STATEMENT OF RESEARCH

Dimitrios Giannakis

Center for Atmosphere Ocean Science, Department of Mathematics
Courant Institute of Mathematical Sciences
New York University

October 5, 2017

Since joining the faculty of the Courant Institute in 2012, my research has focused on data-driven techniques for pattern extraction and prediction in dynamical systems combining ideas from machine learning and operator-theoretic aspects of ergodic theory [1–13]. This approach allows for the construction of nonparametric (“equation-free”) statistical prediction models that rely on time-ordered measurements of the system alone, yet provably converge to the prediction skill of the full first-principles model as the timespan of the training data increases. The framework also enables the decomposition of complex signals into temporal and spatial patterns that are intrinsic to the dynamical system generating the data through the solutions of an eigenvalue problem for the generator of the dynamics. I have worked on applications of these techniques to a number of research topics in atmosphere ocean science (AOS) and fluid dynamics, including studies on low-frequency oceanic variability [14–16], Arctic sea ice [17–22], Antarctic circumpolar waves [23], tropical convective organization [24–31], and Rayleigh-Bénard convection [32, 33]. Collaboration with both junior (students and postdocs) and senior colleagues has played an essential role in these activities.

1 Data-driven techniques for dynamical systems

A central theme in my research has been the use of kernel algorithms from machine learning to perform data-driven approximation of operators governing the evolution of observables and probability measures in dynamical systems. This framework has in its foundation the operator-theoretic description of dynamical systems [34] (dating back to the seminal work of Koopman in the 1930s [35]), whereby a nonlinear dynamical system is represented by a group or semigroup of linear operators acting on a, generally infinite-dimensional, space of observables (functions of the state). It is a remarkable fact that, for an appropriate choice of the space of observables, the associated groups or semigroups of operators (collectively referred to here as Koopman operators) encodes all information of the original system despite it being linear. In particular, if this space has Banach or Hilbert space structure, then one has access to the full machinery of operator theory in such spaces to carry out tasks such as pattern extraction and prediction.

Historically, data-driven techniques taking advantage of this framework were introduced in the late 1990s and early 2000s [36, 37], and have since become recognized as a powerful tool for dynamical systems modeling. Around the same time, another major branch of data analysis algorithms was introduced and saw wide adoption; namely, kernel algorithms for statistical and manifold learning [38–40]. From the point of view of dynamical systems modeling, these techniques have arguably complementary strengths and weaknesses. On the one hand, Koopman operator techniques target the intrinsic operators governing the evolution of observables, but generally do not provide a framework to control the regularity of the solutions. In contrast, kernel methods have the ability to approximate function spaces with a well-defined notion of regularity, yet besides special cases (e.g., stochastic gradient flows [41]), they are not as well suited for approximating systems with non-compact evolution operators such as the systems encountered in AOS. My work in this area has aimed to address these issues through the development of data-driven techniques for pattern extraction and prediction in dynamical systems that take advantage of the strengths of both Koopman operator and kernel methods. That is, the general approach in these methods is to use kernel algorithms to construct bases for the natural function spaces on which Koopman operators act (rather than using these methods to perform dimension reduction or to approximate the dynamics), and employ these bases to construct data-driven approximations of Koopman operators with regularity and convergence guarantees. As an effort to disseminate this research, I have developed and made publicly available (http://cims.nyu.edu/~dimitris) an object-oriented Matlab code implementing a number of the techniques described below, which I plan to continuously update as this work progresses. Most of the numerical results described here were obtained using that code.

Representing bounded evolution operators. As a concrete example, consider a smooth flow $\Phi : M \mapsto M$, $t \in \mathbb{R}$, on a closed manifold $M$, possessing a smooth, ergodic invariant measure $\mu$. Associated with this system is a group of unitary Koopman operators $U^t : H \mapsto H$, $t \in \mathbb{R}$, acting on the Hilbert space $H = L^2(M, \mu)$ associated with the invariant
forecasting that approximates from data. In [4], in collaboration with Tyrus Berry and John Harlim, we developed a technique called diffusion of a time-ordered sequence of measurements of the system, but otherwise no knowledge of the dynamics \( \Phi \) states initialization represented by \( \nu \). In other words, \( g_t(\nu) \) is a statistical forecast of \( f \) that is natural to approximate from data. In [4], in collaboration with Tyrus Berry and John Harlim, we developed a technique called diffusion forecasting that approximates \( g_t(\nu) \) (as well as other quantities of interest in statistical prediction) assuming availability of a time-ordered sequence of measurements of the system, but otherwise no knowledge of the dynamics \( \Phi \) or the structure of the state space manifold \( M \). In particular, we consider that available to us is a sequence \( y_0, y_1, \ldots, y_{N-1} \) of data samples \( y_n = f(x_n) \in \mathbb{R}^d \), taken through a smooth observation map \( F: M \to \mathbb{R}^d \) for the underlying dynamical states \( x_n = \Phi^{\tau t}(x_0) \), where \( x_0 \in M \) is an arbitrary initial condition and \( \tau > 0 \) a fixed sampling interval. Diffusion forecasting creates a data-driven approximation \( \hat{g}_t(\nu) \) of \( g_t(\nu) \), that provably converges in the limit \( N \to \infty \) of large data (in fact, even if the system is stochastic). The method employs the following key ingredients:

- **Representation of bounded operators.** Because \( U \) is bounded (and thus continuous) it can be represented through its elements \( A_{jk}(t) = \langle \phi_j, U^t \phi_j \rangle_H \) in any orthonormal basis \( \{ \phi_j \}_{j=0}^\infty \) of \( H \). In particular, we have \( g(\nu) = \sum_{k=0}^\infty c_k \Phi_{jk}(t) \alpha_k \), where \( \alpha_k = \langle \phi_j, \rho \rangle_H \) and \( c_k = \langle \phi_j, f \rangle_H \) are expansion coefficients of \( \rho \) and \( f \) in the \( \phi_j \) basis. Note that this representation of \( U_t \) holds even if this operator has continuous spectrum (as is necessarily the case in chaotic systems), which has been a challenge in many data-driven prediction techniques.

- **Ergodicity.** This allows us to approximate integrals with respect to the invariant measure by time averages along trajectories. Importantly, along a trajectory \( x_0, x_1, \ldots \), the action of the Koopman operator \( U_t \), with \( t = q \tau \) an integer multiple of the sampling interval \( \tau \), is simply given by a time shift, \( U^t f(x_n) = f(x_{n+q}) \). Thus, we can approximate \( A_{jk}(t) \) by \( A_{N,j}(t) = \sum_{n=0}^{N-1} \phi_j(x_n) \hat{f}(x_n+q)/N \), and as \( N \to \infty \), \( A_{N,j}(t) \) converges to \( A_{jk}(t) \) almost surely by the Birkhoff pointwise ergodic theorem.

- **Data-driven orthonormal basis.** We construct a data-driven orthonormal basis \( \{ \hat{\phi}_j \}_{j=0}^{\infty} \) of an \( N \)-dimensional Hilbert space \( H_N \) associated with the sampling measure on the trajectory \( x_0, \ldots, x_{N-1} \) by applying the diffusion maps algorithm [40] to the observed data \( y_0, \ldots, y_{N-1} \) with a suitably chosen kernel [42], possibly also performing delay-coordinate maps [43] if the observations are incomplete (i.e., \( F \) is not an embedding). For this choice of kernel, the basis functions \( \hat{\phi}_j \) converge (in the sense established in [44]) to the eigenfunctions of a heat operator \( \mathcal{P} = e^{t\Delta}, \, \varepsilon \geq 0, \) on \( H \), generated by a self-adjoint Laplace-Beltrami operator \( \Delta \). Further, these basis functions have optimal regularity in the sense of extremizing a Dirichlet energy functional \( \mathcal{E}(f) = \int_M \| \text{grad} f \|^2 \, d\mu \) measuring the roughness of functions on \( M \). It can be shown that the empirical matrix elements \( \hat{A}_{jk}(t) = \langle \phi_j, U^t \phi_j \rangle_H_N = \sum_{n=0}^{N-1} \hat{\phi}_j(x_n) \hat{\phi}_j(x_{n+q})/N \) of the Koopman operator at \( t = q \tau \) converge as \( N \to \infty \) to the true values \( A_{jk}(t) \). As a result, given a spectral truncation parameter \( l \leq N - 1 \), the data-driven forecast

\[
\hat{g}_t(\nu) = \sum_{j,k=0}^l \hat{\alpha}_j \hat{A}_{jk}(t) \hat{c}_k, \quad \alpha_j = \langle \rho, \hat{\phi}_j \rangle_{H_N}, \quad \hat{c}_k = \langle f, \hat{\phi}_j \rangle_{H_N},
\]

converges to \( g_t(\nu) \) as \( N,l \to \infty \) almost surely. In practice, with work with \( l \ll N - 1 \), which enforces regularity in the predictand \( \hat{g}_t(\nu) \).

**Conditional analog forecasting.** One of the advantages of the diffusion forecasting technique outlined above is its flexibility with respect to the choice of the initial probability measure \( \nu \). That is, the user can supply any \( \nu \), and so long as it has sufficient regularity (at least \( L^2 \) if the prediction observable \( f \) is \( L^2 \)) the method will converge. However, the method does not address the question of how to assign \( \nu \) given a measurement of the system \( y = F(x) \) at forecast initialization associated with an unknown dynamical state \( x \in M \). Moreover, the analysis in [4] was carried out under the assumption that \( F \) is injective (one-to-one), and even though that assumption can be generically relaxed using delay-coordinate maps, the ability to make unbiased forecasts with non-injective observation maps would be desirable.
observables as a directional derivative along the dynamical flow; specifically, the latter is a skew-adjoint, unbounded operator acting on the Hilbert space $\mathcal{H}$ from a dynamical trajectory $x\in\mathbb{R}^3$, obtained via (1) for an initial density $\rho$ obtained by pulling back a Gaussian density in $\mathbb{R}^3$ to the L63 attractor. Dotted lines show forecasts for the variance, computed by applying diffusion forecasting to predict the observable $\sigma^2 = \|U^t F - \bar{\bar{F}}\|^2_{L^2}$, where $\bar{\bar{F}} = \int_M F\,d\mu$ is the mean of $F$ and $\|\cdot\|_{L^2}$ is the canonical Euclidean norm in $\mathbb{R}^3$. The fact that the predicted $\sigma$ tracks fairly well the RMSE indicates that in addition to predicting the mean, diffusion forecasting performs useful uncertainty quantification (UQ). These skill metrics for the L63 model are compared against forecasts made with regression based techniques, as well as ensemble predictions made with the full L63 model. Notice that diffusion forecasting has comparable skill to the full model over short to intermediate lead times ($t \lesssim 2.5$), and converges to the correct equilibrium statistics at late times. Diffusion forecasting also performs more useful UQ than the regression based techniques, which either overestimate or underestimate $\sigma$. In the Niño 3.4 prediction results, the pattern correlation score remains above 0.5 for approximately five months, but increases again to larger than 0.5 values at 12–15 month leads. The non-monotonic time dependence of skill may be due to a seasonal predictability barrier that ENSO is known to exhibit [45].

In work with former postdoc Zhizhen (Jane) Zhao (currently tenure-track assistant professor at the University of Illinois Urbana-Champaign) [6], we developed a data-driven prediction method called kernel analog forecasting that addresses these objectives. The method derives its name from the analog prediction technique introduced by Lorenz in 1969 [46] (which provided the original inspiration of our work), as it can be thought of as a Bayesian probabilistic extension of that technique. This interpretation was established in recent work with PhD student Romeo Alexander [11]. In particular, it is a well known fact from statistical learning theory that the assignment of a probability measure $\nu(y)$ to initial data $y \in Y = F(M)$ that minimizes the expected forecast error over all initial data $y \in Y$ is given by a conditional probability measure associated with the observation map $F$. Specifically, defining the predictand $h_t(y) = g_t(\nu(y))$, this optimal assignment $y \mapsto \nu(y)$ satisfies $h_t(F(x)) = \mathbb{E}[U^t f \mid \sigma_F](x)$, where $\mathbb{E}[U^t f \mid \sigma_F] \in \mathcal{H}$ is the conditional expectation of $U^t f$ with respect to a sub-$\sigma$-algebra $\sigma_F$ induced by $F$. In kernel analog forecasting, $\mathbb{E}[U^t f \mid \sigma_F]$ is approximated by a function $\tilde{h}_t$ constructed from time-ordered pairs of training samples $(F(x_n), f(x_n))$ from a dynamical trajectory $x_0, x_1, \ldots, x_N-1$ on $M$ using a variety of kernel algorithms for supervised learning [47, 48]. For example, one of the techniques we have explored is the Nyström method, which under mild assumptions on the kernel, produces an approximation $\tilde{h}_t$ lying in a reproducing kernel Hilbert space that can be evaluated for arbitrary (previously unseen) initial data, and converges to the true conditional expectation as the number $N$ of training samples tends to infinity. Specific applications of kernel analog forecasting to AOS are illustrated in Figures 6 and 10, and further discussed in Section 2.

Representing unbounded generators. The methods discussed thus far approximate, at some level, the action of an evolution operator $U^t$ on observables, but do not take advantage of any continuity properties (with respect to time $t$) that the group or semigroup $\{U^t\}$ of such operators might possess. For instance, consider a smooth flow $\Phi^t : M \rightarrow M$ on a manifold $M$ possessing an ergodic invariant measure $\mu$ supported on compact invariant set $X \subseteq M$ (not necessarily smooth). Then, it is well known that the associated unitary group of Koopman operators $U^t$ acting on the Hilbert space $H = L^2(X, \mu)$ is strongly continuous, and thus can be completely characterized by its generator. The latter is a skew-adjoint, unbounded operator $V : D(V) \rightarrow H$ with a dense domain $D(V) \subset H$ that acts on observables as a directional derivative along the dynamical flow: specifically, $V f = \lim_{\epsilon \to 0}(U^{t\epsilon} f - f)/\epsilon$ whenever the limit exits. One of the useful properties of $V$ for data analysis purposes is that it leads, through its eigenfunctions, to a
distinguished class of observables, called Koopman eigenfunctions, that evolve under the dynamics by multiplication by a time-dependent phase factor. Specifically, for a Koopman eigenfunction \( V \in D(\mathcal{V}) \) we have \( \mathcal{V}^t z_j = \lambda_j z_j \), where the eigenvalue \( \lambda_j = i \omega_j \) is purely imaginary and corresponds to an intrinsic frequency \( \omega_j \in \mathbb{R} \) of the dynamics, and \( z_j \) evolves according to \( U^t z_j = e^{i \omega_j t} z_j \). That is, \( z_j \) has a highly coherent and predictable temporal evolution despite the fact that the underlying dynamical system may exhibit complex, aperiodic behavior. More generally, the subspace \( H_p = \text{span}\{ z_j \} \subseteq H \) associated with the point spectrum of \( \mathcal{V} \) consists of observables with integrable temporal behavior, which warrant identification from data. These and other attractive properties have led to the development of a broad range of data-driven techniques for approximating Koopman eigenvalues and eigenfunctions [49], where, broadly speaking, the \( z_j \) and the projections of the data onto them play the roles of the temporal and spatial modes, respectively, in covariance-based techniques such as principal components analysis (PCA) and proper orthogonal decomposition (POD) [50]. Yet, the development of data-driven approximation techniques for Koopman eigenvalue problems with rigorous convergence guarantees in the fully general case with mixed (point and continuous) spectrum of \( \mathcal{V} \) has thus far been lacking. Here, the challenges encountered include the facts that:

- \( \mathcal{V} \) is unbounded; thus any representation of this operator must invariably confront with domain issues.
- \( \mathcal{V} \) will generally have non-isolated eigenvalues and/or continuous spectrum (in fact, the only classes of continuous-time deterministic ergodic systems where neither of these cases occurs are periodic or limit-cycling systems).
- The invariant set \( \mathcal{X} \) may be a non-smooth set of zero Lebesgue measure in the ambient space \( \mathcal{M} \) (e.g., in dissipative systems), and in realistic scenarios, measurements of the system will not be taken exactly on \( \mathcal{X} \).

In a series of papers co-authored with Sudhausatwa (Shuddho) Das (currently ONR-funded Courant postdoc), Joanna Slawinska, and Jane Zhao [5, 7, 8] we have developed a data-driven framework for Koopman eigenvalue and eigenfunction approximation that converges despite the issues listed above. The main findings that have enabled the development of this scheme can be summarized as follows:

1. A family of compact Markov operators \( P_\mathcal{Q} : H \rightarrow H \) constructed by applying diffusion maps using a kernel operating on delay-coordinate mapped data (here, \( \mathcal{Q} \) is the number of delays) was shown to commute, in the limit \( \mathcal{Q} \rightarrow \infty \), with the Koopman operators \( U^t \) for all \( t \in \mathbb{R} \). This means that \( \mathcal{V} \) and \( P_\mathcal{Q} \) have common finite-dimensional eigenspaces (which are all subspaces of \( H_p \)), and on these eigenspaces \( \mathcal{V} \) is diagonalizable.

2. It was shown that associated with \( P_\mathcal{Q} \) (in particular, with the eigenvalues and eigenfunctions of this operator) is a Sobolev space \( H_p^2 \subset H_p \) of sufficient regularity such that \( \mathcal{V} : H_p^2 \rightarrow H_p \) is bounded. Moreover, we constructed an advection-diffusion operator \( L = \mathcal{V} - \Delta \), where \( \theta > 0 \) and \( \Delta : H_p^2 \rightarrow H_p \) is a diffusion operator associated with \( P_\mathcal{Q} \), such that (i) \( L \) has the same eigenspaces as \( \mathcal{V} \), and (ii) it has only isolated eigenvalues converging to those of \( \mathcal{V} \) as \( \theta \rightarrow 0 \). A Petrov-Galerkin method to compute eigenfunctions and eigenvalues of \( L \), satisfying appropriate inf-sup conditions for well-posedness, was also developed.

3. Under the assumption that the invariant measure \( \mu \) is physical on a compact, forward-invariant set \( \mathcal{W} \) of positive Lebesgue measure such that \( \mathcal{X} \subseteq \mathcal{W} \subseteq \mathcal{M} \), it was shown that the diffusion maps algorithm provides a data-driven approximation \( P_{\mathcal{Q},\mathcal{N}} : H_N \rightarrow H_N \) of \( P_\mathcal{Q} \) that converges in spectrum, as the number of samples \( \mathcal{N} \) tends to infinity, to \( P_\mathcal{Q} \) (see the description of diffusion forecasting for a definition of \( H_N \)). As a result, the eigenfunctions \( \hat{\phi}_j \) of \( P_{\mathcal{Q},\mathcal{N}} \) can be used as basis functions for a data-driven Petrov-Galerkin method for \( L \), and the numerical eigenvalues and eigenfunctions converge for any initial condition starting in \( \mathcal{W} \) in an appropriate limit of large data and infinitely many delays.

It should be noted that the regularity of observables appropriate for the unbounded operators \( \mathcal{V} \) and \( L \) plays a key role in these techniques. In particular, we build a data-driven basis of \( H_p^2 \) by scaling the eigenfunctions \( \hat{\phi}_j, \ j \geq 1 \), by an appropriate factor \( 1/\eta_j \) (where \( \eta_j \) is an increasing sequence with \( \eta_0 = 0 \), determined from the eigenvalues \( \hat{\Lambda}_j \) corresponding to \( \hat{\phi}_j \)). This enforces \( H^2 \) regularity in expansions of observables of the form \( \sum_j c_j \hat{\phi}_j / \eta_j \). In addition, we order the data-driven Koopman eigenfunctions \( \hat{z}_j = \sum_{j=1}^{N-1} c_j \hat{\phi}_j \) from the Galerkin scheme in order of increasing Dirichlet energy (roughness), \( \delta'(\hat{z}_j) = \sum_{j=1}^{N-1} |\eta_j| |c_j|^2 \). Thus, we favor solutions with high regularity that can be robustly approximated from data.

An application of these techniques to a mixed-spectrum system with a non-smooth attractor is shown in Figure 2.

It is worthwhile noting that result 1 above provides an interpretation of why kernel algorithms operating on data in
delay-coordinate space can recover temporally coherent modes from complex signals containing multiple timescales; see, e.g., the family of nonlinear Laplacian spectral analysis (NLSA) algorithms co-developed with Andrew Majda [1, 2, 51] and independent work by Tyrus Berry and collaborators [52]. In particular, the fact that at large $Q P_Q$ approximately commutes with $U^t$ means that each eigenfunction of $P_Q$ corresponding to a nonzero eigenvalue should be a coherent, narrowband signal with power spectral density concentrated near a particular Koopman eigenfrequency. We will discuss a number of examples consistent with this expected behavior in Section 2.

**Modeling skew-product systems.** The techniques discussed above all assume that the system is ergodic (or they are valid for a single ergodic component of a non-ergodic system). In recent work with Shuddho Das [9], we developed an extension of our data-driven modeling framework to a class of potentially non-ergodic systems with a skew-product structure. The primary class of physical systems motivating this work is passive tracer advection in time-dependent, incompressible fluid flows [53]. From the point of view of a passive (Lagrangian) tracer, the dynamics governing its motion in a domain, $X$, is non-autonomous as the velocity field, $\vec{v}$, is time-dependent. However, if the evolution of $\vec{v}$ is itself governed by autonomous dynamics (“the Navier-Stokes equations”) on a state space $A$, the dynamical system on the product space $M = A \times X$ is autonomous and has a skew-product structure since the state in $A$ affects the dynamics in $X$ but not vice versa. Intuitively, one can think of $M$ as a “space-time” manifold with $X$ playing the role of physical space and $A$ the role of time as it controls the non-autonomous aspect of the dynamics on $X$. In particular, given any initial state $\vec{a} \in A$, a scalar-valued function $f$ on $M$ can be associated with a spatiotemporal pattern $\bar{f}(t,x) = f(\Phi^t(\vec{a}),x)$ where $\Phi^t : A \to A$ is the evolution map for the driving system, $t \in \mathbb{R}$ is time, and $x$ is an arbitrary spatial point in $X$. In [8], we propose a scheme for identifying coherent spatiotemporal patterns formed by Lagrangian tracers through a distinguished set of observables on $M$ given by eigenfunctions of a Koopman generator, $V$, for the skew-product system. Here, $V$ can be expressed as the sum $u + \vec{v}|_a \cdot \vec{\nabla}$ (after an appropriate operator closure), where $u$ is the generator of the dynamics on $A$, $\vec{v}|_a$ is the state-dependent velocity field on $X$ with $a \in A$, and $\vec{\nabla}$ is the gradient operator on $X$. In particular, note that eigenfunctions of $V$ at small corresponding eigenvalue $i\omega$ correspond to approximately conserved observables on the tracers (which intuitively corresponds to a notion of a coherent pattern), but the scheme also allows for the recovery of periodic patterns if $\omega$ is significantly different from zero. In a data-driven context, our method approximates $V$ from a time-ordered sequence of velocity field snapshots $v_0, v_1, \ldots, v_{N-1}$ under the assumption that the dynamics on $A$ is ergodic (this is needed to obtain a dense sampling of $A$ as $N \to \infty$). Note that we do not require the product dynamics on $M$ to be ergodic, and moreover we do not require availability of explicit tracer trajectories. Given this input data, we compute approximate Koopman eigenfunctions through a Galerkin method for a regularized generator $L = V - \theta A$ with only point spectrum in a data-driven basis acquired from diffusion maps via an extension of the techniques described above. As above, we identify coherent patterns through the solutions of this problem with minimal Dirichlet energy; that is, we recover patterns that are coherent with respect to their temporal evolution along tracer trajectories but also have high regularity as functions on $M$. Coherent patterns obtained via this method applied
Figure 3. Illustration of the Koopman operator framework for skew-product systems in [9] in the case of a time-periodic flow featuring a shifting vortex with a Gaussian streamfunction in a two-dimensional periodic domain. Left-hand panels: Snapshots of coherent spatiotemporal patterns associated with regularized Koopman eigenfunctions with small Dirichlet energy. Center panels: A comparison of the time evolution of an initially Gaussian probability density under the time-dependent flow as simulated by our nonparametric model with an ensemble of Lagrangian tracers evolved with the full model. Right-hand panels: Evolution of tracer positions with our nonparametric model compared to an ensemble simulation with the full model. Please see http://cims.nyu.edu/~dimitris for visualizations of the results in [9] in videos.

to a time-periodic flow are shown in Figure 3. Additional examples with aperiodic flows can be found in [8].

In addition to pattern identification, we use our data-driven representation of the generator to perform nonparametric prediction of observables and probability measures under the skew-product system. In this setting, a specification of a probability density on $M$ represents our knowledge about both the position of tracers in $X$ and the state of the fluid flow in $A$. Modeling the evolution of such a density would be of relevance, for instance, when making forecasts of the concentration of a pollutant in a fluid. Unlike the methods described previously, here we cannot approximate the Koopman operators $U^t$ generated by $V$ directly as that would require access to explicit tracer trajectories. However, what we can do is take advantage of the relationship $U^t = e^{tV}$ and compute $U^t$ via exponentiation of the generator. In [9], we relax this problem to an approximation of the evolution semigroup $\{S^t\}_{t \geq 0}$ generated by $L$; by construction of $L$, this is a contraction semigroup and $S^t = e^{tL}$ can be computed stably. In our numerical scheme, we compute the action $S^t f$ on an observable $f$ through the use of Leja polynomial interpolation techniques for matrix exponentials [54]. These schemes do not require the formation of $e^{tL}$ itself (and thus can take advantage of any sparsity the generator matrix might possess), and also allow for large forecast stepsizes $t$. Overall, this approach can be viewed as a data-driven exponential integration scheme for the joint fluid-tracer system. Prediction results with this method for a time-periodic flow are displayed in Figure 3.

Ongoing and future work. Below is a list of topics on data-driven dynamical systems modeling that I am either currently working on, or plan to do so in the near future:

- **Pattern extraction in spaces of vector-valued observables.** In recent work with Abbas Ourmazd, Joanna Slawinska, and Jane Zhao [12], we have developed a technique inspired by the theory of operator-valued kernels for multitask learning [55] for decomposition of signals of vector-valued observables of dynamical systems. The primary example of interest in this work is spatiotemporal signals that can be modeled as functions of the state of a dynamical system taking values in a vector space of functions (spatial patterns) over a domain. An advantage of this approach over conventional data analysis techniques operating on spaces of scalar-valued observables is that the recovered patterns can have manifestly non-separable structure in the spatial and temporal degrees of freedom, enabling efficient representation of signals with intermittency in both space and time. Preliminary applications of this technique on systems exhibiting complex spatiotemporal behavior (e.g., the Kuramoto-Sivashinsky model) have demonstrated significant improvement over conventional techniques.

- **Representing vector fields and higher-order tensors.** In work with Tyrus Berry [10] we have been developing a framework for approximating vector fields, differential forms, and higher-order tensors of manifolds which is
based on spectral representations of these objects, viewed as operators on functions spaces. One of the goals of this work is to eventually develop data-driven spectral decomposition techniques for linearizations of dynamical systems, which, among other applications, can be used to study the growth of instabilities in fluid flows.

- **Representing unbounded operators with continuous spectra.** The techniques described above involving the unbounded generator $V$ of a measure-preserving dynamical system employ (for the purposes of either eigendecomposition or prediction) a regularization of this operator leading to the advection-diffusion operator $L = V - \theta\Delta$. Thus far, our results on this operator (and in particular its behavior in the limit $\theta \to 0$) are mainly related to the point spectrum subspace $H_p$ [8]. A natural, yet challenging, extension of this work would be a characterization of the behavior of $L$ (or the development of alternative methods for representing $V$) in the subspace $H^\perp$ associated with the continuous spectrum of $V$. Progress in this topic could lead, for instance, to the development of forecasting techniques with higher computational efficiency than the original diffusion forecasting method when dealing with systems possessing mixed or purely continuous spectrum.

2 Applications to atmosphere ocean science and fluid dynamics

My work in AOS has focused on four main topics, namely (1) oceanic variability in the Indo-Pacific on seasonal to decadal timescales, (2) sea ice variability in the Arctic, (3) Antarctic circumpolar waves, and (4) convective organization in the tropics. A major component of these studies has been to use the mode decomposition techniques described in Section 1 to obtain reduced representations of the phenomena of interest with high physical interpretability, and use these representations to perform physics-based mechanistic descriptions and statistical forecasts. In particular, due to their close connection with the spectral properties of the dynamical system generating the data, the techniques of Section 1 are well suited to capture the modulating relationships across multiple timescales that many modes of climate variability exhibit, requiring no ad hoc preprocessing of the input data. Besides AOS, I have worked on applications of NLSA and Koopman operator analysis to study the heat transfer and large-scale circulation in direct numerical simulations of Rayleigh-Bénard convection [32, 33].

**Representing the El Niño Southern Oscillation.** The El Niño Southern Oscillation (ENSO) is the dominant mode of climate variability on interannual ($\sim 5$ yr) timescales, manifested as an oscillation between periods of anomalously high and low sea surface temperature (SST) anomalies in the eastern equatorial Pacific called El Niño and La Niña events, respectively [56]. A key feature of ENSO is a phase locking with the seasonal cycle characterized by a preferential occurrence of El Niño and La Niña events (the peaks of the ENSO cycle) in boreal winter [57]. Recently, aspects of the physics underlying this phase locking have been explained on the basis of a family of modulated patterns evolving at the sum and difference between the ENSO frequency, $v_{\text{ENSO}} = 0.25 \text{ yr}^{-1}$, and the annual cycle frequency, $v_A = 1 \text{ yr}^{-1}$, called ENSO combination modes [58–60]. It was also hypothesized that these patterns are members of a family of frequencies $v_{mn} = m v_{\text{ENSO}} + n v_A$ generated from the basic frequencies $v_{\text{ENSO}}$ and $v_A$ for integer $m$ and $n$.

In the spectral theory of dynamical systems, such a frequency family is precisely what would be expected from the eigenvalues of the generator if the attractor has a quasiperiodic component—in that scenario, the Koopman operator and NLSA techniques described in Section 1 would be expected to have high skill in recovering the ENSO-annual cycle frequency family from data. Indeed, in recent work with Joanna Slawinska [15, 16] we successfully recovered several modes in this family from Indo-Pacific SST data in models and observations without performing spectral pre-filtering on the input data. In addition, while classical data-driven approaches based on empirical orthogonal function (EOF) analysis (equivalent to PCA) mix these frequencies in individual modes, the NLSA-derived ENSO combination modes were found to be in very good agreement with the theoretically expected combination frequencies and mode multiplicities. Despite being recovered from SST data alone, the spatiotemporal patterns of SST, surface atmospheric circulation, and thermocline depth anomalies reconstructed from the NLSA modes capture the essential features of the ENSO lifecycle, including a southward shift of equatorial zonal winds in the western Pacific occurring in late boreal winter, which is thought to be responsible for restoring ENSO-neutral conditions from El Niño/La Niña conditions [61]. Representative eigenfunction time series and SST composites from the NLSA-derived ENSO combination mode family are shown in Figure 4. Besides ENSO, another family of interannual modes and their associated combination modes identified in [15, 16] represents the tropospheric biennial oscillation (TBO) [62] (together with the corresponding precipitation and surface wind patterns) associated with the Asian-Australian monsoon. Looking forward, the work in [15, 16] has laid the foundation for a number of studies which we are currently pursuing or planning to pursue in the future, including statistical forecast models for ENSO and its regional impacts on the rest of the climate system...
families

... respectively. Notice the strong activity exhibited by deterministic nature of the relationship between with the southward shift of \( z \) calculation patterns due to these modes are consistent in Part II we will establish that the surface atmospheric cir-

... EOF analysis of surface atmospheric circulation fields. The dominant frequencies of ENSO-A1 and ENSO-

... signals) were represented as a single pair of in-

... \( \nu \), where \( \nu = 1 \text{ yr}^{-1} \) is the annual cycle frequency. Notice also the strong SST anomalies in (f) in the Indian Ocean region west of the Sunda Strait, which is employed in the definition of indices for the Indian Ocean dipole (IOD) [63] (the dominant mode of interannual SST variability in the Indian Ocean). In [15, 16], we established that the ENSO combination mode family from NLSA explains \( \sim 50\% \) of the SST variance in that region. This implies that a significant portion of IOD variability can be attributed to ENSO, and that the ENSO combination modes are useful IOD predictors.

(15 JULY 2017) SLAWINSKA AND GIANNAKIS 5275

Arctic sea-ice: Re-emergence mechanisms and data-driven prediction. Another aspect of climate variability where the data-driven techniques of Section 1 have revealed a spectral behavior analogous to that of ENSO combination modes is Arctic sea ice reemergence—a phenomenon whereby sea ice concentration (SIC) anomalies during a given ice growth (melt) season are positively correlated with anomalies in the subsequent melt (growth) season, despite decay of correlations in the intervening months [64]. In work with Mitch Bushuk (former AOS PhD student and currently permanent research scientist at GFDL) and Andrew Majda [17–20], we found that this process can be represented in a physically consistent manner through NLSA mode families, each consisting of a primary low-frequency (interannual) mode together with the associated combination modes with the annual cycle as described above for ENSO. This decomposition reveals aspects of sea ice reemergence beyond those accessible via global correlation analyses; in particular:

- Sea-ice reemergence can exhibit distinct patterns in space generated by an interannual mode of variability and its coupling with the annual cycle. For example, the manifestation of the leading (with respect to NLSA eigenvalues) such pattern of North Pacific SIC was found to correlate strongly with the North Pacific gyre oscillation (NPGO) pattern [65], and the second low-frequency pattern was found to correlate strongly with the Pacific decadal oscillation (PDO) [66].

- Within each reemergence family, the synchronization with the seasonal cycle (which is an outcome of the physics of the coupled ice-ocean-atmosphere system in response to the annually varying solar insolation) is characterized by modulated patterns (combination modes), with the annual cycle and its harmonics acting as the carrier signal and the low-frequency mode as the modulating envelope. In particular, the amplitude of these modes is slaved to their parent low-frequency mode, meaning that the strength of reemergence due to the family as a whole also
Figure 4. Spatial pattern composites of SIC (%) and SIT (m), computed using the NLSA reemergence family of the control run. These composites are computed over all times in which the leading low-frequency SIC mode is active in positive phase (LSIC\textsubscript{1,2}).

Figure 5. Spatial pattern composites of SST (K) and SLP (Pa), computed using the NLSA reemergence family of the control run. SLP contours are plotted in black. These composites are computed over all times in which the leading low-frequency SIC mode is active in positive phase (LSIC\textsubscript{1,2}).

Figure 5. Monthly composites of sea ice concentration (SIC, %), sea ice thickness (SIT, meters), sea surface temperature (SST, Kelvin), and sea level pressure (SLP, Pascal) computed using an NLSA mode family recovered from a control integration of CCSM4 capturing both growth-to-melt and melt-to-growth reemergence. In melt-to-growth reemergence, SIC anomalies at marginal sea ice zones in September (the sea ice minimum) are collocated with SIT anomalies of the same sign. As the sea ice edge migrates southwards in the ensuing growth season (September–March), SIC anomalies also move southwards and vanish from those regions (since SIC reaches its climatological, 100% value, for those months), but the SIT anomalies persist. As a result, when sea ice begins retreating in March (the sea ice maximum), the SIC anomalies begin to retreat northward and eventually encounter the region of persistent SIT anomalies. If it happens that these SIT anomalies are positive (negative), ice will tend to melt out slower (faster), leading to SIC anomalies of the same sign as the original anomalies in the previous growth season. An analogous process takes place in melt-to-growth reemergence, but with SST playing the role of memory in place of SIT. In this case, SIC anomalies in the beginning of the melt season (March) are collocated with SST anomalies of opposite sign. The SST anomalies persist during the summer months, and in the following growth season ice forms at higher (lower) quantities if the SST anomalies are negative (positive), leading again to reemergence of SIC anomalies. During both of these reemergence processes the low-frequency configuration of surface atmospheric circulation is characterized by an SLP anomaly dipole between North America and Eurasia, whose associated geostrophic winds are consistent with opposite-sign anomalies SIC between the Pacific and Atlantic basins.

exhibits interannual variability. This property may help identify initial conditions with high predictability; an important current problem in sea ice prediction research [67].

- A single reemergence family can exhibit both growth-to-melt and melt-to-growth sea ice reemergence, with SST and sea ice thickness providing the source of memory, respectively. An illustration of these reemergence processes associated with the leading low-frequency NLSA family recovered from a Panarctic domain in CCSM4, together with a mechanistic description, is displayed in Figure 5.

Besides studies focusing on sea-ice reemergence mechanisms, I have also worked on data-driven prediction studies utilizing the kernel analog forecasting techniques described in Section 1 to predict the low-frequency NLSA modes associated with sea-ice reemergence families [21], as well as sea ice cover on regional and Panarctic scales [22]. These
studies have been carried out in collaboration with Darin Comeau (former postdoc, currently research scientist at Los Alamos National Laboratory), Jane Zhao, and Andrew Majda, and have thus far involved hindcast experiments in control integrations of the CCSM3 and CCSM4 models with fixed external forcings (e.g., volcanic and anthropogenic emissions). Prediction experiments with observational datasets, addressing, in particular, the issue of trend extrapolation via kernel methods, are planned for future research. It should be noted that in sea ice prediction “trivial” methods such as persistence and damped persistence forecasts (or linear extrapolation in the presence of climate change trends) are frequently used as benchmarks as they can be challenging to beat by statistical or first-principles models alike [68].

In [21], we demonstrated that kernel analog forecasting can predict the NLSA-derived NPGO and PDO modes with pattern correlation scores exceeding 0.6 by up to 20 months, providing an $\sim 8$ month increase of skill over persistence. This implies that the sea ice reemergence patterns associated with these modes are also predictable over similar intervals. While encouraging, these results do not of course imply that the actual area covered by ice within a given geographical domain (which is the outcome of several dynamic and thermodynamic processes not represented by the NPGO and PDO) is predictable on interannual timescales. Indeed, in the context of sea ice forecasts on either Panarctic or regional scales, skillful forecasts on significantly shorter, seasonal, timescales are challenging even in a perfect-model scenario. Nevertheless, in [21], we demonstrated that in the context of a control integration of CCSM4, kernel analog forecasting predicts total Arctic sea ice cover with skill out to $\sim 6$ months, which is a 3–4 month improvement of skill over the persistence forecast. We also examined the skill of regional-scale forecasts (see Figure 6), and found that the method provides a similar improvement of skill over persistence, with the exception of certain North Atlantic regions exhibiting particularly high persistence (e.g., the Barents and Greenland Seas) where gains over persistence were more moderate. In a number of cases, we observed a non-monotonic dependence of skill on lead time and target month featuring limbs of enhanced correlation characteristic of sea ice reemergence processes. Specifically, in the central Arctic basins (e.g., Beaufort, Chukchi, East Siberian, Laptev, and Kara Seas), we found high predictability occurring during the melt seasons, indicating that growth-to-melt reemergence is aiding skill. Similarly, in the marginal ice zones of the Bering Sea and Sea of Okhotsk, the regions of high predictability are in the growth seasons, where melt-to-growth reemergence is present.

**Antarctic circumpolar waves.** It is known that SST and SIC variability in the Southern Ocean on seasonal to interannual timescales exhibits a significant traveling wave component [69]. These Antarctic circumpolar waves (ACWs) generally propagate eastward along the Antarctic circumpolar current, and exhibit zonal wavenumbers in the range 2–3, but their dynamical origins and detailed propagation characteristics are only partially known. For instance, while some studies report unambiguous detection of zonal (eastward or westward) ACW propagation [70], others have found that the dominant interannual mode of Antarctic sea ice variability exhibits a standing, rather than traveling, dipole pattern with opposite-sign anomalies in the Pacific and Atlantic sectors of the Southern ocean [71]. Another pertinent question is the role of local air-sea interaction versus remote forcing in ACW dynamics; here, the general consensus is that wavenumber-2 ACWs are forced remotely by ENSO, whereas wavenumber-3 ACWs are an outcome of a local air-sea interaction [72]. Motivated by these questions, in a study with Xinyang Wang (AOS PhD student) and Joanna Slawinska [23] we applied NLSA to model and observational datasets of Antarctic SST and SIC, and studied the propagation characteristics of the recovered modes as well as their relationship with ENSO. The leading family of interannual modes recovered in these analyses consists of a primary pair of oscillatory modes at the frequency $\nu_{ACW} \approx 0.25 \text{ yr}^{-1}$ and its associated combination modes with the annual cycle at the frequencies $\nu_{mn} = m
u_{ACW} + n\nu_A$, with integer $m$ and $n$. In terms of SST and SIC anomalies, the primary ACW mode features a clear eastward-propagating, wavenumber-2 disturbance characteristic of the ACW, which attains its maximum strength over the eastern Pacific and western Atlantic sectors of the Southern Ocean, while diminishing significantly in the Indian Ocean sector (see Figure 7). The combination modes in this family feature analogous modulating relationships with the primary modes as those described above for ENSO and Arctic sea-ice reemergence. Intriguingly, the direction of propagation of the ACW combination modes can be either eastward or westward (depending on the sign of $\nu_{mn}$), and moreover these modes feature meridional pulsations in response to the advance/retreat of the ice edge with the annual cycle (which is in some sense reminiscent of the sea-ice reemergence modes recovered in the Arctic). Put together, this mode family leads to an intricate propagation structure for the ACW, exhibiting both propagating and standing components, for which individual members of the family act as “building blocks”. Another noteworthy property of this class of modes is its high correlation with the NLSA ENSO mode family recovered from Indo-Pacific SST [15, 16]. In the case of model (CCSM4) data, we found pattern correlation coefficients between the Indo-Pacific and Antarctic-derived modes upwards of 0.9; in the case of observational (HadISST) data, that value was a smaller, but statistically significant, 0.6. Our study thus establishes a direct link between ENSO combination modes and ACWs. As a next step, we
**Figure 6.** Pattern correlation score of kernel analog forecasting of sea ice cover in a control integration of CCSM4 as a function of target month and lead time by region. Notice that pan-Arctic September shows the lightest color for lag 0, but is over 0.5 for 3–4 months. In general across regions, predicting fall is typically the most skillful, with the exception of the North Pacific basins in marginal ice zones (Bering Sea and Sea of Okhotsk), where higher predictability is in the winter. The North Atlantic adjacent regions have a larger extent of predictive skill, with apparent limbs of high correlation that may be reemergence processes aiding kernel analog forecasting.

In addition to the wavenumber-2 ACW mode family, we also identified in Antarctic data from CCSM4 an oscillatory pair of modes capturing a wavenumber-3, eastward propagating ACW structure at a frequency $\nu_{ACW3} \approx 0.5$ yr$^{-1}$, shown in Figure 7. No evidence of this mode was found in Indo-Pacific SST, suggesting that it is indeed an outcome of local dynamics as opposed to remote ENSO forcing as suggested by previous studies. It should be noted that recovering this mode via objective eigendecomposition techniques such as NLSA is particularly challenging since its frequency of oscillation is close to some of the combination mode frequencies of the wavenumber-2 ACW family. Indeed, recovering a pure wavenumber-3 ACW mode from observational data was not possible in our analysis, though we did recover a mode that appears to be a mixture between a wavenumber-3 ACW mode and a higher-order wavenumber-2 combination mode with $n = 2$. This mode is also displayed in Figure 7. In future work, we plan to use the ACW modes identified in [23] in conjunction with trend modes that NLSA recovers from observational data (not discussed here) to assess the contribution of interannual modes versus externally forced climate response to the observed positive trend of Antarctic sea ice cover over the past decade.

**Recovering decadal variability.** Besides interannual modes such as ENSO and the TBO, the Indo-Pacific is the domain of activity of prominent modes of decadal variability. Among these, the interdecadal Pacific oscillation (IPO)
(a) CCSM4 Antarctic SIC input data

(b) HadISST Antarctic SST and SIC input data

Figure 7. ACW modes recovered via NLSA from (a) SST input data from CCSM4 and (b) joint SST and SIC input data from the HadISST observational dataset. From left to right, the panels show snapshots of reconstructed SIC anomaly fields associated with the primary wavenumber-2 ACW modes, the combination modes between the wavenumber-2 ACW and the annual cycle at the frequencies $\nu_A - \nu_{ACW}$ and $\nu_A + \nu_{ACW}$, and the primary wavenumber-3 ACW modes. In the HadISST case, the wavenumber-3 modes exhibit mixing with higher-order wavenumber-2 ACW combination modes at nearby frequencies.

[73] (which is closely related to the PDO defined on a North Pacific domain [66]) is arguably the most studied pattern, although several studies point to the existence of other decadal modes playing a role in the observational record [74]. Understanding the physical origins, as well as the interactions with the higher-frequency, interannual modes, of such decadal modes is a topic of considerable current interest in AOS. To some extent, research on decadal variability revolves around the question of whether it is an intrinsic dynamical phenomenon or an outcome of statistical data analysis techniques. In particular, many of the classical mode decomposition techniques such as EOF analysis rely on low-pass filtering in order to recover decadal modes (otherwise, the results are dominated by high-variance interannual modes such as ENSO), and there is a risk that the extracted patterns are residuals of interannual modes. As an effort to identify decadal modes of climate variability while avoiding the risk of introducing biases due to low-pass filtering, in [15, 16] we studied a set of decadal modes recovered simultaneously with the ENSO and TBO modes described above from model and observational Indo-Pacific SST data. While one of these modes represents the IPO, the leading decadal mode in the NLSA spectrum is a new pattern of lower frequency, referred to as west Pacific multidecadal mode (WPMM) due to an associated prominent cluster of SST anomalies in the western tropical Pacific (see Figure 8). We found that the WPMM exhibits significant correlation relationships (pattern correlation coefficients equal to 0.65 and 0.53 in CCSM4 and 20th Century Reanalysis, respectively) with the amplitude of ENSO, with the WPMM phases associated with enhanced ENSO activity. Physically, we attributed this behavior to anomalous atmospheric subsidence occurring in the western equatorial Pacific during cold WPMM phases (which can last for several decades), creating a pattern of anomalously strong zonal winds in the central Pacific and a flatter zonal thermocline profile there. Modeling studies, e.g., [75], have established that such background conditions favor increased ENSO activity. Another important characteristic of the WPMM is its high correlation with decadal precipitation over the Australian continent (correlation coefficient as high as 0.67 in both CCSM4 and 20th Century reanalysis), which we justified on the basis of advection of warm moist air over the Australian continent by the atmospheric circulation patterns associated with the WPMM. In current work, we have observed that robust WPMM patterns can be extracted from sparse point measurements of Indo-Pacific SST in climate models. This motivates studies aiming to detect this mode in paleoclimate proxy data, and assess its role in explaining the recent climatic record, including the global warming hiatus observed in the first decade of the 21st century [76]. In addition, the fact that NLSA recovers the WPMM and interannual modes simultaneously (reducing the risk that the former is a residual of the latter), makes this mode a good candidate to study physical aspects of decadal climate variability and its interactions with phenomena such as ENSO.
Extracting and predicting tropical intraseasonal oscillations. Tropical intraseasonal oscillations are planetary-scale modes of convective organization in the tropics and subtropics, spanning thousands of kilometers in space and evolving on $\sim 50$ day timescales with strong influences on weather and climate predictability [77]. In boreal winter, the dominant intraseasonal pattern is the Madden-Julian oscillation (MJO) [78]; an eastward-propagating dipolar envelope of enhanced and suppressed convective activity that originates off the east coast of Africa in the Indian Ocean, and propagates eastwards over the Maritime Continent and into the western Pacific Ocean until dissipating upon reaching the dateline. In boreal summer, as the solar heating shifts to the north of the equator, the dominant intraseasonal mode becomes the boreal summer intraseasonal Oscillation (BSISO) [79], which is characterized by northeastward propagation of convective disturbances over the Indian Ocean and the Indian subcontinent. This mode is closely related to the monsoon intraseasonal oscillation (MISO) [80] associated with the active and dry spells of the Indian monsoon. Due to model errors caused by subgrid-scale representation of moist convection, intraseasonal oscillations are generally poorly represented by current weather and climate models, yet these phenomena may provide an important source of predictability on subseasonal timescales linking weather and climate forecasts.

One of the central questions in current research on intraseasonal oscillations is how to actually define the phenomena themselves [81, 82]. In particular, since the MJO and BSISO are not represented as linear modes in idealized primitive models such as the shallow water equations [82], they are generally defined via indices (temporal patterns) extracted from atmospheric variables through data analysis techniques, sometimes with significant discrepancies in the identification of significant events and their associated spatial patterns. Arguably, these discrepancies are at least partly caused by the data analysis techniques employed; for instance, by any preprocessing to which the data is subjected prior to analysis, or by intrinsic limitations of the methods in representing distinct physical phenomena through individual modes. For example, perhaps the most widely used indices for monitoring the amplitude and phase of the MJO are the real-time multivariate MJO (RMM) indices [83] (shown in Figure 9), which are constructed through multivariate EOF analysis of outgoing longwave radiation (OLR; a proxy for atmospheric deep convection) and lower (850 hPa) and upper (200 hPa) zonal winds, averaged over the latitudes $15^\circ$S–$15^\circ$N, and bandpass-filtered to isolate the intraseasonal timescales of interest. While these indices successfully recover intraseasonal variability in the tropical atmosphere, their temporal evolution is rather noisy as they represent both the eastward-propagating MJO and the northeastward-propagating BSISO in a single pair of principal components (temporal modes), hampering the physical interpretability of statistical analyses utilizing them as well as their predictability with first-principles or statistical models.

In a series of papers [24, 26–28] in collaboration with Andrew Majda, Eniko Székely (former postdoc, currently research scientist at the Swiss Data Science Institute), and Wen-wen Tung, we have explored the use of NLSA as an alternative method for identification of tropical intraseasonal oscillations as well as other modes of convective
organization on interannual to diurnal timescales. By virtue of its connections with the spectral theory of dynamical systems discussed in Section 1, this method requires a minimal amount of data preprocessing and subjective choices from the part of the user. In particular, we found that applied to unfiltered, two-dimensional brightness temperature data (a proxy for atmospheric convection analogous to OLR) from satellite observations, NLSA naturally recovers two pairs of oscillatory eigenfunctions representing the MJO and BSISO, respectively. These eigenfunctions exhibit high temporal coherence and, as shown Figure 9, strong discriminating power between eastward (MJO) propagation in boreal winter and poleward (BSISO) propagation in boreal summer. In [28], we studied the initiation and termination statistics of the NLSA-derived MJO and BSISO modes, and found that these modes exhibit little to no retrograde propagation in the corresponding modal phase spaces, and once active, tend to undergo a complete evolution through their lifecycle before decaying, which suggests that they have favorable predictability properties. In a related study in collaboration with C. T. Sabeerali, R. S. Ajayamohan, and Andrew Majda [31], we recovered MISO modes with a similarly high temporal coherence to the BSISO modes recovered from brightness temperature data by applying NLSA to precipitation data over an Asian summer monsoon domain. We further demonstrated that these modes explain a higher fractional variance of intraseasonal rainfall anomalies than the corresponding modes obtained via extended EOF analysis (a variant of EOF analysis/PCA applied to delay-coordinate mapped data), while also providing a more realistic spatial representation of the regions of monsoon activity.
Besides these studies on mode identification, I have worked on prediction studies utilizing statistical [25, 30] or first-principles [30] models to predict the NLSA-derived MJO, BSISO, and MISO modes. In work in collaboration with Nan Chen and Andrew Majda [25], it was demonstrated that the MJO eigenfunction time series can be predicted with skill out to $\sim 45$ days within the training data using a stochastic oscillator model with a time-periodic damping modeling the active MJO phase during boreal winter and its quiescent phase during boreal summer. Chen and Majda subsequently demonstrated [84] that the same class of oscillator models predicts the NLSA BSISO modes with similar skill. It is worthwhile noting that the ability of such oscillator models to accurately model the evolution of the NLSA eigenfunction time series is consistent with the fact that the NLSA eigenfunctions can be viewed as data-driven approximations of Koopman eigenfunctions (see Section 1), which are theoretically expected to have coherent oscillatory behavior along orbits of the dynamics. In a study in collaboration with Romeo Alexander, Eniko Székely, and Jane Zhao, we demonstrated that a similarly high skill of MJO and BSISO prediction is attained via kernel analog forecasting in a realistic hindcast setting. Results from this study are displayed in Figure 10, where it can be seen that the MJO and BSISO forecasts have skill out to $\sim 50$ days. By comparison, the predictability of the RMM index with conventional parametric regression models is typically 10–15 days [85]. In [31], we demonstrated that the NLSA indices for MISO can also be predicted out to $\sim 3$ weeks in extended-range hindcasts based on NCEP Coupled Forecast System version 2 operational output. In summary, the results in [24–28, 30, 31, 84] demonstrate that the NLSA-based modes for tropical intraseasonal oscillations provide useful representations of these phenomena requiring no prefiltering of the input data, while having both high physical interpretability and predictability. As an extension of our work on prediction of MJO and BSISO indices, we are working on methods for prediction of physical variables, e.g., brightness temperature and precipitation, on regional scales conditioned on those indices. Another topic that we are currently studying is the covariability between convective and SST anomalies associated with the MJO and their role in ENSO initiation.

Figure 10. Kernel analog prediction of one of the two NLSA eigenfunctions representing the MJO for the interval July 1, 2006 to January 31, 2009. (a–d): Running forecasts (orange) with lead times of 15, 30, 45, and 60 days, respectively, along with the true signal (blue). (e) RMSE and (f) pattern correlation (PC) error metrics for individual years as well as for the entire testing period. The amount of time spent above the 0.6-PC threshold is listed in the legend of (f) for each grouping. The PC and RMSE skill scores are calculated by excluding June, July, and August when the MJO is inactive.
Identifying convectively coupled equatorial waves. The planetary-scale MJO and BSISO described above are at the top of a dynamically complex multiscale hierarchy of tropical convective organization, whereby cloud clusters and mesoscale convective systems (with horizontal scales of the order of a few hundred kilometers and lifetimes of the order of a day) organize into synoptic-scale, longer-lived convectively coupled equatorial waves (CCEWs), which in turn organize into planetary-scale modes such as the MJO and BSISO. CCEWs, in particular, are traveling disturbances that generally propagate parallel to the equator (in either eastward or westward directions) on timescales of \( \sim 5–30 \) days [86], and are clearly visible as embedded structures within the MJO convective envelope [87]. These waves are typically classified as eastward-propagating Kelvin waves, westward-propagating equatorial Rossby and mixed Rossby-gravity (MRG) waves, and eastward/westward inertia-gravity waves, based on linear, dry, shallow-water wave theory [88]. It is a remarkable fact that, despite the approximations involved, the dispersion characteristics obtained from that theory agree well with the observed frequency-wavenumber spectrum of convective disturbances of the tropical atmosphere [89]. At the same time, dry shallow-water models ignore certain fundamental aspects of tropical atmospheric dynamics such as moisture and topography, and this precludes a complete characterization of the CCEW spectrum encountered in nature and its interaction with the MJO on the basis of that theory. In particular, a question of significant current interest is whether there exist systematic dependencies in the types of CCEWs that develop within the MJO convective envelope, with some studies indicating that the answer to that question is negative [90].

Perhaps even more so than in the case of the intraseasonal oscillations, data analysis techniques requiring minimal prior assumptions on the frequency-wavenumber characteristics of the recovered signals are expected to be beneficial for improved identification and physical understanding of CCEWs. In particular, unlike intraseasonal oscillations which are dominated by two distinct modes (the MJO and BSISO), there exists a plethora of CCEWs forming a quasi-continuous spectrum in observational data, and isolating these disturbances via spectral filtering techniques is highly non-trivial. In work in collaboration with Joanna Slawinska and Jane Zhao, we have performed a preliminary analysis of CCEW modes recovered from equatorially averaged [5] and two-dimensional [29] brightness temperature data from the CLAUS satellite archive via the Koopman operator techniques described in Section 1, using the output of NLSA as basis functions. The results from these studies, examples of which are displayed in Figure 11, demonstrate that the data-driven Koopman eigenfunctions can successfully capture the temporal and spatial characteristics expected of at least some CCEW types, despite operating on unfiltered data. For example, the pattern in the bottom-left panel in Figure 11 is an MRG-like pattern propagating westward over the western Pacific warm pool and the Maritime Continent with an approximately five-day period, featuring a predominantly antisymmetric structure about the equator and eastward zonal tilts that are not predicted by the theoretical MRG solutions of the shallow water equations. A qualitatively similar tilted structure was also identified in [91] from filtered OLR data. It is worthwhile noting that although it is asymptotically equivalent with Koopman operator techniques [8], NLSA was not able to identify CCEW patterns as distinct modes from the same CLAUS dataset (though the interannual, intraseasonal, and diurnal modes recovered by NLSA are also recovered by the Koopman operator analysis). Indeed, the results in [5, 29] appear to be the first successful identification of CCEWs via objective eigendecomposition techniques. At the same time, it should be emphasized that, compared to our earlier work on intraseasonal oscillations, the post-analysis task of cataloging and examining the physical properties of the recovered CCEW patterns is more challenging due to the sheer number of these modes. Currently, we are in the process of analyzing and assessing the physical significance of our results, and plan to report on them in the coming months. Another possibility we have been exploring is to apply the operator-valued kernel technique [12] mentioned in Section 1 to construct a basis for Koopman operator analysis (as opposed to NLSA), which should have higher efficacy in capturing spatially intermittent patterns such as CCEWs.

Analysis of heat transfer and large-scale circulation in Rayleigh-Bénard convection. Rayleigh-Bénard convection is one of the most widely studied pattern-forming spatiotemporal systems [92]. A general property of this class of systems is that for sufficiently large values of the Rayleigh number (heating strength), the flow develops a large-scale circulation (LSC) evolving on timescales that may be several orders of magnitude longer than the natural convective timescale of the flow, and whose properties depend strongly on the geometry of the domain [93]. The LSC influences in turn the flow’s transport properties, including its heat transfer characteristics.

In work in collaboration with Noah Brenowitz and Andrew Majda [32], we studied aspects of heat transfer and LSC in two-dimensional Rayleigh-Bénard convection in a high-aspect-ratio periodic domain (aspect ratio equal to 10) in a regime (Rayleigh and Prandtl numbers equal to \( 5 \times 10^6 \) and 1, respectively) characterized by the formation of a cellular array of convective rolls which exchange heat and momentum leading to large-scale organization. An intriguing aspect of that regime is intermittent large deviations of the horizontally integrated vertical heat transfer (measured by the Nusselt number) towards anomalously small values. The main goal of our study was to identify the spatiotemporal...
patterns and mechanisms responsible for that behavior. From a data analysis standpoint, a challenge with that objective is caused by the invariance of heat transfer with respect to translations in the horizontal direction—this implies that the spatial and temporal patterns which account for horizontal translations of the flow, extracted by data analysis algorithms agnostic to that symmetry, are irrelevant from the point of view of explaining the observable of interest. To address that issue, we performed a judicious, physically motivated reduction of the input data inspired by the isentropic streamfunction analysis developed in [94]; specifically, our approach was to average the vertical component of the velocity along isothermal lines. In essence, the effect of this averaging is to quotient out the symmetry of the problem without causing loss of information to the observable of interest. Applying NLSA to the isothermally averaged vertical velocity field data led to an identification of a family of modes (examples of which are shown in Figure 12) that accurately capture the variability of the Nusselt number, including its intermittent fluctuations to anomalously small values. In particular, the analysis produced a low-frequency mode which is preferentially active during periods of anomalously low heat transfer, and whose spatial pattern in isothermal coordinate space accounts for reduced convective heat flow. Physically, we associated this mode with periods of anomalously high activity of the convective plumes developing in the domain, which results to increased tilting of the respective convective cells and thus heat transfer reduction. The asymmetry in the heat transfer statistics originates from the fact that a tilted cell can become more strongly tilted, but there is a limit to how “un-tilted” a cell can be.

In more recent work in collaboration with Anastasiya Kolchinskaya, Dmitry Krasnov, and Jörg Schumacher [33], we applied the Koopman operator techniques described in Section 1 to study the LSC in a three-dimensional turbulent Rayleigh-Bénard convection flow ($Ra = 10^7$ and $Pr = 0.7$) in a cubical domain. In this class of flows the geometry and dynamics of the LSC is strongly influenced by the discrete symmetry of the domain. In particular, it is known from both laboratory experiments [95] and large-eddy simulations [96] that the LSC exhibits transitions between four metastable configurations characterized by convective rolls along the diagonals of the square horizontal cross section of the domain (two diagonals times two roll orientations), and these states are mediated by unstable configurations in which the LSC is oriented parallel to the side walls. Our analysis was performed on a long direct numerical simulation spanning 10,000 free-fall times at a lower Rayleigh number than [95, 96]. We found that despite the high dimension ($128^3$ gridpoints) and broadband nature of the input data, the leading data-driven Koopman eigenfunctions, depicted in Figure 13, capture the spatial patterns and regime transitions of both the stable and unstable LSC states without performing any prefiltering. Moreover, the residence times in the stable diagonal states were found to range up to several thousand free-fall times. Intriguingly, we observed an asymmetry in the transitions between some of the stable

![Figure 11. Time series and frequency spectra of Koopman eigenfunctions representing CCEWs recovered from CLAUS brightness temperature data (top) and snapshots of the corresponding spatial patterns (bottom; arbitrary units). Left: 5-day, westward-propagating mixed Rossby-gravity wave. Center: 15-day, westward-propagating equatorial Rossby wave. Right: 10-day, eastward-propagating Kelvin wave.](image-url)
Figure 12. Left-hand panels: Time series, power spectral densities, and probability density functions for the Nusselt number Nu and the leading NLSA eigenfunctions for a portion of the two-dimensional Rayleigh-Bénard dataset. Notice that the variability of the Nusselt number has a regular oscillatory component (here, associated with a sloshing motion of convective cells) which is modulated by a low-frequency envelope preferentially strengthening the negative phase of the high-frequency oscillation. These two distinct components of the Nusselt number variability are captured by the leading three NLSA eigenfunctions \( \phi_1, \phi_2, \phi_3 \), where \( \phi_1, \phi_2 \) comprise a high-frequency oscillatory pair (only \( \phi_1 \) is shown here), and \( \phi_3 \) captures the low-frequency envelope of Nu. Right-hand panel: Phase composite of isothermally averaged vertical velocity \( \bar{w}(z, T) \) as a function of temperature \( T \) (horizontal axis) and vertical coordinate \( z \) (vertical axis) associated with low-frequency eigenfunction \( \phi_3 \), computed over all samples of the dataset with \( Nu \leq 8 \). Here, red (blue) colors occurring at \( T < 0.5 \) (\( T > 0.5 \)) indicate anomalously cold (warm) fluid moving upwards (downwards), leading to a negative net heat transfer anomaly.

LSC states which is not compatible with the cubical symmetry of the domain. We believe that this behavior is evidence for a type of “broken ergodicity” for this system necessitating an even longer simulation to provide a uniform sampling of all transitions. In addition to these primary Koopman eigenfunctions, we identified a set of secondary eigenfunctions describing secondary LSC structures such as corner vortices and swirls. As shown in Figure 13, these eigenfunctions exhibit significant amplitude modulations corresponding to preferential activity during particular diagonal (stable) primary LSC states.

Acknowledgments

This research was supported by ONR grant N00014-14-1-0150, ONR MURI grant 25-74200-F7112, ONR YIP grant N00014-16-1-2649, NSF grant DMS-1521775, DARPA grant HR0011-16-C-0116, and grant MM/SERP/NYU/2014/SSC-01/002 by the Earth System Science Organization, Ministry of Earth Sciences (MoES), Government of India.

References

Figure 13. Data-driven Koopman analysis of turbulent Rayleigh-Bénard flow in a cubical cell. Top-left panels (clockwise from top left): Instantaneous velocity field streamlines, time-averaged velocity field streamlines, time series of the Nusselt number, and vertical profile of convective heat transfer. Top-right panels: Three-dimensional phase space formed by the real and imaginary parts of the leading three data-driven Koopman eigenfunctions (ordered in order of increasing Dirichlet energy as described in Section 1), with a trajectory plot revealing the LSC transitions (center). The outer panels show reconstructions of the vertical velocity field on a horizontal cross-section taken in the middle of the vertical extent of the domain, with red and blue colors corresponding to positive and negative values of vertical velocity anomalies (relative to the time-averaged state). Notice the stable diagonal states labeled 1, 3, 5, 7 and the unstable states labeled 2, 4, 6, 8. Note also that the association of each outer snapshot to a single point in the three-dimensional phase space is for illustration purposes only, as multiple points in that phase space are used for reconstruction in accordance with delay-coordinate embedding; see [33] for more details. Bottom panels: Time series of the real part of the leading Koopman eigenfunctions (top; shown for reference to indicate the primary LSC state transitions) and the real parts of secondary Koopman eigenfunctions with modulating behavior conditioned on the primary LSC state. The secondary Koopman eigenfunctions form doubly degenerate oscillatory pairs (one member from each pair is shown here), which are preferentially active during residence times to a particular stable primary LSC state.


