arXiv:1812.09438v1 [physics.flu-dyn] 22 Dec 2018

Data-driven Spatio-temporal Prediction of High-dimensional Geophysical Turbulence using Koopman Operator Approximation

M. A. Khodkar[1],* Athanasios C. Antoulas[2,3], and Pedram Hassanzadeh[1,4]†
[1]*Department of Mechanical Engineering, Rice University, Houston, TX 77005, USA*
[2]*Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005, USA*
[3]*Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg 39106, Germany*
[4]*Department of Earth, Environmental and Planetary Sciences, Rice University, Houston, TX 77005, USA*

We show the outstanding skills of a data-driven low-dimensional linear model in predicting the spatio-temporal evolution of turbulent Rayleigh-Bénard convection. The model is based on dynamic mode decomposition with delay-embedding, which provides a data-driven finite-dimensional approximation to the system's Koopman operator. The model is built using vector-valued observables from direct numerical simulations, and can provide accurate predictions for 100s of advective time-scales. Similar high prediction skills are found for the Kuramoto-Sivashinsky equation in the strongly-chaotic regimes.

Predicting the spatio-temporal evolution of high-dimensional nonlinear dynamical systems, such as turbulent flows, has been of long-standing interest in science and engineering [1, 2]. For example, forecasting turbulent flows plays a key role in controlling and optimizing various engineering systems (e.g. wind farms) and predicting the state of the atmosphere and/or ocean (e.g. day-to-day weather) [3–6]. For many of these problems, an objective of particular interest and wide-ranging applications is predicting extreme events at some useful lead time [7–9].

Data-driven prediction of chaotic and turbulent flows has received significant attention in recent years, in particular, for problems in which the high-dimensional, nonlinear governing equations cannot be solved fast enough to be useful (e.g. for online control/optimization), or, in which some of the physical processes (and thus the governing equations) are not fully understood but observational data from the past are available (e.g. the weather/climate systems) [10–15]. Rapid advances have been recently made in this area based on using techniques from machine learning or approximating the Koopman operator. These approaches involve using the past data to build/train a model that can produce accurate and fast predictions about the future spatio-temporal evolution of the flow. Promising results for prototypes of chaotic flows, e.g. Lorenz-63/96 and Kuramoto-Sivashinsky (K-S) equations, have been reported using machine learning methods such as recurrent neural networks [16, 17] and reservoir computing [18].

The Koopman operator [19], which is an infinite-dimensional linear operator that shifts observables forward in time, offers a powerful framework for analyzing and tackling nonlinear systems such as fluid flows [20, 21]. Data-driven, finite-dimensional approximations of the Koopman operator, using methods such as dynamic mode decomposition (DMD) and its variants [22–29], have been extensively used to analyze various flows in recent years [21, 30]. Furthermore, in Brunton *et al.* [13] and Giannakis [14], skillful data-driven predictions of chaotic systems are obtained using Koopman-based methods. Building on the work of Tu *et al.* [24] in using DMD on a Hankel matrix containing delay-embedded observables [31], Arbabi and Mezić [26] introduced the Hankel-DMD method and proved its convergence to the Koopman operator. More recently, Arbabi *et al.* [32] used Hankel-DMD in a predictive-control framework to successfully stabilize a two-dimensional (2D) lid-driven cavity flow that has bifurcated to a limit cycle. These promising results for chaotic systems offer hope for data-driven prediction in fully-turbulent flows, which has not been much explored yet.

In this paper, we investigate using data-driven approximations of the Koopman operator for spatio-temporal prediction in 3D Rayleigh-Bénard convection (RBC), which is a widely-used prototype for buoyancy-driven turbulence and geophysical flows. We use data from long direct numerical simulation (DNS) of RBC at the Rayleigh number $Ra = 10^6$, at which the flow is fully turbulent [33, 34]. While using the standard DMD formulation or scalar-valued observables does not lead to any useful prediction skill, using an appropriate vector-valued observable in the Hankel-DMD framework renders remarkable prediction skills for the spatio-temporal evolution of the horizontally-averaged temperature. The approach yields similar prediction skills for the strongly-chaotic K-S systems.

The 3D RBC system consists of fluid between two horizontal walls that have a vertical spacing of H and fixed temperature difference ΔT (the bottom wall is warmer) [33]. The horizontal directions $(x - y)$ are periodic. The flow is modeled using the 3D Boussinesq equations for velocity $\mathbf{v}$, temperature T, and pressure p:

$$\nabla \cdot \mathbf{v} = 0\,, \quad \partial \mathbf{v}/\partial t + \mathbf{v}\cdot\nabla\mathbf{v} = -\nabla p + Pr\,\nabla^2\mathbf{v} + Ra\,Pr\,T\,\hat{z}\,, \quad \partial T/\partial t + \mathbf{v}\cdot\nabla T = \nabla^2 T\,. \tag{1}$$

At the walls ($z/H = \pm 0.5$), $\mathbf{v} = 0$. Here, $Pr = 0.707$ and $Ra = 10^6$ [35], which is ≈ 585 times higher than the critical Ra at the onset of linear instability [36]. The DNS of (1) is conducted using a Fourier-Fourier-Chebyshev pseudo-spectral solver with the resolution $128 \times 128 \times 129$. A DNS dataset is generated that contains multiple independent simulations started with different random initial conditions.

The objective here is to find a low-dimensional model that can provide fast and reasonably-accurate predication of the spatio-temporal evolution of horizontally-averaged temperature in the turbulent flow governed by (1). More specifically, we are interested in the evolution of $\theta(z,t) = \overline{T} - \langle \overline{T} \rangle$ (the overbar and $\langle\ \rangle$ indicate horizontal and long-time averaging, respectively), and aim to find a linear operator A such that $\boldsymbol{\theta}_{i+1} = A\boldsymbol{\theta}_i$, where $\boldsymbol{\theta}_i$ is a vector of θ snapshot over n grid points at time $t = i\Delta t$. To see the rationale behind seeking such low-dimensional model and the physics involved in A, we can average the heat equation in (1) over the horizontal directions and then subtract the long-time mean [33]. We find that the dynamics of $\theta(z,t)$ is governed by vertical heat fluxes due to molecular diffusion (conduction) and turbulent eddy diffusion. If we assume that the vertical eddy heat flux can be expressed as a function of θ alone, and that the linear representation of their functional relationship is sufficiently accurate for our puprose, then one finds the linear ordinary differential equation $\dot{\theta} = L\theta$ [33, 34], or, in the discrete form, $\boldsymbol{\theta}_{i+1} = A\boldsymbol{\theta}_i$. The linear operators L and $A = \exp(L\Delta t)$ represent molecular and eddy heat fluxes, however, due to the lack of a complete theory for turbulence, such a representation of eddy fluxes requires ad hoc closure schemes [34, 37]. Instead, here we seek to find A using a data-driven method without using any closure model.

Using the data-driven Hankel-DMD method, N samples of vector-valued observables $\boldsymbol{\theta}_i$ from the DNS dataset are arranged in the Hankel matrix $\mathcal{H}$

$$\mathcal{H} = \begin{bmatrix} \boldsymbol{\theta}_1 & \boldsymbol{\theta}_2 & \dots & \boldsymbol{\theta}_{N-q+1} \\ \boldsymbol{\theta}_2 & \boldsymbol{\theta}_3 & \dots & \boldsymbol{\theta}_{N-q+2} \\ \vdots & \vdots & \dots & \vdots \\ \boldsymbol{\theta}_q & \boldsymbol{\theta}_{q+1} & \dots & \boldsymbol{\theta}_N \end{bmatrix}, \tag{2}$$

where q is the delay-embedding dimension (part of the dataset used in (2) is referred to as the training set hereafter). The size of $\mathcal{H}$ is $(n \times q) \times (N - q + 1)$. Following the Exact DMD formulation [24], we then construct

$$X = \mathcal{H}(:, 1 : N - q)\,, \quad Y = \mathcal{H}(:, 2 : N - q + 1)\,, \tag{3}$$

and conduct reduced singular value decomposition $X = USV^\dagger$, in the subspace of leading r singular vectors ($\dagger$ indicates conjugate transpose). The data-driven approximation of A using Hankel-DMD is computed as

$$A_{HDMD} = U^\dagger Y V S^{-1}, \tag{4}$$

which has the size $r \times r$.

Once A_{HDMD} is calculated, a future vector-valued observable $\boldsymbol{\theta}_{m+1}$, that was not part of the training set, can be predicted from

$$\boldsymbol{\Theta}^r_{m+1} = A_{HDMD}\boldsymbol{\Theta}^r_m, \tag{5}$$

where $\boldsymbol{\Theta}^r_{m+1}$, a vector of length r, is $\boldsymbol{\Theta}_{m+1} = [\boldsymbol{\theta}_{m-q+2}\ \boldsymbol{\theta}_{m-q+3}\ \cdots\ \boldsymbol{\theta}_{m+1}]^*$, a vector of length $n \times q$, projected onto the subspace ($*$ indicates transpose). All values in $\boldsymbol{\Theta}_m = [\boldsymbol{\theta}_{m-q+1}\ \boldsymbol{\theta}_{m-q+2}\ \cdots\ \boldsymbol{\theta}_m]^*$ (and its projection onto the subspace, $\boldsymbol{\Theta}^r_m$), are either known from the initial condition or already predicted.

To test the predictive capabilities of Hankel-DMD, we have first computed A_{HDMD} using $\boldsymbol{\theta}_i$ from all grid points in the vertical direction (i.e. $n = 129$) and a training set with $N = 188000$, which is of length $t_N = N\Delta t \approx 21900\tau_{adv}$, where τ_{adv} is the advective time-scale [35]. Empirically-found optimal parameters $r = 1100$ and $q = 22$ are used [33]. Figure 1(a) compares the predicted time-series θ_{pred} (from (5)) at a given point against the *true* time-series from DNS (θ_{DNS}). The testing set ($t \geq 0$) and the training set have no overlap and are chosen from separate independent simulations. The time-series of θ_{pred} and θ_{DNS} agree well until $\approx 529\tau_{adv}$, when the time-series of θ_{pred} diverges and decays to zero. Similar high prediction skills and behavior are observed for the principal component (PC) of the leading mode of $\theta(z,t)$ obtained from proper orthogonal decomposition (Figure 1(b)). These comparisons show the remarkable capability of A_{HDMD} in accurately predicting not only the evolution of the DNS time-series at different time-scales, but also the amplitude and time of extreme events that are a few standard deviations large. Figure 2 further shows that A_{HDMD} can in fact accurately predict the evolution of the entire profile of θ. A property of θ_{pred}, which is highly desirable, is that it always satisfies the boundary conditions ($\theta = 0$ at the walls) even though the boundary conditions are not explicitly enforced during prediction.

It should again be emphasized that in the results presented above, no DNS data from $t \geq 0$ have been used during prediction, i.e. solving Eq. (5). In fact, the only $\boldsymbol{\theta}_{DNS}$ used during prediction are the last $q = 22$ snapshots ($\approx 3\tau_{adv}$) from $t < 0$ needed as the initial conditions in (5).

To examine how the prediction skills of A_{HDMD} scale with N, we define the relative error $E(t) = \{\|\boldsymbol{\theta}_{pred} -$

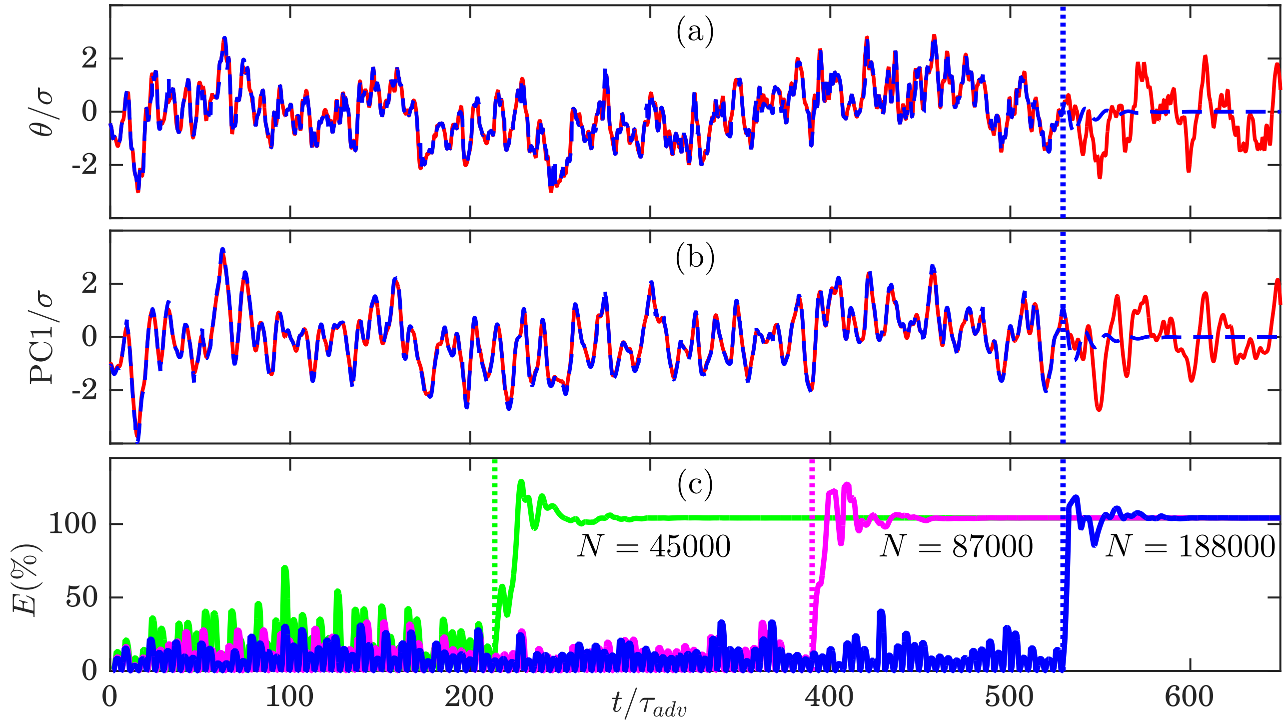

FIG. 1. (a) Time-series of θ at $z/H = -0.18$. The vertical dotted line shows the time t_l at which θ_{pred} (dashed blue line) begins to consistently diverge from θ_{DNS} (solid red line). σ is the standard deviation of the time-series. (b) Same as (a) but for PC1. (c) Relative error E, averaged over 8 independent tests, as N is varied. The longest dataset was used in (a)-(b). Time is scaled by advective time-scale τ_{adv}, which is related to other internal time-scales of the system as $\tau_{adv} \approx 0.36\tau_d \approx 0.04/\Lambda_{max}$, where τ_d is the e-folding decorrelation time of PC1 and Λ_{max} is the largest positive Lyapunov exponent.

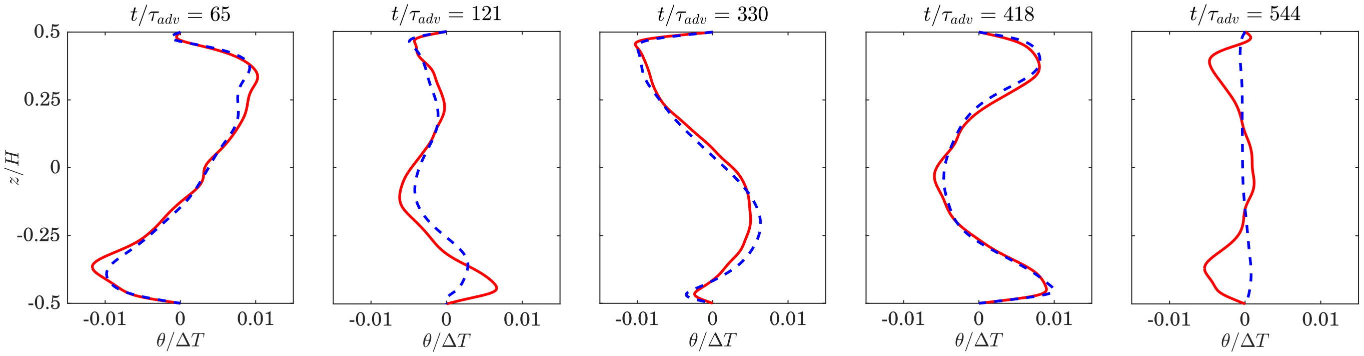

FIG. 2. Snapshots of θ_{DNS} (solid red line) and θ_{pred} (dashed blue line) at five times showing the skills of A_{HDMD} in predicting the profile of $\theta(z,t)$. θ_{pred} is computed using the same A_{HDMD} as in Figure 1(a)-(b). While θ_{pred} might seem to start to deviate from θ_{DNS} at some times, e.g. $t/\tau_{adv} = 121$, the prediction converges back to θ_{DNS}, e.g. $t/\tau_{adv} = 330$ or 418, before it starts to consistently diverge at $\approx 529\tau_{adv}$.

$\boldsymbol{\theta}_{DNS}\|/\,\|\boldsymbol{\theta}_{DNS}\|\}$, where $\|\ \|$ and $\{\ \}$ indicate the Euclidean norm and running average over $3\tau_{adv}$, respectively. From inspecting a number of test results, the prediction horizon t_l, at which θ_{pred} starts to consistently diverge from θ_{DNS}, can be reasonably defined as when E remains above 30% for at least $4\tau_{adv}$. Figure 1(c) shows $E(t)$ for three training datasets with $N = 45000$ ($\approx 5.2 \times 10^3\tau_{adv}$), $N = 87000$ ($\approx 10.1 \times 10^3\tau_{adv}$), and $N = 188000$ ($\approx 21.9 \times 10^3\tau_{adv}$). As N is almost doubled from 45000 to 87000, the prediction horizon nearly doubles too, and E_{ave} (which is E averaged over 8 independent tests and averaged from $t = 0$ to t_l) decreases from 16.3% to 9.72%. However, as N is further (more than) doubled to 188000, t_l increases from $390\tau_{adv}$ to $529\tau_{adv}$ (just a 36% improvement), and E_{ave} decreases only by $\approx 1\%$.

Figure 3 further shows the scaling of the prediction horizon with the size of the training set. We have also

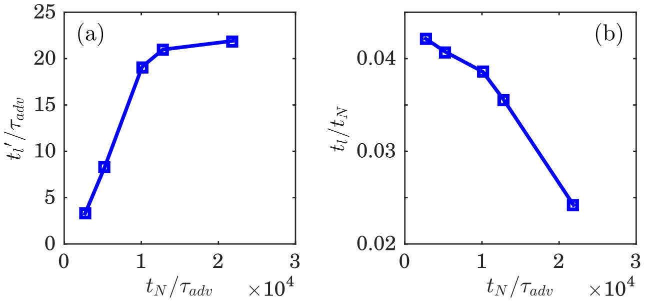

FIG. 3. Scaling of prediction horizon with the length of the training set t_N: (a) t_l', (b) t_l.

defined another measure of prediction horizon t_l' as when E first exceeds 20%, which can be important in practical applications. t_l' increases monotonically but sub-linearly with t_N, reaching $\approx 22\tau_{adv}$ for the largest training set. As alluded to in Figure 1(c), while t_l increases with N, the scaling is sub-linear, e.g. t_l/t_N declines from 4.2% to 2.5% as t_N is increased by a factor of 8.1 (Figure 3(b)). The results indicate reaching a limit of predictability. Using the algorithm of Sano and Sawada [38], we have found the largest positive Lyapunov exponent of θ_{DNS} to be $\Lambda_{max} \approx 0.04/\tau_{adv}$. Therefore, with the longest training set, $t_l \approx 21.1/\Lambda_{max}$. The limit of predictability, in particular in the presence of uncertainties in the initial conditions, should be further investigated in future work.

The skillful predictions with A_{HDMD} presented above hinge on two critical components of the formulation in (2)-(4): the use of i) delay-embedding and ii) vector-valued observables. With $q = 1$, the formulation is reduced to the original DMD method [22, 24], which we found to have very little prediction skill, as θ_{pred} vanishes in a fraction of τ_{adv}. This can be attributed to the observation that the eigenvalues of A_{HDMD}, and in particular the slowest-decaying ones, are much closer to the unit circle compared to those of A_{DMD}; consequently, the DMD modes decay at much faster rates.

Using scalar-valued observables (i.e. $n = 1$) in Hankel-DMD does not lead to any useful prediction skill. For example, for the same training set, A_{HDMD} computed using θ from a single point at $z/H = -0.18$ can predict the time-series at that point only up to $t_l \approx 0.82\tau_{adv}$ (case C1 in Table I), while A_{HDMD} computed using all 129 grid points has remarkable prediction skills at $z/H = -0.18$ (Figure 1(a)) and can predict the time-series up to $t_l \approx 529\tau_{adv}$. Given the spatio-temporal correlations of the turbulent flow, it is not surprising that adding more spatial points enhances the performance of A_{HDMD}; however, it is of practical interest to explore how the prediction skill is affected if the data is subsampled. This is important for problems in which the data is only sparsely available (e.g. due to practical limitations) and/or to keep the size of $\mathcal{H}$ and A_{HDMD} computationally tractable (e.g. in large-scale 3D problems). Table I compares the prediction skills of A_{HDMD} computed using several different n and n_b (number of points within the thermal boundary layer). Testing several subsampling schemes showed that choosing points symmetrically (with respect to $z = 0$) always results in better prediction skills; thus in all cases in Table I except for C1 (discussed earlier), points are chosen symmetrically. As expected, increasing n improves the prediction horizon and accuracy. Increasing n_b, however, does not have a substantial impact on the prediction horizon/accuracy, except for the case with large n (C9), in which t_l' and E_{ave} (but not t_l) are improved over the case C8 (same n but $n_b = 0$). The weak effect of n_b is of practical significance because collecting measurements from thin boundary layers can be challenging. Furthermore, case C9 shows that using only half of the points, A_{HDMD} still has reasonable prediction horizon/accuracy, which encourages seeking an optimal subsampling strategy in future work (it is worth higlighting that optimal sensor placement is currently a major challenge in estimation/control of spato-temporal systems [39]).

It is also informative to examine the performance and prediction characteristics of the data-driven framework (2)-(4) for the K-S equation, which is a commonly-used prototype for chaotic systems. As described in Supplemental Material [33], the K-S equation is numerically integrated for a broad range of chaoticity and attractor dimension. Tests show skillful spatio-temporal predictions by A_{HDMD}, including for high-dimensional and strongly-chaotic cases. Similar to the behavior observed for the turbulent RBC: i) the Hankel-DMD prediction is accurate until the prediction horizon t_l (often around $30/\Lambda_{max}$), after which the prediction rapidly diverges from the true time-series and vanishes, and ii) t_l and E_{ave} improve as N increases, but the scaling is sub-linear and the improvements become increasingly slow at larger N.

In summary, we show the high prediction skills of the data-driven Hankel-DMD framework when used

TABLE I. Comparison between t_l', t_l, and E_{ave} for different subsampling schemes. n_b is the number of grid points inside the thermal boundary layer. Results are from the longest training set ($t_N \approx 21.9 \times 10^3\tau_{adv}$). See the Supplemental Material [33] for the extended version and further discussions.

Case	n	n_b	t_l'/τ_{adv}	t_l/τ_{adv}	$E_{ave}(\%)$
C1	1	0	0.82	0.82	17.4
C2	5	0	2.80	276	15.7
C3	5	2	2.72	277	16.4
C4	11	0	4.86	305	13.7
C5	11	2	4.88	306	13.7
C6	29	0	7.34	361	12.7
C7	29	4	7.35	361	12.2
C8	65	0	13.2	432	11.7
C9	65	30	21.4	432	10.4
Full grid	129	60	21.8	529	8.82

with vector-valued observables for a fully-turbulent flow (RBC) and highly-chaotic system (K-S). While here we focused on RBC, a prototype for geophysical turbulence, the framework is readily applicable to any turbulent flow. In future work, the prediction skills for data with measurement noise, sparsely-available/subsampled data, non-stationary data, and 3D observables (e.g. $(\boldsymbol{v}, T)$) should be further studied.

We highlight that while A_{HDMD} is skillful for short-term forecasts, it cannot accurately predict the long-term response of the system to an external forcing, e.g. heating $\overline{f}(z)$. Khodkar and Hassanzadeh [40] and Khodkar *et al.* [34] have recently shown, using the same RBC system, that two other techniques, one data-driven based on using the fluctuation-dissipation theorem [41, 42], and another based on the Green's function (GRF) method (which involves simulations with many weak, localized forcings applied one at a time [43, 44]) can compute the system's linear response function L that accurately predicts the long-term response as $L^{-1}f$, while not having skills for short-term prediction. Interestingly, the first few slowest-decaying eigenvectors of A_{HDMD} and L_{GRF} are strikingly similar [33], suggesting that a close examination of these two vastly different techniques together might provide a deeper insight into the dynamics that they capture and their prediction skills.

It should be also mentioned that the system identification method *matrix pencil* (which is called Loewner method when used in frequency-domain) in time-domain is mathematically equivalent to Hankel-DMD [45]. This approach has been recently used for data-driven model-reduction of nonlinear systems such as Burgers' equation [46–48]. The underlying connections between the matrix pencil and Hankel-DMD techniques warrant further investigation [49].

Finally, a data-driven method called *linear inverse modeling* (LIM) [50] has been long used for forecasting low-frequency climate phenomena such as El Niño and Madden-Julian Oscillation [12, 51, 52]. LIM is derived from the Fokker-Planck equation, but is mathematically equivalent to standard DMD [24, 40]. The findings of this paper suggest that augmenting LIM with delay-embedding and ensuring rich-enough vector-valued observables might significantly help such data-driven weather/climate forecasting efforts, especially for the highly turbulent midlatitude circulation.

We thank Hassan Arbabi, Matthias Heinkenschloss, and Carl Wunsch for fruitful discussions. This study was partially supported by NASA grant 80NSSC17K0266 and a Faculty Initiative Fund award from the Rice University Creative Ventures (P.H.) and NSF grant CCF-1816219 (A.C.A.). This work used the Stampede2 cluster through XSEDE allocation ATM170020 and the DAVinCI cluster of the Rice University Center for Research Computing.

---

[*] mkhodkar@rice.edu

[†] pedram@rice.edu

[1] E. N. Lorenz, J. Atoms. Sci. **20**, 130 (1963).

[2] G. E. P. Box, J. G. M., G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control* (John Wiley & Sons, 2015).

[3] G. Van Kuik *et al.*, Wind Energ. Sci. **1**, 1 (2016).

[4] T. Duriez, S. L. Brunton, and B. R. Noack, *Machine Learning Control-Taming Nonlinear Dynamics and Turbulence* (Springer, 2017).

[5] T. Palmer and R. Hagedorn, *Predictability of weather and climate* (Cambridge University Press, 2006).

[6] A. J. Majda, Commun. Pure Appl. Math. **65**, 920 (2012).

[7] P. Bauer, A. Thorpe, and G. Brunet, Nature **525**, 47 (2015).

[8] A. Majda and N. Chen, Entropy **20**, 644 (2018).

[9] M. Farazmand and T. P. Sapsis, Sci. Adv. **3** (2017).

[10] C. Wunsch, Bull. Am. Meteorol. Soc. **80**, 245 (1999).

[11] H. Van den Dool, *Empirical methods in short-term climate prediction* (Oxford University Press, 2007).

[12] N. R. Cavanaugh, T. Allen, A. Subramanian, B. Mapes, H. Seo, and A. J. Miller, Clim. Dyn. **44**, 897 (2015).

[13] S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, and J. N. Kutz, Nat. Commun. **8**, 19 (2017).

[14] D. Giannakis, Appl. Comput. Harmon. Anal. (2017).

[15] D. Comeau, Z. Zhao, D. Giannakis, and A. J. Majda, Clima. Dyn. **48**, 1855 (2017).

[16] P. R. Vlachas, W. Byeon, Z. Y. Wan, T. Sapsis, and P. Koumoutsakos, Proc. R. Soc. A **474** (2018).

[17] R. Yu, S. Zheng, A. Anandkumar, and Y. Yue, arXiv:1711.00073 (2017).

[18] U. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, Phys. Rev. Lett. **120** (2018).

[19] B. O. Koopman, Proc. Natl. Acad. Sci. **17** (1931).

[20] I. Mezić, Nonlin. Dyn. **41**, 309 (2005).

[21] I. Mezić, Annu. Rev. Fluid Mech. **45** (2013).

[22] P. J. Schmid, J. Fluid Mech. **656**, 5 (2010).

[23] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, J. Fluid Mech. **641**, 115 (2009).

[24] J. H. Tu, C. W. Rowley, L. D. M., S. L. Brunton, and J. N. Kutz, J. Comp. Dyn. **1**, 391 (2014).

[25] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, J. Nonlin. Sci. **25** (2015).

[26] H. Arbabi and I. Mezić, SIAM J. Appl. Dyn. Syst. **16**, 2096 (2017).

[27] H. Arbabi and I. Mezić, Phys. Rev. Fluids **2**, 124402 (2017).

[28] J. Page and R. R. Kerswell, Phys. Rev. Fluids **3**, 071901 (2018).

[29] M. Korda and I. Mezić, J. Nonlin. Sci. **28**, 687 (2018).

[30] C. W. Rowley and S. T. M. Dawson, Annu. Rev. Fluid Mech. **49**, 387 (2017).

[31] F. Takens, Lect. Notes Math. **898**, 366 (1981).

[32] H. Arbabi, M. Korda, and I. Mezić, arXiv:1804.05291 (2018).

[33] See the Supplemental Material at `http://pedram.web.rice.edu/KhodkarHDMD2018SM.pdf`.

[34] M. A. Khodkar, P. Hassanzadeh, N. Saleh, and P. Grover, Phys. Rev. Fluids (in press) (2018), available at arXiv:1805.01596.

[35] The Prandtl number, Rayleigh number, and advective

time-scale are defined, respectively, as $Pr = \nu/\kappa$, $Ra = g\alpha\Delta TH^3/\nu\kappa$, and $\tau_{adv} = \sqrt{H/(g\alpha\Delta T)}$, where ν is kinematic viscosity, κ is thermal diffusivity, g is gravitational acceleration, and α is thermal expansion coefficient.

[36] P. G. Drazin and W. H. Reid, *Hydrodynamic stability* (Cambridge University Press, 2004).

[37] J. Bailon-Cuba and J. Schumacher, Phys. Fluids **23**, 077101 (2011).

[38] M. Sano and Y. Sawada, Phys. Rev. Lett. **55**, 1082 (1985).

[39] K. Manohar, B. W. Brunton, J. N. Kutz, and S. L. Brunton, IEEE Control Syst. Mag. **38**, 63 (2018).

[40] M. A. Khodkar and P. Hassanzadeh, J. Fluid Mech. **852** (2018).

[41] R. Kubo, Rep. Prog. Phys. **29**, 255 (1966).

[42] P. Hassanzadeh and Z. Kuang, J. Atoms. Sci. **73**, 3441 (2016).

[43] P. Hassanzadeh and Z. Kuang, J. Atoms. Sci. **73**, 3423 (2016).

[44] Z. Kuang, J. Atoms. Sci. **67**, 941 (2010).

[45] A. C. Ionita and A. C. Antoulas, in *Developments in Control Theory: Towards Glocal Control*, Vol. 76, edited by L. Qiu, J. Chen, T. Iwasaki, and H. Fujioka (IET Control Engineering Series, 2012) pp. 79–88.

[46] A. C. Antoulas, I. V. Gosea, and A. C. Ionita, SIAM J. Sci. Comput. **38**, B889 (2016).

[47] A. C. Antoulas, I. V. Gosea, and M. Heinkenschloss, in *Active Flow and Combustion Control 2018* (Springer, 2019) pp. 255–270.

[48] S. Gugercin, C. A. Beattie, and A. C. Antoulas, *Data-driven and interpolatory model reduction* (Society for Industrial and Applied Mathematics, 2019).

[49] L. Pogorelyuk and C. W. Rowley, arXiv:1802.09878 (2018).

[50] C. Penland, Mon. Weather Rev. **117**, 2165 (1989).

[51] C. Penland and T. Magorian, J. Climate **6**, 1067 (1993).

[52] M. A. Alexander, L. Matrosova, C. Penland, J. D. Scott, and P. Chang, J. Climate **21**, 385 (2008).