Stable reproducibility of turbulence dynamics by machine learning

Satoshi Matsumoto ,[1,*] Masanobu Inubushi ,[1,2,3] and Susumu Goto [1]
[1]Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan
[2]Department of Applied Mathematics, Tokyo University of Science, 1-3, Kagurazaka, Shinjuku, Tokyo 162-8601, Japan
[3]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, CB3 0WA, Cambridge, United Kingdom

(Received 22 May 2024; accepted 17 September 2024; published 11 October 2024)

We investigate the stability and accuracy of a machine-learning-based turbulence closure model. To this end, we construct a turbulence closure model for a shell model, which is a toy model of turbulence, based on the inference of sub-grid scale (SGS) variables using a recurrent neural network, and conduct an extensive parameter survey of the constructed model. The model stably and accurately reproduces the statistics of grid-scale variables when the cutoff wave number κ_c is higher than $0.2\eta^{-1}$, where η denotes the Kolmogorov length. This is because in this case, SGS variables are subordinate to grid-scale ones. On the other hand, when κ_c is lower than $0.2\eta^{-1}$, the model becomes stochastically unstable. However, an appropriate regularization in the inference step of SGS variables realizes a sufficiently long lifetime of the model.

DOI: 10.1103/PhysRevFluids.9.104601

I. INTRODUCTION

Since turbulence is ubiquitous, predicting its behavior is crucial in many systems. For a relatively compact system, we may conduct direct numerical simulations (DNS), where we numerically integrate the Navier-Stokes equation without any modeling. However, the number of grid points for DNS drastically increases with the Reynolds number. Hence, for applications with flow at high Reynolds numbers, we often employ large-eddy simulations (LES) to reduce the computational cost by resolving only large-scale fluid motion.

LES requires a turbulence closure model. More precisely, we must describe the dynamics of variables in length scales [i.e., grid-scales, (GS)] larger than a given length r_c or, equivalently, in the wave number range lower than the cutoff wave number κ_c ($=2\pi/r_c$) in a closed form. In other words, the impact of sub-grid scales (SGS) on GS must be characterized using only GS variables. According to Kolmogorov-Richardson phenomenology (Chap. 7 of Ref. [1]), larger-scale eddies generate smaller-scale ones through scale-by-scale energy transfer, i.e., energy cascade. Therefore, the smaller-scale dynamics are governed by the larger-scale ones. Simultaneously, the energy cascade makes the statistics in sufficiently small scales independent of the forcing type and boundary conditions. The dependence of smaller-scale flow on larger-scale ones and the universality in small scales provide the basis of turbulence models. Indeed, various turbulence models have been proposed relying on these properties and demonstrated significant success in numerous situations [2]. However, we must not forget that turbulence models are grounded on a closure theory, and the closure problem remains unresolved despite extensive studies since the pioneering studies by

*Contact author: s_matsumoto@fm.me.es.osaka-u.ac.jp

Heisenberg [3], Kraichnan [4,5], and many others [6]. Therefore, some assumptions are needed to deduce a turbulence model from the Navier-Stokes equation, and ensuring their validity is generally challenging. In contrast, based on turbulence properties, such as the smaller-scale-flow subordination to a large-scale one and the universality, machine learning (ML) can construct a closure with high generalization performance using only the given data without such assumptions. Furthermore, this advantage allows ML to be applied to cases where universality does not hold, such as low-Reynolds number turbulence or turbulence affected by boundary conditions. Indeed, many studies [7–10] have shown that ML is a useful tool for turbulence modeling.

Here, we emphasize the following: when we say that ML works, we mostly mean that it can work in some cases. In fact, it is known that ML-based turbulence models can become unstable without *ad hoc* stabilization [11–14]. For example, Maulik *et al.* [11] conducted turbulence modeling for decaying two-dimensional turbulence using a feed-forward neural network (FNN). More concretely, they inferred the SGS terms in the vorticity equations by FNN and time evolved the GS variables using the inferred values. Although FNN accurately inferred SGS terms, a stable LES was not realized without correcting the inferred SGS terms to prevent enstrophy back scatter. Similar results were reported by Guan *et al.* [12] and Ayapilla and Hattori [13] when performing ML-based turbulence modeling for two-dimensional turbulence. Miyazaki and Hattori [14] used FNN to infer the SGS stress for three-dimensional homogeneous isotropic turbulence. Although the inferred SGS stress correlates well with true values, their model becomes unstable without restraining energy back-scattering.

We also emphasize that it is not necessarily clear whether the cause of model instability is due to the nature of turbulence or ML. The main aim of the present study is to clarify this. However, conducting an extensive parameter survey on high-Reynolds number turbulence governed by the Navier-Stokes equation is impractical due to the enormous computational cost. Since, in the present study, we rely on the Kolmogorov similarity [15] to evaluate the accuracy of the constructed model, we must realize the DNS of turbulence with the clear similarity, but it is still challenging [16]. Therefore, we conduct turbulence modeling for a dynamical system (i.e., the so-called shell model), which exhibits the Kolmogorov similarity despite small degrees of freedom, and evaluate the stability and accuracy of an ML-based turbulence model by varying the Reynolds number, cutoff wave number, and ML hyper-parameters. Here, we refer to an important previous study by Ortali *et al.* [17], who conducted turbulence modeling for the Sabra shell model [18]. They simulated the relatively short GS dynamics by inferring the SGS variables using long short-term memory (LSTM) [19] to conclude that their model accurately reproduced the turbulence statistics. Then, they claimed to prove the feasibility of ML-based turbulence models. While their research focused on the dynamics for the turnover time T_L of the largest eddies, exploring whether the dynamics could be correctly simulated for longer than T_L is essential to discuss the stability of ML-based turbulence models. We therefore reconsider their research focusing on the long (say, $6000T_L$) dynamics and discuss the stability and accuracy of the ML-based turbulence model. Incidentally, we construct the turbulence model for a shell model using reservoir computing (RC) [20–22] instead of LSTM. Since RC requires lower training costs while achieving effective inference, it suits our extensive parameter survey.

II. FRAMEWORK

A. Target system: sparse-coupling shell model

Shell models are low-dimensional dynamical systems that mimic energy cascade in wave number space while reducing the degrees of freedom by representing modes in a wave number shell by a few variables and assuming local nonlinear interactions between shells [18,23–25]. For example, in the Sabra shell model [18], each wave number shell contains only a single variable u_ℓ ($\in \mathbb{C}$), which corresponds to the Fourier coefficients of the velocity at wave number k_ℓ, and u_ℓ is coupled only with $u_{\ell\pm2}$ and $u_{\ell\pm1}$ as

$$\left(\frac{\mathrm{d}}{\mathrm{d}t} + \nu k_\ell^2\right)u_\ell = i(ak_{\ell+1}u_{\ell+2}u_{\ell+1}^* + bk_\ell u_{\ell+1}u_{\ell-1}^* - ck_{\ell-1}u_{\ell-1}u_{\ell-2}) + f_\ell. \tag{1}$$

Here, ν and f_ℓ are the kinematic viscosity and forcing, respectively. The constant coefficients a, b, and c satisfy $a + b + c = 0$ so that energy conserves when $\nu = 0$ and $f_\ell = 0$. It is remarkable that the Sabra shell model (1) reproduces important properties of turbulence, such as not only the $-5/3$ power-law energy spectrum but also its correction due to intermittency effects. However, it does not strictly exhibit the Kolmogorov similarity. In fact, the probability density functions (PDF) and two-time autocorrelation functions normalized according to the Kolmogorov similarity depend on wave numbers (Appendix A). This property makes it difficult to evaluate the statistical difference between the constructed model and the true system.

The violation of the Kolmogorov similarity of the Sabra shell model (1) is due to dense nonlinear coupling. Recall that the Fourier coefficients $u_i(\boldsymbol{k}, t)$ of the velocity field in a periodic cube with side L obeys the Navier–Stokes equation,

$$\frac{\partial}{\partial t} u_i(\boldsymbol{k}, t) = \sum_{j=1}^{3} \sum_{m=1}^{3} M_{ijm}(\boldsymbol{k}) \sum_{\substack{\boldsymbol{p},\, \boldsymbol{q} \\ (\boldsymbol{p}+\boldsymbol{q}+\boldsymbol{k}=\boldsymbol{0})}} u_j(-\boldsymbol{p}, t) u_m(-\boldsymbol{q}, t) - \nu |\boldsymbol{k}|^2 u_i(\boldsymbol{k}, t), \tag{2}$$

with

$$M_{ijm}(\boldsymbol{k}) = -\frac{\mathrm{i}}{2} \left(\frac{2\pi}{L} \right)^3 \left(k_m \delta_{ij} + k_j \delta_{im} - \frac{2 k_i k_j k_m}{k^2} \right). \tag{3}$$

In Eq. (2), $\boldsymbol{u}(\boldsymbol{k}, t)$ interacts with $\boldsymbol{u}(\boldsymbol{p}, t)$ only through $\boldsymbol{u}(-\boldsymbol{k} - \boldsymbol{p}, t)$. Since the number of Fourier modes is huge in turbulence, the nonlinear coupling is very sparse. In contrast, the nonlinear coupling in the Sabra shell model (1) is dense because the degrees of freedom are too reduced. When the nonlinear coupling is dense, variables have long-term autocorrelation [26], resulting in the violation of the Kolmogorov similarity [Fig. 10(c) in Appendix A]. Therefore, in the present study, we employ the sparse-coupling shell model (SSM) [27], which exhibits clear Kolmogorov similarity [see Fig. 1(c) below].

SSM increases the degrees N of freedom in each shell so that the nonlinear coupling becomes sparse. The temporal evolution of the variables $\{X_i^{(\ell)}\ (\in \mathbb{R})\}_{1 \leqslant i \leqslant N}$ in the ℓ-th shell ($\ell = 1, \ldots, L_{\max}$) corresponding to a wave number $k_\ell = \exp(\ell/2)$ follows

$$\left(\frac{\mathrm{d}}{\mathrm{d}t} + \nu k_\ell^2 \right) X_i^{(\ell)} = 2 k_\ell \sum_{j=1}^{N} \sum_{m=1}^{N} c_{ijm} X_j^{(\ell)} X_m^{(\ell+1)} + k_{\ell-1} \sum_{j=1}^{N} \sum_{m=1}^{N} c_{ijm} X_j^{(\ell-1)} X_m^{(\ell-1)} + f_i^{(\ell)}, \tag{4}$$

where $f_i^{(\ell)}\ (= X_i^{(\ell)} \delta_{\ell 1})$ is the forcing. We define the Reynolds number as $\mathrm{Re} = 1/\nu$. The nonlinear coupling coefficient tensor c_{ijm} is sparse such that any pair of variables have a single direct interaction at the most and satisfies the following four conditions: $c_{ijm} = c_{imj}$, $c_{ijm} + c_{jmi} + c_{mij} = 0$, $c_{ijj} = 0$, and $c_{ijm} = c_{\mathrm{mod}\{i+m', N\} \mathrm{mod}\{j+m', N\} \mathrm{mod}\{m+m', N\}}$, where $\mathrm{mod}\{i, N\}$ is the remainder of i divided by N.

Here, we demonstrate that the statistics of SSM obey clear Kolmogorov similarity. We show, for example, results with $N = 10$, $\mathrm{Re} = 10^3$, and $L_{\max} = 14$. Figure 1(a) shows the energy spectrum

$$E(k_\ell) = \frac{1}{2k_\ell} \sum_{i=1}^{N} \left\langle \left(X_i^{(\ell)}(t) \right)^2 \right\rangle_t, \tag{5}$$

where $\langle \cdot \rangle_t$ represents the time average. We can confirm in this figure that the energy spectrum obeys the $-5/3$ power law, $E(k_\ell) \propto k_\ell^{-\frac{5}{3}}$, in the inertial range ($3 \lesssim \ell \lesssim 5$). Figure 1(b) shows PDF $\mathcal{P}(X_i^{(\ell)})$ of the shell variables in the inertial range ($\ell = 3, 4, 5$). They converge to a normal distribution when normalized by the energy dissipation rate $\epsilon = 2\nu \sum_{\ell=1}^{L_{\max}} k_\ell^3 E(k_\ell)$ and wave number k_ℓ. Figure 1(c) shows two-time autocorrelation functions,

$$C_i^{(\ell)}(\tau) = \left\langle X_i^{(\ell)}(t + \tau) X_i^{(\ell)}(t) \right\rangle_t, \tag{6}$$

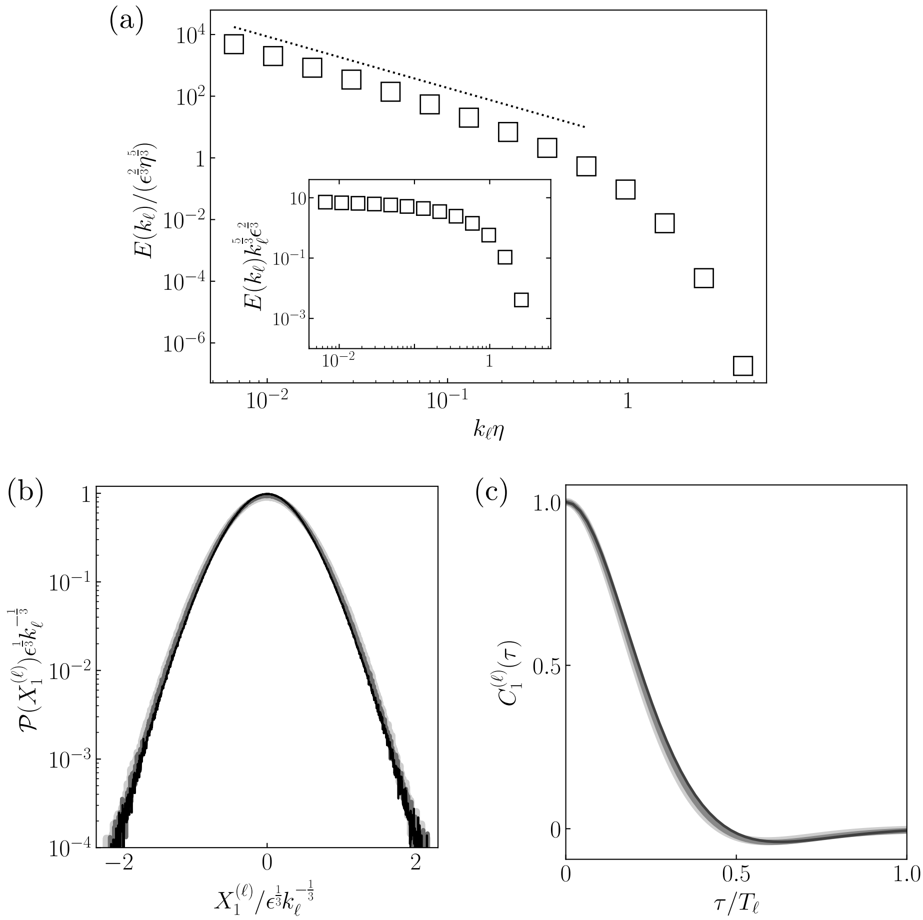

FIG. 1. Statistics of SSM for Re $= 10^3$. (a) Energy spectrum. The dotted line indicates the $-5/3$ power law and the inset shows the compensated spectrum. (b) PDF of the shell variables in the inertial range. From the lighter (and thicker) to darker (and thinner) lines, $\mathcal{P}(X_1^{(3)})$, $\mathcal{P}(X_1^{(4)})$, and $\mathcal{P}(X_1^{(5)})$. (c) Two-time autocorrelation functions of the shell variables in the inertial range. From the lighter (and thicker) to darker (and thinner) lines, $C_1^{(3)}$, $C_1^{(4)}$, and $C_1^{(5)}$. We use the energy dissipation rate ϵ, Kolmogorov length η, and characteristic timescale $T_\ell = \epsilon^{-\frac{1}{3}} k_\ell^{-\frac{2}{3}}$ for the normalization.

of the shell variables in the inertial range ($\ell = 3, 4, 5$). They collapse onto a curve when normalized by the characteristic timescale $T_\ell = \epsilon^{-\frac{1}{3}} k_\ell^{-\frac{2}{3}}$ of each shell. Thus, SSM exhibits clear Kolmogorov similarity, although it does not show any intermittency effect. The perfect collapse of the PDF and two-time autocorrelation functions in the inertial range [Figs. 2(b) and 2(c)] is suitable when we evaluate the error of the constructed ML-based model.

B. Modeling method

In SSM (4), the variables $\boldsymbol{X}^{(\ell)}$ interact directly with $\boldsymbol{X}^{(\ell\pm1)}$ in adjacent shells. Therefore, we need information of $\boldsymbol{X}^{(\ell_c+1)}$ to integrate the governing equations for the variables in the wave number range lower than the cutoff wave number $\kappa_c = k_{\ell_c}$. Here, ℓ_c denotes the cutoff shell number. Since energy transfers from lower to higher wave numbers, the states of $\boldsymbol{X}^{(\ell_c+1)}$ are mostly affected by

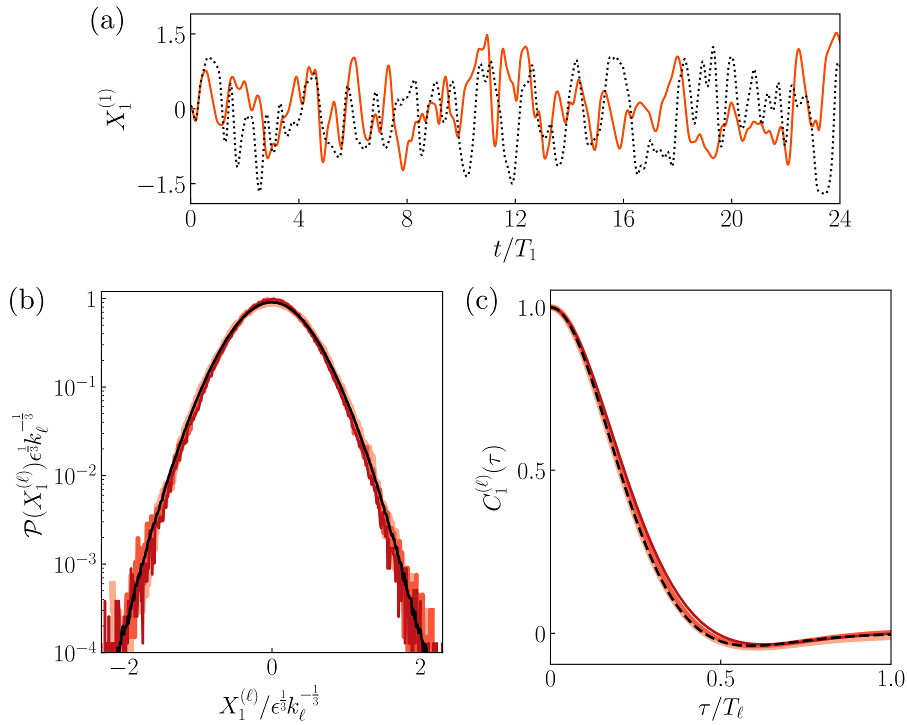

FIG. 2. Modeling results for $\mathrm{Re} = 10^3$ with a relatively high cutoff wave number ($\kappa_c = k_9 \approx 0.36\eta^{-1}$). (a) Temporal evolution of $X_1^{(1)}$. The red solid and black dotted lines are the results of RCTM and truth, respectively. (b) PDF of the shell variables in the inertial range. From the red lighter (and thicker) to red darker (and thinner) lines, $\mathcal{P}(X_1^{(3)})$, $\mathcal{P}(X_1^{(4)})$, and $\mathcal{P}(X_1^{(5)})$ results of RCTM; the black line is the truth. (c) Two-time autocorrelation functions of the shell variables $X_1^{(3)}$, $X_1^{(4)}$, and $X_1^{(5)}$ in the inertial range. From the red lighter (and thicker) to red darker (and thinner) lines, results $C_1^{(3)}$, $C_1^{(4)}$, and $C_1^{(5)}$ of the RCTM; the black dashed line is the truth. We use the characteristic timescale T_ℓ, energy dissipation rate ϵ, and wave number k_ℓ for normalization.

$\boldsymbol{X}^{(\ell_c)}$. Hence, we construct the ML-based turbulence model for SSM by iterating the following two steps:

(i) we infer the SGS variables $\boldsymbol{X}^{(\ell_c+1)}$ by RC with $\boldsymbol{X}^{(\ell_c)}$ as inputs, and

(ii) integrate Eq. (4) using the inferred values $\widehat{\boldsymbol{X}}^{(\ell_c+1)}$.

We standardize the inputs $\boldsymbol{X}^{(\ell_c)}$ using the mean and standard deviation evaluated from the training data. In this paper, we call the model constructed by the above procedure the reservoir computing-based turbulence model (RCTM).

The training and test data for RCTM are generated by integrating Eq. (4) using the second-order Adams-Bashforth method. In the present study, we examine three cases with the Reynolds number Re, resolution $L_{\max}$, and time increment Δt as $(\mathrm{Re}, L_{\max}, \Delta t) = (10^2, 10, 10^{-3})$, $(10^3, 14, 5 \times 10^{-4})$, and $(10^4, 17, 10^{-4})$. In all cases, we set the degrees N of freedom in each shell to 10.

We show the training and test conditions in Table I. We also show in Table I the turnover time $T_L = \langle k_1 \|\boldsymbol{X}^{(1)}\|\rangle_t^{-1}$, where $\|\cdot\|$ denotes the L_2 norm, of the largest eddies. Since we focus on the long-time stability of the constructed model, we set the test time T_{test} much longer than T_L ($T_{\text{test}} \approx 6000T_L$). The details of RC are described in Appendix B. Our framework uses six hyper-parameters: the number N_r of reservoir nodes, the spectrum radius ρ of reservoir weights, the magnitude σ of input weights, the bias magnitude ξ, the leakage rate α, and the regularization parameter β. We determine these parameters as follows. First, we fix $N_r = 500$ and $\rho = 0.90$. Then, σ, ξ, and α are

TABLE I. Training and test conditions: Re, T_L, $\Delta\tau$, $T_{\rm train}$, $Q_{\rm train}$, κ_c, $\Delta T_{\rm test}$, $T_{\rm test}$, and $Q_{\rm test}$ are the Reynolds number, turnover time of the largest eddies, sampling time interval, training data length, number of the training data, cutoff wave number, test time increment, test time, and number of the test data, respectively.

Re	T_L	$\Delta\tau$	$T_{\rm train}$	$Q_{\rm train}$	κ_c	$\Delta T_{\rm test}$	$T_{\rm test}$	$Q_{\rm test}$
10^2	0.39	10^{-3}	2×10^3	2×10^6	$k_1,\dots,k_8$	10^{-3}	2×10^3	2×10^6
10^3	0.32	5×10^{-4}	2×10^3	4×10^6	$k_1,\dots,k_{12}$	5×10^{-4}	2×10^3	4×10^6
10^4	0.31	10^{-4}	2×10^3	2×10^7	$k_1,\dots,k_{15}$	10^{-4}	2×10^3	2×10^7

determined to minimize the maximum element of the inference error vector,

$$e_i^{(\ell_c)} = \frac{1}{Q_{\rm test}} \sum_{q=1}^{Q_{\rm test}} \left[\widehat{X}_i^{(\ell_c+1)}(q\Delta\tau) - X_i^{(\ell_c+1)}(q\Delta\tau)\right]^2, \tag{7}$$

where $Q_{\rm test}$ and $\Delta\tau$ denote the number of the test data and sampling time interval, respectively. We discuss in Sec. III A modeling results with $\beta = 0$, and in Secs. III B and III C results with finite β between 10^{-9} and 10^{-3}. We also show in Appendix C the inference results of the SGS variables, that is the first step (i) of the modeling method.

Even if the RC weights are adequately optimized, the inferred values $\widehat{\boldsymbol{X}}^{(\ell_c+1)}$ for the same input signal differ depending on the random numbers that determine the RC weights. Therefore, we consider the results of 500 trials with different random number seeds below.

III. RESULTS

A. Stability dependence on cutoff wave number without L_2 regularization

In this subsection, we present modeling results and demonstrate that the stability of RCTM depends on the cutoff wave number κ_c. First, as an example, we show the modeling results for $\mathrm{Re} = 10^3$ when the cutoff wave number is relatively high ($\kappa_c = k_9 \approx 0.36\eta^{-1}$). Figure 2(a) shows the temporal evolution of $X_1^{(1)}$ in this case. Although the model result deviates from the true time series after an initial period ($\lesssim T_1$) due to the chaotic nature of the sensitivity to initial conditions, the amplitudes and timescales of fluctuations of the modeled time series agree well with the truth. Similar results are obtained for the temporal evolution of other GS variables (figures are omitted). To demonstrate the evidence of the model's accuracy, we show in Figs. 2(b) and 2(c) PDF and two-time autocorrelation functions of the shell variables $X_1^{(3)}$, $X_1^{(4)}$, and $X_1^{(5)}$ in the inertial range, respectively. Here, we regard the statistics evaluated from a time series of length $100T_{\rm test}$ obtained by direct simulation of Eq. (4) as the truth. In Fig. 2(b), although the tails of the PDF of RCTM fluctuate due to the lack of data, the PDF well coincide with the true ones. The two-time autocorrelation functions are also in excellent agreement with the truth. These results imply that RCTM can accurately reproduce the one- and two-time statistics in the inertial range.

To quantify the reproducibility of RCTM for PDF, we define the Jensen-Shannon divergence (JSD) [28], which measures the distance between the true PDF $\mathcal{P}_{\rm T}$ and the RCTM PDF $\mathcal{P}_{\rm R}$, as

$$\mathcal{J} = \frac{\mathcal{D}(\mathcal{P}_{\rm T}\|\mathcal{M}) + \mathcal{D}(\mathcal{P}_{\rm R}\|\mathcal{M})}{2} \quad \text{with} \quad \mathcal{M} = \frac{\mathcal{P}_{\rm T}+\mathcal{P}_{\rm R}}{2}, \tag{8}$$

where $\mathcal{D}$ is the Kullback-Leibler divergence [29],

$$\mathcal{D}(\mathcal{P}_1\|\mathcal{P}_2) = \int_{-\infty}^{\infty} \mathcal{P}_1(x)\log\left(\frac{\mathcal{P}_1(x)}{\mathcal{P}_2(x)}\right) \mathrm{d}x. \tag{9}$$

We evaluate the mean JSD in each shell as $\mathcal{J}_\ell = \sum_{i=1}^N \mathcal{J}_i^{(\ell)}/N$, where $\mathcal{J}_i^{(\ell)}$ is the JSD of $\mathcal{P}(X_i^{(\ell)})$. Figure 3(a) shows $\mathcal{J}_\ell$ (red squares) for $\mathrm{Re} = 10^3$ and $\kappa_c \approx 0.36\eta^{-1}$. In this figure, we also show the

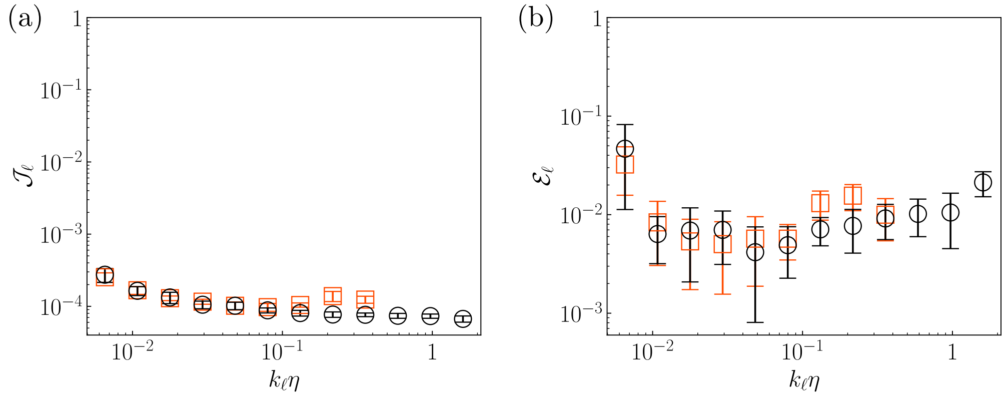

FIG. 3. Quantification of the modeling accuracy for $\mathrm{Re} = 10^3$ and $\kappa_c = k_9 \approx 0.36\eta^{-1}$. (a) Mean JSD of PDF and (b) mean relative error of the two-time autocorrelation function. The red squares are the model results. The black circles are the lower limits of $\mathcal{J}_\ell$ and $\mathcal{E}_\ell$ for data of length T_{test}, which are evaluated by the two directly simulated time series of length T_{test} and a sufficiently long length ($100T_{\mathrm{test}}$). Error bars are standard deviations within each shell. Here, k_ℓ is normalized by the Kolmogorov length η.

JSD (black circles) between the PDF evaluated from time series data of the length $100T_{\mathrm{test}}$ (i.e., the truth) and T_{test}, both obtained by direct simulation of Eq. (4). This value quantifies the JSD due to the different data lengths and it gives the lower limit of the JSD evaluated by the data of length T_{test}. Therefore, even if RCTM ideally reproduces the true dynamics of SSM, the JSD for such RCTM cannot be smaller than this value. Although $\mathcal{J}_\ell$ for $\ell \approx \ell_c$ is slightly larger than the value, $\mathcal{J}_\ell$ is almost identical to it in all other shells. Therefore, we conclude that RCTM reproduces the one-time statistics of GS variables correctly.

Next, we quantify the reproducibility of RCTM for the two-time autocorrelation functions with the average relative error,

$$\mathcal{E}_\ell = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\mathcal{I}_{\mathrm{T}i}^{(\ell)} - \mathcal{I}_{\mathrm{R}i}^{(\ell)}\right|}{\mathcal{I}_{\mathrm{T}i}^{(\ell)}}, \tag{10}$$

of integral time,

$$\mathcal{I}_i^{(\ell)} = \int_0^{T_\epsilon} \frac{C_i^{(\ell)}(t)}{C_i^{(\ell)}(0)}\,\mathrm{d}t, \tag{11}$$

where $\mathcal{I}_{\mathrm{T}}$ and $\mathcal{I}_{\mathrm{R}}$ denote the true value and that with RCTM, respectively, and T_ϵ is the first zero-crossing time of $C_i^{(\ell)}(t)$. Figure 3(b) shows $\mathcal{E}_\ell$ (red squares) for $\mathrm{Re} = 10^3$ and $\kappa_c \approx 0.36\eta^{-1}$ along with the lower limit (black circles) of $\mathcal{E}_\ell$ for data of length T_{test}, which is evaluated as the relative error in the integral time of the autocorrelation functions of the above-mentioned two-time series. Although $\mathcal{E}_\ell$ near ℓ_c is slightly larger than the lower limit, $\mathcal{E}_\ell$ is as small as it is in all other shells. We therefore conclude that RCTM can also reproduce the two-time statistics of GS variables correctly.

Results shown in Figs. 2 and 3 imply that ML can construct an accurate model for high-Reynolds number turbulence when the cutoff wave number is relatively high (i.e., $\kappa_c \approx 0.36\eta^{-1}$). However, for lower κ_c, the temporal evolution of GS variables can diverge before T_{test}. Hereafter, when such divergence occurs, we regard RCTM as unstable. To quantify the stability of RCTM, we count the number N_{F} of cases where the temporal evolution diverges before T_{test}, and define the modeling failure probability P_{F} using N_{F} divided by 500, the total number of trials, i.e., $P_{\mathrm{F}} = N_{\mathrm{F}}/500$. Figure 4 shows the dependence of P_{F} on the cutoff wave number κ_c for the three values of the Reynolds number: $\mathrm{Re} = 10^2$, 10^3, and 10^4. Irrespective of Re, when $\kappa_c \gtrsim 0.2\eta^{-1}$, P_{F} is almost zero, which

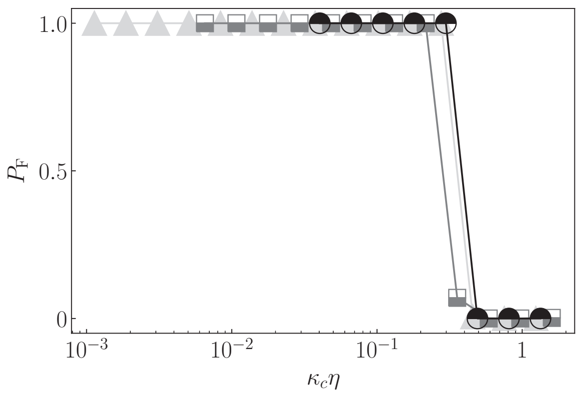

FIG. 4. Dependence of the modeling failure probability P_{F} on the cutoff wave number κ_c for $\beta = 0$. The black circles, gray squares, and light gray triangles are the modeling failure probability for Re $= 10^2$, 10^3, and 10^4, respectively. Here, κ_c is normalized by the Kolmogorov length η.

implies that RCTM stably simulates the dynamics of GS variables up to T_{test} ($\approx 6000 T_L$) in these cases. Recall that for $\kappa_c \gtrsim 0.2\eta^{-1}$, PDF and two-time autocorrelation functions agree well with the true statistics (Figs. 2 and 3). In contrast, $P_{\mathrm{F}} = 1$ for $\kappa_c \lesssim 0.2\eta^{-1}$, meaning that the temporal evolution of GS variables diverges before T_{test} in all trials. These results imply that the stability of RCTM changes at a critical wave number $k^* \approx 0.2\eta^{-1}$.

Previous studies [30–34] have already discussed the importance of the critical wave number k^* as a metric that characterizes the subordination of small- to large-scale dynamics in turbulence. For example, data assimilation studies of turbulence [30–32] examined the possibility of reconstructing the dynamics in the wave number range higher than k_a using data from the lower wave number range. These studies demonstrated that the success or failure of reconstructions depends on whether k_a is higher or lower than k^*. Inubushi *et al.* [33] pointed out that k^* does not depend on data assimilation schemes and used the transverse Lyapunov exponents to show that this is due to the nature of the Navier–Stokes equation. Furthermore, Yoneda *et al.* [34] investigated the energy transfer in turbulence in a periodic cube and showed that eddies smaller than a scale $r^* \sim 1/k^*$ do not transfer energy to further smaller eddies. The present results (Fig. 4) and these known facts are consistent, implying that ML-based turbulence models are stable when $\kappa_c \gtrsim 0.2\eta^{-1}$ regardless of ML methods.

B. Stabilization by L_2 regularization

The previous subsection has shown that RCTM can stably simulate GS dynamics when the cutoff wave number κ_c is higher than $0.2\eta^{-1}$, while it becomes unstable when κ_c is lower than $0.2\eta^{-1}$. In this subsection, we consider a method to stabilize RCTM with low cutoff wave numbers ($\kappa_c \lesssim 0.2\eta^{-1}$).

The instability of ML-based models was also reported in modeling studies [35–41] of nonlinear dynamical systems, such as the logistic map and Kuramoto-Sivasinsky equation and stabilization methods such as regularizations and applying noise to the training data were proposed [35,39–41]. However, RCTM cannot be significantly stabilized even though white noise with a standard deviation of 1 to 10% of that of the input is added to the training data. Therefore, we stabilize RCTM by introducing L_2 regularization in step (i) of the procedure described in Sec. II B.

1. Model stability with L_2 regularization

Figure 5 shows the dependence of the failure probability P_{F} for Re $= 10^3$ on the regularization parameter β. Here, we set the test time $T_{\mathrm{test}} = 2000 \approx 6000 T_L$ as in the previous subsection. The lighter meshes indicate smaller P_{F} and the white ones represent $P_{\mathrm{F}} = 0$. When $\kappa_c \gtrsim 0.2\eta^{-1}$, $P_{\mathrm{F}} \approx 0$

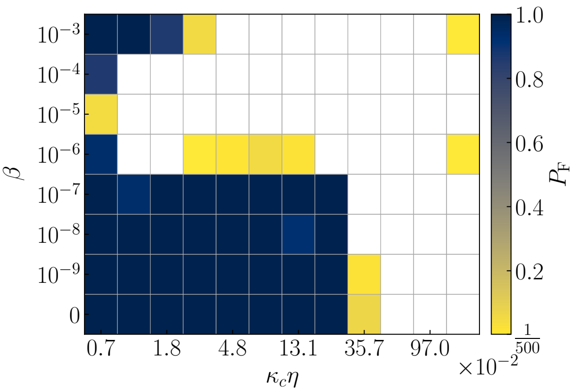

FIG. 5. Modeling failure probability P_F for Re $= 10^{-3}$ and $T_\mathrm{test} \approx 6000 T_L$. The lighter meshes indicate a smaller modeling failure probability and the white meshes represent $P_\mathrm{F} = 0$. The bottom row corresponds to the gray line in Fig. 4. Here, κ_c is normalized by the Kolmogorov length η.

regardless of β. This implies that RCTM can stably simulate the dynamics of GS variables up to T_test for $\kappa_c \gtrsim 0.2\eta^{-1}$ with or without L_2 regularization. In contrast, when $\kappa_c \lesssim 0.2\eta^{-1}$, the choice of β becomes important. Looking at the case for $\kappa_c \lesssim 0.02\eta^{-1}$, we notice that when β is extremely small (i.e., the L_2 regularization effect is too weak) or β is extremely large, P_F becomes larger. For $0.02\eta^{-1} \lesssim \kappa_c \lesssim 0.2\eta^{-1}$, P_F becomes smaller when β is larger than 10^{-6}. Moreover, P_F is almost zero when $\beta \approx 10^{-5}$ for all cutoff wave numbers. Thus, with an appropriate regularization, i.e., $\beta \approx 10^{-5}$ in this case, RCTM can stably simulate GS dynamics up to T_test irrespective of κ_c.

2. Reproducibility of statistics with L_2 regularization

We evaluate the modeling accuracy for the most stable case with $\beta = 10^{-5}$. Figure 6(a) shows the average $\langle \mathcal{J}_\ell \rangle_\mathrm{S}$ of JSD over 500 trials. Here, $\langle \cdot \rangle_\mathrm{S}$ denotes the ensemble average excluding the failed trials. The horizontal and vertical axes represent κ_c and the wave number of the evaluated variable, respectively. In addition, the shell numbers are provided in parentheses on the vertical axis. Lighter meshes indicate smaller $\langle \mathcal{J}_\ell \rangle_\mathrm{S}$, i.e., RCTM more accurately reproduces the PDF. Note that the black meshes correspond to the wave numbers higher than κ_c, which are out of the target of RCTM. We first notice that the meshes for $\ell = \ell_c$ are darker. This is because the RC inference errors directly impact the temporal evolution of $\boldsymbol{X}^{(\ell_c)}$ through Eq. (4). Moreover, for any κ_c, $\langle \mathcal{J}_\ell \rangle_\mathrm{S}$ for $\ell = 1$ is slightly larger than 10^{-3} due to the direct influence of the forcing. Except for these cases, $\langle \mathcal{J}_\ell \rangle_\mathrm{S}$ is as small as 10^{-3}. This result implies that RCTM correctly reproduces the one-time statistics of GS variables, even though introducing the L_2 regularization degrades the inference accuracy of SGS variables (Appendix C).

Next, we show the conditional average $\langle \mathcal{E}_\ell \rangle_\mathrm{S}$ over 500 trials in Fig. 6(b). As in Fig. 6(a), the lighter meshes indicate that RCTM accurately reproduces the two-time autocorrelation functions. The dark range is wider compared to Fig. 6(a). More concretely, $\langle \mathcal{E}_\ell \rangle_\mathrm{S}$ for $\ell = \ell_c - 1$ as well as $\ell = \ell_c$ is large due to the RC inference error. In addition, similar to $\langle \mathcal{J}_\ell \rangle_\mathrm{S}$, $\langle \mathcal{E}_\ell \rangle_\mathrm{S}$ for $\ell = 1$ is larger due to the direct influence of the forcing. However, except in these regions, $\langle \mathcal{E}_\ell \rangle_\mathrm{S} \lesssim 0.01$. This implies that RCTM also reproduces the two-time statistics of GS variables accurately even when κ_c is smaller than $0.2\eta^{-1}$, if we introduce an appropriate L_2 regularization.

We also investigate the dependence of the RCTM accuracy on the Reynolds number. Figure 7 shows $\langle \mathcal{J}_1 \rangle_\mathrm{S}$ and $\langle \mathcal{E}_1 \rangle_\mathrm{S}$ for Re $= 10^2$, 10^3, and 10^4. Here, we choose an appropriate β that minimizes P_F for each Re. More concretely, we set $\beta = 4 \times 10^{-6}$, 10^{-5}, and 6×10^{-5} for Re $= 10^2$, 10^3, and 10^4, respectively. We confirm in Fig. 7 that $\langle \mathcal{J}_1 \rangle_\mathrm{S}$ and $\langle \mathcal{E}_1 \rangle_\mathrm{S}$, as functions of κ_c, are independent of Re when κ_c is normalized by the forcing wave number κ_f and that they are sufficiently small for

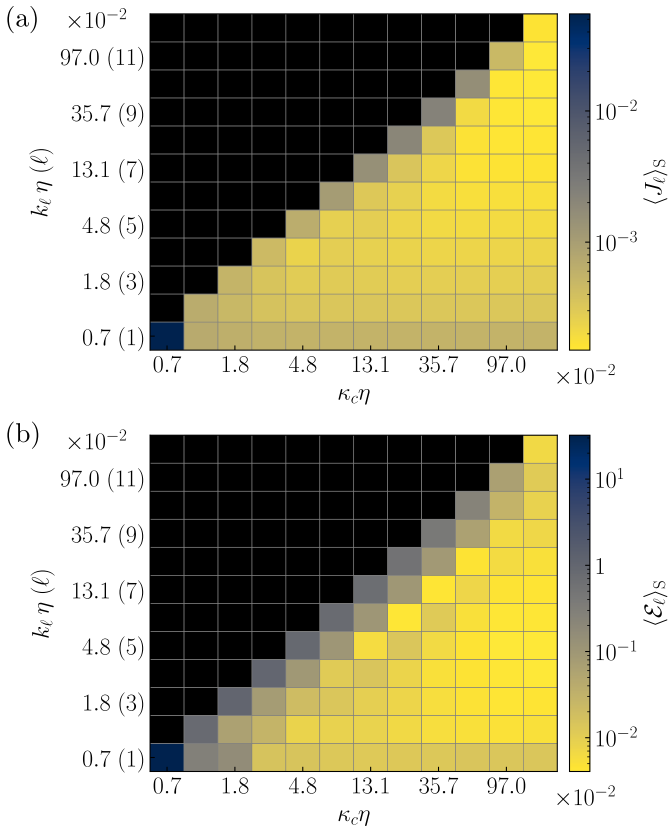

FIG. 6. (a) Mean JSD and (b) mean relative error of the two-time autocorrelation function for $\mathrm{Re} = 10^3$ and $\beta = 10^{-5}$. The lighter meshes indicate a small mean JSD or relative error; the black meshes represent the SGS range. The shell numbers are given in parentheses on the vertical axis. Here, κ_c is normalized by the Kolmogorov length η.

$\kappa_c \gtrsim \kappa_f$ and $\kappa_c \gtrsim 4\kappa_f$, respectively. These results imply that with an appropriate L_2 regularization, ML can construct stable and accurate turbulence models for high-Reynolds number turbulence by setting κ_c sufficiently higher than κ_f even when $\kappa_c \lesssim 0.2\eta^{-1}$.

C. Survival time analysis

The previous subsection has demonstrated that RCTM with an appropriate L_2 regularization can correctly give GS dynamics and reproduce its statistics. However, the test time has been fixed at $T_{\text{test}} \approx 6000 T_L$. If RCTM is unstable stochastically, the test will fail for longer T_{test}. Therefore, we perform a survival time analysis for RCTM. More concretely, we measure time t_S until the temporal evolution diverges and evaluate the survival function $S_{\ell_c}(t_S)$ for each cutoff wave number. Here, $S_{\ell_c}(t_S)$ gives the probability that RCTM stably simulates GS dynamics beyond t_S. Figures 8(a) and 8(b) show the survival functions $S_{\ell_c}(t_S)$ for $\kappa_c = k_1 \approx 0.007\eta^{-1}$ and $k_4 \approx 0.03\eta^{-1}$, respectively, when varying β from 10^{-7} to 10^{-3}. For visibility, we separately show the results for $10^{-7} \leqslant \beta \leqslant 10^{-5}$ in Figs. 8(a1) and 8(b1) and for $10^{-5} < \beta \leqslant 10^{-3}$ in Figs. 8(a2) and 8(b2). In each panel, the darker (thinner) lines indicate larger β, and the black dotted lines are the fitting with exponential functions. Here, we do not plot the survival function for $10^{-6} < \beta < 10^{-3}$ in Fig. 8(b) for $\kappa_c = k_4$ because there is no case of the diverging GS variables in this range of β. Regardless of β and κ_c, the

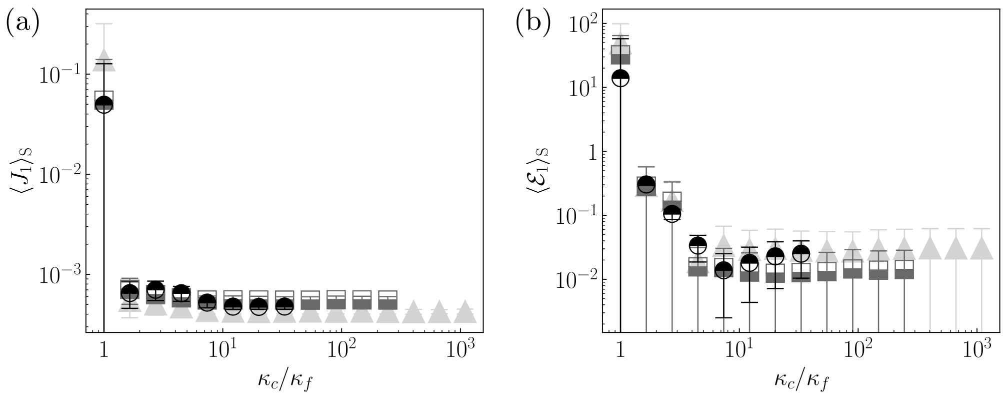

FIG. 7. Dependence of RCTM accuracy on Reynolds number: (a) mean JSD and (b) relative error of the two-time autocorrelation function. The black circles, gray squares, and light gray triangles are the results for Re $= 10^2$, 10^3, and 10^4, respectively. We choose β that minimizes the modeling failure probability P_F for each Re. More concretely, we set $\beta = 4 \times 10^{-6}$, 10^{-5}, and 6×10^{-5} for Re $= 10^2$, 10^3, and 10^4, respectively. Error bars are the standard deviation over the trials, excluding the failed trials. Here, κ_c is normalized by the forcing wave number κ_f.

survival function decays exponentially, implying that RCTM is unstable even with finite β. More concretely, in both cases of κ_c, for $\beta \approx 10^{-7}$, $S_{\ell_c}(t_S)$ decays exponentially after a sharp decrease at $t_S \approx 0$; for $\beta \approx 10^{-6}$, $S_{\ell_c}(t_S)$ decays exponentially from $t_S = 0$. For $\beta \approx 10^{-5}$ and $\kappa_c = k_1$ or $\beta = 10^{-3}$ and $\kappa_c = k_4$, $S_{\ell_c}(t_S)$ remains almost 1, but it agrees with the exponential function. For $10^{-4} \lesssim \beta \leqslant 10^{-3}$ and $\kappa_c = k_1$, while $S_{\ell_c}(t_S)$ decreases very slowly for $t_S \lesssim 70T_L$ and deviates from the fitting with exponential functions for $t_S \gtrsim 150T_L$, it decays according to an exponential function for $70T_L \lesssim t_S \lesssim 150T_L$. Thus, we may conclude that RCTM has finite lifetime even with the L_2 regularization.

Although the above result may negate the feasibility of ML-based turbulence models, when we define the lifetime $\mathcal{T}$ of RCTM by fitting the survival function with an exponential function as $S_{\ell_c}(t_S) \propto \exp(-t_S/\mathcal{T})$, $\mathcal{T}$ drastically increases and becomes sufficiently long for $\beta \approx 10^{-5}$. We show the evidence of this fact in Fig. 9(a), where the black circles and gray squares indicate the results for $\kappa_c = k_1$ and k_4, respectively. When $\kappa_c = k_1$, $\mathcal{T}$ is maximized for $\beta \approx 10^{-5}$ and becomes much longer than the turnover time T_L of the largest eddies. Furthermore, when $\kappa_c = k_4$, $\mathcal{T}$ increases more significantly than the case of $\kappa_c = k_1$, and becomes considerably long to be measured with $T_{\rm test} \approx 6000T_L$. Recall that the PDF and two-time autocorrelation functions for $\beta = 10^{-5}$ coincide with the truth as shown in Sec. III B. These results imply that an appropriate L_2 regularization pushes the RCTM lifetime sufficiently long and enables RCTM to accurately reproduce the statistics of GS variables.

A noteworthy point is that the appropriate β ($\approx 10^{-5}$) where the RCTM lifetime is longest does not give the most accurate inference results of the SGS variables (Appendix C). The lifetime is shorter for smaller β in the range $\beta \lesssim 10^{-5}$, whereas the inference is more accurate for smaller β. These results imply that it is important to introduce a stabilization method such as the L_2 regularization, even at the expense of accuracy in inferring SGS variables.

The amount of training data is also important for the RCTM stability. Figure 9(b) shows the dependence of the RCTM lifetime $\mathcal{T}$ on the regularization parameter β when the training data length is $T_{\rm train} = 50$ ($\approx 150T_L$). Comparing the cases with $T_{\rm train} = 2000$ ($\approx 6000T_L$) [Fig. 9(a)] and $T_{\rm train} \approx 150T_L$ [Fig. 9(b)], we notice that the increase trend of the RCTM lifetime with β is the same in both, while the length of the RCTM lifetime is longer when $T_{\rm train} \approx 6000T_L$. The finding that a large amount of training data stabilizes ML-based turbulence models is consistent with the results by Guan *et al.* [12].

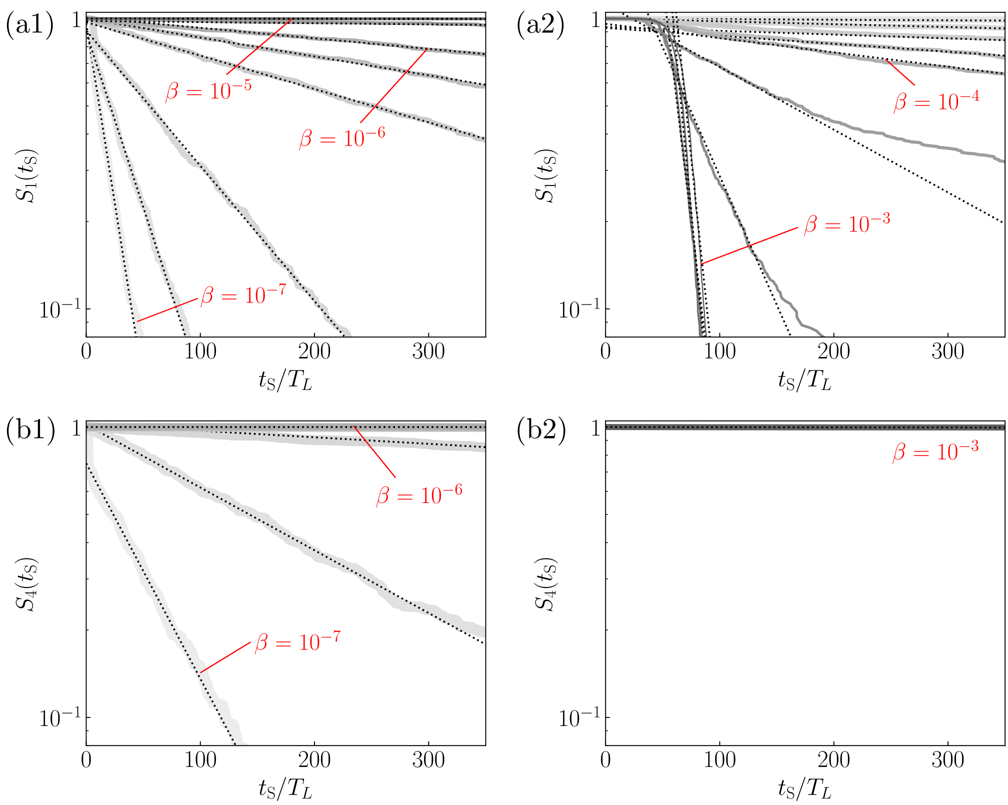

FIG. 8. Survival functions for (a) $\kappa_c = k_1 \approx 0.007\eta^{-1}$ and (b) $\kappa_c = k_4 \approx 0.03\eta^{-1}$. From the lighter (and thicker) to darker (and thinner) lines, (a1) $\beta = 10^{-7}$, 2×10^{-7}, 4×10^{-7}, 6×10^{-7}, 8×10^{-7}, 10^{-6}, 2×10^{-6}, 4×10^{-6}, 6×10^{-6}, 8×10^{-6}, and 10^{-5}; (a2) $\beta = 2 \times 10^{-5}$, 4×10^{-5}, 6×10^{-5}, 8×10^{-5}, 10^{-4}, 2×10^{-4}, 4×10^{-4}, 6×10^{-4}, 8×10^{-4}, and 10^{-3}; (b1) $\beta = 10^{-7}$, 2×10^{-7}, 4×10^{-7}, 6×10^{-7}, 8×10^{-7}, and 10^{-6}; (b2) $\beta = 10^{-3}$. The dotted lines are the fitting with exponential functions. Here, t_S is normalized using the turnover time T_L of the largest eddies.

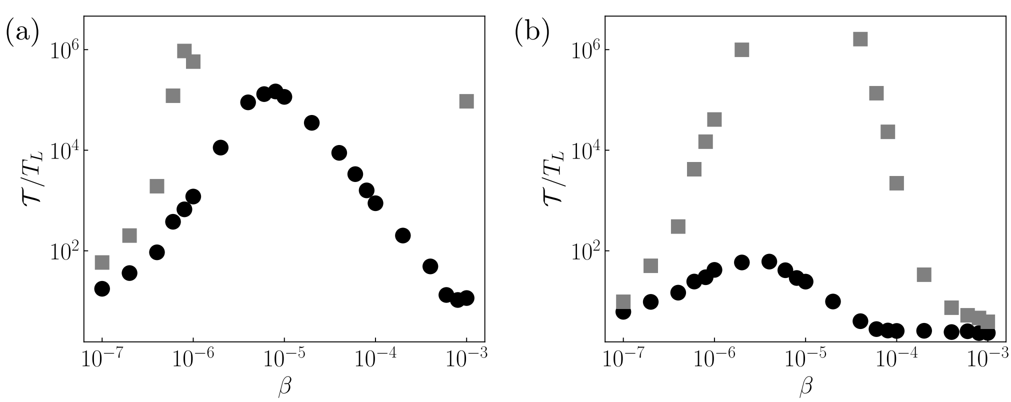

FIG. 9. Dependence of the RCTM lifetime $\mathcal{T}$ on the regularization parameter β for (a) $T_{\text{train}} = 2000 \approx 6000T_L$ and (b) $T_{\text{train}} = 50 \approx 150T_L$. The black circles and gray squares are results for $\kappa_c = k_1 \approx 0.007\eta^{-1}$ and $k_4 \approx 0.03\eta^{-1}$, respectively.

IV. CONCLUSIONS

Large-eddy simulations with turbulence models, based on deductive closure theories on the Navier–Stokes equation, have had tremendous success. However, deductive arguments require assumptions valid only in the universal range of scales. Since we desire a turbulence model applicable to cases where the universality does not hold due to the smallness of the Reynolds number or boundary conditions, expectations for ML-based turbulence models are rising. Although their progress is, in fact, remarkable [7–10], some recent studies [11–14] have reported that ML-based turbulence models can become unstable. Exploring the cause of this instability is the main purpose of the present study. To this end, we have constructed a turbulence model (RCTM, Sec. II B) using RC (Appendix B) and thoroughly evaluated its stability and accuracy.

The target of RCTM in the present study is the sparse-coupling shell model (SSM, Sec. II A), which has considerably fewer degrees of freedom than real turbulence, allowing ease of parameter survey, and retains the Kolmogorov similarity, making it clear to evaluate the accuracy of RCTM.

Although the sensitivity to initial conditions restricts the instantaneous coincidence of temporal evolution within finite time [Fig. 2(a)], RCTM can accurately reproduce the statistics, such as the PDF [Figs. 2(b) and 3(a)] and two-time autocorrelation functions [Figs. 2(c) and 3(b)], of GS variables.

The key findings are that (I) RCTM is stable when the cutoff wave number κ_c is in the viscous range ($\kappa_c \gtrsim 0.2\eta^{-1}$) and unstable when $\kappa_c \lesssim 0.2\eta^{-1}$ (Fig. 4); (II) when κ_c is in the inertial range, the temporal evolution diverges stochastically (Fig. 8). The property of (I) is explained by the nature of turbulence because the dynamics in the wave number range higher than $0.2\eta^{-1}$ are subordinate to those in the lower wave number range [33], which is physically related to the fact that eddies in the wave number range higher than $0.2\eta^{-1}$ do not transfer energy to further lower ones [34].

To overcome the instability of RCTM with κ_c in the inertial range, we have introduced L_2 regularization into step (i) of the procedure described in Sec. II B. An appropriate regularization parameter β does make the RCTM lifetime, evaluated from the exponent of the survival function, longer even for $\kappa_c \lesssim 0.2\eta^{-1}$ (Figs. 5 and 9). Furthermore, RCTM accuracy does not decrease even with the L_2 regularization (Fig. 6), and this result is independent of the Reynolds number (Fig. 7). Although RCTM becomes unstable stochastically even with the appropriate β (Fig. 8), their lifetime can be sufficiently long for the appropriately chosen β (Fig. 9). We therefore conclude that RCTM gives extremely long dynamics of GS variables with an appropriate L_2 regularization.

We reemphasize that the critical wave number $0.2\eta^{-1}$ for the stability of RCTM arises from the nature of turbulence rather than ML. Therefore, we anticipate that the stability of ML-based models for real turbulence governed by the Navier-Stokes equation depends on the cutoff wave number and that we need special treatments for the stability when we set it in the inertial range. Although it is much more challenging to construct an ML-based model for Navier-Stokes turbulence because of the nonlocality of the nonlinear interactions, the investigation of its stability is the target of our near-future study.

ACKNOWLEDGMENTS

This work was partly supported by JSPS Grants-in-Aid for Scientific Research (Grants No. 20K20973 and No. 23KJ1528). Numerical calculations were conducted under the JAXA Supercomputer System Generation 3 (JSS3) and the auspices of the NIFS Collaboration Research Programs (Grants No. NIFS22KISS010 and No. NIFS24KISC007).

APPENDIX A: SABRA SHELL MODEL

In this Appendix, we show the statistics of the Sabra shell model (1) for $\mathrm{Re} = 10^3$ and $L_{\max} = 14$. Here, we set $a = 1$, $b = -1/2$, $c = -1/2$, and the forcing $f_\ell = u_\ell \delta_{\ell 1}$ in Eq. (1). Figure 10(a) shows

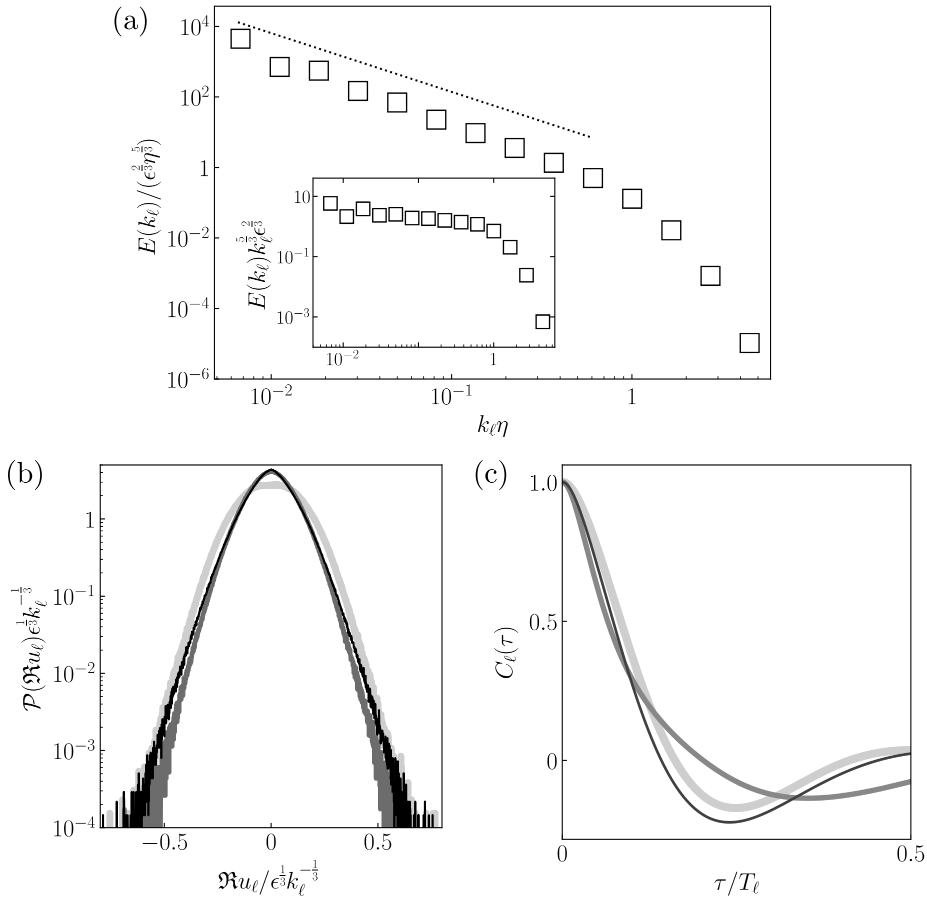

FIG. 10. Statistics of the Sabra shell model (1) for Re $= 10^3$. (a) Energy spectrum. The dotted line indicates the $-5/3$ power law and the inset shows the compensated spectrum. (b) PDF of the real part of shell variables in the inertial range. From the lighter (and thicker) to darker (and thinner) lines, $\mathcal{P}(\Re u_3)$, $\mathcal{P}(\Re u_4)$ and $\mathcal{P}(\Re u_5)$. (c) Two-time autocorrelation functions of the real part of shell variables in the inertial range. From the lighter (and thicker) to darker (and thinner) lines, C_3, C_4, and C_5. Here, we use the energy dissipation rate ϵ, Kolmogorov length η, and characteristic timescale T_ℓ for normalization.

the energy spectrum,

$$E(k_\ell) = \frac{1}{2k_\ell}|u_\ell|^2. \tag{A1}$$

Although we observe the $-5/3$ power law, $E(k_\ell) \propto k_\ell^{-\frac{5}{3}}$, in the inertial range, we can see unnatural zigzag deviations from the power law. Figure 10(b) shows the PDF $\mathcal{P}(\Re u_\ell)$ of the real part of shell variables in the inertial range ($\ell = 3, 4, 5$). They depend on wave numbers even when normalized by the energy dissipation rate ϵ and wave number k_ℓ. Figure 10(c) shows two-time autocorrelation functions,

$$C_\ell(\tau) = \langle \Re u_\ell(t+\tau)\Re u_\ell(t)\rangle_t, \tag{A2}$$

of the real part of shell variables in the inertial range ($\ell = 3, 4, 5$). We notice that the functional shape of the two-time autocorrelation function depends on wave numbers. In other words, a clear collapse cannot be observed even when normalized by the characteristic timescale T_ℓ of each shell.

When we evaluate the RCTM accuracy, these properties make it difficult to consider the cause of differences between the true and RCTM statistics. Therefore, we use SSM, which exhibits clear Kolmogorov similarity (Fig. 1).

APPENDIX B: RESERVOIR COMPUTING

In this Appendix, we describe the overview of reservoir computing (RC) [20–22], a training framework of recurrent neural networks. RC utilizes a random recurrent network to achieve low training costs and effective inference. More precisely, the weights of the input layer and a recurrent layer called the reservoir are determined by random numbers, and only the output weights are optimized. Therefore, the training cost is significantly lower than that of other types of RNN, such as LSTM, which is suitable for the extensive parameter survey.

RC learns the relation between the input signals $\boldsymbol{s}(t)$ $(\in \mathbb{R}^{N_s})$ and the output signals $\boldsymbol{y}(t)$ $(\in \mathbb{R}^{N_y})$ from a finite dataset $\{\boldsymbol{s}(t), \boldsymbol{y}(t)\}$ for $0 \leqslant t \leqslant T_{\text{train}}$. For this, the output weights W_{out} $(\in \mathbb{R}^{N_y \times N_r})$ is optimized so that the output $\widehat{\boldsymbol{y}}(t)$ with the given input approximates $\boldsymbol{y}(t)$, where $\widehat{\boldsymbol{y}}(t)$ is expressed as the linear combination of the reservoir state $\boldsymbol{r}$ $(\in \mathbb{R}^{N_r})$,

$$\widehat{\boldsymbol{y}}(t) = W_{\text{out}}\boldsymbol{r}(t). \tag{B1}$$

We define the loss function as

$$E(W_{\text{out}}) = \frac{1}{Q_{\text{train}}} \sum_{q=1}^{Q_{\text{train}}} \|\boldsymbol{y}(q\Delta\tau) - W_{\text{out}}\boldsymbol{r}(q\Delta\tau)\|^2 + \beta[\text{Tr}(W_{\text{out}}W_{\text{out}}^{\top})], \tag{B2}$$

where Tr denotes the trace. The second term on the right-hand side is the L_2 regularization term, and β adjusts the strength of its effect. We can obtain the optimized output weights W_{out}^* by solving the minimization problem in Eq. (B2) as

$$W_{\text{out}}^* = (R + \beta I)^{-1}S, \tag{B3}$$

where $R_{ij} = \langle r_i(t) r_j(t)\rangle_t$ and $S_{ij} = \langle r_i(t) y_j(t)\rangle_t$.

In the present study, we employ the echo state network [20], which is one of the standard structures of the reservoir [22]. The temporal evolution of the reservoir state $\boldsymbol{r}$ follows

$$\boldsymbol{r}(t) = \alpha \tanh\left[A\boldsymbol{r}(t - \Delta\tau) + W_{\text{in}}\boldsymbol{s}(t) + \boldsymbol{b}\right] + (1 - \alpha)\boldsymbol{r}(t - \Delta\tau), \tag{B4}$$

where A $(\in \mathbb{R}^{N_r \times N_r})$ and W_{in} $(\in \mathbb{R}^{N_r \times N_s})$ are connection weights between nodes in the network. In the present study, we determine them by random numbers sampled from a normal distribution with mean zero and standard deviation $\rho/\sqrt{N_r}$ and a uniform distribution in the range $[-\sigma, \sigma]$, respectively. In Eq. (B4), $\boldsymbol{b}$ $(\in \mathbb{R}^{N_r})$ is the bias, which is determined by random numbers sampled from a uniform distribution in the range $[-\xi, \xi]$, and α $(\in [0, 1])$ is the leakage rate [35].

APPENDIX C: INFERENCE OF SGS VARIABLES

In this Appendix, we show the inference results of SGS variables $\boldsymbol{X}^{(\ell_c+1)}$ for $\text{Re} = 10^3$. Here, the inputs to RC are the true values $\boldsymbol{X}^{(\ell_c)}$ sampled from the test data; that is, we present the test results of the first step (i) of the modeling method described in Sec. II B. Figure 11(a) shows the time series of the inferred variables $\widehat{X}_1^{(11)}$ for $\beta = 10^{-7}$, 10^{-5}, and 10^{-3} with $\kappa_c = k_{10} \approx 0.7\eta^{-1}$. When β is small, the inferred time series coincides with the truth. In fact, that for 10^{-7} correctly captures the true large and short-timescale fluctuations. In contrast, the amplitude of the inferred time series for $\beta = 10^{-5}$ is always smaller than that of the truth, and that for $\beta = 10^{-3}$ is almost zero. This may raise a naive question of whether $\widehat{\boldsymbol{X}}^{(\ell_c+1)} = 0$ would be acceptable because $P_{\text{F}} = 0$ for $0.05\eta^{-1} \lesssim \kappa_c \lesssim 0.2\eta^{-1}$ even when $\beta = 10^{-3}$. However, the results of the closure with setting $\widehat{\boldsymbol{X}}^{(\ell_c+1)} = 0$ are different from the RCTM results with the inference of $\widehat{\boldsymbol{X}}^{(\ell_c+1)}$. More concretely, when $\kappa_c \lesssim 0.2\eta^{-1}$, GS variables in the simple closure with $\widehat{\boldsymbol{X}}^{(\ell_c+1)} = 0$ diverge in time because this

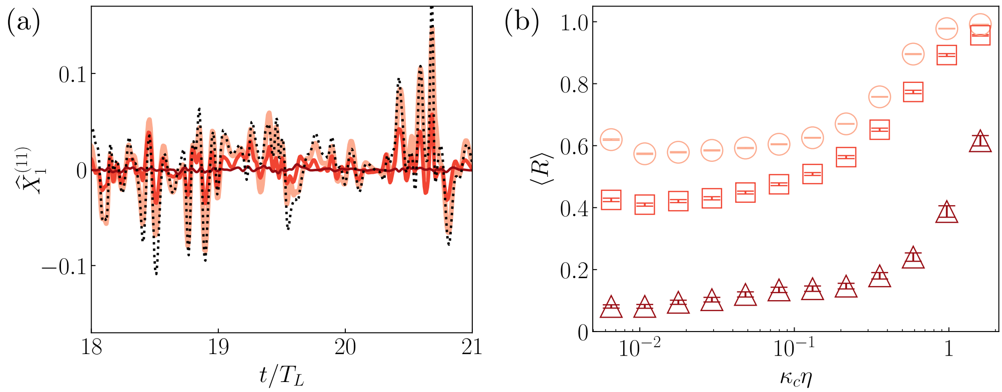

FIG. 11. Inference of SGS variables for Re $= 10^3$. (a) Time series of the inferred variable $\widehat{X}_1^{(11)}$ for $\kappa_c = k_{10} \approx 0.7\eta^{-1}$. From the red lighter (and thicker) to red darker (and thinner) lines, the results for $\beta = 10^{-7}$, 10^{-5}, and 10^{-3}; the black-dotted line is the truth. (b) Dependence of the correlation coefficients between the truth and RC inference on the cutoff wave numbers κ_c. The circles, squares, and triangles are the results for $\beta = 10^{-7}$, 10^{-5}, and 10^{-3}, respectively. Error bars are the standard deviation over 500 trials. We use the turnover time T_L of the largest eddies and Kolmogorov length η for normalization in (a) and (b), respectively.

approximately corresponds to the inviscid truncated system with forcing. On the other hand, when $\kappa_c \gtrsim 0.2\eta^{-1}$, although GS variables do not diverge up to T_{test}, their statistical properties significantly differ from those of the truth.

To quantify the inference accuracy of RC, we evaluate the correlation coefficients between the true and inferred time series. Figure 11(b) shows the ensemble average $\langle R\rangle$ of the correlation coefficients over 500 trials for $\beta = 10^{-7}$, 10^{-5}, and 10^{-3}. Regardless of κ_c, $\langle R\rangle$ is smaller for larger β, implying that the regularization reduces the inference accuracy. Incidentally, except $\kappa_c = k_1$, $\langle R\rangle$ increases with κ_c, and the increase rate for $\kappa_c \gtrsim 0.2\eta^{-1}$ is larger than for $\kappa_c \lesssim 0.2\eta^{-1}$. This may also reflect the subordination of smaller-scale dynamics to larger-scale ones [30–34].

We emphasize that the optimal β for the inference accuracy of SGS variables differs from that for the stability of RCTM. More concretely, for $\kappa_c = k_4 \approx 0.03\eta^{-1}$, among $\beta = 10^{-7}$, 10^{-5}, and 10^{-3}, the highest inference accuracy is achieved with the smallest β ($=10^{-7}$), while the model lifetime is longest for $\beta = 10^{-5}$ (Fig. 9). This result implies that it is crucial to introduce a stabilization method, such as the L_2 regularization, even when sacrificing the inference accuracy of SGS variables.

[1] U. Frisch, *Turbulence: The Legacy of A. N. Kolmogorov* (Cambridge University Press, Cambridge, England, 1995).

[2] C. Meneveau and J. Katz, Scale-invariance and turbulence models for large-eddy simulation, Annu. Rev. Fluid Mech. **32**, 1 (2000).

[3] W. Heisenberg, On the theory of statistical and isotropic turbulence, Proc. R. Soc. London A **195**, 402 (1948).

[4] R. H. Kraichnan, The structure of isotropic turbulence at very high reynolds numbers, J. Fluid Mech. **5**, 497 (1959).

[5] R. H. Kraichnan, Lagrangian-history closure approximation for turbulence, Phys. Fluids **8**, 575 (1965).

[6] Y. Zhou, Turbulence theories and statistical closure approaches, Phys. Rep. **935**, 1 (2021).

[7] M. Gamahara and Y. Hattori, Searching for turbulence models by artificial neural network, Phys. Rev. Fluids **2**, 054604 (2017).

[8] M. P. Brenner, J. D. Eldredge, and J. B. Freund, Perspective on machine learning for advancing fluid mechanics, Phys. Rev. Fluids **4**, 100501 (2019).
[9] C. Xie, Z. Yuan, and J. Wang, Artificial neural network-based nonlinear algebraic models for large eddy simulation of turbulence, Phys. Fluids **32**, 115101 (2020).
[10] S. L. Brunton, B. R. Noack, and P. Koumoutsakos, Machine learning for fluid mechanics, Annu. Rev. Fluid Mech. **52**, 477 (2020).
[11] R. Maulik, O. San, J. D. Jacob, and C. Crick, Sub-grid scale model classification and blending through deep learning, J. Fluid Mech. **870**, 784 (2019).
[12] Y. Guan, A. Chattopadhyay, A. Subel, and P. Hassanzadeh, Stable a posteriori les of 2d turbulence using convolutional neural networks: Backscattering analysis and generalization to higher re via transfer learning, J. Comput. Phys. **458**, 111090 (2022).
[13] A. S. P. Ayapilla and Y. Hattori, A data-driven approach to model enstrophy transfers in large eddy simulation of forced two-dimensional turbulence, Phys. Fluids **35**, 075116 (2023).
[14] S. Miyazaki and Y. Hattori, Improving accuracy of turbulence models by neural network, arXiv:2012.01723.
[15] A. N. Kolmogorov, The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers, C. R. Acad. Sci. URSS **30**, 301 (1941).
[16] T. Ishihara, T. Gotoh, and Y. Kaneda, Study of high–reynolds number isotropic turbulence by direct numerical simulation, Annu. Rev. Fluid Mech. **41**, 165 (2009).
[17] G. Ortali, A. Corbetta, G. Rozza, and F. Toschi, Numerical proof of shell model turbulence closure, Phys. Rev. Fluids **7**, L082401 (2022).
[18] V. S. L'vov, E. Podivilov, A. Pomyalov, I. Procaccia, and D. Vandembroucq, Improved shell model of turbulence, Phys. Rev. E **58**, 1811 (1998).
[19] S. Hochreiter and J. Schmidhuber, Long short-term memory, Neural Comput. **9**, 1735 (1997).
[20] H. Jaeger, The "echo state" approach to analysing and training recurrent neural networks-with an erratum note, GMD Rep. **148**, 13 (2001).
[21] H. Jaeger and H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication, Science **304**, 78 (2004).
[22] K. Nakajima and I. Fischer, *Reservoir Computing* (Springer, Singapore, 2021).
[23] U. Frisch, P.-L. Sulem, and M. Nelkin, A simple dynamical model of intermittent fully developed turbulence, J. Fluid Mech. **87**, 719 (1978).
[24] M. Yamada and K. Ohkitani, Lyapunov spectrum of a chaotic model of three-dimensional turbulence, J. Phys. Soc. Jpn. **56**, 4210 (1987).
[25] L. Biferale, Shell models of energy cascade in turbulence, Annu. Rev. Fluid Mech. **35**, 441 (2003).
[26] S. Goto and S. Kida, Direct-interaction approximation and reynolds-number reversed expansion for a dynamical system, Physica D **117**, 191 (1998).
[27] S. Goto and S. Kida, Sparseness of nonlinear coupling: importance in sparse direct-interaction perturbation, Nonlinearity **15**, 1499 (2002).
[28] D. M. Endres and J. E. Schindelin, A new metric for probability distributions, IEEE Trans. Inf. Theory **49**, 1858 (2003).
[29] S. Kullback and R. A. Leibler, On information and sufficiency, Ann. Math. Stat. **22**, 79 (1951).
[30] K. Yoshida, J. Yamaguchi, and Y. Kaneda, Regeneration of small eddies by data assimilation in turbulence, Phys. Rev. Lett. **94**, 014501 (2005).
[31] C. C. Lalescu, C. Meneveau, and G. L. Eyink, Synchronization of chaos in fully developed turbulence, Phys. Rev. Lett. **110**, 084102 (2013).
[32] Y. Li, J. Zhang, G. Dong, and N. S. Abdullah, Small-scale reconstruction in three-dimensional kolmogorov flows using four-dimensional variational data assimilation, J. Fluid Mech. **885**, A9 (2020).
[33] M. Inubushi, Y. Saiki, M. U. Kobayashi, and S. Goto, Characterizing small-scale dynamics of navier-stokes turbulence with transverse lyapunov exponents: A data assimilation approach, Phys. Rev. Lett. **131**, 254001 (2023).
[34] T. Yoneda, S. Goto, and T. Tsuruhashi, Mathematical reformulation of the Kolmogorov–Richardson energy cascade in terms of vortex stretching, Nonlinearity **35**, 1380 (2022).

[35] H. Jaeger, M. Lukoševičius, D. Popovici, and U. Siewert, Optimization and applications of echo state networks with leaky-integrator neurons, Neural Networks **20**, 335 (2007).

[36] K. Nakai and Y. Saiki, Machine-learning construction of a model for a macroscopic fluid variable using the delay-coordinate of a scalar observable, Discrete Continuous Dynamical Systems-S **14**, 1079 (2021).

[37] M. Hara and H. Kokubu, Learning dynamics by reservoir computing (in memory of prof. Pavol Brunovský), J. Dyn. Differ. Equations **36**, 515 (2022).

[38] A. J. Linot and M. D. Graham, Data-driven reduced-order modeling of spatiotemporal chaos with neural ordinary differential equations, Chaos **32**, 073110 (2022).

[39] A. J. Linot, J. W. Burby, Q. Tang, P. Balaprakash, M. D. Graham, and R. Maulik, Stabilized neural ordinary differential equations for long-time forecasting of dynamical systems, J. Comput. Phys. **474**, 111838 (2023).

[40] A. J. Linot and M. D. Graham, Dynamics of a data-driven low-dimensional model of turbulent minimal couette flow, J. Fluid Mech. **973**, A42 (2023).

[41] A. Wikner, J. Harvey, M. Girvan, B. R. Hunt, A. Pomerance, T. Antonsen, and E. Ott, Stabilizing machine learning prediction of dynamics: Novel noise-inspired regularization tested with reservoir computing, Neural Networks **170**, 94 (2024).