

# Data-driven linearization of turbulence

By Z. Chua Khoo, D. Cooper, Y. Bengana AND Y. Hwang<sup>†</sup>

Turbulence encountered in many engineering applications is commonly found with various forms of external disturbances and perturbations: e.g., surface roughness, wind gust, rain, ice and acoustic noise. The understanding and characterization of how such perturbations affect the dynamics and statistics of turbulence are of crucial importance for the prediction and modeling of the flow, and they offer valuable physical insights into controlling turbulence in a robust way. However, turbulence is a high-dimensional chaotic system, which makes its characterization fundamentally difficult due to the butterfly effect, i.e., the sensitive dependence on initial conditions. To overcome this challenge, this study aims to develop a novel data-driven input–output analysis framework that characterizes the effect of perturbations in turbulence. To this end, two approaches have been investigated for their suitability: (i) stochastic system identification (SSI) and (ii) linear inverse modeling (LIM). SSI requires a further feasibility test for a large number of inputs, while LIM over predicts the transient growth experienced by perturbations applied to the turbulent flow. A novel strategy is therefore currently under development.

---

## 1. Introduction

The main theoretical challenge in the description for the spatiotemporal evolution of a perturbation in turbulent flow originates from the so-called butterfly effect, which refers to the sensitive dependence on the chaotic dynamics to initial conditions. When a small perturbation is added to the initial condition, the evolution of turbulence becomes unpredictable after some time, which is characterized by the Lyapunov exponent. Therefore, the attempt to develop a theoretical description for the instantaneous evolution of perturbations in turbulent flows is fundamentally limited by the Lyapunov time scale. Unfortunately, the Lyapunov time scale has been known to decrease rapidly as Reynolds number increases (Ruelle 1979; Mohan *et al.* 2017).

The difficulties above may be overcome by introducing a suitable averaging procedure in the evolution of the dynamics, as originally proposed by Hussain & Reynolds (1970), who introduced triple decomposition. Consider a dynamical system,  $\frac{\partial \mathbf{q}}{\partial t} = \mathbf{N}(\mathbf{q}) + \mathbf{f}$ , where  $\mathbf{q}$  is the state variable and  $\mathbf{f}$  is a deterministic forcing representing the mechanism to introduce perturbations. Now, consider an ensemble averaging procedure, denoted by  $\langle \bullet \rangle$ , applied to the dynamical system. This results in

$$\frac{\partial \langle \mathbf{q} \rangle}{\partial t} = \langle \mathbf{N}(\mathbf{q}) \rangle + \mathbf{f}. \quad (1.1)$$

By restricting the analysis to small perturbations, the perturbation dynamics can be described by a linear operator opted from linearization about the mean equilibrium. We note that the linearized operator must be linearly stable for chaotic attractors, including turbulence at high Reynolds numbers, thereby providing a well-posed mathematical

<sup>†</sup> Department of Aeronautics, Imperial College London, United Kingdom

description for the perturbation dynamics in turbulent flows. However, as discovered by Hussain & Reynolds (1970), the linearized operator has a closure problem similar to the Reynolds-Averaged Navier-Stokes (RANS) equations. A simple mixing-length model for the closure problem was originally proposed by Hussain & Reynolds (1970). Many subsequent studies (del Álamo & Jiménez 2006; Pujals *et al.* 2009; Hwang & Cossu 2010) have relied on this simple model for the linear analysis of turbulent flow, such as transient growth (Butler & Farrell 1993) and input-output analysis (Jovanović & Bamieh 2005; McKeon & Sharma 2010).

The objective of this study is to identify the linearized operator for perturbations in turbulent flows by adopting a data-driven approach, while examining the validity of the simple closure by Hussain & Reynolds (1970). Two approaches are introduced: (i) stochastic system identification method by Luchini *et al.* (2006); (ii) linear inverse method by Penland (1989). The former models the linearized operator using a stochastic response of the given system, and the latter indirectly models it based on the covariance of the unforced flow system.

## 2. Statistical state dynamics

We start by discussing the idea of applying an averaging procedure to the Navier-Stokes equations. Statistical state dynamics was introduced to the study of fluid mechanics by Hopf (1952). Hopf used the concept of cumulants to study the velocity field of homogeneous isotropic turbulence. The framework has been proved to be fruitful in elucidating the dynamics of turbulence modeling in the context of geophysical fluid dynamics by Farrell & Ioannou (2007) and Marston *et al.* (2008). This concept is also related to the framework of the triple decomposition of a turbulent velocity field, introduced by Hussain & Reynolds (1970), and it decomposes a turbulent velocity field into three components: a time-averaged component (mean), turbulent fluctuations and the velocity perturbation originating from the given external (deterministic) forcing.

### 2.1. Triple decomposition

The evolution of the velocity field,  $\mathbf{U}(\mathbf{x}, t)$ , and the pressure,  $P(\mathbf{x}, t)$ , of an incompressible Newtonian fluid are governed by the Navier-Stokes equations. With a deterministic body forcing,  $\mathbf{F}(\mathbf{x}, t)$ , the equations are written as

$$\begin{aligned} \frac{\partial}{\partial t} \mathbf{U} + (\mathbf{U} \cdot \nabla) \mathbf{U} &= -\nabla P + \frac{1}{Re} \nabla^2 \mathbf{U} + \mathbf{F}, \\ \nabla \cdot \mathbf{U} &= 0. \end{aligned} \quad (2.1)$$

In contrast with Hussain & Reynolds (1970), instead of the phase averaging, the ensemble-averaged triple decomposition is introduced by decomposing the velocity field into

$$\mathbf{U}(\mathbf{x}, t) = \overline{\mathbf{U}}(\mathbf{x}) + \mathbf{u}'(\mathbf{x}, t) + \tilde{\mathbf{u}}(\mathbf{x}, t), \quad (2.2)$$

where  $\overline{\mathbf{U}}(\mathbf{x})$  is the temporal mean of the velocity field,  $\mathbf{u}'(\mathbf{x}, t)$  is the turbulent fluctuations, and  $\tilde{\mathbf{u}}(\mathbf{x}, t)$  is the perturbation field associated with the forcing  $\mathbf{F}$ . For a statistically stationary flow, the following relation,  $\langle \mathbf{U} \rangle(\mathbf{x}, t) = \overline{\mathbf{U}}(\mathbf{x}) + \tilde{\mathbf{u}}(\mathbf{x}, t)$ , is satisfied. From the relation above, it can be further inferred that  $\langle \mathbf{u}' \rangle(\mathbf{x}, t) = 0$  and  $\langle \tilde{\mathbf{u}} \rangle(\mathbf{x}, t) = 0$ . Using the triple decomposition introduced above, the statistical state dynamics evolution equations

for the ensemble averaged mean state are written as

$$\begin{aligned} \frac{\partial}{\partial t} \langle \mathbf{U} \rangle + (\langle \mathbf{U} \rangle \cdot \nabla) \langle \mathbf{U} \rangle + \nabla \cdot \langle \mathbf{u}' \otimes \mathbf{u}' \rangle &= -\nabla \langle P \rangle + \frac{1}{Re} \nabla^2 \langle \mathbf{U} \rangle + \mathbf{F}, \\ \nabla \cdot \langle \mathbf{U} \rangle &= 0, \end{aligned} \quad (2.3)$$

which is the unsteady RANS equation with the well-known closure issue for the Reynolds stress term,  $\langle \mathbf{u}' \otimes \mathbf{u}' \rangle$ .

### 2.2. Statistical state dynamics for small perturbations

In a statistically stationary flow, an ensemble average is equivalent to a time average. Therefore, the time-averaged velocity field, denoted by  $\overline{(\bullet)}$ , is an equilibrium state of the mean statistical state equations [Eq. (2.3)]. By linearizing around this equilibrium state, the linearized mean statistical state equations is derived. Consider a small perturbation under the effect of  $\mathbf{F} = \epsilon \mathbf{f}$ ,

$$\left\langle \begin{bmatrix} \mathbf{U} \\ P \\ \mathbf{u}' \\ \mathbf{u}' \otimes \mathbf{u}' \end{bmatrix} \right\rangle (\mathbf{x}, t) \approx \overline{\begin{bmatrix} \mathbf{U} \\ P \\ 0 \\ \mathbf{u}' \otimes \mathbf{u}' \end{bmatrix}} (\mathbf{x}) + \epsilon \begin{bmatrix} \widetilde{\mathbf{u}} \\ p \\ 0 \\ (\mathbf{u}' \otimes \mathbf{u}') \end{bmatrix} (\mathbf{x}, t),$$

for  $\epsilon \ll 1$ . At  $O(1)$ , the mean equilibrium state adheres to the time-averaged Navier-Stokes equations, albeit augmented with the related Reynolds stress term,  $\overline{\mathbf{u}' \otimes \mathbf{u}'}$ , resulting in

$$\begin{aligned} (\overline{\mathbf{U}} \cdot \nabla) \overline{\mathbf{U}} + \nabla \cdot \overline{(\mathbf{u}' \otimes \mathbf{u}')} &= -\nabla \overline{P} + \frac{1}{Re} \nabla^2 \overline{\mathbf{U}}, \\ \nabla \cdot \overline{\mathbf{U}} &= 0. \end{aligned} \quad (2.4)$$

The dynamics of the perturbation terms,  $\widetilde{(\bullet)}$ , is described at  $O(\epsilon)$ , such that

$$\begin{aligned} \frac{\partial}{\partial t} \widetilde{\mathbf{u}} + \nabla \cdot (\mathbf{u}' \widetilde{\mathbf{u}}) &= -\nabla \widetilde{p} + \frac{1}{Re} \nabla^2 \widetilde{\mathbf{u}} \\ &\quad - (\overline{\mathbf{U}} \cdot \nabla) \widetilde{\mathbf{u}} - (\widetilde{\mathbf{u}} \cdot \nabla) \overline{\mathbf{U}} + \mathbf{f}, \\ \nabla \cdot \widetilde{\mathbf{u}} &= 0. \end{aligned} \quad (2.5)$$

In the absence of the Reynolds stress perturbation term, Eq. (2.5) is the linearized Navier-Stokes equations around the temporal mean state. However, for the ensemble-averaged state, the addition of a small perturbation is expected to yield the resulting Reynolds stress term,  $\nabla \cdot (\mathbf{u}' \widetilde{\mathbf{u}})$ . This Reynolds stress term leads to a similar closure problem as the standard RANS equations. As discussed, to overcome this issue, Hussain & Reynolds (1970) introduced a simple mixing length model used for the Reynolds stress term in Eq. (2.4). The key task of this study is to identify the role of this term in the evolution of the ensemble-averaged velocity field,  $\widetilde{\mathbf{u}}$ , by identifying Eq. (2.5) in a data-driven manner.

### 3. Data-driven identification of linearized statistical state dynamics

Here, we introduce two mathematical frameworks to identify Eq. (2.5) for a general form of forcing in a data-driven manner. For this purpose, let us assume that the perturbation dynamics is modeled by an autonomous linearly stable dynamical system with inputs

$$\dot{\mathbf{u}}(\mathbf{x}, t) = \mathbf{L}\mathbf{u}(\mathbf{x}, t) + \mathbf{f}(\mathbf{x}, t), \quad (3.1)$$

where  $\mathbf{u}(\mathbf{x}, t)$  is the state vector representing  $\tilde{\mathbf{u}}$  in Eq. (2.5), and  $\mathbf{L}$  the linear time-invariant operator for the temporal dynamics of the system. The general solution of Eq. (3.1) is given as

$$\mathbf{u}(\mathbf{x}, t) = e^{t\mathbf{L}}\mathbf{u}(\mathbf{x}, 0) + \int_0^t e^{(t-s)\mathbf{L}}\mathbf{f}(\mathbf{x}, s) ds. \quad (3.2)$$

### 3.1. SSI

The first approach was originally proposed by Luchini *et al.* (2006), which we shall refer to as SSI, since it identifies  $\mathbf{L}$  by applying a stochastic forcing to the system. In this instance, let  $\mathbf{f}(t)$  be governed by a real-valued multivariate Gaussian distribution,  $\mathbf{f} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , with the mean,  $\boldsymbol{\mu} = \mathbf{0}$ , and the covariance matrix,  $\boldsymbol{\Sigma} = \mathbf{I}$ . For the given modeling ansatz in Eq. (3.1) for the small perturbation dynamics, the computation of the cross-correlation between the state vector and the input vector,  $\mathbf{C}_{\mathbf{u},\mathbf{f}}^\tau(\mathbf{x}, \mathbf{x}')$ , identifies the linear operator using the following property

$$\mathbf{C}_{\mathbf{u},\mathbf{f}}^\tau(\mathbf{x}, \mathbf{x}') \equiv \langle \mathbf{u}(\mathbf{x}, t) \otimes \mathbf{f}^*(\mathbf{x}', t - \tau) \rangle = e^{\tau\mathbf{L}}, \quad (3.3)$$

with  $(\bullet)^*$  being the complex conjugate and derived from Eq. (3.2) using Itô isometry.  $\mathbf{L}$  is then obtained as

$$\mathbf{L} = \frac{1}{\tau} \ln[\mathbf{C}_{\mathbf{u},\mathbf{f}}^\tau(\mathbf{x}, \mathbf{x}')]. \quad (3.4)$$

We note that SSI is an intrusive approach in the sense that the forcings representing the perturbations are directly applied to the system. In this approach, the system response to a broadband stochastic forcing is directly measured to reconstruct the system in a data-driven manner. The main challenge here is that the amplitude of the forcing with the degree of freedom of the given system needs to be small enough to satisfy the linearity assumed, while many samples are required to ensure the convergence of  $\mathbf{C}_{\mathbf{u},\mathbf{f}}^\tau(\mathbf{x}, \mathbf{x}')$  by overcoming the background turbulence and the stochastic noise.

### 3.2. LIM

The second approach examined in the present study was proposed by Penland (1989) in the context of meteorology, with some promising results for the Lorenz system (Penland 1989) and for the spatial mean in turbulent channel flow (Farrell & Ioannou 2024). This approach, referred to as LIM, assumes that turbulent fluctuations,  $\mathbf{u}'$ , may be modeled by

$$\dot{\mathbf{u}}' = \mathbf{L}\mathbf{u}'(\mathbf{x}, t) + \boldsymbol{\eta}(\mathbf{x}, t), \quad (3.5)$$

where  $\boldsymbol{\eta}(t)$  characterizes a suitable stochastic forcing term derived to mimic the random nature of turbulent fluctuations. Although Penland (1989) and Farrell & Ioannou (2024) formulated their problem by assuming that  $\boldsymbol{\eta}(t)$  is a white noise, this is not technically necessary for the modeling procedure. In this approach, the key assumption is that the small perturbation field are affected by the background turbulence similar to the way that turbulent fluctuation field is; thus, the coherent perturbation field shares an identical operator,  $\mathbf{L}$ , as the fluctuation field. In other words, the coherent perturbation field is modeled by adding a perturbation to Eq. (3.5), such that

$$\dot{\mathbf{u}} = \mathbf{L}\mathbf{u}(\mathbf{x}, t) + \mathbf{f}(\mathbf{x}, t), \quad (3.6a)$$

where

$$\mathbf{L} = \frac{1}{\tau} \ln[\mathbf{C}_{\mathbf{u}',\mathbf{u}'}^\tau(\mathbf{C}_{\mathbf{u}',\mathbf{u}'}^0)^{-1}] \quad \text{with} \quad \mathbf{C}_{\mathbf{u}',\mathbf{u}'}^\tau = \langle \mathbf{u}'(\mathbf{x}, t) \otimes \mathbf{u}'^*(\mathbf{x}', t - \tau) \rangle. \quad (3.6b)$$

Here, we note that  $\mathbf{L}$  is now obtained from Eq. (3.5) rather than from Eq. (3.1), unlike the SSI approach. To complete the description, once  $\mathbf{L}$  is obtained from Eq. (3.6b), the noise term in Eq. (3.5) is characterized by the following Lyapunov equation

$$\mathbf{0} = \mathbf{L}\mathbf{C}_{\mathbf{u}',\mathbf{u}'}^0 + \mathbf{C}_{\mathbf{u}',\mathbf{u}'}^0\mathbf{L}^H + \mathbf{Q} \quad \text{with} \quad \mathbf{Q} = \langle \mathbf{u}' \otimes \boldsymbol{\eta} \rangle + \langle \boldsymbol{\eta} \otimes \mathbf{u}' \rangle, \quad (3.7)$$

where  $(\bullet)^H$  is the complex conjugate transpose.

The main benefit of the LIM approach is that it is not intrusive. In other words, the linear operator  $\mathbf{L}$  is actually obtained without applying any forcing. This is a distinct advantage over SSI, where a small-amplitude white noise forcing is applied to the system. However, the validity of the key assumption that turbulent perturbations would be affected by the background turbulence similar to the turbulent fluctuation dynamics deserves some careful scrutiny, especially given the encouraging observation by Penland (1989) and Farrell & Ioannou (2024).

## 4. Operator identifications

### 4.1. Turbulent channel flow

We consider plane channel flow to test SSI and LIM in the previous section. The streamwise direction is denoted by  $x$ , the wall-normal direction by  $y$  and the spanwise direction by  $z$ . Two parallel walls are located  $2h$  apart in the wall-normal direction at  $y = 0, 2h$ , with  $h$  being the half-height of the channel. The flow is driven by a pressure gradient, while keeping its mass flow rate constant. The streamwise and spanwise directions are set to be periodic. To sample the data required, direct numerical simulation (DNS) is performed with the code used by Hwang (2013), where the Fourier-Galerkin method is employed in the streamwise and spanwise directions and the second-order central difference method is adopted in the wall-normal direction. The flow investigated here is at  $Re_\tau (\equiv u_\tau h / \nu) = 100$  ( $u_\tau$  is the friction Reynolds number) with the domain  $(L_x, L_z) = (12.6h, 4.2h)$ . The mesh discretization is selected to be  $(N_x, N_y, N_z) = (66, 65, 66)$ , leading to a grid spacing of  $(\Delta x^+, \Delta z^+) = (30, 10)$ , where  $(\bullet)^+$  denotes inner units, normalized by the inner viscous length scale,  $\delta_\nu = \nu / u_\tau$ .

### 4.2. SSI

For the SSI approach, initial studies on a laminar channel flow at  $Re_h = u_l h / \nu = 500$  based on centerline streamwise velocity,  $u_l$ , was carried out for wave number  $(k_x h, k_z h) = (0, 1)$ . The velocity covariance of the stochastically forced simulation converged to the analytical solution of the Lyapunov equation after  $2 \times 10^6$  samples, with the flow field sampled every  $tu_l / h = 0.1$ . The SSI methodology is deemed too computationally expensive to be extended to turbulent flows especially with a large number of parameters (e.g., wave numbers) and forcing inputs. In subsequent sections, only the LIM approach is investigated.

### 4.3. Linear inverse modeling

Given the large number of degrees of freedom of the system, it is highly beneficial to consider low-rank approximations of the linear operator to be identified. The flow field is therefore projected onto a subspace composed of a set of orthonormal basis modes,  $\Phi_{\mathbf{u}}(\mathbf{x})$ , such that

$$\mathbf{u}(\mathbf{x}, t) = \Phi_{\mathbf{u}}(\mathbf{x})\boldsymbol{\chi}(t) = \sum_{j=1}^m \phi_{\mathbf{u},j}(\mathbf{x})\chi_j(t), \quad (4.1)$$

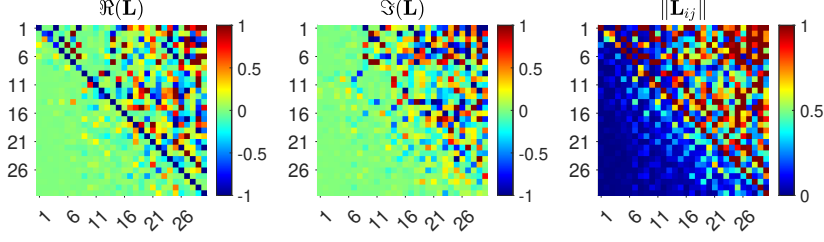


FIGURE 1. The linear operator identified using LIM in the subspace spanned by 30 controllability modes for  $(k_x h, k_z h) \simeq (0, 3)$ : (left) real component; (middle) imaginary component; and (right) magnitude of each element of the operator.

with  $\Phi_{\mathbf{u}}(\mathbf{x}) = [\phi_{\mathbf{u},1}(\mathbf{x}) \phi_{\mathbf{u},2}(\mathbf{x}) \dots \phi_{\mathbf{u},m}(\mathbf{x})]$  and  $\chi(t) = [\chi_1(t) \chi_2(t) \dots \chi_m(t)]^T$ , where  $\chi(t)$  is the temporal coefficient for each basis. Hence, the time-lagged covariance tensor,  $\mathbf{C}_{\mathbf{u}',\mathbf{u}'}^\tau$ , can be transformed into  $\chi$  coordinates via

$$\mathbf{C}_{\chi,\chi}^\tau = \Phi_{\mathbf{u}}^H \mathbf{C}_{\mathbf{u}',\mathbf{u}'}^\tau \Phi_{\mathbf{u}}. \quad (4.2)$$

In this study, the orthonormal basis,  $\Phi_{\mathbf{u}}$ , is chosen to be the controllability mode from the linearized Navier-Stokes operator about the temporal mean flow without the addition of eddy viscosity at  $Re_\tau = 100$ . The linear operator governing the small perturbation dynamics may be modeled with the correlations introduced in Section 3.2 with an appropriate low-rank approximation. However, in practice, these correlations are noisy due to the averaging required with a limited number of samples. Therefore, a least-squares optimization problem is formulated to identify the best fit of the linear operator from the computed correlations. For example, in the case of the LIM approach, the following least-squares problem is considered in the low-rank subspace

$$\underset{\mathbf{L}}{\operatorname{argmin}} \sum_{i=1}^{N_\tau} \left\| \frac{1}{\tau_i} \ln[\mathbf{C}_{\chi,\chi}^{\tau_i} (\mathbf{C}_{\chi,\chi}^0)^{-1}] - \mathbf{L} \right\|_F^2, \quad (4.3)$$

where  $\tau_i$  is the correlation time lag at a discretized point ( $i = 1, 2, \dots, N_\tau$ ) with the total number of the discretized points  $N_\tau$ . The analytical solution to Eq. (4.3) is given by

$$\mathbf{L} = \frac{1}{N_\tau} \sum_i^{N_\tau} \frac{1}{\tau_i} \ln[\mathbf{C}_{\chi,\chi}^{\tau_i} (\mathbf{C}_{\chi,\chi}^0)^{-1}]. \quad (4.4)$$

Since plane channel flow is homogeneous in the streamwise and spanwise directions, we focus on studying the linear operator for a plane Fourier mode with  $(k_x h, k_z h) \simeq (0, 3)$  [or  $(\lambda_x/h, \lambda_z/h) = (\infty, 2.1)$ ] and  $(k_x h, k_z h) \simeq (0, 6)$  [or  $(\lambda_x^+, \lambda_z^+) = (\infty, 100)$ ]. The Fourier mode was sampled from  $5 \times 10^5$  snapshots for both cases. Figure 1 shows the linear operator identified using LIM in the subspace defined by the leading 30 controllability modes for the  $(k_x h, k_z h) \simeq (0, 3)$  wavenumber. It is interesting to note that the higher-order controllability modes exhibit strong coupling to the leading controllability modes, as observed from the large values of nonzero upper triangular terms. Conversely, the leading controllability modes are less coupled to the dynamics of the less dominant controllability modes as seen from the near zero values of the lower triangular off-diagonal terms in the linear operator. Similar properties are also observed for the linear operator identified by LIM for the plane Fourier mode with  $(k_x h, k_z h) \simeq (0, 6)$  [or  $(\lambda_x^+, \lambda_z^+) = (\infty, 100)$ ].

## 5. Operator evaluations

### 5.1. Eigenspectra

We first evaluate the relevance of the linear operator identified through the LIM. The eigenvalues of the computed linear operator are all found to be linearly stable. We note that this feature must be observed if the linear operator identified by the LIM approach is correctly implemented. Indeed, from Eq. (4.3), the identified linear operator  $\mathbf{L}$  is approximated as

$$e^{\tau\mathbf{L}} \simeq \mathbf{C}_{\mathbf{x},\mathbf{x}}^{\tau} (\mathbf{C}_{\mathbf{x},\mathbf{x}}^0)^{-1}. \quad (5.1)$$

We note that, for any turbulent flow,  $\mathbf{C}_{\mathbf{x},\mathbf{x}}^{\tau} \rightarrow \mathbf{0}$  as  $\tau \rightarrow \infty$ . This property implies that all the eigenvalues of  $\mathbf{L}$  must be stable.

### 5.2. Transient energy amplification

Now, we compare the linear operator identified using LIM approach with those previously used in the literature (Hussain & Reynolds 1970; Pujals *et al.* 2009; Hwang & Cossu 2010) for the time evolution of small perturbations in turbulent flow. In particular, we perform an optimal transient growth analysis (Farrell & Ioannou 1993; Schmid & Henningson 2001) to quantify the amplification of initial conditions through the stable linear operator,  $\mathbf{L}$ . The related optimization problem is formulated as follows

$$\max_{\hat{\mathbf{u}}(0)} G(t) = \frac{\|\hat{\mathbf{u}}(t)\|^2}{\|\hat{\mathbf{u}}(0)\|^2}, \quad \text{where} \quad \|\hat{\mathbf{u}}(t)\|^2 = \int_0^{2h} \hat{\mathbf{u}}^H(t) \hat{\mathbf{u}}(t) dy, \quad (5.2)$$

where  $(\hat{\bullet})$  denotes the plane Fourier mode.

For the purpose of comparison with the existing linear model describing the evolution of small perturbations in turbulent flows, we further consider optimal transient growth from two linear operators: One is the linearized Navier-Stokes operator about turbulent mean (LNS model; Butler & Farrell 1993), and the other is the same operator augmented with the eddy viscosity model of Cess (1958) (LNS $_{\nu}$  model; Pujals *et al.* 2009). Transient growth of these operators has been computed using the full operator and also by restricting them within the subspace that is spanned by the 30 leading controllability modes. Furthermore, an ensemble DNS is carried out to validate the transient growth predictions of the LIM model. The ensemble consists of 300 snapshots of initial conditions, sampled at  $Re_{\tau} = 100$ , and the ensemble mean of the Fourier velocity fluctuations,  $\langle \hat{\mathbf{u}} \rangle$ , is verified to be of the order  $\mathcal{O}(10^{-4})$  for both  $(k_x h, k_z h) = (0, 3)$  and  $(0, 6)$ . Velocity perturbations are applied to the ensemble of initial conditions to investigate the effect of the perturbation's transient growth under the effect of background turbulence, which is described by its Fourier velocity component,  $\langle \hat{\mathbf{u}} \rangle(t)$ .

The amplified velocity fields are shown in Figure 2. We observe that the LNS $_{\nu}$  and LIM models have the peak streamwise velocity closer to the wall than the LNS model. This suggests that the LIM operator sampled from simulation data more closely resembles the eddy viscosity augmented LNS $_{\nu}$  model. Comparing the velocity contours between the LNS $_{\nu}$  model and LIM model more closely, we observe a sharper decrease in streamwise velocity as we move away from the wall for the LIM model. Furthermore, the more pronounced cross-stream circulation about  $y/h = 0.5$  and  $y^+ = 50$  from the LIM results, compared to a more centreline position,  $y/h = 1.0$  and  $y^+ = 100$ , as in LNS and LNS $_{\nu}$  models, indicate that the Reynolds stress experienced by the LIM model may have different anisotropic effects from the isotropic LNS $_{\nu}$  eddy viscosity model. Hence, from an eddy viscosity modeling perspective, progress may be made by relaxing an isotropic eddy

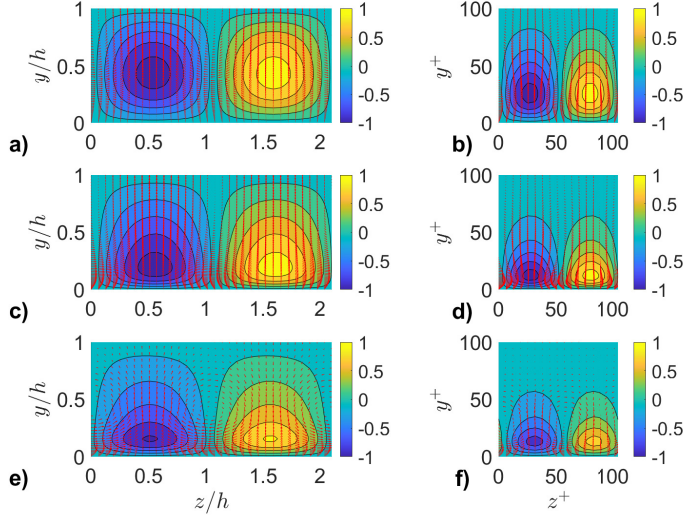


FIGURE 2. Velocity flow field at maximum amplification: The left columns (a, c and e) are for  $(\lambda_x^+, \lambda_z^+) = (\infty, 2)$ , while the right columns (b, d and f) are for  $(\lambda_x^+, \lambda_z^+) = (\infty, 100)$ ; (a, b) Contour plots from the LNS model optimal amplified state. (c, d) Contour plots from the  $\text{LNS}_\nu$  model optimal amplified state. (e, f) Contour plots from the LIM model optimal amplified state. The contour represents the streamwise velocity field, while the vector field represents the crossflow velocity field. The streamwise velocity is normalized by  $\max u_\tau$ , and contour lines are plotted in steps of 0.2 units. The wall is located at  $y/h = 0$  ( $y^+ = 0$ ), and the centerline is at  $y/h = 1$  ( $y^+ = 100$ ). The crossflow velocity vectors are scaled for visualization.

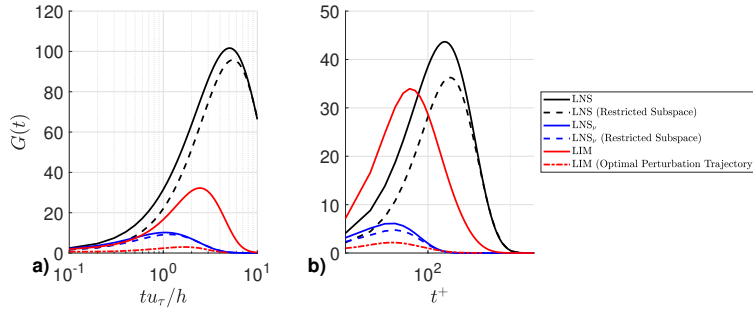


FIGURE 3. Transient energy growth from the optimal initial condition for a target time horizon: (a)  $(\lambda_x^+, \lambda_z^+) = (\infty, 2)$ , where the peak of  $G(t)$  for LNS is at  $tu_\tau/h = 5.1$ , for the  $\text{LNS}_\nu$  model at  $tu_\tau/h = 1.1$ , and for the LIM model at  $tu_\tau/h = 2.7$ ; (b)  $(\lambda_x^+, \lambda_z^+) = (\infty, 100)$ , where the peak of  $G(t)$  for LNS is at  $t^+ = 160$ , for the  $\text{LNS}_\nu$  model, at  $t^+ = 40$ , while for the LIM model, at  $t^+ = 70$ . LIM (Optimal Perturbation Trajectory) represents the evolution of the perturbed DNS ensemble using the optimal velocity perturbation from the LIM model.

viscosity assumption as an attempt to replicate the cross-stream velocity features of the LIM result.

Figure 3(a,b) shows the optimal transient energy growth computed from three different linear operators and the DNS ensemble DNS. A reasonable agreement is found between the full degree-of-freedom linear operator and linear operator from the restricted subspace when only the leading 30 controllability modes are retained. This suggests that the 30 controllability modes retained for the LIM model should sufficiently resolve the

transient growth for both wave numbers considered. We also see that the LNS operator experiences the largest amplification, sharply contrasting with the  $\text{LNS}_\nu$  operator for both wave numbers. However, the LIM transient growth dynamics are more akin to the  $\text{LNS}_\nu$  operator for  $(\lambda_x/h, \lambda_z/h) = (\infty, 2)$  while being closer to the LNS operator for  $(\lambda_x^+, \lambda_z^+) = (\infty, 100)$ . In contrast with the analytical modes, the optimal perturbations that lead to maximum gain predicted by the LIM operators in Figure 3 do not lead to a similar gain when applied to the DNS ensemble. These observations suggest that the LIM operator does not reliably capture the perturbation effect on turbulence, which differs from the conclusions by Penland (1989) and Farrell & Ioannou (2024). In terms of the transient amplification of the ensemble-averaged perturbations, the best model is still the  $\text{LNS}_\nu$  model of Hussain & Reynolds (1970), indicating that caution is required for the LIM approach.

## 6. Conclusion

In this work, two data-driven methodologies were investigated for the possibility of extracting a linear operator that is capable of representing the dynamics of small-amplitude perturbations applied to a turbulent flow. The SSI approach is deemed computationally impractical to be applied to the system identification for turbulent flows, as  $\mathcal{O}(10^6)$  samples are required for convergence of the covariance velocity statistics for a laminar setting. Unlike laminar flows, no analytical predictions can be made about the steady-state velocity covariance from the Lyapunov equation as turbulence is nonlinear, while many forcing modes are required to capture the degrees of freedom of the turbulent response. Hence, if extended to turbulent flows at higher Reynolds numbers, there is no good reference for the validation of statistical convergence and they are expected to be noisy due to the large number of forcing modes. The expected non-normal nature of the underlying linear operator will only complicate the SSI approach.

By contrast, only a DNS is required for the LIM approach, where  $\mathcal{O}(10^5)$  samples are required to achieve reasonable statistical convergence in the velocity covariance. The transient growth analysis carried out shows that the LIM approach predicts gains that are between the predictions of the LNS and  $\text{LNS}_\nu$  models, while the most amplified structures more closely resemble the  $\text{LNS}_\nu$  model. However, the comparison of the LIM-predicted gains to the ground truth DNS ensemble simulation shows non-negligible disparity between the predicted and observed energy growth.

Through the observed computational efficiency of the ensemble DNS carried out and the robust convergence of its results, a direct data-driven approach is currently under consideration for extraction of a linear operator that could model the perturbation dynamics of a turbulent flow.

## Acknowledgements

We are very grateful to Prof. Parviz Moin, Prof. Beverley McKeon, Dr. Salvador Gomez and Mr. Ahmed Elnahas for their support and hospitality during the CTR Summer Program. This work was also supported by the European Office of Aerospace Research and Development (EOARD) (FA8655-23-1-7023; Program Manager: Dr. Douglas Smith) (Z. C. K. and Y. H.) and by the Engineering and Physical Science Research Council (EPSRC) [D. C. (UROP bursary scheme); Y. B. and Y. H. (EP/T009365/1)] in the UK.

## REFERENCES

- BEWLEY, T. R., MOIN, P. & TEMAM, R. 2001 DNS-based predictive control of turbulence: an optimal benchmark for feedback algorithms. *J. Fluid Mech.* **447**, 179–225.
- BUTLER, K. M. & FARRELL, B. F. 1993 Optimal perturbations and streak spacing in wall-bounded turbulent shear flow. *Phys. Fluids* **5**, 774–777.
- CESS, R. D. 1958 A survey of the literature on heat transfer in turbulent tube flow. *Res. Rep.* 8–0529–R24.
- DEL ÁLAMO, J. C. & JIMÉNEZ, J. 2006 Linear energy amplification in turbulent channels. *J. Fluid Mech.* **559**, 205–213.
- FARRELL, B. F. & IOANNOU, P. J. 1993 Optimal excitation of three-dimensional perturbations in viscous constant shear flow. *Phys. Fluids* **5**, 1390–1400.
- FARRELL, B. F. & IOANNOU, P. J. 2007 Structure and spacing of jets in barotropic turbulence. *J. Atmos. Sci.* **64** (10), 3652–3665.
- FARRELL, B. F. & IOANNOU, P. J. 2024 Statistical state dynamics-based study of the stability of the mean statistical state of wall-bounded turbulence. *Phys. Rev. Fluids* **9**, 024605.
- HOPF, E. 1952 Statistical Hydromechanics and Functional Calculus. *J. Ration. Mech. Anal.* **1**, 87–123.
- HUSSAIN, A. K. M. F. & REYNOLDS, W. C. 1970 The mechanics of an organized wave in turbulent shear flow. *J. Fluid Mech.* **41** (2), 241–258.
- HWANG, Y. 2013 Near-wall turbulent fluctuations in the absence of wide outer motions. *J. Fluid Mech.* **723**, 264–288.
- HWANG, Y. & COSSU, C. 2010 Amplification of coherent streaks in the turbulent Couette flow: an input-output analysis at low Reynolds number. *J. Fluid Mech.* **643**, 333–348.
- JOVANOVIĆ, M. R. & BAMIEH, B. 2005 Componentwise energy amplification in channel flow. *J. Fluid Mech.* **543**, 145–183.
- LUCHINI, P., QUADRIO, M. & ZUCCHER, S. 2006 The phase-locked mean impulse response of a turbulent channel flow. *Phys. of Fluids* **18**, 121702.
- MARSTON, J. B., CONOVER, E. & SCHNEIDER, T. 2008 Statistics of an unstable barotropic jet from a cumulant expansion. *J. Atmos. Sci.* **65**, 1955–1966.
- MCKEON, B. J. & SHARMA, A. S. 2010 A critical-layer framework for turbulent pipe flow. *J. Fluid Mech.* **658**, 336–382.
- MOHAN, P., FITZSIMMONS, N. & MOSER, R. D. 2017 Scaling of Lyapunov exponents in homogeneous isotropic turbulence. *Phys. Rev. Fluids* **2**, 114606.
- PENLAND, C. 1989 Random Forcing and Forecasting Using Principal Oscillation Pattern Analysis. *Monthly Weather Review*, **117**, 2165–2185.
- PUJALS, G., GARCÍA-VILLALBA, M., COSSU, C. & DEPARDON, S. 2009 A note on optimal transient growth in turbulent channel flows. *Phys. Fluids* **21**, 015109.
- RUELLE, D. 1979 Microscopic fluctuations and turbulence. *Phys. Lett. A* **72**, 81–82.
- SCHMID, P. J. & HENNINGSON, D. S. 2001 *Stability and Transition in Shear Flows*. New York: Springer.