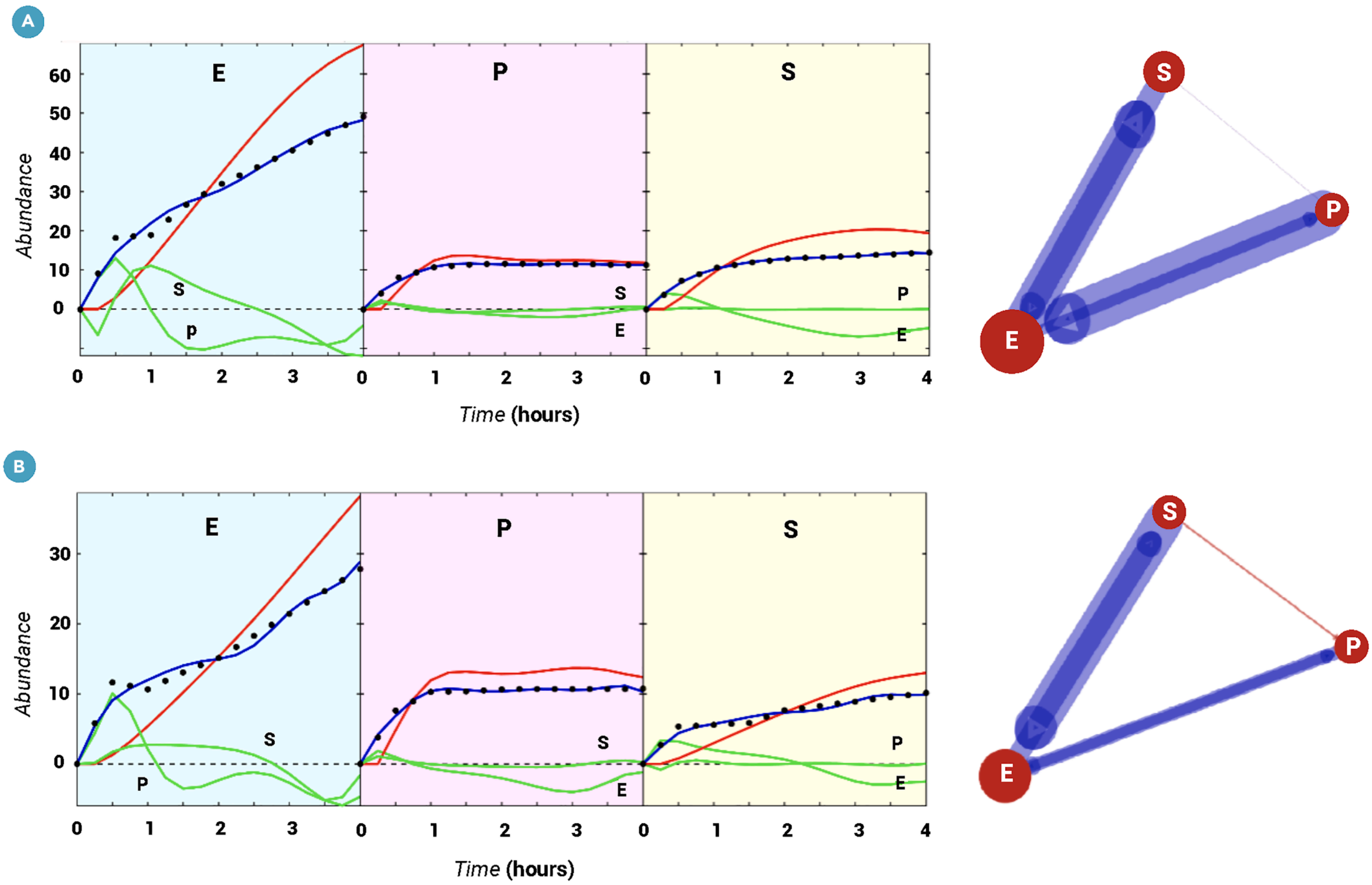


Figure 2. Deterministic interaction networks inferred from decomposed abundance dynamics using mODEs Deterministic idopNetworks of tri-cultured bacterial species (*E. coli*, *S. aureus*, and *P. aeruginosa*) from triads #64 (A) and #91 (B) by mODEs. Left: overall abundance trajectories of each species (*E. coli*, *S. aureus*, and *P. aeruginosa*) (blue) decomposed into independent component curves (red) and dependent component curves (green). The independent component arises from the intrinsic capacity of a species that is fully expressed when this species is assumed to be in isolation, whereas the dependent component is the consequence of directed influence of other species on this species. Right: three-node bidirectional, signed, and weight interaction networks. Red and blue arrowed lines denote promotion and inhibition, respectively, with the thickness of lines proportional to the strength of interaction.

species type and tri-culture. It appears that species *E. coli* and *S. aureus* each grow in an S-shaped curve in triad #64, which displays a different pattern from species *P. aeruginosa* in the same tri-culture. We find that estimated growth trajectories fluctuate during ontogeny, but with a considerably more frequent and larger fluctuation for species *S. aureus* and *P. aeruginosa* than *E. coli* (Figure 3).

The Yau-Yau filter contains a nonlinear observation equation that can characterize how different species act and interact with ach other to mediate the dynamics of microbial communities. Using the Yau-Yau nonlinear filter algorithm, we reconstruct stochastic Yau-Yau idopNetworks (Figure 4). In general, the Yau-Yau idopNetworks are in broad agreement with deterministic idopNetworks (Figure 2) in microbial interaction structure and dynamics, but the former allow the nodes and edges to fluctuate around their actual states at each time (Video S2). We find that both microbial abundance and microbial interactions fluctuate strikingly during microbial growth, implying that bacterial species can better adapt to their co-existing counterparts by adjusting their states.

Monte Carlo simulation

We perform computer simulation studies to investigate the statistical properties of the Yau-Yau nonlinear filter algorithm in complex system dissection. We simulate data under three assumed systems (linear sensor system, nonlinear sensor system, and SINDy system), which are analyzed by the Yau-Yau algorithm. The linear sensor system is described by the following equations:

$$\begin{cases} dx(t) & = Ax(t)dt + dV(t) \\ dy(t) & = h(x(t))dt + dW(t), \end{cases} \qquad \text{(Equation 8)}$$

where $x(0) \sim N(0, I_{d_x})$, A is a matrix with size $d_x \times d_x$ defined as

$$A_{jj'} = \begin{cases} 0.1, & \text{if } j' = j - 1 \\ 0.5, & \text{if } j' = j \\ 0, & \text{else} \end{cases},$$

and $h(x(t))$ is set as $h(x(t)) = Px(t) + Qx^2(t)$. To improve the generality of Yau-Yau filter in real data analysis, we orthogonalize matrix P, including the following procedure:

$$\begin{aligned} \tilde{p}_1 & = P_1; \\ \tilde{p}_2 & = P_2 - (P_2, \tilde{p}_1)\frac{\tilde{p}_1}{\|\tilde{p}_1\|_2}; \\ \tilde{p}_3 & = P_3 - (P_3, \tilde{p}_1)\frac{\tilde{p}_1}{\|\tilde{p}_1\|_2} - (P_3, \tilde{p}_2)\frac{\tilde{p}_2}{\|\tilde{p}_2\|_2}; \\ \cdots & \;\; \cdots \\ \tilde{P}_j & = \frac{\tilde{p}_j}{\|\tilde{p}_j\|_2}\left(1 \le j \le d_y\right). \end{aligned}$$

which can be written in matrix form as $\tilde{P} = TP$, where T is a reversible lower triangular matrix when P is a full rank matrix. Similarly, we perform the same linear transform T to orthogonalize matrix Q as $\tilde{Q} = TQ$. Then, we obtain the orthogonalized observation function $\tilde{h}(x(t)) = Th(x(t))$. In a three-dimensional linear sensor system, we implement the Yau-Yau algorithm to estimate the real state process of three variables (Figure 5A). It can be seen that the Yau-Yau algorithm can capture the general trend of the real state process of each variable with increasing accuracy when the observations are orthogonalized.

The nonlinear sensor system is formulated as

$$\begin{cases} dx(t) & = (Ax(t) + cos(x(t)))dt + dV(t) \\ dy(t) & = h\left(x^3(t)\right)dt + dW(t), \end{cases} \qquad \text{(Equation 9)}$$

where A is the same in Equation 8. We orthogonalized the observation equation as $h(x^3(t)) = Px^3(t) + Qx^6(t)$. Figure 5B illustrates the real state process and its estimation by the Yau-Yau algorithm. In the nonlinear system, orthogonalization can strikingly increase the estimation process of the real state.

For a SINDy system, we choose the sensor system from the library in Equation 4, which is expressed as

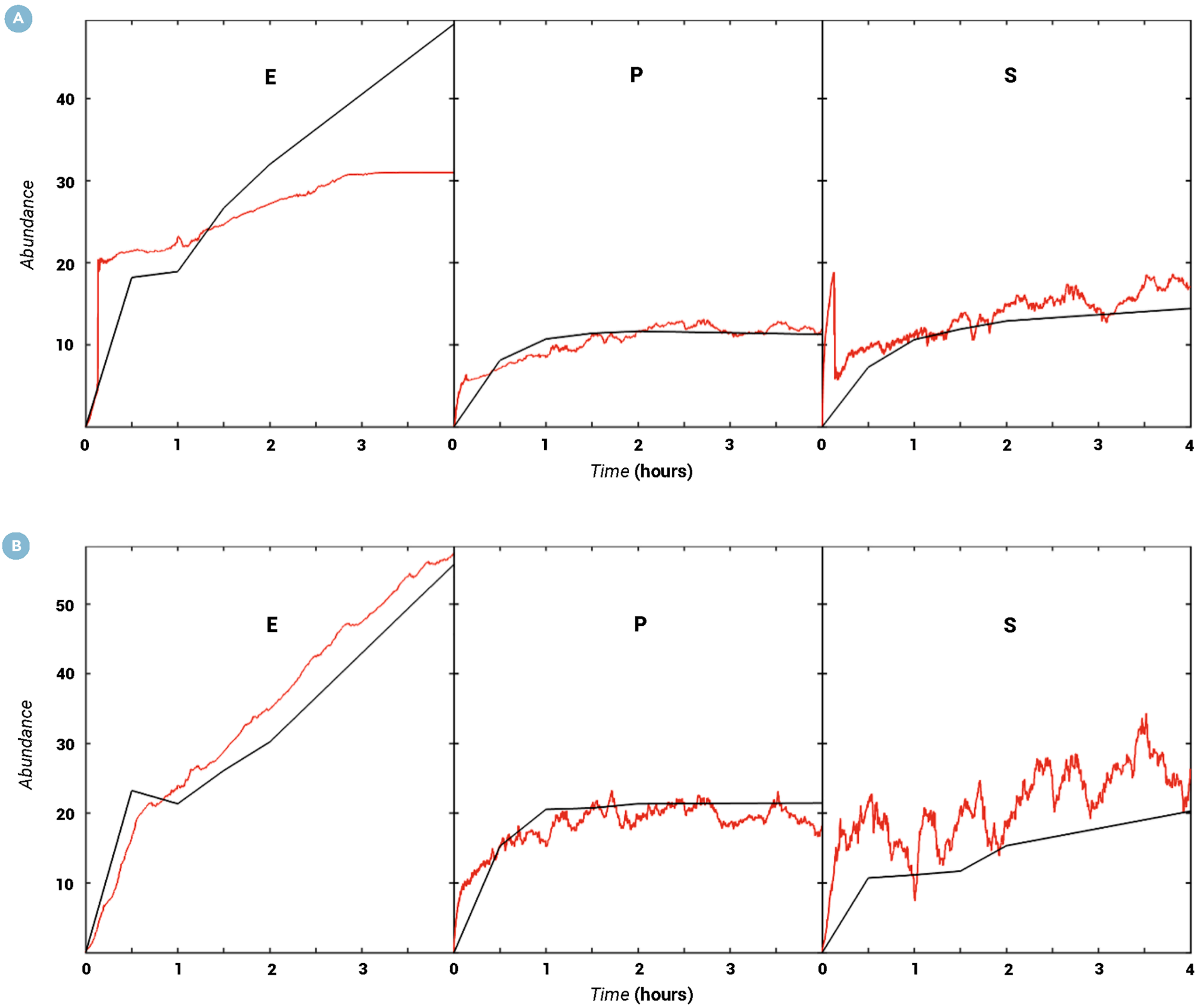


Figure 3. Comparison of real and estimated microbial abundance states Shown are the real state (black) and estimated states (red) of microbial abundance trajectories for tri-cultured bacterial species (*E. coli*, *S. aureus*, and *P. aeruginosa*) from triads #64 (A) and #91 (B).

$$\begin{cases} dx(t) &= S\mathbf{L}(x(t))dt + dV(t) \\ dy(t) &= h(x(t))dt + dW(t) \end{cases}, \quad \text{(Equation 10)}$$

where $\mathbf{L}(\cdot)$ is the basis library with m triangle functions, i.e., $\mathbf{L}(\cdot) = [sin(\cdot), sin(2\cdot), sin(3\cdot),\cdots, cos(\cdot), cos(2\cdot), cos(3\cdot),\cdots]$, and S is a sparse coefficient matrix, $\in \mathbb{R}^{n\times m}$. As can be seen, the Yau-Yau algorithm can accurately estimate the real state of the system based on orthogonalized observations (Figure 5C).

We assess the precision of the Yau-Yau algorithm by estimating the mean square errors. As expected, under the same condition, a simpler linear sensor system can be estimated with greater precision than a more complex nonlinear system, both of which can be much more precisely estimated as compared with a SINDy system (Table 1). However, when the observation process is orthogonalized, the three systems can be estimated by the Yau-Yau algorithm almost with the same precision.

We perform additional simulation studies to validate the Yau-Yau algorithm for analyzing a complex system simulated by linear, nonlinear, and SINDy algorithms under various levels of noise (0.8, 1.0, 1.5, and 2.0). In general, the actual state of the noisy system can be well estimated (Figure S1), although, as expected, the root-mean-square errors (RMSEs) of the system estimation increase with an increasing level of noise (Table S1). Given that the noise value of 2.0 is quite large relative to the system size, a reasonably low RMSE estimate indicates the robustness of the Yau-Yau algorithm.

Materials availability

Ecological experiment of microbial culture. We design and conduct an ecological experiment to validate the biological value of the model. We collected 100 bacterial strains of each species, *E. coli*, *S. aureus*, and *P. aeruginosa*, from the National Infrastructure of Microbial Resources, China. These strains were pre-cultivated with a procedure described by He et al.[54] The culture experiment includes 100 tri-cultures each for a set of three strains (triad) at a 1:1:1 ratio from different species. To assure the independence of triads, none of the same strains were used in any two co-cultures or any two tri-cultures. All cultures were taken in flasks of the same size, filled with 25 mL of twice-diluted brain heart infusion medium (Oxoid, Basingstoke, England). The flasks were incubated at 30°C and shaken at 130 rpm.

Each strain in the flasks was sampled once every 0.5 h during the first 2 h of cultivation, followed by once every 2 h after 2 h and once every 4 h after 12 h, with a total culture period of 36 h. Quantitative PCR (qPCR) measurements of *E. coli* and *S. aureus* were performed according to the previous protocol. To detect *P. aeruginosa* strains, 159 bp of the gyrA gene encoding DNA gyrase subunit A were amplified by forward primer CAAGCCCTACAA GAAATCCG and reverse primer TCCACCGAACCGAAGTTG. The qPCR counts of each strain at a time point, averaged over three replicates, were used as the time-varying abundance level of this strain.

Data preprocessing. We use Box-Cox transformation to assure that microbial abundance data follow a normal distribution. A data imputation method is applied to obtain

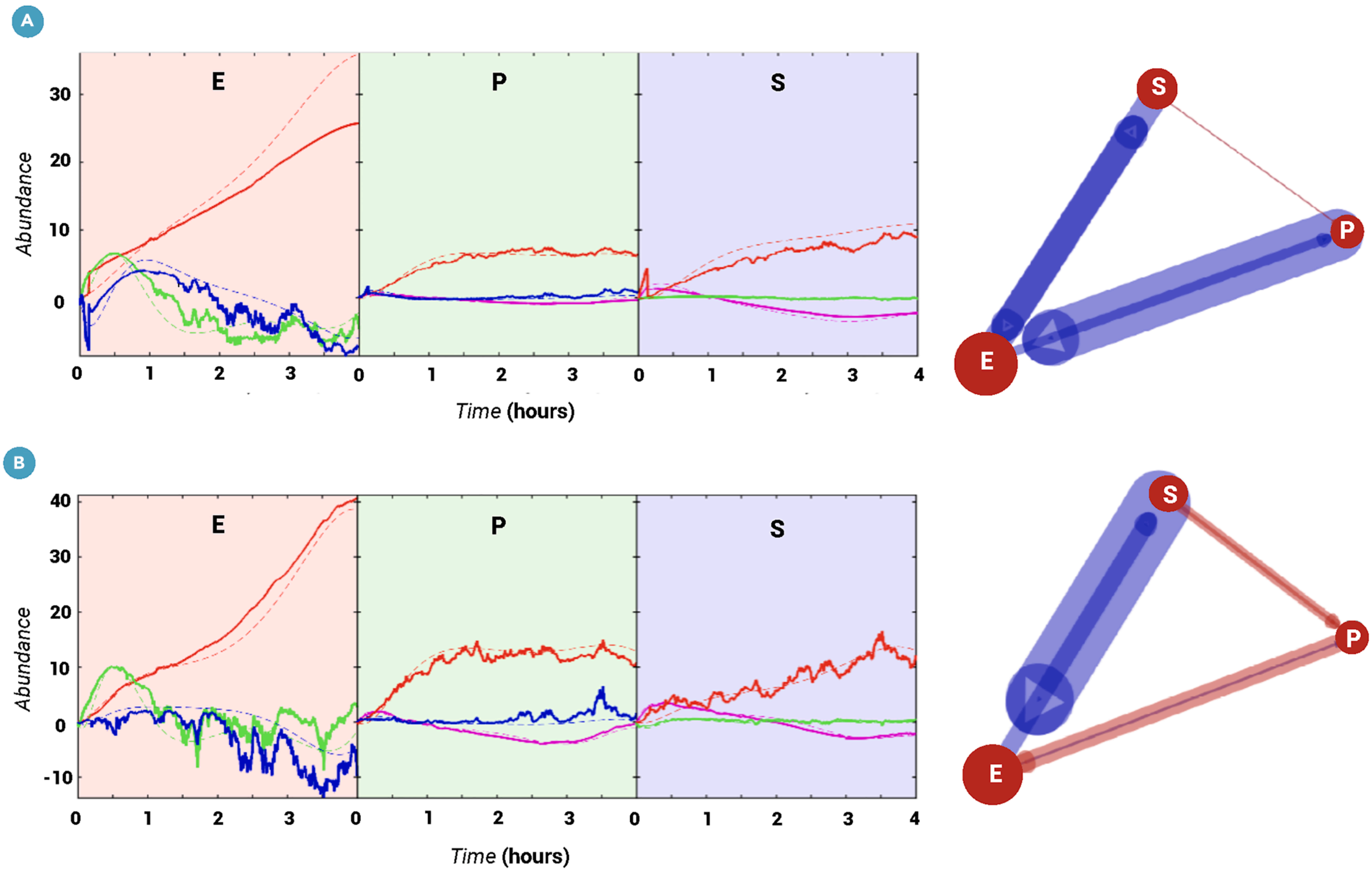


Figure 4. Stochastic interaction networks from perturbation-based analysis using the Yau-Yau nonlinear filter algorithm Shown are Yau-Yau networks of tri-cultured bacterial species (*E. coli*, *S. aureus*, and *P. aeruginosa*) from triads #64 (A) and #91 (B) by the Yau-Yau nonlinear filter algorithm. Left: perturbed overall abundance trajectories of each species (*E. coli*, *S. aureus*, and *P. aeruginosa*) (blue) into perturbed independent component curves (red) and perturbed dependent component curves (green). Right: three-node bidirectional, signed, and weighted interaction stochastic networks with nodes and interactions fluctuating around their real states.

dense and equally spaced abundance data. Here, we use linear interpolation to fill in the observations at the newly added observation time points. The advantage of using linear interpolation is that the newly filled observations will also obey a normal distribution due to the additivity of the normal distribution.

Algorithm availability

The Yau-Yau algorithm. We convert the problem of interaction network evolution into a nonlinear filtering problem with a form described by Equation 2. We aim to obtain the accurate and instantaneous estimation of the state process $\mathbf{x}(t)$ through the noisy observation history $\mathbf{y}(t) := \{\mathbf{y}(s): 0 \le s \le t\}$ by calculating the conditional expectation $E(\mathbf{x}(t)|\mathbf{y}(t))$. Per the Yau-Yau nonlinear filter algorithm, we calculate the entire conditional density $\rho(t, \mathbf{x})$ of $\mathbf{x}(t)$ given the observation history $\mathbf{y}(t) := [\mathbf{y}(s): 0 \le s \le t]$ in a general form of Equation 2. In the 1960s, Duncan, Mortensen, and Zakai proposed $\sigma(\mathbf{x},t)$, the unnormalized density function $\rho(t,\mathbf{x})$,[44–46] which satisfies the DMZ equation:

$$\begin{cases} d\sigma(\mathbf{x},t) = L\sigma(\mathbf{x},t)dt + d\sigma(\mathbf{x},t)\mathbf{h}^T(\mathbf{x},t)d\mathbf{y}(t) \\ \sigma(\mathbf{x},0) = \sigma_0(\mathbf{x}) \end{cases},$$

where $\sigma_0(\mathbf{x})$ is the probability density of the initial state, and

$$L(*) \equiv \frac{1}{2}\Delta C * - \sum_{j=1}^{m} \frac{\partial(f_j *)}{\partial x_j}.$$

Next, we construct robust state estimators from any observed sample paths. For each given observation $\mathbf{y}(t)$, let

$$u(\mathbf{x},t) = \exp\left[d\mathbf{h}^T(\mathbf{x})\mathbf{y}(t)\right]\sigma(\mathbf{x},t),$$

which yields the "pathwise-robust" DMZ equation.

In the Yau-Yau algorithm,[47,48] the robust DMZ equation is reformulated as

$$\begin{cases} \frac{\partial u}{\partial t}(t,\mathbf{x}) = \frac{1}{2}\Delta C u(t,\mathbf{x}) + (-\mathbf{f}(\mathbf{x}) + \nabla K(t,\mathbf{x}))\cdot\nabla u(t,\mathbf{x}) \\ \quad + \left(-\nabla\cdot\mathbf{f}(\mathbf{x}) - \frac{d}{2}|\mathbf{h}(\mathbf{x})|^2 + \frac{1}{2}\Delta C K(t,\mathbf{x}) \right. \\ \quad \left. - f(\mathbf{x})\cdot\nabla K(t,\mathbf{x}) + \frac{1}{2}|\nabla K(t,\mathbf{x})|^2\right)u(t,\mathbf{x}), \\ u(0,\mathbf{x}) = \sigma_0(\mathbf{x}), \end{cases}$$

where $K = \sum_{j=1}^{m} dy_j(t)h_j(\mathbf{x}), \mathbf{f} = (f_1,\ldots,f_m)^\top, \mathbf{h} = (h_1,\ldots,h_m)^\top$. We further have

$$\begin{cases} \frac{\partial u_j}{\partial t}(t,\mathbf{x}) = \frac{1}{2}\Delta C u_j(t,\mathbf{x}) \\ \quad + \sum_{\ell=1}^{m}\left(-f_\ell(\mathbf{x}) + \sum_{j=1}^{m} y_j(\tau_t)\frac{\partial h_j}{\partial x_\ell}(\mathbf{x})\right)\frac{\partial u_j}{\partial x_\ell}(t,\mathbf{x}) \\ \quad - \left(\sum_{\ell=1}^{m}\frac{\partial f_L}{\partial x_\ell}(\mathbf{x}) + \frac{1}{2}\sum_{\ell=1}^{m} h_\ell^2(\mathbf{x}) - \frac{1}{2}\sum_{j=1}^{m} V_j(\tau_t)\Delta C h_j(\mathbf{x}) \right. \\ \quad + \sum_{j=1}^{m}\sum_{\ell=1}^{m} y_j(\tau_t) f_\ell(\mathbf{x})\frac{\partial h_j}{\partial x_\ell}(\mathbf{x}) \\ \quad \left. -\frac{1}{2}\sum_{p=1}^{m}\sum_{j=1}^{m}\sum_{\ell=1}^{m} y_j(\tau_t)V_\ell(\tau_t)\frac{\partial h_j}{\partial x_p}(\mathbf{x})\frac{\partial h_e}{\partial x_p}(\mathbf{x})\right)u_j(t,\mathbf{x}) \\ u_j(\tau_{t-1},\mathbf{x}) = u_{j-1}(\tau_{t-1},\mathbf{x}). \end{cases}$$

Let

$$u_j(\mathbf{x},\tau_t) = exp\left(-\sum_{j=1}^{n} dy_j(\tau_{t-1})h_j(\mathbf{x})\right)\tilde{u}_t(\mathbf{x},\tau_t).$$

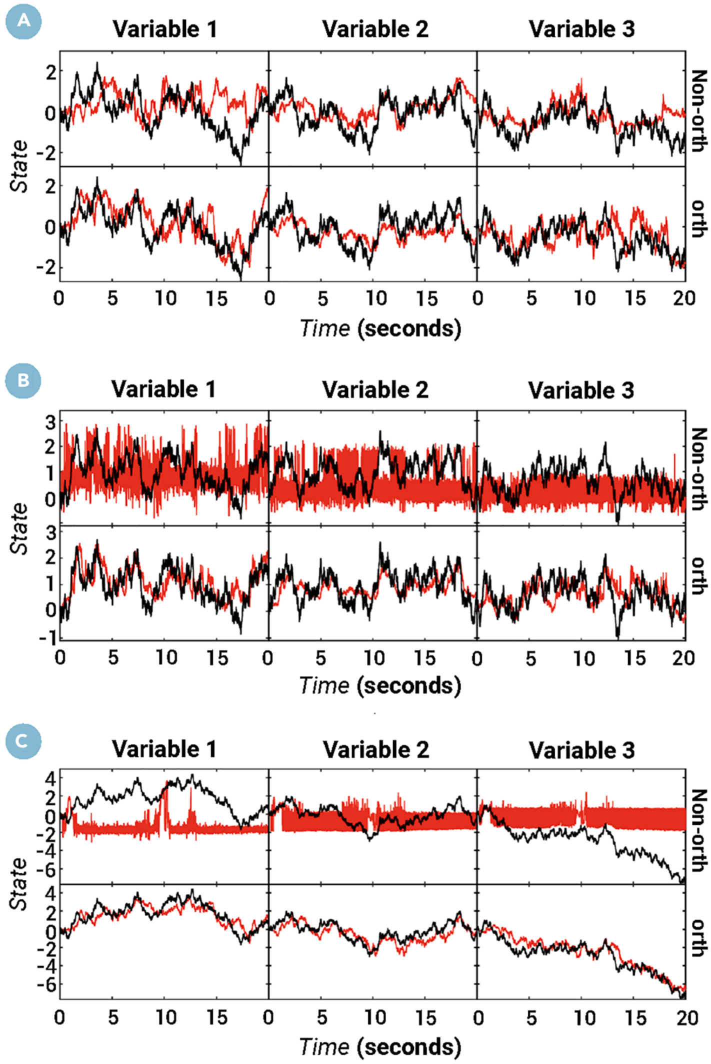


Figure 5. State estimation performance with versus without observation orthogonalization across system types Shown are simulation results for linear (A), nonlinear (B), and SINDy sensor systems (C) from non-orthogonalized observations and orthogonalized observations estimated by the Yau-Yau filter algorithm. Black and red curves denote the real state and estimated states of the systems, respectively.

Then, $u_j(\mathbf{x},\tau_t)$ can be computed by $\tilde{u}_j(\mathbf{x},\tau_t)$, where $\tilde{u}_j(\mathbf{x},t)$ for $\tau_{t\text{-}1} \le t \le \tau_t$ satisfies the following Kolmogorov equation, expressed as

$$\begin{cases} \frac{\partial \tilde{u}_j}{\partial t}(\mathbf{x},t) &= \frac{1}{2}\Delta C\tilde{u}_j(\mathbf{x},t) - \sum_{j=1}^{m} f_j(\mathbf{x})\frac{\partial \tilde{u}_j}{\partial x_j}(\mathbf{x},t) \\ & \quad - \left(divf(\mathbf{x}) + \frac{d}{2}\sum_{j=1}^{m} h_j^2(\mathbf{x})\right)\tilde{u}_j(\mathbf{x},t), \\ \tilde{u}_j(\mathbf{x},\tau_{t-1}) &= exp\left(\sum_{j=1}^{m} d\left(y_j(\tau_{t-1}) - y_j(\tau_{t-2})\right)h_j(\mathbf{x})\right)\tilde{u}_{j-1}(\mathbf{x},\tau_{t-1}). \end{cases} \qquad \text{(Equation 11)}$$

In Equation 11, the PDE itself does not contain observations, which means that, for any moment of observational updating, the equations that need to be solved are the same, only the initial conditions are different, and with the appropriate PDE solution, the updating at each moment can be transformed into matrix multiplication so that the online updating is hardly time consuming. At the same time, the initial state contains only the observations of that moment and the previous moment, and there are very little data to be saved. The method of solving Partial Differential Equation (PDE) here mainly uses quasi-implicit Euler method (QIEM).

At the end, we substitute $E(\mathbf{x}(t)|\mathbf{y}(t))$ into $h_{j\leftarrow j'}(\cdot)$ to obtain a microbial interaction network that takes into account stochasticity, observation errors, and the pattern of interaction evolution over time.

Table 1. Precision analysis and computation cost of the Yau-Yau algorithm to analyze linear, nonlinear, and SINDy sensor systems based on orthogonalized observations in a comparison with those based on non-orthogonalized observations shown in paratheses

	Linear	Nonlinear	SINDy
RMSE	0.8421 (0.97103)	0.5797 (1.2687)	0.9052 (4.0451)
Mean error	0.7633 (0.81728)	0.5074 (1.1815)	0.8201 (3.7653)
Time cost, s	13.8775 (14.0072)	12.3792 (12.3732)	75.9961 (76.8672)

RESULTS AND DISCUSSION

Stochastic networks are the domain of a wide variety of complex systems in which components act and interact stochastically across spatiotemporal scales, subject to a structural rearrangement described by a graph.[36,55] While previous studies have focused on deterministic regular expressions of complex networks, it is largely unexplored to what extent nonlinearities, unclear causal chains, and unpredictable perturbations jointly govern system behavior. Analysis and interpretation of such stochastic networks require developing and combining concepts from graph theory to characterize inter-agent interactions and stochastic modeling to reveal how these interactions shape the long-term dynamics of the systems. Several studies have begun to infer SDEs that describe complex systems from observed data.[32,56–59] However, they fail to characterize how real states of nodes and interactions within networks fluctuate due to unpredictable errors and how these real states can be detected from noisy data.

In this study, we implement the Yau-Yau nonlinear filter into a statistical mechanics framework for reconstructing stochastic idopNetworks (i.e., Yau-Yau networks) that are fundamental to complex systems. To demonstrate the statistical properties of the Yau-Yau algorithm, we conducted a series of computer simulations by assuming a variety of filtering systems. Results from simulation studies show that the Yau-Yau algorithm performs reasonably well in these simulation scenarios, especially after the observation data are orthogonalized. Beyond previous state-of-the-art deterministic networks, Yau-Yau networks can identify real states of how each agent acts and how it interacts with every other agent in response to stochastic errors to shape system dynamics. In theory, it is impossible to measure real states of natural phenomena, but the Yau-Yau networks can extract fundamental information or principles from observation data. All of these features make the Yau-Yau networks a powerful tool to automatize scientific discovery from chaotic disorders.

We design and conduct an ecological experimental by jointly culturing different strains from three bacterial species in the same medium to create a small-scale society. Deterministic models coalesce three species into idopNetworks in which bDSW interactions between each pair of species are characterized. Although idopNetworks can illustrate the time-varying trajectories of microbial abundance and interactions, they fail to capture their stochastic fluctuations at species time points. In natural and living systems (e.g., the soil microbiota, the gut microbiota, or the ocean microbiota), microbial activities always tend to constantly perturb in order to better adapt to random errors during growth ontogeny. The Yau-Yau networks leverage idopNetworks to distill such concealed stochastic perturbation from observed data, which helps to unveil the biological mechanisms and rules that govern natural and social phenomena.

The Yau-Yau networks described in this study are reconstructed from continuous time series data, but in practice, such data may not be available. We introduce and integrate allometric scaling law to convert static data into its quasi-dynamic representation by which Yau-Yau networks can be inferred. This integration not only showcases the applicability and flexibility of Yau-Yau networks but also generates novel insights for understanding the mechanisms hidden in various complex system scenarios using only snapshots rather than expensive temporal data.

One important feature of many complex systems is that they are composed of many agents and processes that interact with each other in interdependent manners and whose structure and behavior can be described at several nested, coupled scales by different principles and laws. By integrating RNNs, Chen et al.[60] proposed an algorithm to solve high-dimensional filtering problems. It shows that their algorithm can estimate 60–100 dimensions of filtering systems, remarkably enhancing the ability of the Yau-Yau filter to analyze high-dimensional data. To disentangle the ultrahigh dimensionality of complex systems, we may classify high-dimensional data into distinct modules, each of a smaller size, by implementing functional clustering.[61,62] A small-size module can be readily analyzed by the Chen et al.[60] algorithm.

Higher-order interactions have been increasingly recognized to be fundamental part of complex systems.[63–65] Feng et al.[66] proposed a general model for characterizing inter-agent interactions at any order across a wide spectrum of scenarios by which to reconstruct hypernetworks filled with nodes and directed, signed, and weighted edges and hyperedges. The Yau-Yau networks can be extended to accommodate high-order interactions by incorporating additional terms into the Yau-Yau filter. While many complex networks have been successfully mapped in the past, dissecting the topological architecture of these networks may not always be feasible in certain scenarios. More recently, Yau and his team developed a GLMY model for computing the path homology of digraphs.[29] The integration of GLMY theory and Yau-Yau filters through idopNetworks will not only more precisely unveil the hidden patterns of complex systems but, more importantly, open up a unique gateway for contextualizing these two distinct disciplines into a unified mathematical framework. We pack the algorithm for Yau-Yau networks into Yau-YauAL, a computing platform for solving nonlinear filtering problems.[67]

RESOURCE AVAILABILITY

Data and code availability

The microbial data and code for data analysis and computer simulation in this study have been deposited in the public GitHub repository (http://GitHub/BIMSA_YauYauNetwork).

FUNDING AND ACKNOWLEDGMENTS

We thank many of our colleagues for stimulating discussions of the Yau-Yau filter, which made it possible to write this manuscript. Part of this work is supported by the Interdisciplinary Project at the Shanghai Institute for Mathematics and Interdisciplinary Sciences (SIMIS-ID-2024-WN).

AUTHOR CONTRIBUTIONS

S.X. and Y. Wang (first) derived the model and performed computation. S.W., Y. Wang (second), and A.D. participated in the research. S.S.-T.Y. and S.-T.Y. supervised the project. R.W. conceived the idea. R.W. and S.X. wrote the paper. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interest.

SUPPLEMENTAL INFORMATION

It can be found online at https://doi.org/10.1016/j.xinn.2026.101267.

REFERENCES

1. Amaral, L.A.N., Díaz-Guilera, A., Moreira, A.A. et al. (2004). Emergence of complex dynamics in a simple model of signaling networks. *Proc. Natl. Acad. Sci. USA* **101**:15551–15555. DOI:10.1073/pnas.0404843101
2. Holland, J.H. (2014). Complexity: A Very Short Introduction (Oxford University Press). DOI:10.1093/actrade/9780199662548.001.0001
3. Yang, K., Li, J., Liu, M. et al. (2023). Complex systems and network science: a survey. *J. Syst. Eng. Electron.* **34**:543–573. DOI:10.23919/jsee.2023.000080
4. Dragicevic, A.Z. (2019). Spacetime Discounted Value of Network Connectivity. *Adv. Complex Syst.* **21**:1850018. DOI:10.1142/s0219525918500182
5. Dragicevic, A.Z. and Gurtoo, A. (2022). Stochastic control of ecological networks. *J. Math. Biol.* **85**:7. DOI:10.1007/s00285-022-01777-5
6. Mensah, P., Katerere, D., Hachigonta, S. et al. (2018). Systems Analysis Approach for Complex Global Challenges (Springer), pp. 179–276. DOI:10.1007/978-3-319-71486-8
7. Levin, S. (2002). Complex adaptive systems: Exploring the known, the unknown and the unknowable. *Bull. Amer. Math. Soc.* **40**:3–19. DOI:10.1090/s0273-0979-02-00965-5
8. Carmody-Bubb, M. (2023). What Is a Complex Adaptive System? In Cognition and Decision Making in Complex Adaptive Systems, pp. 7–11. DOI:10.1007/978-3-031-31929-7_2
9. Laughlin, C.D. (2023). Consciousness as an intelligent complex adaptive system: A neuro-anthropological perspective. *Anthropol. Conscious* **35**:15–41. DOI:10.1111/anoc.12213
10. He, J. (2011). Complexity in Adaptive Systems. In Encyclopedia of Machine Learning (Springer). DOI:10.1007/SpringerReference_178860
11. Corning, P.A. (2002). The re-emergence of "emergence": A venerable concept in search of a theory. *Complexity* **7**:18–30. DOI:10.1002/cplx.10043
12. Yuan, B., Zhang, J., Lyu, A. et al. (2024). Emergence and Causality in Complex Systems: A Survey of Causal Emergence and Related Quantitative Studies. *Entropy* **26**:108. DOI:10.3390/e26020108
13. Ernesto, E. (2012). Community structure of networks. In The Structure of Complex Networks (Oxford University Press)), pp. 187–213. DOI:10.1093/acprof:oso/9780199591756.003.0010
14. Barabási, A.-L. (2016). Network Science (Cambridge University Press)
15. Barzel, B. and Barabási, A.-L. (2013). Network link prediction by global silencing of indirect correlations. *Nat. Biotechnol.* **31**:720–725. DOI:10.1038/nbt.2601
16. Yang, Y., Han, L., Yuan, Y. et al. (2014). Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. *Nat. Commun.* **5**:3231. DOI:10.1038/ncomms4231
17. Castillo, E., Gutiérrez, J.M. and Hadi, A.S. (1997). Learning Bayesian Networks. In Expert Systems and Probabilistic Network Models, pp. 481–527. DOI:10.1007/978-1-4612-2270-5_11
18. Girvan, M. and Newman, M.E.J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci.* **99**:7821–7826. DOI:10.1073/pnas.122653799
19. Battiston, F., Cencetti, G., Iacopini, I. et al. (2020). Networks beyond pairwise interactions: Structure and dynamics. *Phys. Rep.* **874**:1–92. DOI:10.1016/j.physrep.2020.05.004
20. Gosak, M., Markovič, R., Dolenšek, J. et al. (2018). Network science of biological systems at different scales: A review. *Phys. Life Rev.* **24**:118–135. DOI:10.1016/j.plrev.2017.11.003
21. Yang, D., Jin, Y., He, X. et al. (2021). Inferring multilayer interactome networks shaping phenotypic plasticity and evolution. *Nat. Commun.* **12**:5304. DOI:10.1038/s41467-021-25086-5
22. Boccaletti, S., Latora, V., Moreno, Y. et al. (2006). Complex networks: Structure and dynamics. *Phys. Rep.* **424**:175–308. DOI:10.1016/j.physrep.2005.10.009
23. Boccaletti, S., Bianconi, G., Criado, R. et al. (2014). The structure and dynamics of multilayer networks. *Phys. Rep.* **544**:1–122. DOI:10.1016/j.physrep.2014.07.001
24. Chen, C., Jiang, L., Fu, G. et al. (2019). An omnidirectional visualization model of personalized gene regulatory networks. *NPJ Syst. Biol. Appl.* **5**:38. DOI:10.1038/s41540-019-0116-1
25. Sun, L., Dong, A., Griffin, C. et al. (2021). Statistical mechanics of clock gene networks underlying circadian rhythms. *Appl. Phys. Rev.* **8**:021313. DOI:10.1063/5.0029993
26. Wu, R. and Jiang, L. (2021). Recovering dynamic networks in big static datasets. *Phys. Rep.* **912**:1–57. DOI:10.1016/j.physrep.2021.01.003
27. Dong, A., Wu, S., Che, J. et al. (2023). idopNetwork: A network tool to dissect spatial community ecology. *Methods Ecol. Evol.* **14**:2272–2283. DOI:10.1111/2041-210x.14172
28. Wang, H., Ye, M., Fu, Y. et al. (2021). Modeling genome-wide by environment interactions through omnigenic interactome networks. *Cell Rep.* **35**:109114. DOI:10.1016/j.celrep.2021.109114
29. Grigor'yan, A.A., Lin, Y., Muranov, Y.V. et al. (2020). Path Complexes and their Homologies. *J. Math. Sci.* **248**:564–599. DOI:10.1007/s10958-020-04897-9
30. Wu, S., Liu, X., Dong, A. et al. (2023). The metabolomic physics of complex diseases. *Proc. Natl. Acad. Sci.* **120**:e2308496120. DOI:10.1073/pnas.2308496120
31. Gong, H., Wang, H., Wang, Y. et al. (2024). Topological change of soil microbiota networks for forest resilience under global warming. *Phys. Life Rev.* **50**:228–251. DOI:10.1016/j.plrev.2024.08.001
32. Gao, T.-T., Barzel, B. and Yan, G. (2024). Learning interpretable dynamics of stochastic complex systems from experimental data. *Nat. Commun.* **15**:6029. DOI:10.1038/s41467-024-50378-x
33. Holland, J.H. (1998). Emergence From Chaos to Order (OUP Oxford), pp. 1–15. DOI:10.1093/oso/9780198504092.001.0001
34. Raducha, T. and San Miguel, M. (2020). Emergence of complex structures from nonlinear interactions and noise in coevolving networks. *Sci. Rep.* **10**:15660. DOI:10.1038/s41598-020-72662-8
35. Holland, J.H. (1992). Adaptation in Natural and Artificial Systems (MIT press), p. 529. DOI:10.7551/mitpress/1090.001.0001
36. Rolls, E.T. and Deco, G. (2010). The Noisy BrainStochastic Dynamics as a Principle of Brain Function (Oxford University Press), pp. 1015–1016. DOI:10.1093/acprof:oso/9780199587865.001.0001
37. Mendonça, P.R.F., Vargas-Caballero, M., Erdélyi, F. et al. (2016). Stochastic and deterministic dynamics of intrinsically irregular firing in cortical inhibitory interneurons. *eLife* **5**:e16475. DOI:10.7554/eLife.16475
38. Jazwinski, A.H. (2007). Stochastic Processes and Filtering Theory (Analytical Mechanics Associates, Inc.)
39. Alan Bain, D.C. (2009). Fundamentals of Stochastic Filtering (Springer). DOI:10.1007/978-0-387-76896-0
40. Kalman, R.E. (1960). A New Approach to Linear Filtering and Prediction Problems. *J. Basic Engin.* **82**:35–45. DOI:10.1115/1.3662552
41. Kalman, R.E. and Bucy, R.S. (1961). New Results in Linear Filtering and Prediction Theory. *J. Basic Engin.* **83**:95–108. DOI:10.1115/1.3658902
42. Anderson, J.L. and Anderson, S.L. (1999). A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts. *Mon. Weather Rev.* **127**:2741–2758. DOI:10.1175/1520-0493(1999)127<2741:Amciot>2.0.Co;2
43. Stensrud, S.L.a.D.J. (2009). Ensemble Kalman filter. *IEEE Control Syst.* **29**:34–46. DOI:10.1109/mcs.2009.932225
44. Duncan, T.E. (1967). Probability Density for Diffusion Processes with Applications to Nonlinear Filtering Theory and Diffusion Theory (Stanford University)
45. Mortensen, R.E. (1966). Optimal Control of Continuous-Time Stochastic Systems (University of California)

46. Zakai, M. (1969). On the optimal filtering of diffusion processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **11**:230–243. DOI:10.1007/bf00536382
47. Yau, S.-T. and Yau, S.S.T. (2000). Real time solution of nonlinear filtering problem without memory I. *Math. Res. Lett.* **7**:671–693. DOI:10.4310/MRL.2000.v7.n6.a2
48. Yau, S.-T. and Yau, S.S.T. (2008). Real Time Solution of the Nonlinear Filtering Problem without Memory II. *SIAM J. Control Optim.* **47**:163–195. DOI:10.1137/050648353
49. Shi, J., Chen, X. and Yau, S.S.-T. (2021). A Novel Real-Time Filtering Method to General Nonlinear Filtering Problem Without Memory. *IEEE Access* **9**:119343–119352. DOI:10.1109/access.2021.3050572
50. Yueh, M.-H., Lin, W.-W. and Yau, S.-T. (2014). An efficient algorithm of Yau-Yau method for solving nonlinear filtering problems. *Commun. Inf. Syst.* **14**:111–134. DOI:10.4310/CIS.2014.v14.n2.a3
51. Liu, Z., Dong, F. and Ding, L. (2012). Numerical Results of Nonlinear Filtering Problem from Yau-Yau Method. *J. Comput.* **7**:971–976. DOI:10.4304/jcp.7.4.971-976
52. Sun, Z., Chen, X. and Yau, S.S.-T. (2024). On the Convergence Analysis of Yau-Yau Nonlinear Filtering Algorithm: from a Probabilistic Perspective. Preprint at *arXiv*. DOI:10.48550/arXiv.2405.06162
53. Kang, J., Salmon, A. and Yau, S.S.T. (2023). Log-Concave Posterior Densities Arising in Continuous Filtering and a Maximum A Posteriori Algorithm. *SIAM J. Control Optim.* **61**:2407–2424. DOI:10.1137/22m1508352
54. He, X., Jin, Y., Ye, M. et al. (2017). Bacterial Genetic Architecture of Ecological Interactions in Co-culture by GWAS-Taking Escherichia coli and Staphylococcus aureus as an Example. *Front. Microbiol.* **8**:2332. DOI:10.3389/fmicb.2017.02332
55. Palmer, T.N. (2019). Stochastic weather and climate models. *Nat. Rev. Phys.* **1**:463–471. DOI:10.1038/s42254-019-0062-2
56. Fatemi, M., Setoodeh, P. and Haykin, S. (2016). Observability of stochastic complex networks under the supervision of cognitive dynamic systems. *J. Complex Netw.* **5**:cnw021. DOI:10.1093/comnet/cnw021
57. Zhu, Y., Tang, Y.-H. and Kim, C. (2023). Learning stochastic dynamics with statistics-informed neural network. *J. Comput. Phys.* **474**:111819. DOI:10.1016/j.jcp.2022.111819
58. Casert, C., Tamblyn, I. and Whitelam, S. (2024). Learning stochastic dynamics and predicting emergent behavior using transformers. *Nat. Commun.* **15**:1875. DOI:10.1038/s41467-024-45629-w
59. Course, K. and Nair, P.B. (2023). State estimation of a physical system with unknown governing equations. *Nature* **622**:261–267. DOI:10.1038/s41586-023-06574-8
60. Chen, X., Sun, Z., Tao, Y. et al. (2025). A Uniform Framework of Yau–Yau Algorithm Based on Deep Learning With the Capability of Overcoming the Curse of Dimensionality. *IEEE Trans. Automat. Contr.* **70**:339–354. DOI:10.1109/tac.2024.3424628
61. Kim, B.-R., Zhang, L., Berg, A. et al. (2008). A Computational Approach to the Functional Clustering of Periodic Gene-Expression Profiles. *Genetics* **180**:821–834. DOI:10.1534/genetics.108.093690
62. Wang, Y., Xu, M., Wang, Z. et al. (2012). How to cluster gene expression dynamics in response to environmental signals. *Brief. Bioinform.* **13**:162–174. DOI:10.1093/bib/bbr032
63. Lambiotte, R., Rosvall, M. and Scholtes, I. (2019). From networks to optimal higher-order models of complex systems. *Nat. Phys.* **15**:313–320. DOI:10.1038/s41567-019-0459-y
64. Battiston, F., Amico, E., Barrat, A. et al. (2021). The physics of higher-order interactions in complex systems. *Nat. Phys.* **17**:1093–1098. DOI:10.1038/s41567-021-01371-4
65. Malizia, F., Corso, A., Gambuzza, L.V. et al. (2024). Reconstructing higher-order interactions in coupled dynamical systems. *Nat. Commun.* **15**:5184. DOI:10.1038/s41467-024-49278-x
66. Feng, L., Gong, H., Zhang, S. et al. (2024). Hypernetwork modeling and topology of high-order interactions for complex systems. *Proc. Natl. Acad. Sci.* **121**:e2412220121. DOI:10.1073/pnas.2412220121
67. Wang, Y., Xu, S., Wei, X. et al. (2025). Yau-YauAL: A computer tool for solving nonlinear filtering problems. *Data Anal. Topol.* **1**:77–84. DOI:10.4310/daat.250617012448