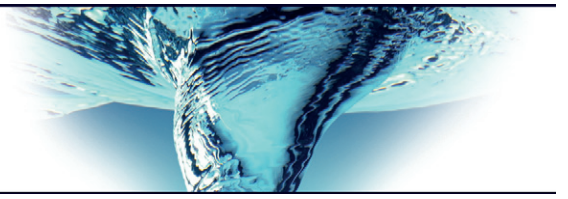
J. Fluid Mech. (2018), *vol.* 852, R3, doi:10.1017/jfm.2018.586



JFM RAPIDS
journals.cambridge.org/rapids

Data-driven reduced modelling of turbulent Rayleigh–Bénard convection using DMD-enhanced fluctuation–dissipation theorem

M. A. Khodkar[1] **and Pedram Hassanzadeh**[1,2,†]

[1]Department of Mechanical Engineering, Rice University, Houston, TX, USA

[2]Department of Earth, Environmental, and Planetary Sciences, Rice University, Houston, TX, USA

(Received 25 May 2018; revised 21 July 2018; accepted 22 July 2018;
first published online 6 August 2018)

A data-driven model-free framework is introduced for the calculation of reduced-order models (ROMs) capable of accurately predicting time-mean responses to external forcings, or forcings needed for specified responses, e.g. for control, in fully turbulent flows. The framework is based on using the fluctuation–dissipation theorem (FDT) in the space of a limited number of modes obtained from dynamic mode decomposition (DMD). Use of the DMD modes as the basis functions, rather than the commonly used proper orthogonal decomposition modes, resolves a previously identified problem in applying FDT to high-dimensional non-normal turbulent flows. Employing this DMD-enhanced FDT method (FDT_{DMD}), a linear ROM with horizontally averaged temperature as state vector is calculated for a 3D Rayleigh–Bénard convection system at a Rayleigh number of 10^6 using data obtained from direct numerical simulation. The calculated ROM performs well in various tests for this turbulent flow, suggesting FDT_{DMD} as a promising method for developing ROMs for high-dimensional turbulent systems.

Key words: low-dimensional models, nonlinear dynamical systems, turbulent convection

1. Introduction

The development of accurate reduced-order models (ROMs) for high-dimensional and complex turbulent systems is the subject of ever-growing interest and extensive research (Mezić 2013; Rowley & Dawson 2017). For example, reduced-order modelling of buoyancy-driven turbulence, which is prevalent in many engineering flows (e.g. energy systems) and natural flows (e.g. atmospheric/ocean circulations), has been actively pursued by the fluid dynamics and climate science communities in

† Email address for correspondence: pedram@rice.edu

© Cambridge University Press 2018

https://doi.org/10.1017/jfm.2018.586 Published online by Cambridge University Press

the past few decades; see below, and also Tu *et al.* (2014) and Khodkar *et al.* (2018) and references therein.

In many reduced-order modelling efforts, an alternative to the computationally prohibitive high-dimensional systems of the nonlinear partial differential equations governing the turbulent fluid flow is sought in the form of low-dimensional systems of ordinary differential equations (ODEs), such as the linear ROM

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{L}\boldsymbol{x}(t) + \boldsymbol{f}(t). \tag{1.1}$$

Here, $\boldsymbol{x}$ and $\boldsymbol{L}$ are respectively the state vector of the system and the evolution operator or linear response function. The term $\boldsymbol{f}(t)$ may include external forcings/actuations (e.g. controlling inputs) and stochastic representation of unresolved scales/physics. The calculation of accurate $\boldsymbol{L}$ for high-dimensional nonlinear systems such as fully turbulent flows using data-driven methods is the goal of many reduced modelling studies, including the present one.

In recent years, significant effort, particularly in the fluid dynamics community, has been focused on calculating $\boldsymbol{L}$ using some variant of dynamic mode decomposition (DMD) (e.g. Rowley *et al.* 2009; Schmid 2010; Tu *et al.* 2014; Williams, Kevrekidis & Rowley 2015; Arbabi & Mezić 2017; Brunton *et al.* 2017; Korda & Mezić 2018), which provides a finite-dimensional data-driven approximation (see § 2) to the system's Koopman operator, which is infinite-dimensional (Koopman 1931; Mezić 2005). DMD-based methods have been applied to a variety of fluid flows (Mezić 2013; Tu *et al.* 2014; Rowley & Dawson 2017), including buoyancy-driven turbulence (e.g. Kramer *et al.* 2017). Although these studies have produced promising results, particularly not far from the onset of linear instability, application of these methods to fully turbulent flows is currently the subject of extensive research.

In climate science, the focus has mainly been on using the fluctuation–dissipation theorem (FDT) (Leith 1975; Majda, Abramov & Grote 2005). The FDT, a powerful tool from statistical physics (Nyquist 1928; Kubo 1966), provides a data-driven approximation of $\boldsymbol{L}$ for nonlinear systems from the Fokker–Planck equation; see § 2. The $\boldsymbol{L}$ calculated using the FDT ($\boldsymbol{L}_{FDT}$ hereafter) is of particular interest because it is, theoretically, expected to predict long-time-mean responses to external forcings or forcings needed for a specified mean response in a nonlinear system via (1.1) (Majda *et al.* 2005). The FDT has been found to work well when applied to very simple models of geophysical turbulence such as the Lorenz equations. However, the calculation of accurate $\boldsymbol{L}$ for more complex systems such as the quasi-geostrophic equations or large-scale atmospheric turbulence has been found to be challenging (Gritsun & Branstator 2007; Cooper & Haynes 2011; Fuchs, Sherwood & Hernandez 2015; Lutsko, Held & Zurita-Gotor 2015; Hassanzadeh & Kuang 2016*b*). The latter study showed that a commonly used step that involves employing the leading (orthogonal) modes obtained from proper orthogonal decomposition (POD) as basis functions for truncating the data can lead to significant inaccuracy in $\boldsymbol{L}$ if the system is non-normal, which is common in geophysical flows. This step is necessary when the dataset is short, as is often the case for high-dimensional systems; see § 2 for further discussion.

As a result, it is worthwhile to further examine the performance of $\boldsymbol{L}_{FDT}$ in the context of a canonical fully turbulent flow system and explore whether basis functions other than POD modes can improve the performance of the FDT for developing ROMs for high-dimensional systems. Along these lines, the purpose of this study is twofold.

https://doi.org/10.1017/jfm.2018.586 Published online by Cambridge University Press

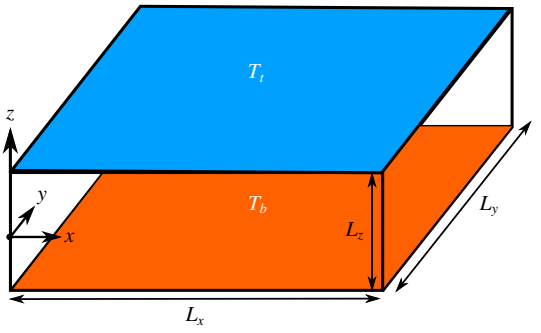


FIGURE 1. The 3D RBC system. The temperature at the top (bottom) wall is held at T_t (T_b), and $\Delta T = T_b - T_t > 0$. The horizontal directions (x–y) are periodic. The no-slip boundary condition is enforced at the walls. Here, $L_x = L_y = \pi L_z$, the Prandtl number is $Pr = 0.707$ and $Ra = 10^6$.

(1) To examine the performance of the FDT in calculating $\boldsymbol{L}$ for a fully turbulent flow, i.e. 3D Rayleigh–Bénard convection (RBC) at a Rayleigh number of $Ra = 10^6$ (figure 1). Direct numerical simulation (DNS) of RBC, a fitting prototype for buoyancy-driven turbulence, is used to generate the data for the FDT.

(2) To show that the use of DMD modes, rather than the commonly used POD modes (also known as EOF modes), as the basis functions in the FDT calculation can resolve the problem previously identified in Hassanzadeh & Kuang (2016*b*) and remarkably improve the performance of the FDT applied to high-dimensional turbulent systems.

Furthermore, this work aims to better connect the seemingly independent advances in the fluid dynamics and climate science communities. It is worth mentioning here, and further discussing in § 2, that the FDT and DMD are not unrelated. In fact, another method, called linear inverse modelling (LIM; Penland 1989), that is also derived from the Fokker–Planck equation and is closely related to the FDT is, as pointed out by Tu *et al.* (2014), mathematically equivalent to DMD, although LIM and DMD are derived using different concepts. These connections are not surprising given that the Koopman operator is the adjoint of the Perron–Frobenius operator (Klus, Koltai & Schütte 2016) and that the latter is connected to the Fokker–Planck equation (Lasota & Mackey 2013).

This paper is structured as follows. The formulations of the FDT and DMD are discussed in § 2. The 3D RBC system, its ROM and the DNS solver are described in § 3. In § 4, the accuracy of $\boldsymbol{L}$ in predicting the time-mean response to forcing is examined for the FDT with basis functions of POD modes (FDT_{POD}) and DMD modes (FDT_{DMD}) using DNS of RBC and stochastic ODEs (SDEs). A summary and future work are discussed in § 5.

2. The FDT and DMD

Let $\boldsymbol{x}_t \in \mathbb{R}^m$ be time-mean-removed measurements (e.g. from DNS data) of the state vector (which might involve velocity and temperature) over m grid points at time t, and

$$\boldsymbol{X}_o = \{\boldsymbol{x}_{\Delta t}\ \boldsymbol{x}_{2\Delta t}\ \ldots\ \boldsymbol{x}_{N\Delta t}\}, \quad \boldsymbol{X}_\tau = \{\boldsymbol{x}_{\tau+\Delta t}\ \boldsymbol{x}_{\tau+2\Delta t}\ \ldots\ \boldsymbol{x}_{\tau+N\Delta t}\}, \tag{2.1a,b}$$

https://doi.org/10.1017/jfm.2018.586 Published online by Cambridge University Press

where Δt is the sampling interval and N is the number of samples. Therefore, $\boldsymbol{X}_0$ and $\boldsymbol{X}_\tau$ are $m \times N$ matrices. Below, we present the mathematical formulation and numerical procedure for calculating $\boldsymbol{L}$ from matrices like $\boldsymbol{X}_o$ and $\boldsymbol{X}_\tau$ using the FDT, LIM and DMD. It is more convenient to start with the latter.

2.1. *The DMD and LIM*

Following the exact DMD formulation of Tu *et al.* (2014), the operator $\boldsymbol{A}_{DMD} = \exp(\boldsymbol{L}_{DMD}\tau)$ is calculated as

$$\boldsymbol{A}_{DMD} = \boldsymbol{X}_\tau \boldsymbol{X}_o^{+}. \tag{2.2}$$

Here, $+$ denotes the pseudoinverse. The DMD modes (values) are the eigenvectors (values) of $\boldsymbol{A}_{DMD}$. In practice, one often uses $\tau = \Delta t$ and calculates the reduced singular value decomposition (SVD) $\boldsymbol{X}_o = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\dagger}$ and then $\boldsymbol{A}_{DMD} = \boldsymbol{U}^{\dagger}\boldsymbol{X}_\tau \boldsymbol{V}\boldsymbol{S}^{-1}$, where $\dagger$ denotes the adjoint. It should be noted that some of the DMD modes might be complex; when we later choose a subset of DMD modes as basis functions, we ensure that the complex conjugate of any chosen complex DMD mode is also included (see § 4.3 for more details).

Penland (1989) showed that the operator $\boldsymbol{A}_{LIM} = \exp(\boldsymbol{L}_{LIM}\tau)$ can be calculated, from the Fokker–Planck equation, as

$$\boldsymbol{A}_{LIM} = \boldsymbol{C}_\tau \boldsymbol{C}_o^{-1}, \tag{2.3}$$

where $\boldsymbol{C}_\tau = \boldsymbol{X}_\tau \boldsymbol{X}_o^{\dagger}$ is the lag-τ covariance matrix. In practice, for high-dimensional systems, covariance matrices are nearly singular (i.e. ill-conditioned), because often not enough data are available to accurately characterize a system that has a large number of degrees of freedom (i.e. the dataset is short). To overcome this problem, a common regularization strategy is to first project $\boldsymbol{x}_t$ onto the leading r POD/EOF modes (obtained from SVD of $\boldsymbol{X}_o$) and perform the calculations in (2.3) in this reduced space. The value of r is chosen such that the retained POD modes represent at least 95 % (or even 99 %) of the variance (Penland 1989; Ring & Plumb 2008; Lutsko *et al.* 2015).

It should be noted that because $\boldsymbol{X}_\tau \boldsymbol{X}_o^{+} = \boldsymbol{X}_\tau(\boldsymbol{X}_o^{\dagger}(\boldsymbol{X}_o \boldsymbol{X}_o^{\dagger})^{-1}) = \boldsymbol{C}_\tau \boldsymbol{C}_o^{-1}$, (2.2) and (2.3) are equivalent; see Tu *et al.* (2014) for further discussion. It should be pointed out that the Koopman operator, which describes the evolution of observables, and Perron–Frobenius operator, which describes the transition density function, are adjoints (e.g. Klus *et al.* 2016), and that if the stochastic noise vanishes, the Fokker–Planck operator reduces to the Perron–Frobenius operator; see, e.g., Lasota & Mackey (2013, chap. 11) and Giannakis (2017).

2.2. *The FDT*

According to the FDT (Kubo 1966; Leith 1975), the linear response function can be calculated as

$$\boldsymbol{L}_{FDT} = -\left[\int_0^{\infty} \boldsymbol{C}_\tau \boldsymbol{C}_0^{-1}\, \mathrm{d}\tau\right]^{-1}. \tag{2.4}$$

It should be noted that the integrand is basically $\boldsymbol{A}_{DMD}$ or $\boldsymbol{A}_{LIM}$, consistent with integrating $\boldsymbol{A} = \exp(\boldsymbol{L}\tau)$ over τ from 0 to ∞ if $\boldsymbol{L}$ only has decaying modes. For

https://doi.org/10.1017/jfm.2018.586 Published online by Cambridge University Press

details of the derivation of (2.4) from the Fokker–Planck equation see Majda *et al.* (2005) and Gritsun & Branstator (2007). It is of particular interest to find $\boldsymbol{L}_{FDT}$, because it allows calculation of the time-mean response to an imposed forcing or the forcing needed for a specified response via $\boldsymbol{L}_{FDT}\langle \boldsymbol{x}\rangle = -\langle \boldsymbol{f}\rangle$, where $\langle\,\rangle$ denotes long-time averaging. It should be noted that the key underlying assumption for the applicability of $\boldsymbol{L}_{FDT}$ is not that the system is linear, but that the forcing is weak enough such that the response of the nonlinear system changes linearly with the forcing (Gritsun & Branstator 2007; Cooper & Haynes 2011).

In practice, similarly to LIM, the calculations required to evaluate the expression appearing in (2.4) are typically carried out in the reduced space of the leading r POD modes to avoid singular covariance matrices. The upper bound of the integral is also replaced with a finite limit τ_∞. The reason(s) behind the inaccuracy of $\boldsymbol{L}_{FDT}$ calculated for high-dimensional systems (see § 1) is not fully understood, and often attributed to a number of potential fundamental and practical issues. For example, (2.4) (and (2.3)) is exact only if the statistics of $\boldsymbol{x}$ is Gaussian (Majda *et al.* 2005; Gritsun & Branstator 2007), which is not the case for turbulent flows such as atmospheric circulation (Cooper & Haynes 2011; Hassanzadeh, Kuang & Farrell 2014). Examples of practical issues include unsuitable choices of r or τ_∞, short datasets and shortcomings of POD modes as basis functions (Cooper, Esler & Haynes 2013; Hassanzadeh & Kuang 2016*b*). The latter issue is particularly significant and is addressed in the current study. However, first, we describe in § 3 the DNS dataset that is used for calculating matrices in (2.1).

3. The 3D RBC mathematical model, DNS solver and 1D ROM

The RBC system of figure 1 is modelled using the 3D Boussinesq equations. Choosing the height L_z, temperature $\Delta T = T_b - T_t$ and diffusive time scale $\tau_{diff} = L_z^2/\kappa$ (κ is the thermal diffusivity) as characteristic scales, the dimensionless equations are

$$\nabla^* \cdot \boldsymbol{u}^* = 0, \tag{3.1}$$

$$\frac{\partial \boldsymbol{u}^*}{\partial t^*} + (\boldsymbol{u}^* \cdot \nabla^*)\boldsymbol{u}^* = -\nabla^* p^* + Pr\,\nabla^{*2}\boldsymbol{u}^* + Ra\,Pr(T^* - T^*_{cond})\hat{\boldsymbol{e}}_z, \tag{3.2}$$

$$\frac{\partial T^*}{\partial t^*} + (\boldsymbol{u}^* \cdot \nabla^*)T^* = \nabla^{*2}T^*, \tag{3.3}$$

where $\boldsymbol{u}^*$, T^* and $T^*_{cond} = 1/2 - z^*$ are the 3D velocity field, temperature and conduction temperature profile respectively. The superscript $*$ indicates dimensionless variables and operators. We define the Rayleigh and Prandtl numbers as $Ra = (g\alpha\Delta T L_z^3)/(\nu\kappa)$ and $Pr = \nu/\kappa$, where g, α and ν are the gravitational acceleration, and thermal expansion coefficient and kinematic viscosity of the fluid respectively.

Equations (3.1)–(3.3) for the system shown in figure 1 are simulated (at $Ra = 10^6$) using a pseudospectral Fourier–Fourier–Chebyshev DNS solver with a resolution of $128 \times 128 \times 129$ (see Khodkar *et al.* (2018) for more detail). The spatio-temporal analysis of the DNS data in Khodkar *et al.* (2018) shows that the flow is fully turbulent at this Rayleigh number, which is around 585 times higher than the critical Rayleigh number for the onset of linear instability.

As shown in Khodkar *et al.* (2018), a 1D ROM in the form of (1.1) can be formulated for this 3D RBC system,

$$\dot{\bar{\theta}} = \boldsymbol{L}\,\bar{\theta} + \bar{f}, \tag{3.4}$$

https://doi.org/10.1017/jfm.2018.586 Published online by Cambridge University Press

where the overbar indicates horizontal (x–y) averaging. Here, the state vector $\boldsymbol{x} = \overline{\theta}(z,\ t)$ is the response of the horizontally averaged temperature to external forcing $\overline{f}(z,\ t)$, i.e. the deviation from the long-time-mean horizontally averaged temperature of the unforced system (hereafter, 'unforced' systems refer to (3.1)–(3.3); in the forced systems, an external forcing f is added to (3.3)). Here, '1D' highlights that the ROM describes the system in only one spatial direction, z, as the state vector is horizontally averaged. The linear response function $\boldsymbol{L}$ in (3.4) includes the vertical heat flux by molecular diffusion as well as vertical eddy heat flux. Because of the latter, $\boldsymbol{L}$ cannot be derived directly from (3.1)–(3.3). Khodkar *et al.* (2018) showed that $\boldsymbol{L}$ can be accurately calculated using the Green's function (GRF) method (Kuang 2010; Hassanzadeh & Kuang 2016*a*), which requires many forced DNS of (3.1)–(3.3). As demonstrated in § 4, using the FDT, $\boldsymbol{L}$ in (3.4) can be accurately calculated from a dataset obtained from unforced DNS.

4. Results

4.1. *The DNS of RBC at* $Ra = 10^6$*:* FDT_{POD}

From DNS of the unforced system, after the flow reaches quasi-equilibrium, $N = 1.1 \times 10^5$ samples of $\overline{T}(z,\ t) - \langle \overline{T} \rangle (z)$ are collected every $\Delta t = 0.12\tau_{adv}$, where $\tau_{adv} = \sqrt{L_z/(g\alpha\Delta T)}$ is the advective time scale. As stated in Khodkar *et al.* (2018), in this system, $\tau_{adv} = 0.4\tau_d = 0.0012\tau_{diff}$, where τ_d is the decorrelation time of the leading POD mode (POD1). Using these data, $\boldsymbol{L}_{FDT_{POD}}$ is calculated from (2.4) for various values of r and τ_∞. Several tests involving prediction of the time-mean response to an external forcing or the forcing needed for a specific time-mean response are used to evaluate the accuracy of the calculated $\boldsymbol{L}_{FDT_{POD}}$. The 'true' responses or forcings for these tests are obtained using DNS of forced systems. Figure 2 depicts the results for four of these tests.

As shown in figure 2(*a*–*c*), $\boldsymbol{L}_{FDT_{POD}}$ predicts the pattern of the time-mean responses well, but generally over- or underestimates the amplitudes. Figure 2(*d*) demonstrates the accuracy of $\boldsymbol{L}_{FDT_{POD}}$ for the inverse problem (i.e. flow control): prediction of the forcing needed to produce a specified change in the time-mean flow (i.e. a target response). As before, the FDT-predicted forcing can produce the pattern of the target reasonably well, but the amplitude is incorrect. The results presented in this figure are calculated using the optimal (r, τ_∞), obtained from exploring the accuracy of the predicted responses/forcings in each case over a range of these two parameters (figure 3). We find that for all tests, ($r = 20$, $\tau_\infty = 0.83\tau_d$) is optimal and leads to the closest agreement with the truth (i.e. DNS). These results suggest that for the best accuracy at $N = 1.1 \times 10^5$, independent of the forcing, the spatial dimension of the original samples $\boldsymbol{x}_t$ ($m = 129$) should be reduced to $r = 20$, and that τ_∞, which the accuracy is notably sensitive to, should be chosen slightly less than the decorrelation time of POD1 (τ_d). The latter is consistent with the findings of Gritsun & Branstator (2007) and Hassanzadeh & Kuang (2016*b*) in climate models. Figure 3(*c*,*f*) shows how the accuracy improves by increasing N.

The results shown in figures 2 and 3 (and more tests, not shown) are promising, particularly given that the flow is complex and fully turbulent. However, the performance of $\boldsymbol{L}_{FDT_{POD}}$ is still not fully satisfactory as the predicted amplitudes are inaccurate and the FDT_{POD} is substantially outperformed by the accurate but computationally demanding (and not model-free) GRF method. An analysis by Khodkar *et al.* (2018) based on evaluation of the ϵ-pseudospectrum showed that the RBC system under consideration here is moderately non-normal (see their figure 10*b*),

https://doi.org/10.1017/jfm.2018.586 Published online by Cambridge University Press

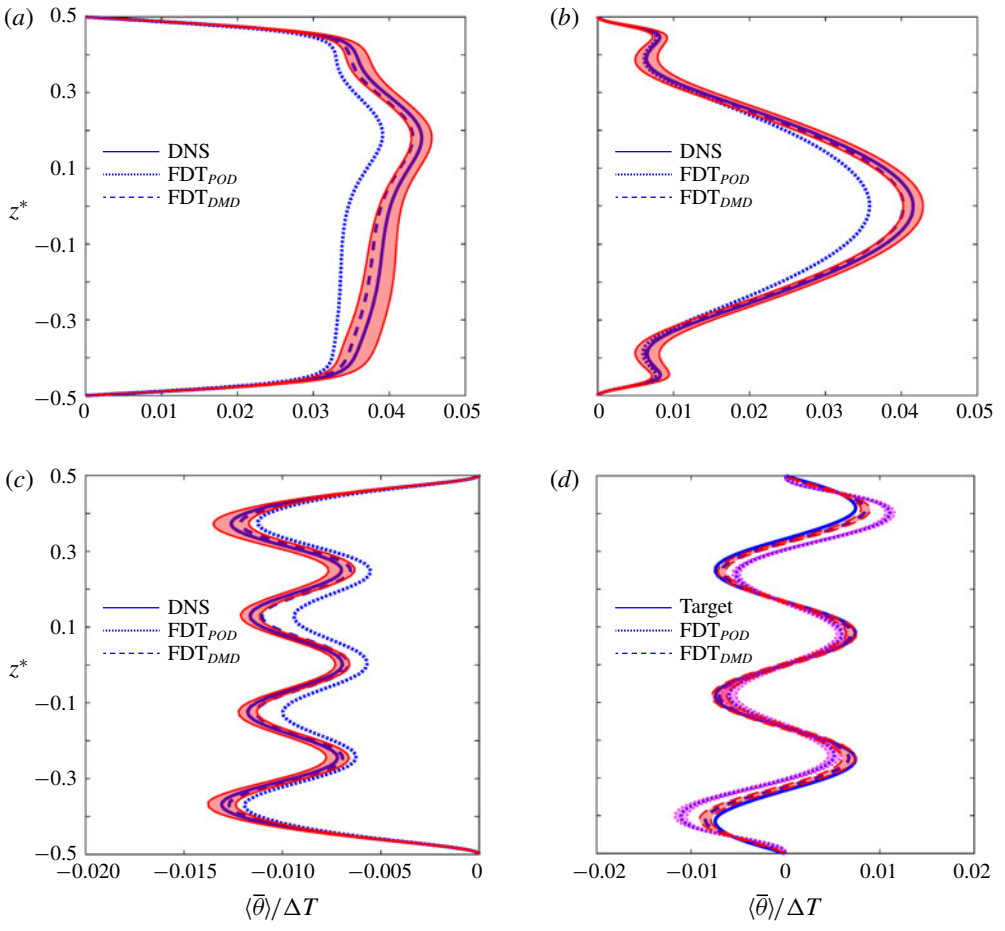


FIGURE 2. Time-mean responses $\langle\overline{\theta}\rangle$ to forcings in the form of $f_i = (\Delta T/\tau_{diff})\hat{f}_i$: (*a*) $\hat{f}_1 = 10\exp[-(z^*-0.2)^2/0.1^2]$, (*b*) $\hat{f}_2 = 20\cos(2\pi z^*)$, (*c*) $\hat{f}_3 = 20\cos(8\pi z^*)$. To quantify the 'true' responses and their uncertainty, each long forced DNS dataset is divided into eight equal segments and solid lines (red shading) show their mean (± 1 standard deviation). Dotted (dashed) lines show $-\boldsymbol{L}^{-1}_{FDT_{POD}} f_i$ ($-\boldsymbol{L}^{-1}_{FDT_{DMD}} f_i$) where the optimal values of $(r, \tau_\infty/\tau_d) = (20, 0.833)$ are used in (2.4) (see the text). (*d*) The inverse problem: the forcing needed for a specified time-mean response $\langle\overline{\theta}\rangle_{target}$ (solid line) is calculated as $-\boldsymbol{L}_{FDT_{POD}}\langle\overline{\theta}\rangle_{target}$ and $-\boldsymbol{L}_{FDT_{POD}}\langle\overline{\theta}\rangle_{target}$. To examine the accuracy, long DNS with these forcings is conducted, and the dashed/dotted lines show the calculated $\langle\overline{\theta}\rangle$.

which suggests that the performance of $\boldsymbol{L}_{FDT_{POD}}$ might be suffering from the same problem as identified in Hassanzadeh & Kuang (2016*b*): use of the leading r POD modes, which are orthogonal, can significantly degrade the performance of FDT$_{POD}$ if the system is non-normal, even if the $r\,(<m)$ modes explain a large percentage of the variance. This problem, and a potential remedy based on using DMD rather than POD modes for basis functions, is best seen by considering simple 2×2 systems of SDEs. This is done below, followed by application of FDT$_{DMD}$ to the same DNS dataset in § 4.3.

https://doi.org/10.1017/jfm.2018.586 Published online by Cambridge University Press

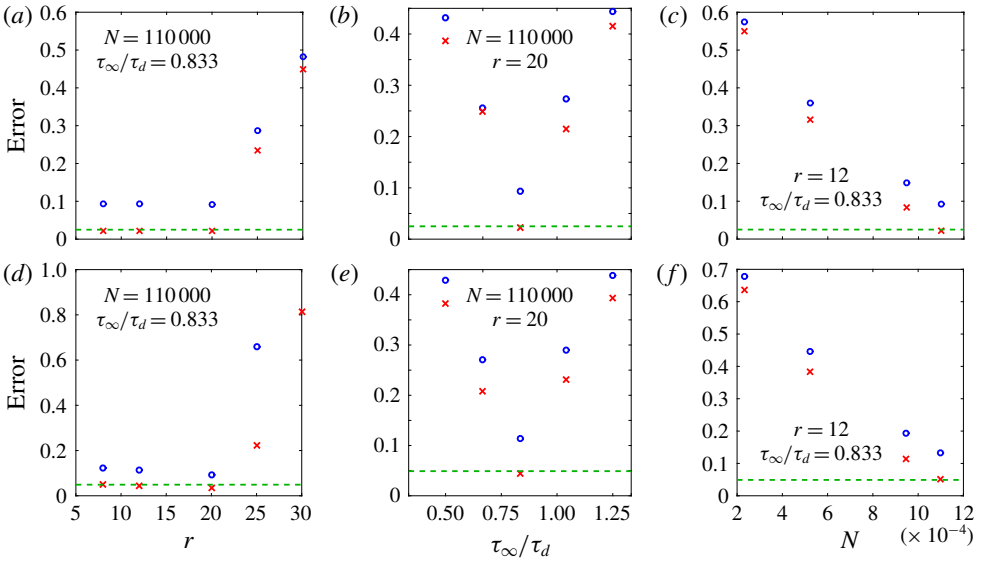


FIGURE 3. The accuracy of $\boldsymbol{L}_{FDT}$ in predicting the response $\langle\overline{\theta}\rangle$ to forcings f_1 (*a–c*) and f_2 (*d–f*), as functions of the number of leading POD or DMD modes used in the projection r (*a*,*d*), τ_∞/τ_d (*b*,*e*) and the length of the dataset N (*c*,*f*). The values of other parameters that are maintained constant can be seen in each panel. The blue circles and red crosses correspond to FDT_{POD} and FDT_{DMD} respectively. Since for the short datasets, the number of well-captured POD or DMD modes does not exceed 12, we have used $r = 12$ in (*c*) and (*f*) for the sake of a fair comparison. The horizontal dashed lines show the errors of $\boldsymbol{L}_{GRF}$ from Khodkar *et al.* (2018). Errors are calculated as $\|\langle\overline{\theta}\rangle_{\boldsymbol{L}} - \langle\overline{\theta}\rangle_{DNS}\|_2 / \|\langle\overline{\theta}\rangle_{DNS}\|_2$.

4.2. *Normal, non-normal and nonlinear SDEs:* FDT_{POD} *and* FDT_{DMD}

We consider a two-dimensional SDE,

$$\dot{z} = \boldsymbol{A}z + \boldsymbol{\xi} + \boldsymbol{f}, \tag{4.1}$$

where $z^{\mathrm{T}} = [z_1\ z_2]$, $\boldsymbol{\xi}$ is Gaussian white noise and $\boldsymbol{f}$ is a constant forcing. We use three test cases that are respectively normal, non-normal and nonlinear, with $\boldsymbol{A}$ being

$$\boldsymbol{A}_1 = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}, \quad \boldsymbol{A}_2 = \begin{bmatrix} -1 & 5 \\ 0 & -2 \end{bmatrix}, \quad \boldsymbol{A}_3 = \begin{bmatrix} -1 & -4z_2 + 2 \\ 0 & -2 \end{bmatrix}. \tag{4.2a–c}$$

We remark that the same matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ were used in Hassanzadeh & Kuang (2016*b*) for a similar analysis which was focused only on POD modes. Setting $\boldsymbol{f} = 0$, the SDE for each test case is integrated using the Euler–Maruyama method to generate datasets with 30 000 samples of z_t. The POD and DMD modes are calculated from these datasets following § 2 and shown in figure 4. As expected, for $\boldsymbol{A}_1$, the POD and DMD modes and eigenvectors are all identical (and each orthogonal), while for $\boldsymbol{A}_2$, the DMD modes and eigenvectors are the same (and non-orthogonal) but different from the POD modes (which are orthogonal). For $\boldsymbol{A}_3$, the DMD and POD modes differ.

Time-mean responses to an external forcing $\boldsymbol{f}$ that is mostly in the direction of POD1 but has a small projection onto POD2 are predicted using $\boldsymbol{L}_{FDT}$ when no truncation is performed (FDT_{full}) and when the data are truncated onto POD1

https://doi.org/10.1017/jfm.2018.586 Published online by Cambridge University Press

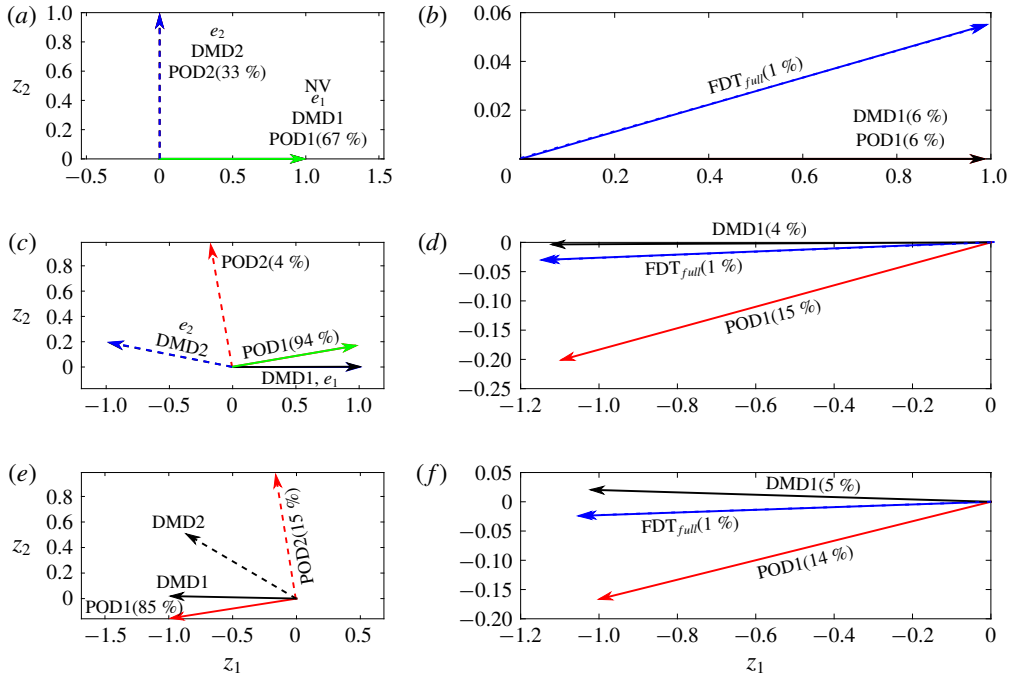


FIGURE 4. (*a*,*c*,*e*) Eigenvectors $\boldsymbol{e}$ (blue lines), DMD modes (black lines) and POD modes (red lines) for the systems with $\boldsymbol{A}_1$ (*a*,*b*), $\boldsymbol{A}_2$ (*c*,*d*) and $\boldsymbol{A}_3$ (*e*,*f*). Here, $\boldsymbol{e}_1$ and DMD1 are the slower-decaying modes. The percentages show the variance explained by each POD mode. (*b*,*d*,*f*) Time-mean response to the forcing $(0.9\text{POD1} + 0.1\text{POD2})/\|0.9\text{POD1} + 0.1\text{POD2}\|_2$, calculated using the FDT (2.4) with no projection (FDT_{full}; dashed blue lines), and projection onto the leading POD (FDT_{POD1}; red lines) or DMD (FDT_{DMD1}; black lines) mode. The solid blue lines show the analytically or numerically calculated true responses. The percentages show errors computed as $\|\boldsymbol{z}_{FDT} - \boldsymbol{z}_{true}\|_2 \times 100/\|\boldsymbol{z}_{true}\|_2$. It should be noted that the ranges of the axes vary among panels.

(FDT_{POD1}). For all test cases, FDT_{full} has an error of $\sim$1 %. While for the normal system FDT_{POD1} is relatively accurate (error $\sim$6 %), for the non-normal system the error is around 15 % even though POD1 explains 94 % of the variance. To explain the source of this inaccuracy, following Hassanzadeh & Kuang (2016*b*), we transfer (4.1) to the basis function space,

$$\dot{\boldsymbol{a}} = [\boldsymbol{B}^{-1}(\boldsymbol{E}\boldsymbol{\Lambda}\boldsymbol{E}^{-1})\boldsymbol{B}]\boldsymbol{a} + \boldsymbol{B}^{-1}\boldsymbol{f}, \tag{4.3}$$

where $\boldsymbol{a}^{\mathrm{T}} = [a_1 \;\; a_2]$ are the projection coefficients, columns of $\boldsymbol{B}$ contain the basis functions (e.g. POD modes), $\boldsymbol{E}$ ($\boldsymbol{\Lambda}$) contain the eigenvectors (values) of $\boldsymbol{A}$ and $\boldsymbol{\xi}$ is ignored for convenience. For a normal system, the POD modes are the same as the eigenvectors, and the matrix in the brackets reduces to the diagonal matrix $\boldsymbol{\Lambda}$, decoupling a_1 and a_2. Projections of $\boldsymbol{f}$ onto POD2 cannot be captured if $\boldsymbol{L}_{POD}$ is calculated only in the space of POD1; however, the accuracy of a_1 will not be affected, leading to the small error in figure 4(*b*). For non-normal systems, the POD modes and eigenvectors can be significantly different, leading to a coupling between a_1 and a_2 that strengthens with non-normality (Hassanzadeh & Kuang 2016*b*). Hence, even

https://doi.org/10.1017/jfm.2018.586 Published online by Cambridge University Press

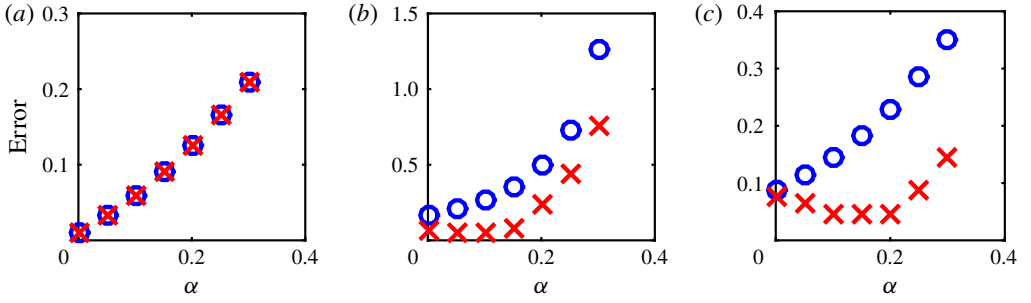


FIGURE 5. For systems with (*a*) $\boldsymbol{A}_1$, (*b*) $\boldsymbol{A}_2$ and (*c*) $\boldsymbol{A}_3$, errors in time-mean response predictions to forcing $[(1-\alpha)\text{POD1} + \alpha\text{POD2}]/\|(1-\alpha)\text{POD1} + \alpha\text{POD2}\|_2$. The blue circles and red crosses indicate FDT_{POD1} and FDT_{DMD1} respectively. The error is computed as $\|\boldsymbol{z}_{FDT} - \boldsymbol{z}_{true}\|_2/\|\boldsymbol{z}_{true}\|_2$. It should be noted that the range of the *y*-axis varies among panels.

small projections of $\boldsymbol{f}$ onto POD2 can substantially degrade the accuracy of a_1 (and thus the FDT prediction) if $\boldsymbol{L}_{FDT}$ is calculated only in the space of POD1 (figure 4*d*).

The above analysis suggests that the use of basis functions that approximate the system eigenvectors might improve the accuracy of $\boldsymbol{L}_{FDT}$. The discussion in § 2 and the results in figure 4 point to DMD modes as potential options. Indeed, use of the slower-decaying DMD mode as the basis function (FDT_{DMD1} hereafter) improves the accuracy compared with FDT_{POD1} by a factor of four for the non-normal system. Similarly, in the nonlinear system, the error of FDT_{DMD1} is 5 %, three times lower than the 15 % error of FDT_{POD1}. To further demonstrate the advantage of using the leading DMD rather than POD mode as the basis function, figure 5 shows that as the projection of the forcing onto POD2 increases in the non-normal and nonlinear systems, the accuracy of FDT-predicted responses rapidly degrades for FDT_{POD1} while FDT_{DMD1} shows a much better performance.

4.3. *The DNS of RBC at* $Ra = 10^6$*:* FDT_{DMD}

Using the unforced system DNS, we have calculated, following § 2.1 for $\tau = \Delta t$, all of the $m = 129$ DMD modes, and chosen the r slowest-decaying ones as basis functions for FDT_{DMD}. If the rth DMD mode is complex, we ensure that its complex conjugate, also a DMD mode, is included in the basis function space as well. Figures 2 and 3 show that $\boldsymbol{L}_{FDT_{DMD}}$ accurately predicts the pattern and amplitude of the time-mean responses and remarkably outperforms $\boldsymbol{L}_{FDT_{POD}}$ in all cases. The linear response function $\boldsymbol{L}_{FDT_{DMD}}$ has an accuracy that is equal to (or in some cases better than) that of $\boldsymbol{L}_{GRF}$. It should be noted that while DMD modes provide suitable basis functions for the FDT, we have found that $\boldsymbol{L}_{DMD}$ (or $\boldsymbol{L}_{LIM}$) cannot accurately predict the time-mean responses/forcings for tests similar to those in figure 2 (not shown).

5. Discussion and conclusions

The DMD-enhanced FDT method is shown to accurately predict the time-mean response to an external forcing, or the forcing needed for a specified response in a canonical buoyancy-driven turbulent flow, RBC at $Ra = 10^6$. Tests using the DNS of RBC and simple non-normal and nonlinear SDEs demonstrate the advantage in using

https://doi.org/10.1017/jfm.2018.586 Published online by Cambridge University Press

a limited number of leading (slowest-decaying) DMD modes over the commonly used POD modes as basis functions for the FDT calculations in (2.4).

This approach suggests a potential remedy to overcome the challenge identified by Hassanzadeh & Kuang (2016*b*) in applying the FDT to high-dimensional non-normal turbulent flows. Nonetheless, while the 1D linear ROM calculated using FDT_{DMD} remarkably outperforms FDT_{POD} and accurately predicts the pattern and amplitude of time-mean responses/forcings for the Rayleigh–Bénard turbulence considered here, how the DMD-enhanced FDT performs for computing 2D and 3D ROMs and for more complex higher-dimensional turbulent systems remains to be examined in future work. Large-scale atmospheric circulation, for which FDT_{POD} has been extensively attempted with mixed outcomes, will be a test case of particular interest.

Finally, given the non-normality of the system, the sensitivity of the calculated responses to the initial condition in initial-value integrations of (3.4) and to small changes in the forcing (in both DNS and (3.4)) should be investigated in subsequent studies.

Acknowledgements

We thank T. Antoulas, H. Babaee, M. Farazmand, P. Grover, M. Heinkenschloss and S. Nabi for insightful discussions, three anonymous reviewers for helpful comments, NSF grant AGS-1552385, NASA grant 80NSSC17K0266 and Rice University's Faculty Initiative Fund for support, and XSEDE Stampede2 (via allocation ATM170020) and Rice's DAVinCI cluster for providing computing resources.

References

Arbabi, H. & Mezić, I. 2017 Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the Koopman operator. *SIAM J. Appl. Dyn. Syst.* **16** (4), 2096–2126.

Brunton, S. L., Brunton, B. W., Proctor, J. L., Kaiser, E. & Kutz, J. N. 2017 Chaos as an intermittently forced linear system. *Nat. Commun.* **8** (1), 1–9.

Cooper, F. C., Esler, J. G. & Haynes, P. H. 2013 Estimation of the local response to a forcing in a high dimensional system using the fluctuation–dissipation theorem. *Nonlinear Process. Geophys.* **20** (2), 239–248.

Cooper, F. C. & Haynes, P. H. 2011 Climate sensitivity via a nonparametric fluctuation–dissipation theorem. *J. Atmos. Sci.* **68** (5), 937–953.

Fuchs, D., Sherwood, S. & Hernandez, D. 2015 An exploration of multivariate fluctuation dissipation operators and their response to sea surface temperature perturbations. *J. Atmos. Sci.* **72** (1), 472–486.

Giannakis, D. 2017 Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Appl. Comput. Harmon. Anal.* (in press).

Gritsun, A. & Branstator, G. 2007 Climate response using a three-dimensional operator based on the fluctuation–dissipation theorem. *J. Atmos. Sci.* **64** (7), 2558–2575.

Hassanzadeh, P. & Kuang, Z. 2016*a* The linear response function of an idealized atmosphere. Part I. Construction using Green's functions and applications. *J. Atmos. Sci.* **73** (9), 3423–3439.

Hassanzadeh, P. & Kuang, Z. 2016*b* The linear response function of an idealized atmosphere. Part II. Implications for the practical use of the fluctuation–dissipation theorem and the role of operator's non-normality. *J. Atmos. Sci.* **73** (9), 3441–3452.

Hassanzadeh, P., Kuang, Z. & Farrell, B. F. 2014 Responses of midlatitude blocks and wave amplitude to changes in the meridional temperature gradient in an idealized dry GCM. *Geophys. Res. Lett.* **41** (14), 5223–5232.

Khodkar, M. A., Hassanzadeh, P., Saleh, N. & Grover, P. 2018 Reduced-order modeling of fully turbulent buoyancy-driven flows using the Green's function method. arXiv:1805.01596.

https://doi.org/10.1017/jfm.2018.586 Published online by Cambridge University Press

KLUS, S., KOLTAI, P. & SCHÜTTE, C. 2016 On the numerical approximation of the Perron–Frobenius and Koopman operator. *J. Comput. Dyn.* **3** (1), 51–79.

KOOPMAN, B. O. 1931 Hamiltonian systems and transformation in Hilbert space. *Proc. Natl Acad. Sci. USA* **17** (5), 315–318.

KORDA, M. & MEZIĆ, I. 2018 On convergence of extended dynamic mode decomposition to the Koopman operator. *J. Nonlinear Sci.* **28** (2), 687–710.

KRAMER, B., GROVER, P., BOUFOUNOS, P., NABI, S. & BENOSMAN, M. 2017 Sparse sensing and DMD-based identification of flow regimes and bifurcations in complex flows. *SIAM J. Appl. Dyn. Syst.* **16** (2), 1164–1196.

KUANG, Z. 2010 Linear response functions of a cumulus ensemble to temperature and moisture perturbations and implications for the dynamics of convectively coupled waves. *J. Atmos. Sci.* **67** (4), 941–962.

KUBO, R. 1966 The fluctuation–dissipation theorem. *Rep. Prog. Phys.* **29** (1), 255–284.

LASOTA, A. & MACKEY, M. C. 2013 *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, vol. 97. Springer Science and Business Media.

LEITH, C. E. 1975 Climate response and fluctuation dissipation. *J. Atmos. Sci.* **32** (10), 2022–2026.

LUTSKO, N. J., HELD, I. M. & ZURITA-GOTOR, P. 2015 Applying the fluctuation–dissipation theorem to a two-layer model of quasigeostrophic turbulence. *J. Atmos. Sci.* **72** (8), 3161–3177.

MAJDA, A., ABRAMOV, R. V. & GROTE, M. J. 2005 *Information Theory and Stochastics for Multiscale Nonlinear Systems*, vol. 25. American Mathematical Society.

MEZIĆ, I. 2005 Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dyn.* **41** (1–3), 309–325.

MEZIĆ, I. 2013 Analysis of fluid flows via spectral properties of the Koopman operator. *Annu. Rev. Fluid Mech.* **45**, 357–378.

NYQUIST, H. 1928 Thermal agitation of electric charge in conductors. *Phys. Rev.* **32** (1), 110–113.

PENLAND, C. 1989 Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Weath. Rev.* **117** (10), 2165–2185.

RING, M. J. & PLUMB, R. A. 2008 The response of a simplified GCM to axisymmetric forcings: applicability of fluctuation–dissipation theorem. *J. Atmos. Sci.* **65** (12), 3880–3898.

ROWLEY, C. W. & DAWSON, S. T. M. 2017 Model reduction for flow analysis and control. *Annu. Rev. Fluid Mech.* **49**, 387–417.

ROWLEY, C. W., MEZIĆ, I., BAGHERI, S., SCHLATTER, P. & HENNINGSON, D. S. 2009 Spectral analysis of nonlinear flows. *J. Fluid Mech.* **641**, 115–127.

SCHMID, P. J. 2010 Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.* **656**, 5–28.

TU, J. H., ROWLEY, C. W., LUCHTENBURG, D. M., BRUNTON, S. L. & KUTZ, J. N. 2014 On dynamic mode decomposition: theory and applications. *J. Comput. Dyn.* **1**, 391–421.

WILLIAMS, M. O., KEVREKIDIS, I. G. & ROWLEY, C. W. 2015 A data-driven approximation of the Koopman operator: extending dynamic mode decomposition. *J. Nonlinear Sci.* **25** (6), 1307–1346.

https://doi.org/10.1017/jfm.2018.586 Published online by Cambridge University Press