Data-driven learning of Mori–Zwanzig operators for isotropic turbulence

Yifeng Tian, Yen Ting Lin, Marian Anghel, et al.

Data-driven learning of Mori–Zwanzig operators for isotropic turbulence

Yifeng Tian,[1,a] (iD) Yen Ting Lin,[2,b] (iD) Marian Anghel,[2,c] (iD) and Daniel Livescu[1,d] (iD)

AFFILIATIONS

[1]Computational Physics and Methods Group, Computer, Computational and Statistical Sciences Division (CCS-2), Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

[2]Information Sciences Group, Computer, Computational and Statistical Sciences Division (CCS-3), Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

[a]Author to whom correspondence should be addressed: yifengtian@lanl.gov
[b]Electronic mail: yentingl@lanl.gov
[c]Electronic mail: manghel@lanl.gov
[d]Electronic mail: livescu@lanl.gov

ABSTRACT

Developing reduced-order models for turbulent flows, which contain dynamics over a wide range of scales, is an extremely challenging problem. In statistical mechanics, the Mori–Zwanzig (MZ) formalism provides a mathematically exact procedure for constructing reduced-order representations of high-dimensional dynamical systems, where the effects due to the unresolved dynamics are captured in the memory kernel and orthogonal dynamics. Turbulence models based on MZ formalism have been scarce due to the limited knowledge of the MZ operators, which originates from the difficulty in deriving MZ kernels for complex nonlinear dynamical systems. In this work, we apply a recently developed data-driven learning algorithm, which is based on Koopman's description of dynamical systems and Mori's linear projection operator, on a set of fully resolved isotropic turbulence datasets to extract the Mori–Zwanzig operators. With data augmentation using known turbulence symmetries, the extracted Markov term, memory kernel, and orthogonal dynamics are statistically converged and the generalized fluctuation–dissipation relation can be verified. The properties of the memory kernel and orthogonal dynamics, and their dependence on the choices of observables are investigated to address the modeling assumptions that are commonly used in MZ-based models. A series of numerical experiments are then constructed using the extracted kernels to evaluate the memory effects on prediction. The results show that the prediction errors are strongly affected by the choice of observables and can be further reduced by including the past history of the observables in the memory kernel.

I. INTRODUCTION

Direct and accurate computations of complex nonlinear dynamical systems in physical sciences and engineering applications, such as turbulent flows, are, in general, prohibitively expensive due to the existence of wide range of length and time scales. This challenge has motivated the development of reduced order models (ROM) to achieve fast and efficient solutions in practical applications. In the field of turbulence simulations, Reynolds-Averaged Navier–Stokes (RANS) and large eddy simulations (LES) have been widely adopted as alternatives for direct numerical simulation (DNS). The computational complexity is reduced through coarse-graining, which effectively narrows the range of scales that need to be resolved.[1] High-order moments and their dynamical equations, which account for the effects of unresolved scales, naturally emerge during the coarse-graining process. The high-order moments are then truncated, and surrogate models, commonly referred to as subgrid-scale models, have been developed to account for the effects from the high-order moments contributions. Such surrogate models are usually developed under the assumptions of scale similarity or universality of fully developed turbulence, and the resulting simplified models are usually Markovian in nature. Kraichnan[2,3] introduced Direct Interaction Approximation (DIA) as a non-Markovian closure model for turbulence statistics. The evolution of a new quantity, the infinitesimal response function, is employed to model the response of turbulent flows to infinitesimal perturbations. However, this method suffers from certain theoretical (inability of reproducing inertial range behavior) and practical (difficulty in calculating long-time statistics) weaknesses.[4]

The Mori–Zwanzig (MZ) formalism was originally developed in statistical physics for constructing projection-based low-dimensional, non-Markovian models for high-dimensional nonlinear dynamical systems.[5,6] It provides a mathematically exact procedure for developing reduced-order models for high-dimensional systems. This reduction of dimensionality is achieved by applying a projection operator to the evolution operator of the original dynamical systems. The resulting lower-dimensional model, referred to as the generalized Langevin equation (GLE), consists of a Markovian term, a memory term, and a noise term. The GLE, as derived in the framework of MZ formalism, is an exact representation of the dynamics of the reduced-order model. In the context of RANS and LES modeling, the truncated high-order moments can be accounted for in the memory integral and noise terms under the Mori–Zwanzig formalism. However, in most previous efforts of subgrid-scale modeling, the evolutionary equations are usually treated as Markovian.

Modeling turbulence under the MZ formalism is extremely challenging due to the limited understanding of the memory kernels and orthogonal dynamics. Obtaining the structure of the memory kernel requires the solution of the unresolved orthogonal dynamics, which is another high-dimensional nonlinear dynamical system. Despite this difficulty, the optimal prediction framework developed in Chorin *et al.*,[7] Givon *et al.*,[8] Chorin and Stinis[9] provides a formal procedure for analyzing the memory effects and developing surrogate models based on the MZ formalism. One of the major modeling difficulties using the MZ formalism is determining the structure and length of the memory term. The widely used t-model approximates the memory length as equal to simulation time t and has been used by Bernstein,[10] Hald and Stinis,[11] and Chandy and Frankel[12] for prediction of Burger's equations, Euler equations, and Navier–Stokes (NS) equations. Parish and Duraisamy[13] later proposed a dynamic-τ model to approximate the memory length using the similarity between two coarse-graining levels. The renormalized MZ models[14,15] embed a larger class of models that share similar functional forms with MZ formalism but with different coefficients to approximate the memory integral. Stinis,[16] and Parish and Duraisamy[17] further used finite order expansion of the orthogonal dynamics and cast it to a set of differential equations that represent the effects of memory integral with finite memory length. These MZ-based turbulence models rely on simplified assumptions and observations of the turbulent flow, most of which have not been verified due to the difficulty in deriving or extracting the MZ operators. Gouasmi *et al.*[18] proposed a method for the estimation of the memory integrals using pseudo orthogonal dynamics, which is only exact for linear dynamics. In previous studies on MZ-based turbulence modeling, the nonlinear projection operator[19] is used to formulate the GLE, so the resulting Markov and memory kernels are nonlinear functions of resolved variables. Mori's linear projection operator, or the finite rank projection operator, is another commonly used projection operator for MZ formulation. The procedure for evaluating Markov operator and memory kernel from numerical data without additional assumptions based on Mori's linear projection operator was first proposed in the study by Chorin *et al.*,[19] although the evaluated memory kernel is not explicitly shown. Also, the finite-rank projection operator is defined in Chorin *et al.*[19] on a set of orthonormal basis functions, which are not as flexible as the general, arbitrary observables employed in modern Koopman analysis.[20] Okamura,[21] and Mori and Okamura[22] employed an iterative approach

to construct the memory kernel from the correlation functions of the numerical data. Meyer *et al.*,[23] Meyer *et al.*[24] proposed an iterative numerical procedures for evaluating memory effects using the auto-correlation function. Maeyama and Watanabe[25] revisited the method proposed by Chorin *et al.*[19] for extracting the memory kernel and modeled the effects of small-scale fluctuations on large-scale fluctuations based on the MZ formalism. Importantly, in these recently proposed methods, the reduced-order space is always restricted to be one-dimensional, allowing only one "coarse-grained" variable. Such a severe model reduction could lead to undesirably poor results—there is only so much information one dynamic variable contains.

The above-mentioned models approach the dimensional reduction starting with the original nonlinear systems. The challenges in approximating the memory kernel come from the nonlinearity of the equations. For the same dynamical system, there exists another formulation, as proposed by Koopman,[26] Koopman and Neumann.[27] In Koopman's description, the system is characterized by a collection of observables which are functions of the original phase space coordinates. The Koopmanian formulation describes how observables evolve in an infinite-dimensional Hilbert space, which is composed of all the possible observables. The advantage of this formulation is that the evolution of the observables, which form a vector in the infinite dimensional Hilbert space, is always linear, even for systems that are nonlinear in the phase-space picture. The disadvantage of this formulation is that the state space of the system, which consists of all possible observables, is infinite dimensional. Based on the Koopman representation of dynamical system, approximate learning methods, such as dynamic mode decomposition (DMD)[28] and extended dynamic mode decomposition (EDMD),[20] have been developed for data-driven modeling of dynamical systems. By combining the Koopman description with MZ formalism, it is possible to perform a dimensional reduction of the infinite dimensional Koopmanian linear formulation to a finite, low-dimensional dynamical system with memory kernels and orthogonal dynamics. Since the observables evolve in a linear space, the learning problem is convex, which can greatly simplify the learning of MZ operators. Lin *et al.*[29] proposed a generalized data-driven learning framework for extracting MZ memory kernel and orthogonal dynamics from high-dimensional data under the generalized Koopman formulation, and analyzed the properties of these terms for a Lorenz '96 model. Theoretically, Lin *et al.*[29] established the formal connection between Koopman operators and MZ formalism in both continuous- and discrete-time formulations. Numerically, motivated by the recent approximate Koopman learning methods, the algorithms presented in Lin *et al.*[29] learn jointly from time series of a finite number of dynamic variables, in contrast to the one-dimensional projection in Meyer *et al.*[23] and Maeyama and Watanabe.[25] Learning jointly from multivariate time series is beneficial especially for those systems with strong interactions among these dynamical variables, e.g., local interactions in fluid dynamical systems.

In this work, we take the first step to apply the learning algorithm in Lin *et al.*[29] to a homogeneous isotropic turbulence DNS database to extract the Markov, memory, and orthogonal (noise) terms for the coarse-grained Navier–Stokes system. To the authors' best knowledge, there has been no study using data-driven methods to accurately extract MZ terms for Navier–Stokes turbulence, despite the fact that understanding the properties of the memory kernels and orthogonal dynamics is crucially important not only to quantitatively address the assumptions in MZ-based turbulence models but, more generally,

understand the past memory effects for the coarse-grained NS dynamics. The manuscript is organized as follows: in Sec. II, the MZ formalism, as a generalized Koopman learning framework, is introduced. The data-driven learning framework for MZ kernels is explained in Sec. III. The DNS database and post-processing procedures are then described in Sec. IV. The results and discussions are presented in Sec. V, and the conclusions are drawn in Sec. VI.

II. MORI–ZWANZIG FORMALISM

Consider the following semi-discrete high-dimensional ordinary differential equation (ODEs) for the full set of state variables $\mathbf{\Phi}(t) = [\phi_1(t), \ldots, \phi_N(t)]^T \in \mathbb{R}^N$:

$$\frac{d\mathbf{\Phi}(t)}{dt} = R(\mathbf{\Phi}(t)), \quad \mathbf{\Phi}(0) = \mathbf{x}, \tag{1}$$

where $R : \mathbb{R}^N \to \mathbb{R}^N$ is a N-dimensional vector of nonlinear real functions defined on $\mathbb{R}^N$ and $\mathbf{x} \in \mathbb{R}^N$. Due to the difficulties in simulating and analyzing high-dimensional nonlinear dynamical systems, it is generally desirable to develop low-dimensional representations of the same system. In order to achieve this reduction of complexity, we consider the evolution of a set of observables $\mathbf{u}(\mathbf{x}, t) := \mathbf{g}(\mathbf{\Phi}(\mathbf{x}, t))$, where $\mathbf{g} : \mathbb{R}^N \to \mathbb{R}^D$ is a D-dimensional vector of observables [functions of the phase space variables $\mathbf{\Phi}(\mathbf{x}, t)$] and, in general, $D < N$. Here, we use $\mathbf{\Phi}(\mathbf{x}, t)$ to denote the solution to Eq. (1) with the initial conditions $\mathbf{\Phi}(0) = \mathbf{x}$, and $\mathbf{g}(\mathbf{x}, t)$ to denote the observables $\mathbf{g}(\mathbf{\Phi}(\mathbf{x}, t))$ at time t. One way to define the observables is to decompose the state variables $\mathbf{\Phi}$ into resolved/relevant variables $\hat{\mathbf{\Phi}}(\mathbf{x}, t) = [\phi_1(t), \ldots, \phi_D(t)]^T \in \mathbb{R}^D$, and unresolved ones $\tilde{\mathbf{\Phi}} = [\phi_{D+1}(t), \ldots, \phi_N(t)]^T \in \mathbb{R}^{N-D}$, and evolutionary equations for the resolved variables $\hat{\mathbf{\Phi}}$ are developed to reduce the dimension of the original nonlinear set of ODEs. In Sec. II B, we will describe a set of coarse-grained observables, which are derived by applying a spatial filter to the velocity field of the incompressible Navier–Stokes equations. For a system with sufficient separation of scales, it might be possible to decompose the system into slow and fast dynamics; the latter might be modeled as function of the former and some simple (white) noise. However, in many physical systems, there exists a continuous spectrum of scales so that the dynamics of the resolved variables are nonlinearly coupled to the unresolved ones. To formally solve this problem, Mori,[5] and Zwanzig[6] developed the projection-based method to express the effects of the unresolved variables in terms of the resolved ones. The key result of Mori–Zwanzig's formulation of the reduced-dimensional system is the generalized Langevin equation (GLE), which is characterized by the emergence of the memory-dependent dynamics—represented as the convolutional integral of the past history of the resolved variables and a memory kernel, and orthogonal dynamics—describing the evolution of unresolved variables in the orthogonal functional space.

A. Derivation of the generalized Langevin equation: Key construct of Mori–Zwanzig Formalism

In this section, we provide a formal derivation of the Mori–Zwanzig formalism. Consider the nonlinear ODE system as shown in Eq. (1), with the initial conditions $\mathbf{\Phi}(0) = \mathbf{x}, \mathbf{x} \in \mathbb{R}^N$. The evolution of a set of observables $\mathbf{u}(\mathbf{x}, t) := \mathbf{g}(\mathbf{\Phi}(\mathbf{x}, t))$, where $\mathbf{g} : \mathbb{R}^N \to \mathbb{R}^D$ is a D-dimensional vector of observables, can be posed as a linear partial differential equation (PDE) in the Liouville form

$$\frac{\partial}{\partial t}\mathbf{u}(\mathbf{x}, t) = \mathscr{L}\mathbf{u}(\mathbf{x}, t), \quad \mathbf{u}(\mathbf{x}, 0) = \mathbf{g}(\mathbf{\Phi}(\mathbf{x}, 0)) = \mathbf{g}(\mathbf{x}), \tag{2}$$

where $\mathscr{L}$ is the Liouville operator

$$\mathscr{L} := \sum_{i=1}^{N} R_i(\mathbf{x})\partial_{x_i}. \tag{3}$$

Thus, the PDE (2) becomes the ODE (1) along the characteristics curves. A special choice of $\mathbf{g}$ in Eq. (2) is $g_i(\mathbf{x}) := x_i$, $i \in 1, \ldots, D$, which extracts the ith component of the state of the system, $\mathbf{x}$. Using the semigroup notation, the solution of the linear PDE can be written as $\mathbf{u}(\mathbf{x}, t) = e^{t\mathscr{L}}\mathbf{g}(\mathbf{x})$, where $e^{t\mathscr{L}}$ is referred to as the evolution operator. It can be shown that $e^{t\mathscr{L}}\mathbf{g}(\mathbf{x}) = \mathbf{g}(e^{t\mathscr{L}}\mathbf{x}) = \mathbf{g}(\mathbf{\Phi}(\mathbf{x}, t))$ since $e^{t\mathscr{L}}\mathbf{x} = \mathbf{\Phi}(\mathbf{x}, t)$. Alternatively, this can be written as $e^{t\mathscr{L}}\mathbf{\Phi}(\mathbf{x}, 0) = \mathbf{\Phi}(\mathbf{x}, t)$ and we recognize that the operator $\mathscr{K}_t = e^{t\mathscr{L}}$ is the one-parameter family of Koopman operators. We further remark that the evolution operator (Koopman operator) $e^{t\mathscr{L}}$ and the Liouville operator $\mathscr{L}$ commute, that is $\mathscr{L}e^{t\mathscr{L}} = e^{t\mathscr{L}}\mathscr{L}$. Hence, Eq. (2) becomes

$$\frac{\partial}{\partial t}e^{t\mathscr{L}}\mathbf{g}(\mathbf{x}) = \mathscr{L}e^{t\mathscr{L}}\mathbf{g}(\mathbf{x}) = e^{t\mathscr{L}}\mathscr{L}\mathbf{g}(\mathbf{x}). \tag{4}$$

In order to construct reduced-order representation of the linear PDE using the reduced D-dimensional vector of observables $\mathbf{u}$, with the initial condition $\mathbf{u}(\mathbf{x}, 0) = \mathbf{g}(\mathbf{x})$, a projection operator, P, needs to be specified. In this work, we denote $L^2(\mu)$ the Hilbert space of functions endowed with the inner product defined as

$$\langle f, g \rangle := \int f(\mathbf{x})g(\mathbf{x})\, d\mu(\mathbf{x}), \quad f, g \in L^2(\mu), \tag{5}$$

where f and g are L^2-integrable functions with respect to the probability distribution μ. In this work, we adopt a stationary measure for $\mathbf{x}$. The projection operator P that maps functions $f(\mathbf{x}) \in L^2(\mu)$ into the subspace $Span\{g_1(\mathbf{x}), \ldots, g_D(\mathbf{x})\}$ is then defined. The particular choice of projection operator determines the functional form of the Mori–Zwanzig formulation. Examples of projection operators include nonlinear projection operator that relies on the marginalization of the under-resolved observables[6] and finite rank projection operator that relies on the inner product in the Hilbert space.[5] After the projection operator P is defined, its complement Q is denoted as $Q = I - P$ and satisfies $PQ = QP = 0$, where I is the identity operator. We then substitute the Dyson identity[30]

$$e^{t\mathscr{L}} = e^{t(P+Q)\mathscr{L}} = e^{tQ\mathscr{L}} + \int_0^t e^{(t-s)\mathscr{L}}P\mathscr{L}e^{sQ\mathscr{L}}ds \tag{6}$$

in Eq. (4) and arrive at

$$\frac{\partial}{\partial t}\left[e^{t\mathscr{L}}\mathbf{g}(\mathbf{x})\right] = e^{t\mathscr{L}}\mathscr{L}\mathbf{g}(\mathbf{x}) = e^{t\mathscr{L}}P\mathscr{L}\mathbf{g}(\mathbf{x}) + e^{t\mathscr{L}}Q\mathscr{L}\mathbf{g}(\mathbf{x})$$
$$= e^{t\mathscr{L}}P\mathscr{L}\mathbf{g}(\mathbf{x}) + e^{tQ\mathscr{L}}Q\mathscr{L}\mathbf{g}(\mathbf{x})$$
$$+ \int_0^t e^{(t-s)\mathscr{L}}P\mathscr{L}e^{sQ\mathscr{L}}Q\mathscr{L}\mathbf{g}(\mathbf{x})ds, \tag{7}$$

which can be written in terms of the observables

$$\frac{\partial}{\partial t}\mathbf{g}(\mathbf{x}, t) = \mathbf{M}(\mathbf{g}(\mathbf{x}, t)) + \mathbf{F}(\mathbf{x}, t) - \int_0^t \mathbf{K}(\mathbf{g}(\mathbf{x}, t-s), s)ds. \tag{8}$$

The specific selection of the projection operator will be discussed below. The above equation is the generalized Langevin equation (GLE), which contains a Markov transition term, orthogonal dynamics, and a memory kernel that are defined as

$$\mathbf{M}(\mathbf{g}(\mathbf{x}, t)) := e^{t\mathscr{L}} P \mathscr{L} \mathbf{g}(\mathbf{x}) \tag{9a}$$

$$\mathbf{F}(\mathbf{x}, t) := e^{tQ\mathscr{L}} Q \mathscr{L} \mathbf{g}(\mathbf{x}) \tag{9b}$$

$$\mathbf{K}(\mathbf{g}(\mathbf{x}, t - s), s) := -e^{(t-s)\mathscr{L}} P \mathscr{L} e^{sQ\mathscr{L}} Q \mathscr{L} \mathbf{g}(\mathbf{x}), \tag{9c}$$

with the orthogonality condition $PF(\mathbf{x}, t) = 0$. These three components represent the key construct of Mori–Zwanzig formalism, and the GLE is exact in describing the dynamics of observables $\mathbf{g}$. Equation (9c) is also referred to as the nonlinear generalized fluctuation-dissipation (GFD) relation.[6] Note that here we use a negative sign in front of the memory term following the convention in Mori[5] and Zwanzig.[6]

We remark that Eq. (8) is the general form of the MZ formulation and is not specific to any projection operator. In practice, the choice of the projection operator is central for constructing the MZ formalism, because the functional form of the components may vary drastically for different projection operators. Here, we give some examples of projections operators that have been employed in the literature:

- In Mori's formulation,[5] the projection operator relies on the inner product defined in Eq. (5). With this inner product, the Mori's projection operator,[5] or the finite rank projection operator, can be defined onto the span of a set of linearly independent basis functions $g_i(\mathbf{x})$, $i \in 1, \ldots, D$:

$$Pf(\mathbf{g}(\mathbf{x})) := \sum_{i,j=1}^{D} \langle f, g_i \rangle \left[\mathbf{C}_0^{-1} \right]_{i,j} g_j(\mathbf{x}), \tag{10}$$

where $\mathbf{C}_0^{-1}$ is the inverse of the covariance matrix $[C_0]_{i,j} = \langle g_i, g_j \rangle$, $i, j \in 1, \ldots, D$. For a special set of orthonormal basis functions $h_i(\mathbf{x})$, the covariance matrix becomes the identity matrix so that the projection operator can be simplified:

$$Pf(\mathbf{h}(\mathbf{x})) := \sum_{i=1}^{D} \langle f, h_i \rangle h_i(\mathbf{x}). \tag{11}$$

In general, one can use the Gram–Schmidt (G–S) procedure to identify the set of orthonormal functions $h_i(x)$ from $g_i(x)$.

- In Zwanzig's formulation,[6] the observables are chosen to be a subset of the variables $\mathbf{g}(\mathbf{x}) = \hat{\mathbf{x}}$ and the projection operator is defined using direct marginalization of the un-resolved variables. If the probability distribution μ for phase-space variable $\mathbf{x}$ is written for resolved/unresolved variables as a density function $\rho(\hat{\mathbf{x}}, \tilde{\mathbf{x}})$, the projection operator is then defined as

$$Pf(\hat{\mathbf{x}}) := \frac{\int f(\hat{\mathbf{x}}, \tilde{\mathbf{x}}) \rho(\hat{\mathbf{x}}, \tilde{\mathbf{x}}) d\tilde{\mathbf{x}}}{\int \rho(\hat{\mathbf{x}}, \tilde{\mathbf{x}}) d\tilde{\mathbf{x}}}. \tag{12}$$

The resulting function Pf is generally nonlinear in $\widehat{\mathbf{vx}}$, so this projection operator is also termed as nonlinear projection[19] or infinite rank projection.[31] This nonlinear projection operator has been adopted in Parish and Duraisamy[13,17] to construct

MZ-based models for turbulence, where the initial conditions were assumed to be fully resolved (i.e., $\tilde{\mathbf{x}} = 0$ at $t = 0$) and the unresolved observables were assumed to remain centered at 0 and delta distributed.

- A recently proposed Wiener projection is used to link the Nonlinear Auto-Regressive Moving Average with eXogenous input (NARMAX) to MZ formalism,[32] where the basis functions $\mathbf{g}$ also embed information from past history. Let f_n and g_n be two discrete-time zero mean wide-dense stationary processes, where subscript n denotes the index of time steps, and then the Wiener projection operator can be written as

$$Pf_n := \sum_{i=1}^{n} h_i g_{n-i}, \tag{13}$$

where the sequence h_i is the Wiener filter.

In this work, we focus on the Mori's finite rank projection operator and the corresponding constructed MZ kernels, which is the foundation of the data-driven algorithms proposed in Lin et al.[29]

1. A discrete-time Mori–Zwanzig formalism

Even though the dynamical system discussed above is formulated in continuous-time, it is common that in high-resolution simulations or experimental measurements, the outputs are discrete-time snapshots, where the temporal derivative is not readily available. Darve et al.[33] formulated the MZ formalism for discrete-time systems. Lin and Lu[32] established the link between NARMAX and MZ formulation of discrete-time system based on Wiener projections. In this work, we employ a similar discrete-time description, but we construct the MZ formulation based on the finite rank projection operator. For completeness, we introduce the discrete formulation of dynamical system and corresponding MZ formulation following Lin and Lu,[32] Darve et al.[33]

We write the dynamical equation for the full set of discrete solution vector $\boldsymbol{\Phi}(n\Delta) \in \mathbb{R}^N$ as

$$\boldsymbol{\Phi}((n+1)\Delta) = S_\Delta(\boldsymbol{\Phi}(n\Delta)), \quad \boldsymbol{\Phi}(0) = \mathbf{x}, \tag{14}$$

where n and Δ are the time step and time interval of the discrete-time snapshots, respectively. Similar to the continuous time derivation, we define the observables as $\mathbf{g}(\boldsymbol{\Phi}(n\Delta)) \in \mathbb{R}^D$, where the components g_i are functions in $L^2(\mu)$ that map the original solution $\boldsymbol{\Phi}$ to a physically observed quantity of interest: $\mathbb{R}^N \to \mathbb{R}$. For simplicity, we use $\mathbf{g}(n\Delta)$ to denote $\mathbf{g}(\boldsymbol{\Phi}(n\Delta))$. To describe the evolution of the observables $\mathbf{g}$, we introduce discrete-time Koopman operator $\mathscr{K}_\Delta$ that satisfies $[\mathscr{K}_\Delta \mathbf{g}](\boldsymbol{\Phi}) = (\mathbf{g} \circ S_\Delta)(\boldsymbol{\Phi})$, where the symbol $\circ$ is used to denote composite functions. By operating the Koopman operator on functions $\mathbf{g}$ and applying to the solution variable at the current state, we can obtain the observables at the future time step. We apply the Koopman operator $n + 1$ times and derive the evolution of observables:

$$\left[\mathscr{K}_\Delta^{n+1}\mathbf{g}\right](\mathbf{x}) = \left[\mathscr{K}_\Delta^n (\mathbf{g} \circ S_\Delta)\right](\mathbf{x}) = \left[\mathscr{K}_\Delta^n \mathbf{g}\right](\boldsymbol{\Phi}(\Delta))$$

$$= \mathbf{g}(\boldsymbol{\Phi}((n+1)\Delta)) = \mathbf{g}((n+1)\Delta). \tag{15}$$

With a given projection operator P on Hilbert space $L^2(\mu)$ and its complement $Q = I - P$, we can then write the Dyson identity for the Koopman operator[32]

$$\mathscr{K}_{\Delta}^{n+1} = \sum_{l=0}^{n} \mathscr{K}_{\Delta}^{n-l} P \mathscr{K}_{\Delta} (Q\mathscr{K}_{\Delta})^{l} + (Q\mathscr{K}_{\Delta})^{n+1}. \quad (16)$$

We substitute Eq. (16) in Eq. (15) and arrive at the evolutionary equation for observables **g**,

$$
\begin{aligned}
\mathbf{g}((n+1)\Delta) &= \left[\mathscr{K}_{\Delta}^{n+1}\mathbf{g}\right](\mathbf{x}) \\
&= \sum_{l=0}^{n}\left[\mathscr{K}_{\Delta}^{n-l}P\mathscr{K}_{\Delta}(Q\mathscr{K}_{\Delta})^{l}\mathbf{g}\right](\mathbf{x}) + \left[(Q\mathscr{K}_{\Delta})^{n+1}\mathbf{g}\right](\mathbf{x}), \\
&= \left[\mathscr{K}_{\Delta}^{n}P\mathscr{K}_{\Delta}\mathbf{g}\right](\mathbf{x}) + \sum_{l=1}^{n}\left[\mathscr{K}_{\Delta}^{n-l}P\mathscr{K}_{\Delta}(Q\mathscr{K}_{\Delta})^{l}\mathbf{g}\right](\mathbf{x}) \\
&\quad + \left[(Q\mathscr{K}_{\Delta})^{n+1}\mathbf{g}\right](\mathbf{x}) \\
&= \boldsymbol{\Omega}_{\Delta}^{(0)}(\mathbf{g}(n\Delta)) + \sum_{l=1}^{n}\boldsymbol{\Omega}_{\Delta}^{(l)}(\mathbf{g}((n-l)\Delta)) + \mathbf{W}_{n+1}(\mathbf{x}).
\end{aligned}
$$
$$(17)$$

The above equation is the discrete-time GLE and can be understood as the discrete counterpart of Eq. (7). The corresponding Mori–Zwanzig operators can then be identified for the three components in Eq. (17),

$$\boldsymbol{\Omega}_{\Delta}^{(0)}(\mathbf{g}(n\Delta)) = \left[\mathscr{K}_{\Delta}^{n}P\mathscr{K}_{\Delta}\mathbf{g}\right](\mathbf{x}) \quad (18)$$

$$\mathbf{W}_{n+1}(\mathbf{x}) = \left[(Q\mathscr{K}_{\Delta})^{n+1}\mathbf{g}\right](\mathbf{x}) \quad (19)$$

$$\boldsymbol{\Omega}_{\Delta}^{(l)}(\mathbf{g}((n-l)\Delta)) = \left[\mathscr{K}_{\Delta}^{n-l}P\mathscr{K}_{\Delta}(Q\mathscr{K}_{\Delta})^{l}\mathbf{g}\right](\mathbf{x}), \quad (20)$$

with the orthogonality condition $P\mathbf{W}_{n+1} = 0$, $\forall\, n \in \mathcal{N}$. Equation (17) is also general, and the specific forms of the MZ operators depend on the choice of projection operator.

B. Reduced-order construction of Navier–Stokes equations based on MZ formalism

In Sec. II A, we have derived and discussed the key components of MZ formalism and their dependence on the projection operator. In this section, we first demonstrate their application to Navier–Stokes turbulence modeling and establish the link between MZ formulation with the nonlinear projection operators and classical turbulence modeling approach, i.e., LES. Due to the challenges in extracting the properties of the corresponding nonlinear MZ operators, we propose the MZ formulation based on Mori's finite rank projection operator as a generalization of Koopman learning framework, which greatly simplifies the learning task and allows us to quantitatively address the assumptions in modeling.

1. LES and MZ formulation based on nonlinear projection operator

Consider the three-dimensional discretized velocity field in a fully resolved numerical simulation $v_i(t, n_x, n_y, n_z)$, where $i \in 1, 2, 3$ is the direction of velocity and $n_x, n_y, n_z \in 1, .., N_x, N_y, N_z$ the discrete spatial coordinates of the velocity field. We can stack the discretized solution of the velocity field at time t into a $N \times 1$ vector $\mathbf{v}(t) \in \mathbb{R}^N$, $N = 3 \times N_x \times N_y \times N_z$, and write the discretized incompressible Navier–Stokes equations with any given numerical scheme into a general form ODE, which follows Eq. (1),

$$\frac{d\mathbf{v}(t)}{dt} = R(\mathbf{v}(t)), \quad (21)$$

where R is the nonlinear function that can be viewed as representing the spatially discretized form of the right-hand side of the Navier–Stokes equations for given a numerical scheme.

For the spatially discretized NS equations, we can then derive the discrete-time formulation for the temporally discretized velocity vector $\mathbf{v}(n\Delta)$ and arrive at Eq. (14). Here, S_{Δ} encodes information of the temporal scheme, for example, $S_{\Delta}(\mathbf{v}) = I + \Delta \cdot R(\mathbf{v})$ for the Euler method if we assume that the time interval Δ is very small.

Fully resolving the dynamics of the Navier–Stokes equations requires prohibitively large amounts of computational resources due to the wide range of scales for practically relevant problems. In the classic approach of reduced-order modeling for turbulence, the velocity field is coarse-grained by applying a spatial filter to reduce the range of scales that need to be resolved. Here, we denote the solution vector of the filtered discretized velocity field as $\bar{\mathbf{v}}(t) := [\overline{v_i}(t, n_x, n_y, n_z)]^T$, $n_x, n_y, n_z \in 1, \ldots, N_{x,c}, N_{y,c}, N_{y,c}$, where the overline denotes the spatial filtering. Commonly used spatial filters include Gaussian filter, box filter, and spectral filter. The size of the computational mesh required to fully resolve $\bar{\mathbf{v}}$ is significantly reduced because of the reduced range of scales, so that $N_{x,c} \ll N_x$, $N_{y,c} \ll N_y$, $N_{z,c} \ll N_z$ and the resolved solution vector $\bar{\mathbf{v}} \in \mathbb{R}^D$ has a reduced dimension $D = 3 \times N_{x,c} \times N_{y,c} \times N_{z,c} \ll N$. As a result of spatially filtering the nonlinear NS equations, high-order moments emerge and the system for the resolved variable $\bar{\mathbf{v}}$ is not closed. Dynamical equations for the filtered velocities can be written as

$$\frac{d}{dt}\bar{\mathbf{v}}(t) = \bar{R}(\bar{\mathbf{v}}(t)) + \tau_{sgs}(\mathbf{v}(t)), \quad (22)$$

where $\bar{R}(\bar{\mathbf{v}}(t))$ takes the same form as the original NS equations but numerically discretized on a coarser grid and is fully closed, while $\tau_{sgs}(\mathbf{v}(t))$ denotes the unclosed sub-grid scale contributions. Transport equations for the higher-order moments can also be derived, such as equations for the sub-grid stress (SGS) in LES, but they depend on even higher-order unclosed terms. In practice, the resulting infinite dimensional system is truncated to include only the resolved variable $\hat{\boldsymbol{\Phi}} = \bar{\mathbf{v}}$, and a sub-grid model is used to compensate for the effects from the unresolved moments/scales, which is usually Markovian: $\tau_{sgs}(\bar{\mathbf{v}}(t))$.

In the framework of MZ formalism, we can employ a nonlinear projection operator similar to that used in Parish and Duraisamy[34] and write an evolution equation for the reduced-order variable $\bar{\mathbf{v}}$ as a GLE that consists of a Markov term, a memory term, and orthogonal dynamics:

$$\frac{d}{dt}\bar{\mathbf{v}}(t) = M(\bar{\mathbf{v}}(t)) - \int_0^t K(\bar{\mathbf{v}}(t-s), s)ds + F(t). \quad (23)$$

We remark that there may exist a projection operator that one can apply to the filtered NS equations and establish the connection between the terms in Eq. (22) and the terms in MZ formulation Eq. (23), but it is challenging to perform quantitative analyses of the sub-grid model in the MZ framework. In this section, we hope to shed some light on turbulence modeling from the perspective of MZ formalism. The traditional sub-grid scales models that are based on assumptions of scale similarity, universality of fully developed

turbulence, etc. have been shown to be inadequate in many complex flow problems, such as transitional flows, flows with separation, etc. Alternatively, we can develop turbulence models from the MZ framework, where more complex flow phenomena, such as transition to turbulence, can be incorporated into the memory kernel and orthogonal dynamics. This will alleviate the need for strong assumptions based on universality of fully developed turbulence to relate the small-scale fluctuations to velocities at the resolved scale $\bar{\mathbf{v}}$. With the addition of memory kernel and orthogonal dynamics, MZ-based turbulence models have the potential of resolving the issues in traditional turbulence modeling. On the other hand, it is generally difficult to accurately model the Markov term, memory kernel, and orthogonal dynamics due to limited knowledge of their properties and the challenge in extracting MZ operators using data/observations.

We also remark that the difficulty of extracting the memory kernel and orthogonal dynamics originates from two aspects: nonlinearity of the Navier–Stokes equations and that of the projection operator. In Sec. II B 2, we solve this issue by introducing Mori's finite rank projection operator and discuss its relation to the Koopman learning framework, which lays the foundation for the extraction of MZ kernels using a data-driven algorithm. We also remark that we only expressed Smagorinsky-type sub-grid models[35] in Eq. (22) for simplicity, while there exists a wider range of models, such as the one-equation model,[36] the dynamic model,[37] etc., in which certain memory effects are incorporated, even though indirectly.

2. MZ formulation based on Mori's projection operator

In the Koopman learning framework,[26] the same dynamical system in Eq. (1) can be characterized by a collection of observables $\mathbf{g}$, which are functions of the physical-space variables $\mathbf{\Phi}$. The system can then be cast from a finite-dimensional system of nonlinear ODEs describing the physical variables to an infinite-dimensional system of linear ODEs that describes all possible observables. In Koopman's formulation, the observables evolve on an infinite-dimensional Hilbert space $\mathcal{H}$, which is composed of all possible observables.

In the Koopman framework, deriving a closed-form solution is equivalent to identifying a set of observables whose dynamics are invariant in a subspace, which is linearly spanned by the set of the observables. In general, it is very challenging to identify the finite set of observables that close the dynamics and one has to resort to approximation methods to close the dynamics. Naturally, we can leverage the Mori–Zwanzig formalism and the inner product in the Hilbert space to construct the dynamical equations for the finite set of observables. Lin et al.[29] showed that by using Mori's finite rank projection operator [Eq. (10)], which depends on the inner product of two functions, both the Koopman[27] and MZ formulations operate in a shared Hilbert space. The advantage of using Mori's projection operator is that the projected low-dimensional functions are linear, which significantly simplifies the derivation/learning of the MZ kernels. This is in contrast to the MZ construction based on nonlinear projection operators as discussed in Sec. II B 1. Following Lin et al.,[29] the MZ formulations when using the Mori's projection operator with linearly independent basis functions $\mathbf{g}$ can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{g}(t) = \mathbf{M} \cdot \mathbf{g}(t) - \int_0^t \mathbf{K}(t-s) \cdot \mathbf{g}(s)\,\mathrm{d}s + \mathbf{F}(t), \quad (24)$$

where $\mathbf{M}$ and $\mathbf{vK(t)}$, $t \,\forall \geq 0$, are $D \times D$ matrices. Similarly, the discrete counterpart of MZ formulation based on Mori's projection operator for the discrete-time observable $\mathbf{g}(n\Delta) = \mathbf{g}(\mathbf{\Phi}(n\Delta))$ can be written as[29]

$$\mathbf{g}((n+1)\Delta) = \mathbf{\Omega}_\Delta^{(0)} \cdot \mathbf{g}(n\Delta) + \sum_{\ell=1}^{n} \mathbf{\Omega}_\Delta^{(\ell)} \cdot \mathbf{g}((n-l)\Delta) + \mathbf{W}_{n+1}$$

$$= \sum_{\ell=0}^{n} \mathbf{\Omega}_\Delta^{(\ell)} \cdot \mathbf{g}((n-l)\Delta) + \mathbf{W}_{n+1}. \quad (25)$$

Note that in the discrete form, $\mathbf{\Omega}_\Delta^{(0)}$ is the Markov operator, $\mathbf{\Omega}_\Delta^{(l)}$, $l \in 1, 2, 3\ldots$ are the memory kernels, $\mathbf{W}_{n+1}$ is the orthogonal dynamics, and Δ represents the discrete time step.[29] There is also a switch of sign in the memory kernel between the two types of formulation in Eqs. (24) and (25), which is because of the conventions in continuous and discrete formulations. In the rest of the paper, the MZ formulation will be mainly discussed using the discrete form, due to the fact that it simplifies the calculation by replacing integral with summation and temporal derivatives with numerical values.

We remark that Eq. (24) [or Eq. (25) for discrete-time formulation] is a special case of the general MZ formulation [Eqs. (7) and (17)], where the projection operator is chosen to be the finite rank projection operator. The basis functions $\mathbf{g}$ are not limited to the original physical-space variables, but can be any linearly independent functions of $\mathbf{\Phi}$. Naturally, we can construct an MZ formulation for the Navier–Stokes equations using Mori's projection operator. The form of the MZ formulation follows that in Eq. (24).

So far, we have presented two different Mori–Zwanzig formulations for the NS equations [Eqs. (24)/(25) and (23)/(17)], and their differences can be understood from two different perspectives. First, *the projection operators are different*: Eq. (23) employs a projection operator based on truncation, which results in a nonlinear Markov operator and memory kernels, while Eq. (24) is based on Mori's projection operator, which results in linear Markov and memory kernels. We have discussed the relation between Eq. (23) and LES modeling and difficulties of deriving/extracting corresponding memory kernels and orthogonal dynamics in Sec. II B 1. This difficulty can be alleviated using the linear MZ formulation in Eq. (24), but there is no direct link to classical LES modeling. Second, *the solution vectors are different*: Eq. (23) describes the evolution of physical-space variables (the filtered velocity field in the context of LES), while Eq. (24) describes the observables $\mathbf{g}$, which can be nonlinear functions of physical-space variables. The observables can also include the physical-space variables themselves in their set.

Given the simplicity of the Markov operator and memory kernels when using Mori's linear projection operator, we show that the Mori–Zwanzig operators can be extracted in a relatively straightforward manner using this formulation, unlike the one based on a nonlinear projection operator. We also continue the discussions using the discrete-time formulation because: (i) observations and simulation results are usually discrete and fully resolved time derivatives are usually not readily available, and (ii) discrete-time formulation can avoid the errors induced by numerically integrating the continuous-time counterpart.

C. The evolution of the two-time correlation matrix

In this section, we derive the evolution equation of the two-time correlation matrix $\mathbf{C}$, which is the foundation of the learning

algorithm. For Mori's projection operator, we choose to use the initial condition $\mathbf{g}(0)$ as the basis of the projected linear subspace. We then apply an inner product $l\langle \mathbf{v}\mathbf{g}^T(0)r\rangle$ to Eq. (25) and obtain an evolutionary equation for the two-time correlation function, $\mathbf{C}(n\Delta)$,

$$\mathbf{C}(n\Delta) = \langle \mathbf{g}(n\Delta), \mathbf{v}\mathbf{g}^T(0)\rangle, \qquad (26)$$

$$\mathbf{C}((n+1)\Delta) = \boldsymbol{\Omega}_\Delta^{(0)} \cdot \mathbf{C}(n\Delta) + \sum_{\ell=1}^{n} \boldsymbol{\Omega}_\Delta^{(\ell)} \cdot \mathbf{C}((n-l)\Delta)$$

$$= \sum_{\ell=0}^{n} \boldsymbol{\Omega}_\Delta^{(\ell)} \cdot \mathbf{C}((n-l)\Delta). \qquad (27)$$

Note that with this procedure, we exploit the orthogonality between the basis function $\mathbf{g}(0)$ and the noise $\mathbf{W}_{n+1}$ to remove the complex orthogonal dynamics in the evolution equation of $\mathbf{C}$. The resulting formula [Eq. (27)] builds the foundation for the data-driven learning algorithm.

D. Generalized fluctuation-dissipation relation

The relation between the memory kernel and the orthogonal dynamics, i.e., Eq. (9c), is commonly referred to as the generalized fluctuation-dissipation relation (GFD). There exist different interpretations of this relation but, in general, it imposes a structural relation between the memory kernel and orthogonal dynamics. Thus, these operators cannot be approximated using models in an arbitrary manner. This relation has been used to estimate the memory kernel when a model for orthogonal dynamics is proposed.[18]

When constructing the MZ formulation using Mori's projection operator, a specific form of the GFD can be derived for the two-time correlation matrix, if the Liouville operator $\mathscr{L}$ is anti-self-adjoint with respect to the chosen inner product,

$$\langle f, \mathscr{L}h\rangle = -\langle \mathscr{L}f, h\rangle, \qquad (28)$$

for any functions f and h of the physical-space variable $\boldsymbol{\Phi}$. Note that the anti-self-adjoint property depends on the choice of the inner product. Lin et al.[29] showed that the Liouville operator of a dynamical system is anti-self-adjoint if the inner product is defined as the temporally averaged value of the product of the test functions evaluated on a long trajectory, provided the observables are bounded along the trajectory. The specific GFD for the discrete-time MZ formulation with Mori's projection operator is[29]

$$\boldsymbol{\Omega}_\Delta^{(l)} = -\langle \mathbf{W}_{l+1}, \mathbf{v}\mathbf{W}_1^T\rangle \mathbf{C}^{-1}(-\Delta), \quad \forall l \in 1, 2, 3..., \qquad (29)$$

where $\mathbf{C}(-\Delta) = \mathbf{C}^T(\Delta)$. This non-trivial relation should be satisfied if the kernels are correctly extracted from the data and will be verified using numerical data.

III. DATA-DRIVEN LEARNING OF THE MZ OPERATORS AND ORTHOGONAL DYNAMICS

In this section, we briefly describe the learning algorithm proposed in Lin et al.[29] The learning procedure is based on Eq. (27) and starts by calculating the two-time correlation matrix $\mathbf{C}(n\Delta) = \langle \mathbf{g}(n\Delta), \mathbf{g}(0)^T\rangle$. As mentioned in Sec. II A, the evaluation of the inner product requires taking the expectation value against the stationary distribution $d\mu$ or temporally and uniformly sampling/averaging along a long trajectory. For non-stationary systems, the distribution of

the system state is time-dependent. Thus, the inner product needs to be performed over an ensemble of simulations, whose initial conditions are drawn from the initial distribution of the states for each specific problem. In this work, we focus on stationary turbulent flows.

Given a long and evenly spaced trajectory of a physical space variable (velocity field) from the fully resolved direct numerical simulation $\boldsymbol{\Phi}(n\Delta), n \in 0, 1...N_t - 1$, the two-time correlation matrix $\mathbf{C}(n\Delta)$ can be calculated as

$$\mathbf{C}(n\Delta) = \frac{1}{N_t - n} \sum_{i=0}^{N_t-n-1} \mathbf{g}(\boldsymbol{\Phi}((n+i)\Delta)) \cdot \mathbf{g}^T(\boldsymbol{\Phi}(i\Delta)), \qquad (30)$$

where N_t is the total number of snapshots and $N_t\Delta \gg T_l$, where T_l is the integral timescale of turbulence. The calculation of the two-time correlation matrix consumes majority of the computational time of the proposed algorithm. To speed up the computation for large dynamical systems, one can impose sparsity constraints on the correlation matrix. In this work, we implement known symmetries of the physical system for data augmentation, which can further improve the accuracy in extracting Mori–Zwanzig operators without generating more data. For isotropic turbulence with triply periodic boundaries, periodicity and rotational symmetries are satisfied and can be used to facilitate data augmentation. This can be implemented in the calculation of the two-time correlation matrix. Suppose there exist N_s symmetric representations of the same physical-space variable $\mathscr{S}_{n_s}(\boldsymbol{\Phi}), n_s = 1, ...N_s$, where $\mathscr{S}_{n_s}$ is the symmetry operator that preserve the statistics of the original physical-space variables $\boldsymbol{\Phi}$. One of such operator could be rotating the velocity field such that $u_1 \to u_2, u_2 \to u_3, u_3 \to u_1$, and the dynamics are the same for all the symmetric representations. Naturally, Eq. (30) can then be modified to

$$\mathbf{C}(n\Delta) = \frac{1}{N_s} \frac{1}{N_t - n} \sum_{n_s=1}^{N_s} \sum_{i=0}^{N_t-n-1} \mathbf{g}(\mathscr{S}_{n_s}(\boldsymbol{\Phi}((n+i)\Delta)))$$

$$\cdot \mathbf{g}^T(\mathscr{S}_{n_s}(\boldsymbol{\Phi}(i\Delta))). \qquad (31)$$

After the calculation of two-time correlation matrix $\mathbf{C}(n\Delta)$, we can set $n = 0$ in Eq. (27) to obtain the Markov operator $\boldsymbol{\Omega}_\Delta^{(0)}$,

$$\boldsymbol{\Omega}_\Delta^{(0)} = \mathbf{C}(\Delta) \cdot \mathbf{C}^{-1}(0). \qquad (32)$$

We can then recursively solve for the memory kernel $\boldsymbol{\Omega}_\Delta^{(n)}$ using two-time correlation matrix $\mathbf{C}(n\Delta)$, Eq. (27) and previously solved low-order $\boldsymbol{\Omega}_\Delta^{(l)}, l < n$,

$$\boldsymbol{\Omega}_\Delta^{(n)} = \left(\mathbf{C}((n+1)\Delta) - \sum_{l=0}^{n-1} \boldsymbol{\Omega}_\Delta^{(l)} \mathbf{C}((n-l)\Delta)\right) \cdot \mathbf{C}^{-1}(0). \qquad (33)$$

After we obtain the memory kernel, the orthogonal dynamics can then be extracted for each section in the trajectory using Eq. (25),

$$\mathbf{W}_{n+1}^{(i)} = \mathbf{g}(\boldsymbol{\Phi}((i+n+1)\Delta)) - \sum_{l=0}^{n-1} \boldsymbol{\Omega}_\Delta^{(l)} \mathbf{g}(\boldsymbol{\Phi}((i+n-l)\Delta)). \qquad (34)$$

After obtaining $\mathbf{W}_{n+1}^{(i)}$, the properties of the orthogonal dynamics can be studied, such as the two-time correlation $\langle \mathbf{W}_{n+1}, \mathbf{v}\mathbf{W}_1^T\rangle$, etc.

IV. GROUND TRUTH TURBULENCE DATA

The "ground-truth" data are generated from the Eulerian DNS solution of the incompressible Navier–Stokes equations on a $(2\pi, 2\pi, 2\pi)$ domain,

$$\frac{\partial v_i}{\partial t} + \frac{\partial v_i v_j}{\partial x_j} = -\frac{\partial p}{\partial x_i} + \nu \frac{\partial^2 v_i}{\partial x_j x_j}, \qquad (35)$$

where pressure p is obtained by solving the Poisson equation which enforces the zero velocity divergence constraint, and ν is the Kinematic viscosity. The isotropic turbulence is generated on a 128^3 grid using the pseudo-spectral method. A large-scale forcing term is applied to prevent turbulence from decaying. Time advancement is achieved through Adams–Bashforth–Moulton method. The Taylor Reynolds number when the turbulence reaches a statistically steady state is approximately 100. See Petersen and Livescu,[38] and Daniel et al.[39] for more details on the numerical method.

After the turbulent flow becomes fully developed, the 3D snapshots of the flow field are stored in consecutive time steps to generate a long trajectory of turbulence data. The total length of the trajectory is approximately $3000T_l$, where T_l is the integral timescale. We then apply the post-processing procedure to obtain the low-dimensional coarse-grained observables. The choices of observables could significantly affect the properties of the Markov operator, memory kernels, and the noise; therefore, in this study, we select several sets of observables that are closely related to the canonical turbulence modeling approach (LES) and turbulence theory. To summarize, the following procedures are used in this work to obtain the observables: (1) similar to LES, we first apply spatial filters to the velocity field at each time step with a wide range of filter sizes l_Δ to obtain the filtered velocity $\overline{v_i}$, (2) various types of observables, such as filtered velocity, pressure, subgrid stress, kinetic energy, etc., are computed, (3) the selected observables on the fine mesh ($128 \times 128 \times 128$) are then uniformly sampled onto a coarse mesh ($4 \times 4 \times 4$), and (4) the resulting observables at each time step are stacked into a single vector. Following these coarse-graining steps, a dimension reduction from 3×128^3 physical-space variables to $n_{obs} \times 4^3$ observables can be achieved, where n_{obs} is total number of function/variable types. We include the filtered velocities in all sets of observables. In this work, we consider the following four sets of observables for extracting MZ kernels:

- Observable set 1 ($n_{obs} = 3$): $\overline{v_1}, \overline{v_2}, \overline{v_3}$.
- Observable set 2 ($n_{obs} = 15$): $\overline{v_1}, \overline{v_2}, \overline{v_3}, \overline{v_1}\,\overline{v_1}, \overline{v_2}\,\overline{v_2}, \overline{v_3}\,\overline{v_3}, \overline{v_1}\,\overline{v_2}, \overline{v_1}\,\overline{v_3}, \overline{v_2}\,\overline{v_3}, \overline{v_1 v_1} - \overline{v_1}\,\overline{v_1}, \overline{v_2 v_2} - \overline{v_2}\,\overline{v_2}, \overline{v_3 v_3} - \overline{v_3}\,\overline{v_3}, \overline{v_1 v_2} - \overline{v_1}\,\overline{v_2}, \overline{v_1 v_3} - \overline{v_1}\,\overline{v_3}, \overline{v_2 v_3} - \overline{v_2}\,\overline{v_3}$,
- Observable set 3 ($n_{obs} = 14$): $\overline{v_1}, \overline{v_2}, \overline{v_3}, \overline{v_1}\,\overline{v_1} + \overline{v_2}\,\overline{v_2} + \overline{v_3}\,\overline{v_3}, S_{11}, S_{22}, S_{12}, S_{13}, S_{23}, W_{12}, W_{13}, W_{23}, S_{ij}S_{ij}, W_{ij}W_{ij}$, where $S_{ij} = \frac{1}{2}\left(\frac{\partial \overline{v_i}}{\partial x_j} + \frac{\partial \overline{v_j}}{\partial x_i}\right), W_{ij} = \frac{1}{2}\left(\frac{\partial \overline{v_i}}{\partial x_j} - \frac{\partial \overline{v_j}}{\partial x_i}\right)$
- Observable set 4 ($n_{obs} = 15$): $\overline{v_1}, \overline{v_2}, \overline{v_3}, \frac{\partial \overline{v_1 v_1}}{\partial x_1}, \frac{\partial \overline{v_2 v_2}}{\partial x_2}, \frac{\partial \overline{v_3 v_3}}{\partial x_3}, \frac{\partial \overline{v_1 v_2}}{\partial x_1}, \frac{\partial \overline{v_1 v_2}}{\partial x_2}, \frac{\partial \overline{v_1 v_3}}{\partial x_1}, \frac{\partial \overline{v_1 v_3}}{\partial x_3}, \frac{\partial \overline{v_2 v_3}}{\partial x_2}, \frac{\partial \overline{v_2 v_3}}{\partial x_3}, \frac{\partial \overline{p}}{\partial x_1}, \frac{\partial \overline{p}}{\partial x_2}, \frac{\partial \overline{p}}{\partial x_3}$.

Note that in observable set 3, not all components of the strain rate tensor S_{ij} are chosen as basis functions for observables because there exists a linear dependence of the diagonal components from the incompressibility condition. In addition to the above-mentioned functions for each set of the observables, a constant function $g_0 = 1$ is added to every set of observables, making the total number of

observables $n_{obs} \times 4^3 + 1$. In Fig. 1, we show the 2D velocity vector field and contours of observables in different observable sets to help the readers understand the flow field.

V. RESULTS

In this section, we present the results and properties of the extracted MZ kernels. We first address the statistical convergence of the learned MZ operators from "ground-truth" data. The non-trivial GFD relation between the memory kernels and orthogonal dynamics is also verified. The properties of the Markov operator, memory kernel, and noise, and their dependence on different choices of observables are then analyzed. Finally, we conduct numerical experiments using the extracted kernels to investigate the effects of adding the memory kernel on prediction.

A. Statistical convergence

The statistical convergence of the computed two-time correlation matrix $\mathbf{C}$ and the learned kernel is an important factor in the proposed learning algorithm[29] and needs to be confirmed in order to reduce the effects of statistical variability on the analysis. The accurate computation of the two-time correlation matrix $\mathbf{C}$ in the ergodic system requires averaging over a long trajectory. This requires a large amount of data samples, which represent the distribution of the stationary system. As described in Sec. IV, we performed fully resolved simulations to generate a long trajectory of 3D turbulence data and stored total N_t 3D snapshots of velocity fields. In the convergence test, three different sampling methods are used to sample from the database:

- Method 1: The convergence test data are randomly sampled from the total N_t snapshots.
- Method 2: The total N_t snapshots are first coarsely sampled in time to make sure that any two snapshots are at least one integral timescale apart. An integral timescale $T_l \approx N_{int}dt$ corresponds to approximately N_t/N_{int} snapshots. The convergence test data are then randomly sampled from the N_t/N_{int} snapshots. This procedure ensures that the data are not temporally correlated and are truly independent samples from the stationary distribution.
- Method 3: Similar to Method 2, but the time differences between two snapshots are reduced to half integral timescale. This works as an intermediate case between Method 1 and Method 2.

After obtaining the samples from the database, we apply the post-processing procedure as discussed in Sec. IV to the 3D simulation data in order to generate the vectors of observables $\mathbf{g}$. In the convergence test, the observable set 1 is chosen with the a spatial filtering length $\pi/8$. Data augmentation-based periodicity and rotational symmetry are also performed for the statistical convergence test. The discrete time step Δ is chosen to be $10dt$. Figure 2 shows the percentage changes of the Frobenius norm of the learned Markov operator and memory kernels with different number of samples, where the percentage change is calculated using formula: $\frac{\|\|\mathbf{\Omega}_\Delta^{(l)}\|_{F,n+1} - \|\mathbf{\Omega}_\Delta^{(l)}\|_{F,n}\|}{\|\mathbf{\Omega}_\Delta^{(l)}\|_{F,n}}$. Here, the subscript F denotes the Frobenius norm and n denotes the number of samples used for calculating the Markov operator $\mathbf{\Omega}_\Delta^{(0)}$ and memory kernels $\mathbf{\Omega}_\Delta^{(l)}, l \in 1, 2, 3....$. The percentage error fluctuates as the number of sample increases, but there exists a converging trend for the upper bound of the fluctuations. By using the largest number of
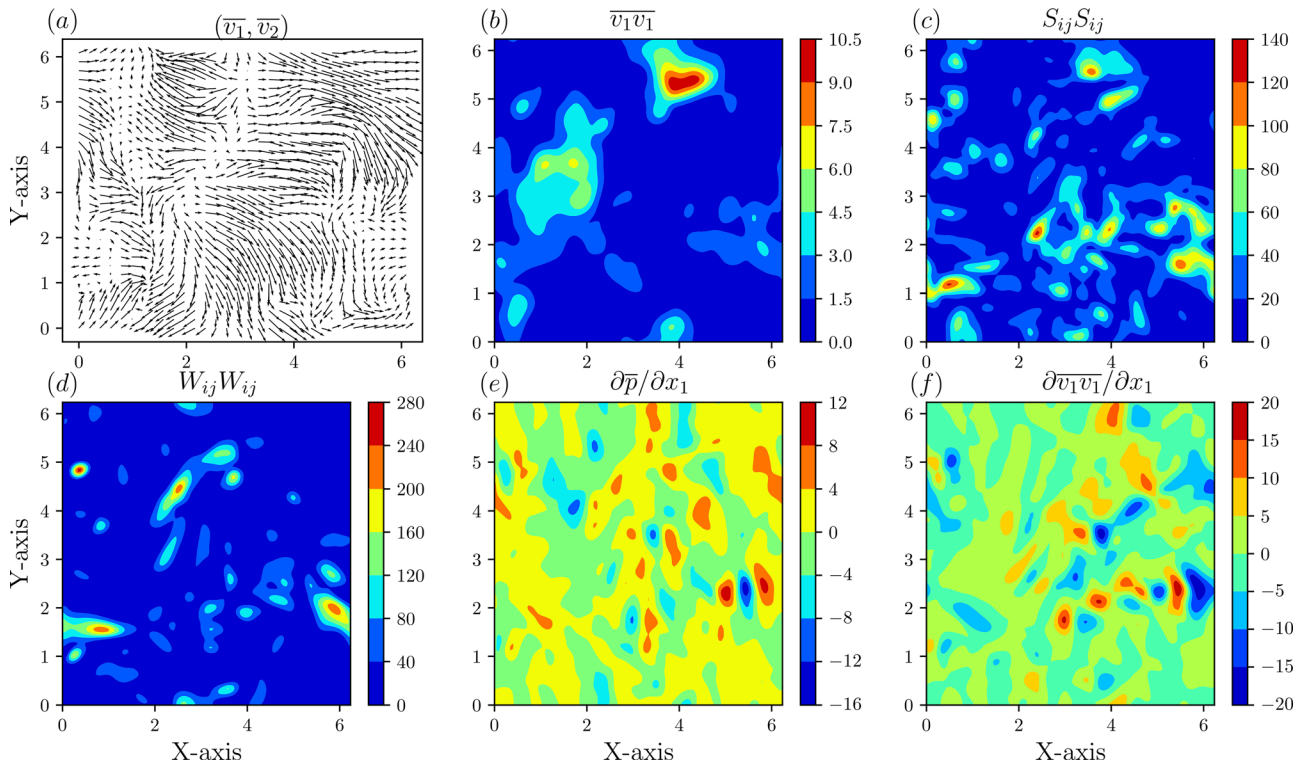
FIG. 1. 2 D snapshots of the flow field and observables from the DNS data. (a) Filtered velocity vector field (found in all observable sets), (b) sub-filter stress (found in observable set 2), (c) magnitude of strain rate tensor (found in observable set 3), (d) magnitude of rotation tensor (found in observable set 3), (e) filtered pressure gradient (found in observable set 4), and (f) gradient of sub-filter stress (found in observable set 4).

samples from the database, the upper bound of the fluctuations can be reduced to $10^{-7} \sim 10^{-6}$, implying that the final percentage error is less than 10^{-6}. Further increasing the number of samples only yields negligible improvements in the accuracy of the learned kernels. It is

also interesting to note that there is no difference in the rate of convergence among the three sampling methods. Considering this, we will use Method 1 for sampling, as it uses the largest amount of data, thus resulting in the smallest statistical error in the learned kernel.
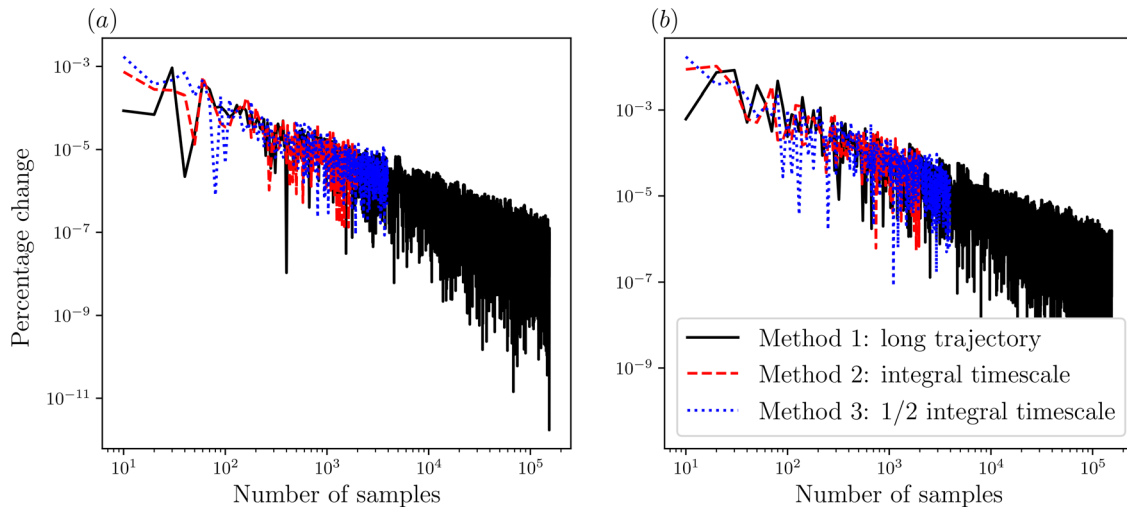
FIG. 2. The percentage error of Frobenius norm of the learned (a) Markov operator $\Omega_\Delta^{(0)}$ and (b) memory kernel $\Omega_\Delta^{(l)}$ for a time delay of approximately 5 Kolmogorov time scales ($\tau_\eta \approx 5.7 \times 10^{-2}$) as the number of samples increases is shown to verify the statistical convergence. The percentage error is calculated as $\frac{|\|\Omega_\Delta^{(l)}\|_{F,n+1} - \|\Omega_\Delta^{(l)}\|_{F,n}|}{\|\Omega_\Delta^{(l)}\|_{F,n}}$.

Considering that for large matrices like the Markov term and memory kernel the convergence of their Frobenius norm does not necessarily translate to the convergence of individual components of the matrices, we provide further test to ensure the full convergence of each component. Figure 3 shows the convergence of the learned individual components of the Markov operator and memory kernels as the number of samples increases. Three components, $[\mathbf{\Omega}_\Delta^{(0)}]_{1,1}$, $[\mathbf{\Omega}_\Delta^{(0)}]_{1,2}$, and $[\mathbf{\Omega}_\Delta^{(0)}]_{1,4}$, are shown in Fig. 3(a), which correspond to Markovian contributions from different observables. It can be seen that both $[\mathbf{\Omega}_\Delta^{(0)}]_{1,1}$ (representing contributions from $\overline{v_1}$ to $\overline{v_1}$) and $[\mathbf{\Omega}_\Delta^{(0)}]_{1,4}$ (represents contributions from $\overline{v_1}$ to $\overline{v_1}$ at neighboring points) values achieve convergence when the number of samples is larger than 10^5. On the other hand, the component $[\mathbf{\Omega}_\Delta^{(0)}]_{1,2}$ (representing contributions from $\overline{v_1}$ to $\overline{v_2}$) shows a decreasing trend as the number of sample increases and reaches almost three orders of magnitude smaller than the dominant components in the matrix. Intuitively, the correlation between velocity components should vanish in an isotropic turbulent field, which implies that $[\mathbf{\Omega}_\Delta^{(0)}]_{1,2}$ is trivial in the Markov operator, and further increasing samples would result in negligible improvements. Ideally, one could inject known physics into the learning framework to impose constraints on the learned kernels. However, in our first attempt in extracting MZ operators, we only apply the physical symmetries in the learning and let the data to inform/reveal the structure of the MZ operators. Figure 3(b) shows similar results for the memory kernel $\mathbf{\Omega}_\Delta^{(1)}$, which provide evidence for the convergence of the learning procedure.

B. Generalized fluctuation–dissipation relation

The generalized fluctuation–dissipation relation (GFD) refers to the subtle self-consistent relationship between the learned memory kernel $\mathbf{\Omega}_\Delta^{(\ell)}$ and orthogonal dynamics $\mathbf{W}_{l+1}$ with a suitable choice of projection operator. The specific GFD has been derived in Lin *et al.*[29] for Mori's projection operator. This relation can be used to verify the correctness of the learned memory kernel and orthogonal dynamics. In Fig. 4, we present the comparison between the left-hand-side (LHS) and the RHS of Eq. (29). The computed LHS aligns well with the RHS, confirming that the GFD relation is satisfied.

C. Properties of the learned Mori–Zwanzig operators

There have been very few attempts to perform quantitative analysis of the memory kernels and orthogonal dynamics of Navier–Stokes equations, due to the difficulty in developing tools for accurately and efficiently extracting them with nonlinear projection operators. The lack of knowledge makes it difficult to justify various assumptions on turbulence models based on MZ. In this section, we unveil some of the important properties of the extracted memory kernels and orthogonal dynamics with the current learning framework, in order to lay foundations for future turbulence model development.

In the discrete Mori–Zwanzig formulation, a hyper-parameter that is not related to the physical system is the discrete time interval Δ. Given a database for a long trajectory with fixed temporal step dt, one can extract the corresponding Markov operator and memory kernels with the discrete time interval Δ being an integer number of the temporal step $\Delta = n \, dt$. Lin *et al.*[29] showed that for the Lorenz '96 system, the memory kernels extracted with different Δ values collapse onto each other after applying a normalization factor Δ^{-2}. Figure 5 shows the components of the extracted memory kernel (observable set 1) with different Δ values as a function of the time delay (normalized by Kolmogorov timescale $\tau_\eta \approx 5.7 \times 10^{-2}$) for the NS turbulence system. After proper normalization, the memory kernels extracted with different Δ values also collapse onto the same curve, despite a minute smoothing at the largest value ($\Delta = 0.32$). This shows that the structure of the memory kernel and memory length do not depend strongly on the discrete time interval Δ after proper scaling.
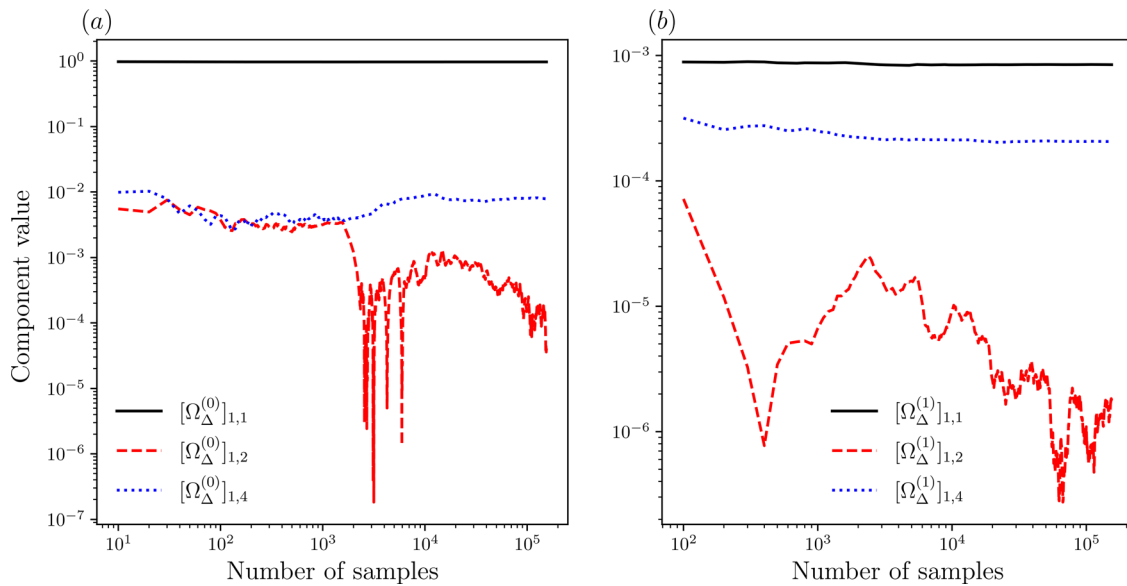

FIG. 3. The convergence of components in the learned Markov operator (a) and memory kernel (b) as the number of samples increases.
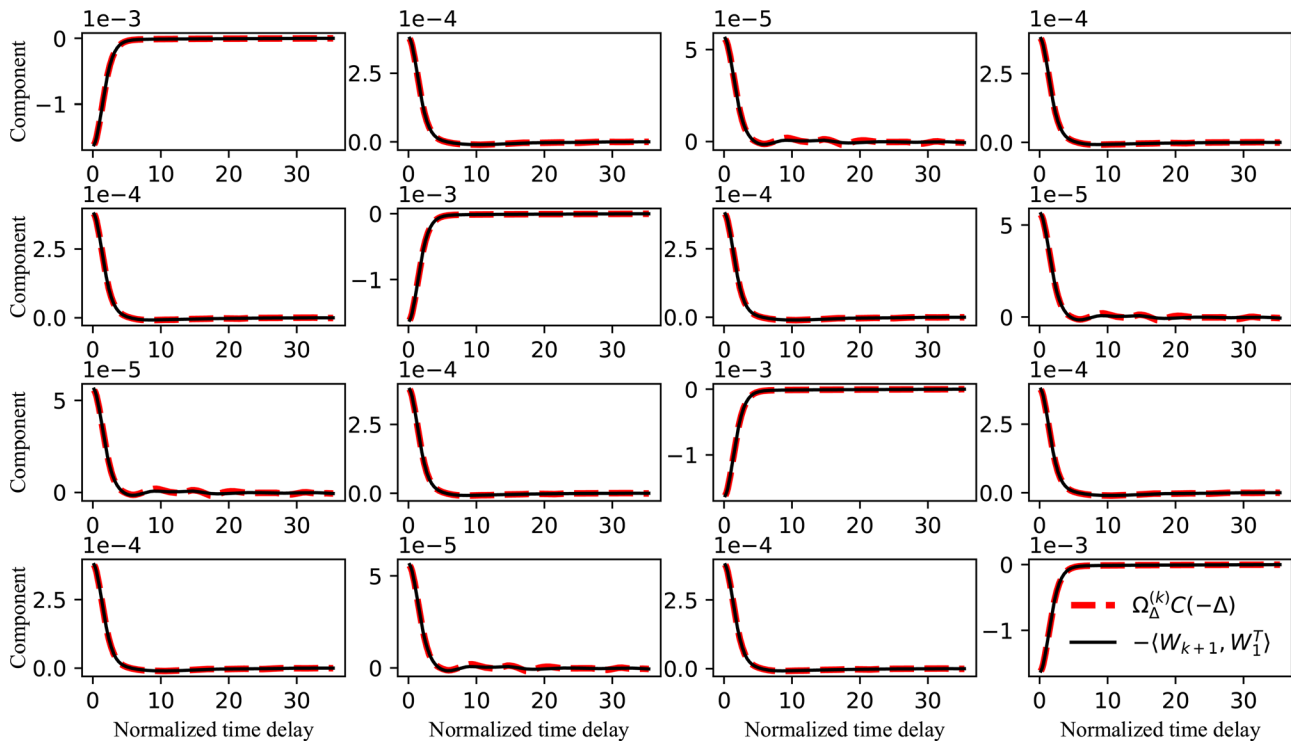
FIG. 4. Numerical validation of the discrete-time GFD relation for $\Delta = 10dt$. The individual components of the matrices in the GFD relation [Eq. (29)], $[\boldsymbol{\Omega}_\Delta^{(l)} \cdot \mathbf{C}(-\Delta)]_{ij}$ and $-[\langle \mathbf{W}_{l+1}, \mathbf{W}_1^T \rangle]_{ij}$, $i, j \in 1, 2, 3, 4$ are shown as a function of normalized time delay using Kolmogorv timescale. The LHS of Eq. (29) is computed using the learned memory kernel, $\boldsymbol{\Omega}_\Delta^{(l)}$ and two-time correlation matrix $\mathbf{C}(-\Delta)$. It is in very good agreement with the RHS: two-time correlation of the learned orthogonal dynamics $-\langle \mathbf{W}_{l+1}, \mathbf{W}_1^T \rangle$.

1. The effects of spatial filters

One of the most important assumptions in MZ-based models is about the length of the memory kernel. In the popular t-model,[11] which has been applied to many nonlinear dynamical systems and achieved certain success, the convolution integral of the memory term is carried out using a simple left-hand quadrature rule, which can be interpreted as assuming an infinite long memory length. Other MZ-based models[13,16,17,40] also make various assumptions and approximations on the length and shape of the memory kernel.

In Fig. 6, we present the Frobenius norm of Markov operator and the temporal decay of memory kernel. Two types of spatial filters that are commonly used in LES are also used here to compute the coarse-grained observables, in order to understand the effects of the filter type on memory length. In addition, the coarse-graining length scale, as reflected by the filtering length l_Δ, is also examined. Figure 6(a) shows the dependence of Markovian contribution on the spatial filter sizes (normalized by Kolmogorov length scale η). It is noted that the Markovian contribution decreases as the filtering size increases. In Fig. 6(b), the Frobenius norm of memory kernel (normalized by its corresponding Markov operator) is plotted against the normalized time delay. From Fig. 6(b), we can make a few important observations. First, the Frobenius norm of the memory kernel does not decrease to zero with a finite time delay; however, it becomes 2–3 orders of

magnitude smaller at a time delay around several Kolmogorov timescales. This indicates that using finite support in the memory integral can be a reasonable modeling assumption because the contributions from large time delays are generally negligible. Second, the difference between the two different spatial filter types is small, but the effects of the filtering length scale are significant. With larger filtering sizes, the temporal decay of the memory kernel becomes slower, making the finite memory length longer, which indicates a shift of dynamical contributions from Markov term to memory integral. Generally, when the filtering length scale increases, the range of scales that can be resolved by the coarse-grained observables becomes smaller. In this case, more past history of the observables needs to be included in the dynamics and, as discussed later, in the prediction, because the memory kernel formally characterizes the interactions between the coarse-grained dynamical variables and the under-resolved degrees of freedom. These observations suggest a qualitative statement that the more we coarse-grain the observables, the less Markovian the coarse-grained model should be.

So far, we have shown that the finite memory length assumption in the MZ-based models for NS turbulence is generally reasonable. However, the quantitative estimation of such memory length/timescale can be challenging. Various approximation methods have been proposed to calculate the finite memory length to construct MZ-based models for turbulence. Parish and Duraisamy[13] used a dynamical
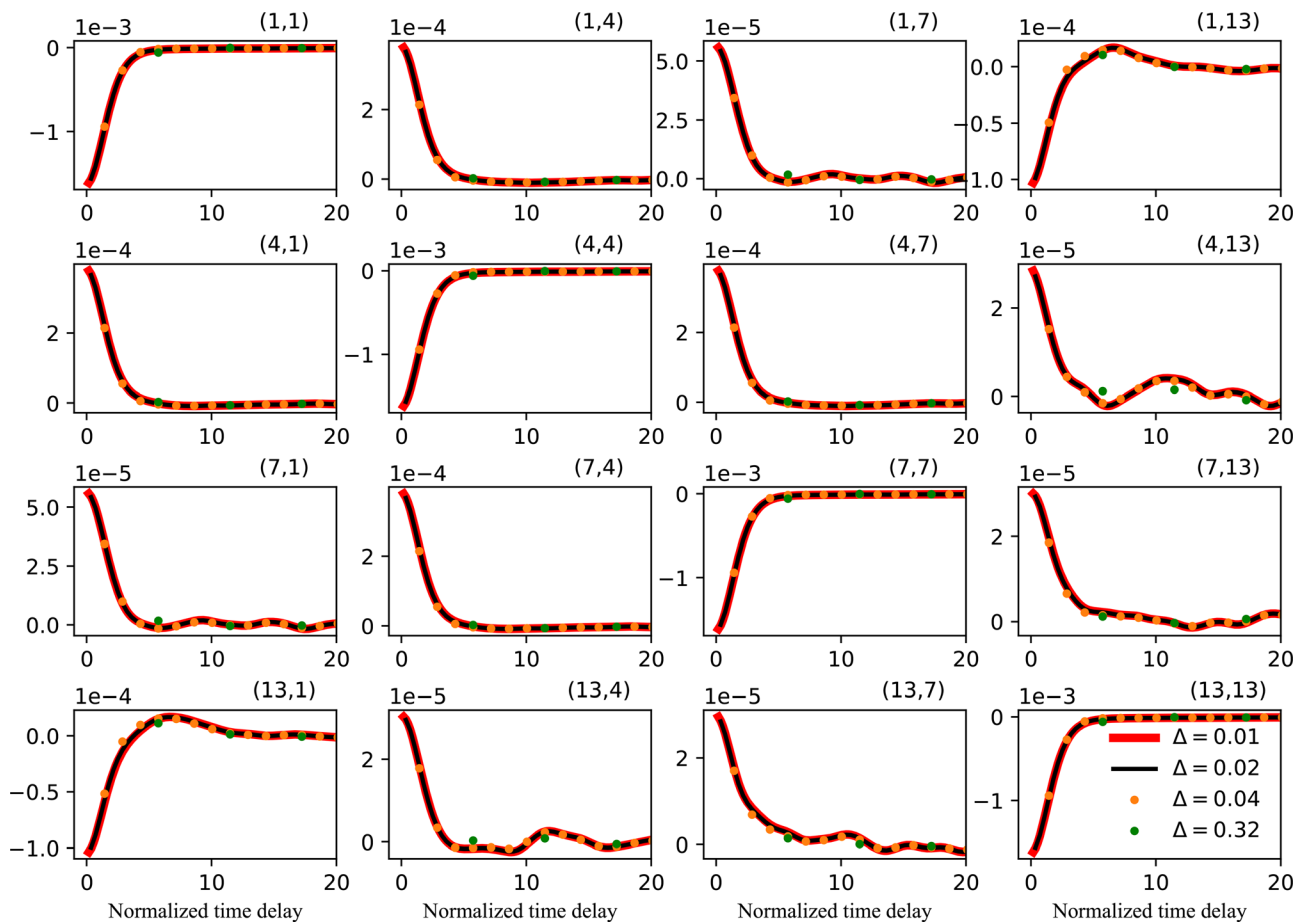
FIG. 5. The components of the learned memory kernel for different discrete time intervals $\Delta = 0.01, 0.02, 0.04, 0.032$ as a function of normalized time delay. The normalization is performed using Kolmogorov timescale, $\tau_\eta \approx 5.7 \times 10^{-2}$. The observable set 1 is used for computing the memory kernels.

procedure based on Germano identity[41] and Parish and Duraisamy[17] used the spectral radius of the Jacobian of resolved variables to estimate the finite memory length. With the extracted memory kernels in the current study, we are able to calculate the timescales of the memory kernel and use them to verify the assumptions in these models. Figure 7 shows the extracted timescales (normalized by Kolmogorov timescale) for various spatial filtering sizes (normalized by Kolmogorov scale). Two methods are employed for the calculation: (1) the integral of the memory kernel divided by the value at the smallest time delay and (2) the time delay when the memory kernel dropped to 10% of the maximum value. Note that we used a mean timescale based on the Frobenius norm to quantify the finite memory length, but the timescales may vary for different components of the memory kernel. We observe that the memory length is generally short, within the range of several Kolmogorov timescales. The memory length increases with the increase in the coarse-graining size, in agreement with those estimated by previously proposed models. There exists a weak dependence on the type of filters, which should be taken into account for modeling.[13] We remark that it may seem that including memory integral may significantly increase the storage overhead,

but when conducting reduced-order simulations using MZ-based models, it may not be necessary to store all the past information to calculate the memory kernel. Methods have been proposed based on the quadrature rule to model the memory integral as an additional set of ODEs,[40] which are solved alongside the main coarse-grained equations. One of the aims of this work is to provide validation for constructing MZ-based memory closure models.

Addressing the noise/orthogonal dynamics is another essential component in developing models in Mori–Zwanzig formalism. Similar to the memory kernel, it is extremely difficult to theoretically model the high-dimensional orthogonal dynamics, especially in the absence of data. To circumvent this difficulty, previous models based on MZ formalism have been focused on the projected image, where the noise term vanishes due to its orthogonality to the projection operator. Under the current learning framework, the noise term can be numerically extracted using the DNS database. By analyzing its properties, we hope to shed some light on the modeling of the orthogonal dynamics.

Figure 8(a) shows the temporal two-time correlation of the noise, as well as the PDFs of the orthogonal dynamics for the same set of
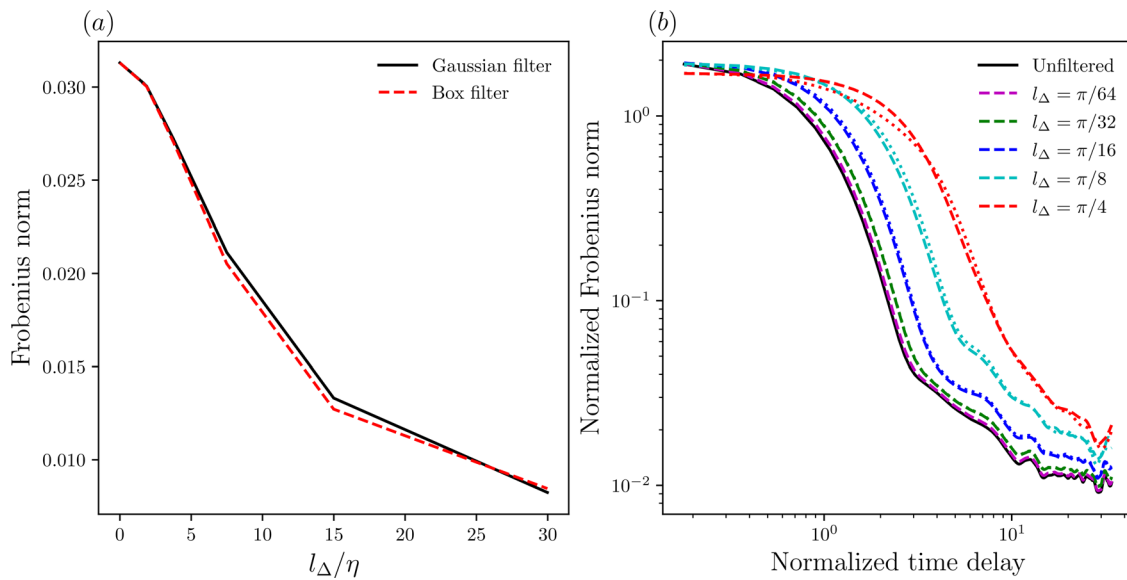
FIG. 6. (a) The Frobenius norm of Markov operator as a function of normalized filter sizes (normalized by Kolmogorov length scale, $\eta \approx 2.62 \times 10^{-2}$) and (b) normalized Frobenius norm of the learned memory kernel for observable set 1, $||\Omega_\Delta^{(\ell)}||_F$ as a function of normalized time delay. Two types of spatial filters, Gaussian and box filters, with various filtering length scales are applied to the physical-space variables.

observables as in Fig. 6. We observe that noise term has a similar temporal decay to the memory kernel. This behavior is expected, due to the GFD relation. Thus, Eq. (29) states that the temporal autocorrelation matrix of orthogonal dynamics $\langle \mathbf{W}_{l+1}, \mathbf{W}_1^T \rangle$ with a time delay $l\Delta$ should be equal to the memory kernel with same time delay multiplied by a constant matrix $\mathbf{C}(-\Delta)$, so it is reasonable for them to have
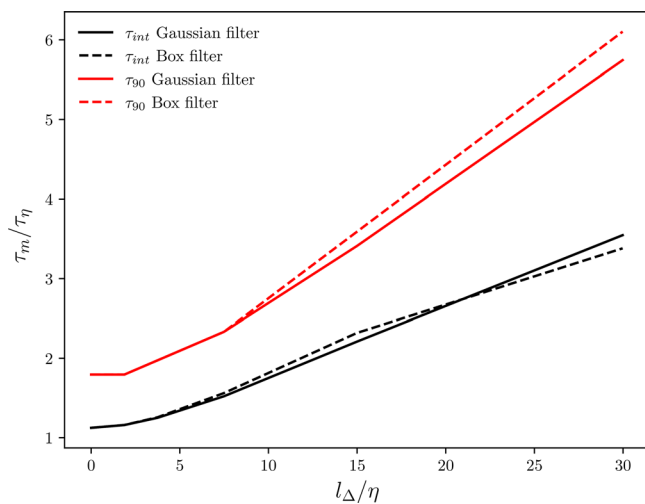

FIG. 7. The dependence of timescale of the extracted memory kernel on the length scale of the filter for observable set 1. Two different methods are used to estimate the timescale: 1. τ_{int}: timescale estimated using the integral of the memory kernel and 2. τ_{90} timescale estimated based on the location where the memory kernel drops to 10% of the value at smallest time delay.

similar temporal decay. That is to say, if one seeks to derive a model for the noise with short temporal correlation, the corresponding memory kernel should have a similar temporal length. This property has been used to develop consistent MZ-based reduced-order models. On the other hand, the decay of the Frobenius norm of the two-time correlation of noise is not monotonic, different from that of the memory kernel. Between the time delay of $5\tau_\eta \sim 15\tau_\eta$, there is a transient behavior in the temporal decorrelation, which contains the nonlinear dynamics that cannot be fully represented by the chosen observables in the linear space. This transient behavior becomes less significant with the increase in the filter size. If the correlation decays fast enough and the transient behavior happens after the fast decay and becomes trivial, the modeling can be greatly simplified by considering the correlation on a shorter timescale. Figure 8(b) shows the normalized PDFs of the orthogonal dynamics with different filter sizes and a reference Gaussian distribution. The PDFs of the noise do not depend on the time delay (results not shown), so we only include in Fig. 8 the PDFs of the shortest time delay. With the increase in the filter size, it can be seen that the noise becomes more Gaussian-like and there is a trade-off between the timescale of the correlation and the "Gaussian-like" shape of the PDF.

2. The effects of observable choices on the extracted operators

Similar to the Koopman learning, choosing the appropriate set of observables that can best close the original dynamics systems in the linear space is an important task in the current MZ learning framework, with another layer of complexity due to the existence of memory kernel and orthogonal dynamics. Here, we investigate the effects of different sets of observables on the properties of the learned kernels.
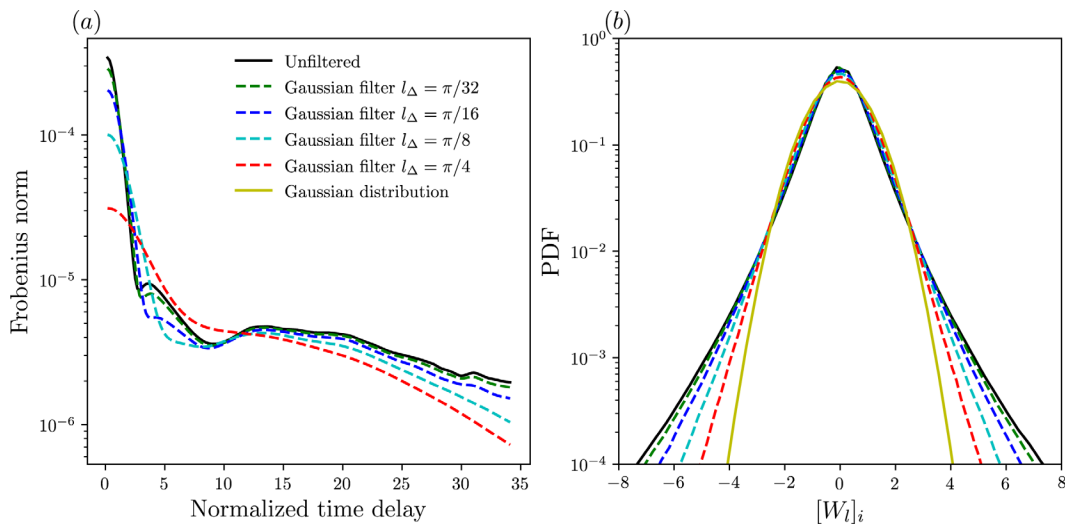
FIG. 8. (a) The Frobenious norm of the learned temporal correlation of the noise for observable set 1, $||\langle \mathbf{W}_{l+1}, \mathbf{W}_1^T \rangle||_F$ as a function of normalized time delay and (b) PDFs of the orthogonal dynamics. The time delay is normalized by Kolmogorov timescale, $\tau_\eta \approx 5.7 \times 10^{-2}$.

These sets are described in Sec. IV and correspond to different aspects of NS equations and turbulence variables. Figure 9 shows the Frobenius norm of the learned memory kernels for four sets of observables and the Frobenius norm of the corresponding orthogonal dynamics. It is obvious from Fig. 9 that the choices of observables for constructing MZ-based models can significantly influence the properties of the extracted operators. The observable sets 2 and 4 have a similar magnitude of the memory kernel and the rate of decay as compared to observable set 1. On the other hand, observable set 3, which contains variables that are related to small-scale turbulence phenomena, exhibits a transient process of the memory effect: the faster early decay of the memory kernel is related to variables that are short-time correlated; the later increase then shows the intermittent behavior of the chosen observables. From the modeling perspective, it is more challenging to devise an accurate model that can reproduce the transient memory decay of the observable set 3. As for the properties of the orthogonal dynamics, all four sets of observables show a certain level of transient behavior rather than exponential decay.

The memory kernels shown in Fig. 9 contain components from all the observables in each set. For observables with different
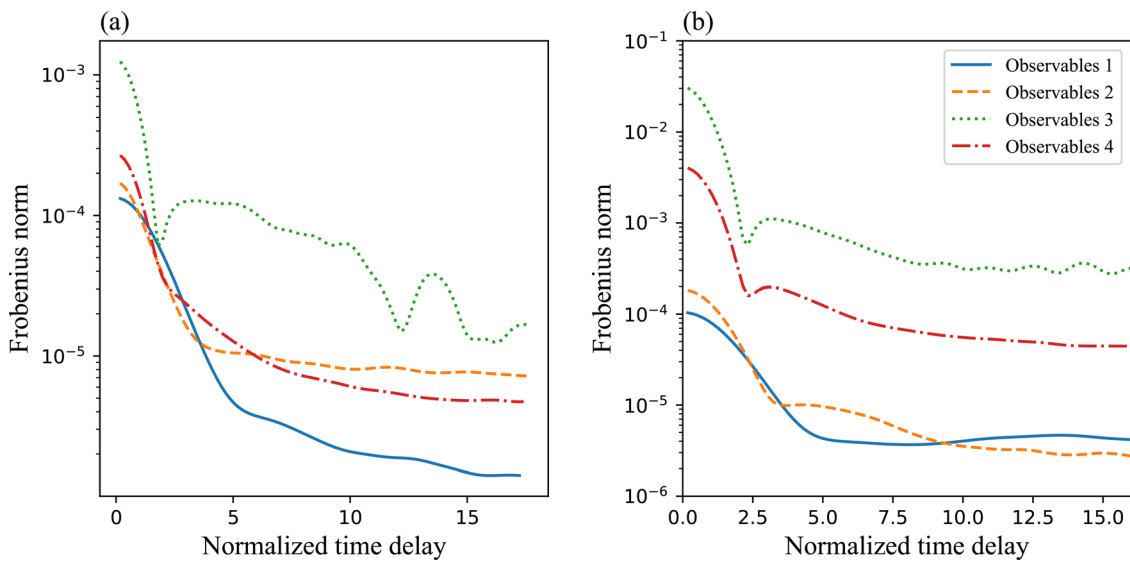

FIG. 9. (a) The Frobenius norm of the learned memory kernel and (b) two-time correlation of orthogonal dynamics (right) as a function of time delay for different sets of observables. The time delay is normalized by Kolmogorov timescale, $\tau_\eta \approx 5.7 \times 10^{-2}$, and the discrete time interval is $\Delta = 0.01$.

dimensions and magnitudes, the Frobenius norm may hide important information about the properties of the extracted kernels. This makes it difficult to study how the quantities of interest, e.g., filtered velocities, are affected by the different choices of observables. To understand this, we extract from the learned kernels the subset of the matrices that correspond to the filtered velocity, which has the same dimension in all observable sets. The Frobenius norms of the subset of the extracted kernels are compared across different sets of observables and shown in Fig. 10. We observe that the properties of the memory kernel for the filtered velocity are not significantly modified by including more observables in sets 2 and 3, even though the memory kernels for the full set of observables exhibit significant differences. On the other hand, for the observable set 4, the contribution from memory kernel is reduced and the decay of the memory kernel becomes smoother. A similar trend is observed for the noise, with a smaller magnitude and smoother rate of decay. We conclude from these observations that by using observable set 4, which contains the RHS of the governing equations, the contribution to dynamics shifts from memory kernel to the Markov term. By comparing Figs. 9 and 10, we note that different sets of observables may exhibit different memory kernel structure and length. This implies that the modeling strategy may need to be changed for different observables. We also remark that simply adding more observables is not necessarily guaranteed to improve the properties of memory kernel for modeling, as shown by the results of observable set 3.

D. Numerical experiments on memory effects

In this section, we conduct numerical experiments to illustrate the advantage of including memory kernels for prediction. Consider the following procedure: (a) after we apply the learning algorithm for a set of observables $\mathbf{g}$ with a time interval Δ, we generate additional

samples ($>2 \times 10^4$) on the trajectories of the same observables using the fully resolved simulation. These samples had not been used in the learning of the corresponding kernels. Within each sample, the snapshots are also evenly spaced in time with the same time interval Δ. (b) The total length of the trajectories is longer than the timescale of the memory kernel. (c) We then use the additional snapshots to predict $\delta = n\Delta$ into the future using the following formula recursively:

$$\mathbf{g}_{pred}((n+1)\Delta) = \mathbf{\Omega}_\Delta^{(0)} \cdot \mathbf{g}(n\Delta) + \sum_{\ell=1}^{n} \mathbf{\Omega}_\Delta^{(\ell)} \cdot \mathbf{g}((n-l)\Delta), \quad (36)$$

which is the discrete GLE [Eq. (25)] with the assumption that the orthogonal dynamics $\mathbf{W}_{n+1} = 0$, and calculate the errors on the prediction. (d) The errors are then averaged over samples to reduce statistical variability. In this work, we choose the L^2-norm as the measure of prediction error, which is calculated as

$$\varepsilon^2 = ||\mathbf{g}_{pred} - \mathbf{g}_{DNS}||^2, \quad (37)$$

where $\mathbf{g}_{pred}$ and $\mathbf{g}_{DNS}$ denote the observables from predictions and the "ground-truth" DNS simulations.

Figure 11 shows the L^2-norm of the prediction errors using the discrete Mori–Zwanzig formulation with different memory lengths for prediction. The memory length is normalized using Kolmogorov timescale. We also consider different discrete time intervals $\Delta = 0.01, 0.02, 0.05, 0.1$ for learning the MZ operators. The chosen prediction horizons δ are 0.1 and 0.2, which are multiples of the discrete time intervals. We point out that the first point on the plot has memory length 0.0 so that it corresponds to Markovian models and can be viewed as the Koopman prediction. It is evident from Fig. 11 that the prediction errors decrease when past histories are included in the memory kernel. As the memory length further increases past one
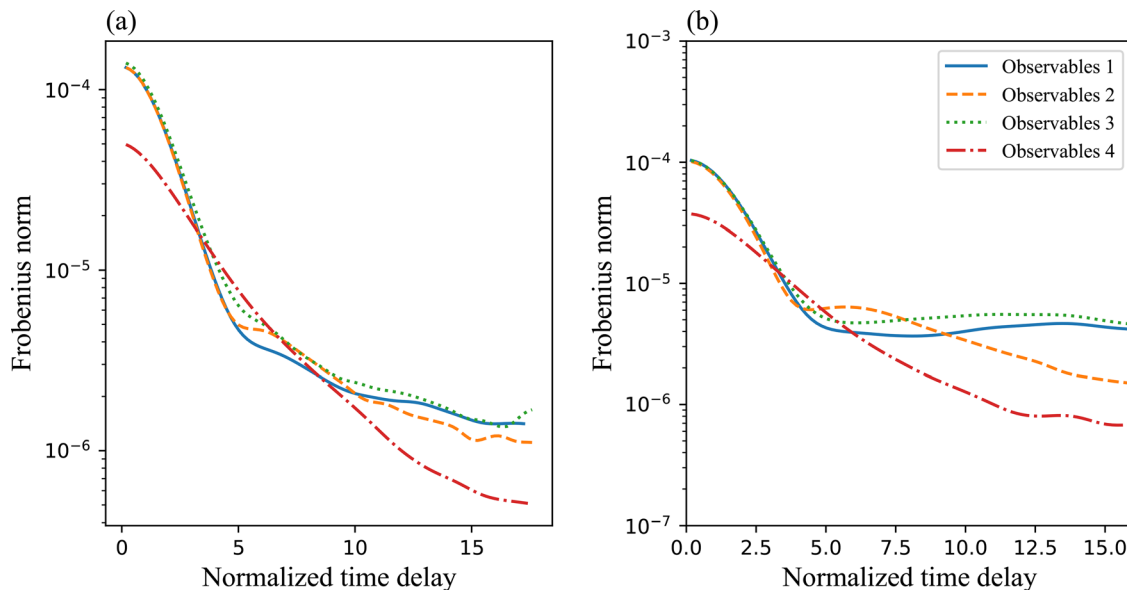

FIG. 10. (a) The Frobenius norm of the learned memory kernel and (b) two-time correlation of orthogonal dynamics for the shared observables among the four sets (filtered velocity) as a function of time delay. The time delay is normalized by Kolmogorov timescale, $\tau_\eta \approx 5.7 \times 10^{-2}$, and the discrete time interval is $\Delta = 0.01$.
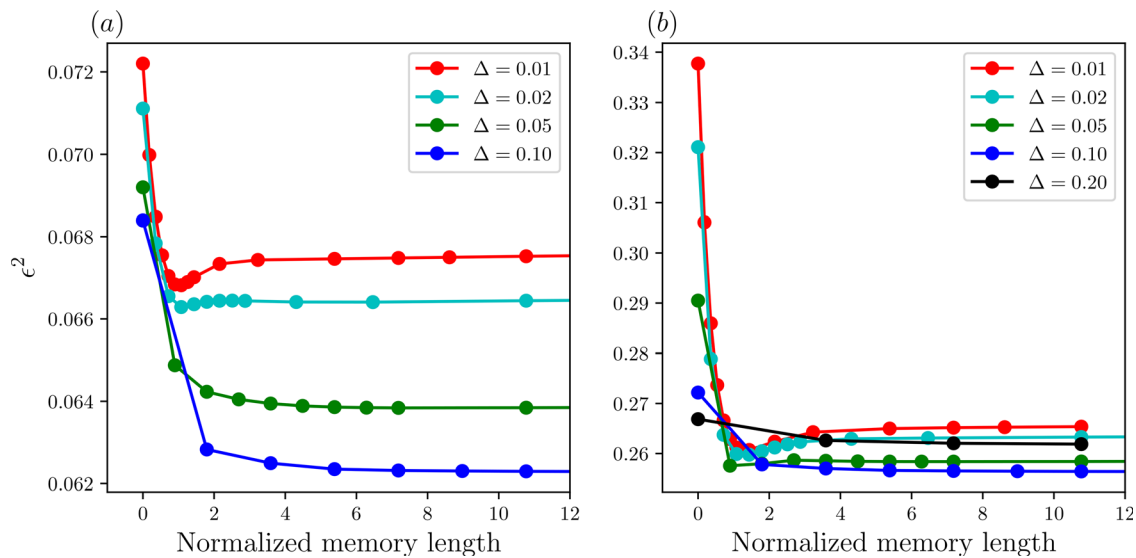
FIG. 11. Comparison of prediction errors of different discrete time intervals $\Delta = 0.01, 0.02, 0.05, 0.10, 0.2$ as a function of the normalized memory length for observable set 1. The memory length is normalized by Kolmogorov timescale, $\tau_\eta \approx 5.7 \times 10^{-2}$. The prediction horizon is (a) 0.1 and (b) 0.2, as is a multiple of the discrete time intervals. The error is calculated using the observables at the final prediction time.

Kolmogorov timescale, their effects on the prediction errors vary for different discrete time intervals Δ.

For the smallest time interval, the prediction error increases and then saturates when the memory length is larger than $4\tau_\eta$. Our numerical results suggest the existence of an optimal memory length when the MZ operators are extracted with small time intervals Δ. When the discrete time interval increases, the optimal memory length disappears and the prediction errors decrease and saturate after a certain memory length. The smallest prediction errors can be achieved when the discrete time interval Δ matches with the prediction horizon.

For a longer prediction horizon $\delta = 0.2$, similar observations can be made on the improvement of prediction errors by including past history. On the other hand, when the discrete time interval Δ matches with the prediction horizon, the error actually becomes larger than that of the smaller discrete time interval, which shows that there also exists an optimal discrete time interval for the selected prediction horizon. For the larger prediction horizon, the overall improvement by including past history is around 23%. Note that the prediction errors depend on the accumulated magnitude of the orthogonal dynamics, which is neglected in the current prediction method. A possible explanation is that when the discrete time interval Δ is small, we need more steps to advance the observables so that the magnitude of the accumulated orthogonal dynamics that are missed in the prediction is larger. On the other hand, when the discrete time interval is too large (compared to the timescale of memory kernel), the projected image across such a large step can become small, which in turn will increase the magnitude of orthogonal dynamics. This may explain why the prediction error is large for discrete time interval $\Delta = 0.2$. Additionally, one should also take into account the different timescales of chosen observables, which may further complicate the process of choosing the optimal discrete interval.

The effects on prediction by choosing different sets of observables are explored next. Figure 12 compares the prediction errors of the quantities of interest (filtered velocity) across the four sets of observables presented above. The prediction horizon is $\delta = 0.05$ and it is the same as discrete time interval, so we only need to solve the discrete MZ formulation (without a model for orthogonal dynamics) for one step. In addition to the prediction improvement by including past history, the prediction errors vary drastically across the four sets of observables. Including turbulence physics-based observables like strain rate tensor, vorticity, and kinetic energy leads to negligible improvement on the prediction error for the observables of interest, namely, filtered velocities. By including the sub-grid stress, a commonly used observable in traditional turbulence modeling approaches, we can observe a marginal improvement. The largest improvement can be seen for the observable set 4, where the observables are the terms in the filtered governing equations. There is over 50% improvement on the prediction error by using observable set 4 compared to the other sets. The majority of the improvement by selecting observable set 4 comes from the Markov term, which is consistent with the results shown above. Further improvement in prediction can be achieved by including the past history in the memory term. The percentage improvement over the corresponding Markovian prediction method is also different for different sets of observables: around 4% for observable set 1–3 and 8.5% for the observable set 4. Overall, we can conclude that the choice of observables significantly affects the prediction capability of the learned MZ kernels.

In the current numerical experiments for prediction, in order to focus on the importance of the memory effects and observable selection, we neglect an important component of the MZ formalism, namely, the orthogonal dynamics/noise. The orthogonal dynamics encode the unresolved initial conditions of the dynamics and are crucial to reproducing the correct statistical properties of the original system. In practice, it is desirable to select a set of observables with a smaller magnitude of the orthogonal dynamics and shorter temporal correlation, which adds another layer of complexity for selecting the
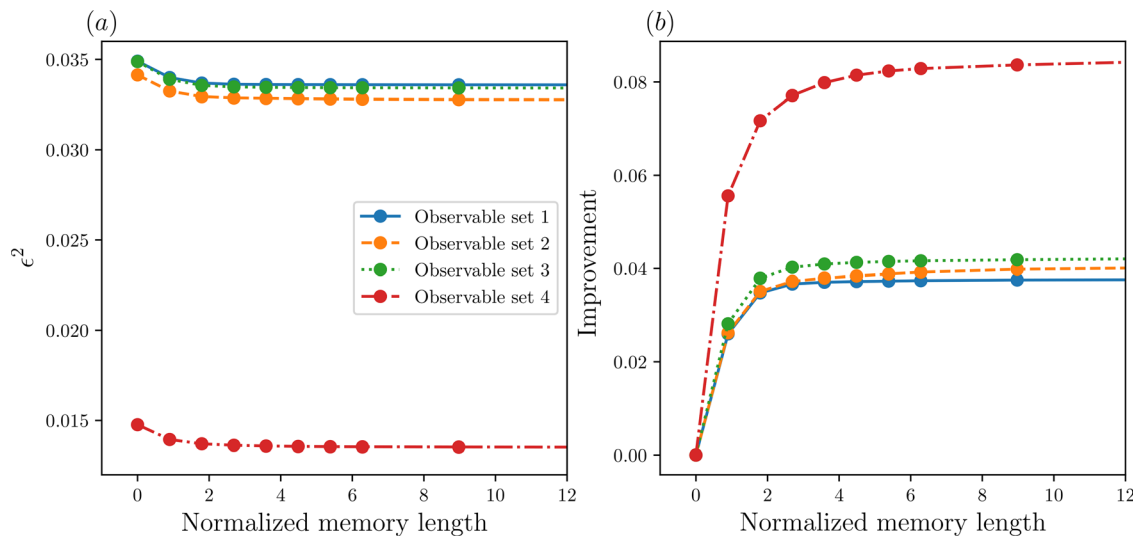
FIG. 12. Comparison of (a) L^2-norm of prediction errors and (b) percentage improvement in prediction for different sets of observables as a function of the normalized memory length. The L^2-norm of prediction errors The memory length is normalized by Kolmogorov timescale, $\tau_\eta \approx 5.7 \times 10^{-2}$. The prediction horizon is 0.05, the same as the discrete time interval Δ. The error is calculated using the shared observables (filtered velocity) at the final prediction time.

observables. Our extracted noise exhibits complicated/nontrivial two-time correlation, meaning that the orthogonal dynamics can only be modeled by highly nontrivial colored noise.

VI. CONCLUSIONS

Developing reduced-order models for turbulence is a challenging problem due to the existence of a wide range of scales with complicated nonlinear interactions. Traditional modeling strategies based on filtering or moment closure (LES, RANS) are derived mostly based on physical intuitions. On the other end of the spectrum, Mori–Zwanzig (MZ) formalism provides a formal mathematical procedure for derivation of low-dimensional representations of high-dimensional nonlinear dynamical systems. The outcome of applying the MZ formalism is the emergence of a memory term and orthogonal dynamics. To reduce the computational complexity compared to the original system, models for these terms need to be developed, which requires a comprehensive understanding of their mathematical properties. However, efforts of directly extracting these terms in turbulent flows have been scarce, so the understanding of their properties is limited. In this work, we are the first to apply a data-driven algorithm to a fully resolved turbulence simulation dataset to extract the memory kernel and orthogonal dynamics and analyze their properties. This provides a foundation for developing accurate MZ-based turbulence models including contributions from memory and orthogonal dynamics.

With data augmentation using known turbulence symmetries, the Markov, memory kernels, and orthogonal dynamics can be successfully extracted using a reasonable amount of data and are shown to be statistically converged. The subtle generalized fluctuation-dissipation relation, which is a natural outcome of MZ formalism, is verified numerically using the extracted kernels and two-time correlation function of the orthogonal dynamics. This confirms the accuracy and correctness of the learning procedure and the learned MZ kernels. The memory kernels are shown to have a strong dependence on the

spatial filtering sizes and weak dependence on filtering type. The Frobenius norm of the memory kernel exhibits a fast decay, indicating that the finite memory length modeling assumption is reasonable. The timescales of the memory kernels are then calculated and qualitative agreement can be observed with previous studies. The two-time correlation matrix of the orthogonal dynamics exhibits similar dependence on the spatial filtering size as the memory kernel. With a larger filter size, the PDF of orthogonal dynamics becomes more Gaussian-like. The effects of different observables choices on the extracted kernels are then examined. We observe that expanding the set of observables using nonlinear functions may significantly influence the decay of the memory kernel. On the other hand, the memory lengths of the added nonlinear observables are different, implying that a multi-timescale model may be needed for the expanded observable set. By using the observable set 4, which includes the terms on right-hand-side of the filtered NS equations, the magnitude of the memory kernel decreases. This is explained as a shift of contributions from memory term to the Markov term.

The advantages of including past history in prediction using MZ-based models are studied by comparing the prediction error with only the Markovian term. The results show that the L^2 prediction error is lowered by including the memory integral, especially for longer predicting horizon. When using kernels extracted on a smaller discrete time interval, there exists an optimal memory length for calculating the memory integral. On the other hand, for larger discrete time interval, the improvement on prediction saturates when long enough past history is used. The optimal discrete time interval may also change for different prediction horizons, which is related to the relative magnitude of discrete time interval and memory length. Finally, the influence of observable choices on prediction is investigated. It is shown that the improvement on prediction of filtered velocity is marginal when including physics-based observables. By using equation-based observables (observable set 4), the improvement of the Markovian

model is significant. In addition, including the memory integral further improves the prediction accuracy and the percentage increase in accuracy is also larger for observable set 4. In this study, we only show the improvements in prediction due to the addition of memory effects, but without the orthogonal dynamics. When proper models for orthogonal dynamics are proposed and used for prediction, the prediction should significantly improve. Future work will be dedicated to developing MZ-based turbulence models from the following perspectives: (a) discovering observables with suitable properties for modeling (finite memory length and simple profile of the memory kernel) and (b) devising stochastic models for orthogonal dynamics that satisfy the learned statistical properties.

In this work, we limit the scope of our research to homogeneous isotropic turbulence (HIT), which is the simplest type of turbulence flow. This is because the understanding of the MZ operators for turbulence is very limited and starting from HIT enables us to investigate the properties of MZ operators for fundamental turbulent dynamics without additional effects. However, the MZ learning algorithm can also be applied to more complex turbulent flows. When using the proposed MZ learning algorithm, one needs to pay attention to the nature of the problem. For non-stationary flows, since the distribution of the system state is time-dependent, one performs averaging over an ensemble of simulations, whose initial conditions are drawn from the initial distribution of the states for each specific problem. This is different to the sampling method used in the paper for an ergodic system, where the underlying distribution of the state is stationary. For this case, one can simply sample from a long trajectory of data. In addition, for flows with heterogeneity and anisotropy, special attention should be paid to not enforce the type of symmetry that is not satisfied by the dynamics. Another factor that may complicate the learning of MZ operators for complex turbulent flows is the choices of observables. One needs to identify observables that can represent the flow field of interest and also embed the dynamics of the system in the linear space. This selection procedure might rely on the domain knowledge of specific applications.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts of interest to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] P. Sagaut, *Large Eddy Simulation for Incompressible Flows: An Introduction* (Springer Science & Business Media, 2006).

[2] R. H. Kraichnan, "Decay of isotropic turbulence in the direct-interaction approximation," Phys. Fluids **7**, 1030–1048 (1964).

[3] R. H. Kraichnan, "Lagrangian-history closure approximation for turbulence," Phys. Fluids **8**, 575–598 (1965).

[4] W. J. Bos and J.-P. Bertoglio, "Lagrangian Markovianized field approximation for turbulence," J. Turbul. **14**, 99–120 (2013).

[5] H. Mori, "Transport, collective motion, and Brownian motion," Prog. Theor. Phys. **33**, 423–455 (1965).

[6] R. Zwanzig, "Nonlinear generalized Langevin equations," J. Stat. Phys. **9**, 215–220 (1973).

[7] A. J. Chorin, O. H. Hald, and R. Kupferman, "Optimal prediction and the Mori–Zwanzig representation of irreversible processes," Proc. Natl. Acad. Sci. U. S. A. **97**, 2968–2973 (2000).

[8] D. Givon, R. Kupferman, and O. H. Hald, "Existence proof for orthogonal dynamics and the Mori-Zwanzig formalism," Isr. J. Math. **145**, 221–241 (2005).

[9] A. Chorin and P. Stinis, "Problem reduction, renormalization, and memory," Commun. Appl. Math. Comput. Sci. **1**, 1–27 (2006).

[10] D. Bernstein, "Optimal prediction of Burgers's equation," Multiscale Modeling Simul. **6**, 27–52 (2007).

[11] O. H. Hald and P. Stinis, "Optimal prediction and the rate of decay for solutions of the Euler equations in two and three dimensions," Proc. Natl. Acad. Sci. U. S. A. **104**, 6527–6532 (2007).

[12] A. J. Chandy and S. H. Frankel, "The t-model as a large eddy simulation model for the Navier–Stokes equations," Multiscale Modeling Simul. **8**, 445–462 (2010).

[13] E. J. Parish and K. Duraisamy, "A dynamic subgrid scale model for large eddy simulations based on the Mori–Zwanzig formalism," J. Comput. Phys. **349**, 154–175 (2017).

[14] P. Stinis, "Renormalized reduced models for singular PDES," Commun. Appl. Math. Comput. Sci. **8**, 39–66 (2013).

[15] P. Stinis, "Renormalized Mori–Zwanzig-reduced models for systems without scale separation," Proc. R. Soc. A **471**, 20140446 (2015).

[16] P. Stinis, "Higher order Mori–Zwanzig models for the Euler equations," Multiscale Modeling Simul. **6**, 741–760 (2007).

[17] E. J. Parish and K. Duraisamy, "Non-Markovian closure models for large eddy simulations using the Mori–Zwanzig formalism," Phys. Rev. Fluids **2**, 014604 (2017).

[18] A. Gouasmi, E. J. Parish, and K. Duraisamy, "A priori estimation of memory effects in reduced-order models of nonlinear systems using the Mori–Zwanzig formalism," Proc. R. Soc. A **473**, 20170385 (2017).

[19] A. J. Chorin, O. H. Hald, and R. Kupferman, "Optimal prediction with memory," Physica D **166**, 239–257 (2002).

[20] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, "A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition," J. Nonlinear Sci. **25**, 1307–1346 (2015).

[21] M. Okamura, "Validity of the essential assumption in a projection operator method," Phys. Rev. E **74**, 046210 (2006).

[22] H. Mori and M. Okamura, "Dynamic structures of the time correlation functions of chaotic nonequilibrium fluctuations," Phys. Rev. E **76**, 061104 (2007).

[23] H. Meyer, P. Pelagejcev, and T. Schilling, "Non-Markovian out-of-equilibrium dynamics: A general numerical procedure to construct time-dependent memory kernels for coarse-grained observables," Europhys. Lett. **128**, 40001 (2020).

[24] H. Meyer, S. Wolf, G. Stock, and T. Schilling, "A numerical procedure to evaluate memory effects in non-equilibrium coarse-grained models," Adv. Theory Simul. **4**, 2000197 (2021).

[25] S. Maeyama and T.-H. Watanabe, "Extracting and modeling the effects of small-scale fluctuations on large-scale fluctuations by Mori–Zwanzig projection operator method," J. Phys. Soc. Jpn. **89**, 024401 (2020).

[26] B. O. Koopman, "Hamiltonian systems and transformation in Hilbert space," Proc. Natl. Acad. Sci. U. S. A. **17**, 315 (1931).

[27] B. Koopman and J. v Neumann, "Dynamical systems of continuous spectra," Proc. Natl. Acad. Sci. U. S. A. **18**, 255 (1932).

[28] P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," J. Fluid Mech. **656**, 5–28 (2010).

[29]Y. T. Lin, Y. Tian, D. Livescu, and M. Anghel, "Data-driven learning for the Mori–Zwanzig formalism: A generalization of the Koopman learning framework," SIAM J. Appl. Dyn. Syst. (in press, 2021).

[30]D. J. Evans and G. P. Morriss, *Statistical Mechanics of Nonequilbrium Liquids* (ANU Press, 2007).

[31]S. K. Falkena, C. Quinn, J. Sieber, J. Frank, and H. A. Dijkstra, "Derivation of delay equation climate models using the Mori–Zwanzig formalism," Proc. R. Soc. A 475, 20190075 (2019).

[32]K. K. Lin and F. Lu, "Data-driven model reduction, Wiener projections, and the Mori–Zwanzig formalism," J. Comput. Phys. 424, 109864 (2021).

[33]E. Darve, J. Solomon, and A. Kia, "Computing generalized Langevin equations and generalized Fokker–Planck equations," Proc. Natl. Acad. Sci. U. S. A. 106, 10884–10889 (2009).

[34]E. Parish and K. Duraisamy, "Reduced order modeling of turbulent flows using statistical coarse-graining," in *46th AIAA Fluid Dynamics Conference* (AIAA, 2016) p. 3640.

[35]J. Smagorinsky, "General circulation experiments with the primitive equations: I. the basic experiment," Mon. Weather Rev. 91, 99–164 (1963).

[36]A. Yoshizawa and K. Horiuti, "A statistically-derived subgrid-scale kinetic energy model for the large-eddy simulation of turbulent flows," J. Phys. Soc. Jpn. 54, 2834–2839 (1985).

[37]M. Germano, U. Piomelli, P. Moin, and W. H. Cabot, "A dynamic subgrid-scale eddy viscosity model," Phys. Fluids A 3, 1760–1765 (1991).

[38]M. R. Petersen and D. Livescu, "Forcing for statistically stationary compressible isotropic turbulence," Phys. Fluids 22, 116101 (2010).

[39]D. Daniel, D. Livescu, and J. Ryu, "Reaction analogy based forcing for incompressible scalar turbulence," Phys. Rev. Fluids 3, 094602 (2018).

[40]P. Stinis, "Mori–Zwanzig reduced models for uncertainty quantification i: Parametric uncertainty," preprint arXiv:1211.4285 (2012).

[41]M. Germano *et al.*, "Turbulence- the filtering approach," J. Fluid Mech. 238, 325–336 (1992).