

Kernel Analog Forecasting of Tropical Intraseasonal Oscillations

ROMEO ALEXANDER, ZHIZHEN ZHAO,^a ENIKO SZÉKELY, AND DIMITRIOS GIANNAKIS

Courant Institute of Mathematical Sciences, New York University, New York, New York

(Manuscript received 12 May 2016, in final form 18 October 2016)

ABSTRACT

This paper presents the results of forecasting the Madden–Julian oscillation (MJO) and boreal summer intraseasonal oscillation (BSISO) through the use of satellite-obtained global brightness temperature data with a recently developed nonparametric empirical method. This new method, referred to as kernel analog forecasting, adopts specific indices extracted using the technique of nonlinear Laplacian spectral analysis as baseline definitions of the intraseasonal oscillations of interest, which are then extended into forecasts through an iterated weighted averaging scheme that exploits the predictability inherent to those indices. The pattern correlation of the forecasts produced in this manner remains above 0.6 for 50 days for both the MJO and BSISO when 23 yr of training data are used and 37 days for the MJO when 9 yr of data are used.

1. Introduction

The Madden–Julian oscillation (MJO; Madden and Julian 1972) and boreal summer intraseasonal oscillation (BSISO; Wang et al. 2009; Kikuchi et al. 2012; Lee et al. 2013), are the two most dominant intraseasonally varying patterns of the tropical atmosphere. The former, an eastward-propagating envelope of convective activity originating in the Indian Ocean during boreal winter, is known to modulate Pacific and Caribbean cyclogenesis, affect rainfall variability on multiple coastal regions, and alter the strength of the ENSO cycle, among other effects (Zhang 2005). The BSISO, meanwhile, is closely related to intraseasonal oscillations of the South Asian monsoon (Goswami 2011, and references therein) and influences its onset and active and dry spells. More broadly speaking, improved understanding of these two intraseasonal modes would go a long way in filling the gap that is beyond the reach of short-term weather

forecasts and below the resolution of long-term climate models (Waliser 2011; Zhang et al. 2013). Yet, despite the understood importance of these modes of variability and several decades of research, there are still significant challenges to improving the real-time monitoring and forecasting of these two modes.

Among these challenges is the still-open problem of defining intraseasonal oscillations in a consistent and objective manner. Although the real-time multivariate MJO index (RMM; Wheeler and Hendon 2004) has emerged as the most commonly used standard for MJO measurement, drawbacks such as biased sensitivity toward wind velocity data and overreliance on preprocessed data have meant that the development and testing of alternative indices has yet to cease (Kiladis et al. 2014). No comparably popular measure has emerged for the BSISO, in no small part because of the relatively greater complexity of monsoon dynamics (Lee et al. 2013). One well-recognized common source of these difficulties, among several, is that the majority of current techniques are not ideally suited for multiscale phenomena like organized tropical convection, as they generally require ad hoc data preprocessing to isolate the temporal and spatial scales of interest. The recently developed technique of nonlinear Laplacian spectral analysis (NLSA; Giannakis and Majda 2011, 2012, 2013, 2014) seeks to redress this mismatch by producing indices based on time-lagged embedding and local measures of data similarity that more sensitively capture nonlinear dynamics than standard eigendecomposition techniques, such as empirical

 Denotes content that is immediately available upon publication as open access.

^a Current affiliation: Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign, Urbana, Illinois.

Corresponding author e-mail: Romeo Alexander, romeo@cims.nyu.edu

orthogonal function (EOF) analysis. This technique has been used to extract families of modes of variability from equatorially averaged (Giannakis et al. 2012; Tung et al. 2014) and two-dimensional (2D) (Székely et al. 2016a,b) brightness temperature (T_b) data spanning interannual to diurnal time scales without prefiltering the input data. These mode families include representations of the MJO and BSISO with higher temporal coherence (Székely et al. 2016b) and stronger discriminating power between eastward and poleward propagation (Székely et al. 2016a) than patterns extracted through comparable linear approaches. Insofar as improved accuracy of representing tropical intraseasonal oscillations (ISOs) can bolster predictability, we explore in this paper the extent to which NLSA-derived indices can be used as a basis for forecasting the MJO and BSISO.

A second challenge is determining how to incorporate information about both the past and present into predictions of any given tropical ISO pattern's future evolution. There will always be, of course, some unpredictability to tropical ISOs that cannot be overcome without numerical simulations, especially since tropical ISOs are affected by global warming (Subramanian et al. 2014), the precise future spatiotemporal characteristics and magnitude of which remain uncertain. Nevertheless, the MJO and BSISO have shown enough regularity and consistency over the past 40 yr to suggest that the past can serve as a guide to the future. One of the simplest empirical forecasting techniques is the classical analog forecasting method of Lorenz (1969), which first identifies, via Euclidean distances, a moment in the past that is most similar to the present and then casts the historical evolution from that moment as the forecast for the future. In the context of tropical intraseasonal oscillations, analog techniques have been employed in real-time forecasting of Indian monsoons with skill extending to 20–25 days (Xavier and Goswami 2007). Generalizations of analog forecasting based on modifications, such as taking weighted averages of multiple historical paths, varying the type of distance measure by which analogs are chosen, and iterating multiple times in order to account for multiple scales, have recently been developed in a framework called kernel analog forecasting (KAF; Zhao and Giannakis 2016; Comeau et al. 2017), which will be employed later in this paper.

Analog forecasting methods can preserve many of the attractive qualities of NLSA-derived indices. The compatibility of these otherwise two distinct techniques lies in their mutual reliance on dynamics-dependent geometric measures of data similarity. In particular, NLSA indices are the eigenvectors of a kernel operator (Belkin and Niyogi 2003; Coifman and Lafon 2006; Berry and Sauer 2016), which can be loosely thought of as a local

covariance matrix. Much of the character of these indices is due to a specific choice of a smooth, data-dependent Gaussian-like kernel that takes dynamics into account through Takens delay-coordinate embeddings (Takens 1981; Packard et al. 1980; Broomhead and King 1986; Sauer et al. 1991). Meanwhile, KAF methods produce forecasts by taking weighted averages of historical data, with weights determined by a measure of similarity between the present and any prior moment in time. By letting these analog weights be determined by kernels of the same class as that used in the extraction of the NLSA indices, the resulting algorithm is more sensitive to oscillations in the intraseasonal range and can thus produce more faithful forecasts than otherwise (Zhao and Giannakis 2016). As shown here, this pairing of NLSA and kernel analog techniques can yield nearly 6 weeks' worth of skill in forecasting the MJO and BSISO. This result is comparable to those of the stochastic oscillator models of Chen et al. (2014) and Chen and Majda (2015a) for predicting the NLSA-derived MJO and BSISO modes, respectively.

Some advantages that KAF can be said to have over other recent skillful forecasting methods of tropical ISOs are its nonparametric nature, which helps avoid dynamical model error, and its ability to operate in real time. Dynamical model errors, in particular, have historically been a significant obstacle to forecasting ISOs whether through numerical models or parametric statistical models, with skill of earlier models limited to 10–15 days (Waliser 2011, and references therein). More recently, however, advances in MJO simulation by coupled models have led to skill as high as 27 days for forecasting the RMM indices (Miyakawa et al. 2014; Neena et al. 2014; Vitart 2014; Xiang et al. 2015). Similar improvement of skill has also been attained by recent parametric empirical models (Kondrashov et al. 2013; Chen and Majda 2015b). It is important to note that the higher predictive skill reported in this paper is at least partly due to the higher intrinsic predictability of the NLSA-derived modes compared to the RMM. Forecast skill aside, while both coupled models and parametric empirical models are important for elucidating the physical processes underlying tropical ISOs, such models often require iterative tuning of numerous parameters, which in many ways can introduce significant biases. Kernel analog forecasting, on the other hand, is purely empirical in the sense that none of its parameters rely on any physical model. This nonparametric quality allows the method to both avoid model error and be automatic, at least in the sense that no manual intervention is required after initial data entry. Moreover, KAF can operate in real time, which is a feature that is sometimes absent in statistical models.

The dataset of interest in this paper, to which the KAF technique is applied, is tropical 2D T_b data collected under the Cloud Archive User Service project (CLAUS; Hodges et al. 2000), which is often viewed as a good proxy for tropical convective activity. Our primary experiment utilizes the first 24 yr of available CLAUS T_b data as training data, from which MJO and BSISO forecasts for the period from 2006 to 2009 are constructed. A secondary experiment with only 9 yr of training data, from 1983 to 1992, is also performed, and we additionally perform a suite of sensitivity tests to assess the influence of factors such as the training time series length and the delay embedding window on forecast skill. In particular, standard measures of skill, such as root-mean-square error (RMSE) and Pearson correlation (PC) are presented for lead forecast times varying from 15 to 60 days. The chief result of this paper is that the PC score for kernel analog forecasts of the MJO and BSISO can stay above a 0.6 threshold for ~ 40 days of lead time. Context for how this result fares against other prediction methods and ISO indices is provided in the discussion and conclusions of this paper.

The plan for the rest of the paper is as follows. An overview of the KAF method is provided in section 2. The specific data of interest are described in section 3. The results of applying KAF to this data, as well as a sensitivity analysis, are presented in sections 4 and 5, respectively. A discussion of these results constitutes section 6, and broader context, possible future research directions, and other concluding remarks are given in section 7.

2. Kernel analog forecasting

We provide here a brief overview of ensemble kernel analog forecasting, a more complete description of which can be found in Zhao and Giannakis (2016). The first subsection outlines the general framework of the method, illustrating how the final forecast is the result of multiple iterations of weighted averages. The subsequent subsections describe the key components of the method, including time-lagged embedding, creation of training data via NLSA extraction, and kernel specification. The final subsection describes the ways in which the resulting forecasts are assessed.

a. Preliminaries

The forecasting problem consists of predicting what the value of some physically meaningful quantity f will be at some lead time τ in the future. We refer to f as the forecast observable and represent it as a map $t \rightarrow f(t)$, where $t \in \mathbb{R}$ represents a moment in time, and $f(t)$ lies in some output space Y . For instance, in sections 3 and 4, f will be the NLSA-based MJO or BSISO index, taking

values in $Y = \mathbb{R}$. It is often useful, however, to focus attention instead on the time-shifted forecast observable $U_\tau f : t \rightarrow f(t + \tau)$. With this new observable, the empirical forecasting task can then be stated as one of constructing, through empirical means only, a forecasting function $F_\tau : \mathbb{R} \rightarrow Y$ that reasonably approximates $U_\tau f$ on the entire time domain and for arbitrary lead time τ .

Empirical methods generalize a forecast from a training period $T \subset \mathbb{R}$ over which $U_\tau f$ has already been observed and recorded. We seek to extend this restricted function, denoted by $\widehat{U_\tau f}$, to a function that approximates $U_\tau f$ on the entire domain $\mathbb{R}$. Once the training period T has been fixed, the rest of the method is completely determined by the manner in which $\widehat{U_\tau f}$ is extended to the full domain $\mathbb{R}$. As such, it is often convenient to focus attention just on domain-extending operators P , dependent on an empirical observation map $t \rightarrow x(t)$ of some data space X , and defining the forecasting function as

$$F_\tau = P \circ \widehat{U_\tau f}, \quad (1)$$

where $\circ$ denotes composition of maps. The space X corresponds to the space of initial data for the forecast. In the application presented in subsequent sections, X will be a space of brightness temperature fields in time-lagged embedding space.

It will be useful later on to keep in mind the slightly more abstract interpretation of the forecasting problem in the context of dynamical flows within a hidden state space A (Zhao and Giannakis 2016), which, in the present case, is interpreted as Earth's entire climate system. In particular, all relevant dynamical laws are represented abstractly by a function Φ_t for which $\Phi_t(a)$ is the state that occurs t time units after state a in A is attained. A single realization $\{a_t\}_{t \in \mathbb{R}}$ of flow corresponds to fixing a state a_0 at some reference time $t = 0$ and defining $a_t = \Phi_t(a_0)$. For any such realization, there is a sequence of observations $\{x_t\}_{t \in \mathbb{R}}$ and values of the forecast observable $\{f_t\}_{t \in \mathbb{R}}$. These sequences are alternatively represented by the maps $f(t) = \pi_Y a_t$ and $x(t) = \pi_X a_t$, where π_Y and π_X are nonlinear projections onto Y and X , respectively. Finally, the quantity of interest is represented as $U_\tau f(t) = \pi_Y \Phi_\tau(a_t)$. In dynamical systems theory, the operators U_τ governing the evolution of observables are known as Koopman operators and have several useful properties for mode identification and prediction (Mezić 2005; Budisić et al. 2012; Giannakis 2016).

b. General analog forecasting

Kernel analog forecasting is characterized by having P act as a weighted average over training data acquired

from past observations of the system. In particular, the action of P on any map $G : T \rightarrow \mathbb{R}$ is defined by

$$P \circ (s) = \frac{1}{N} \sum_{t \in T} G(t) \rho(x(s), x(t)), \quad (2)$$

where $s \in \mathbb{R}$ denotes the present time, and $\rho : X \times X \rightarrow [0, \infty)$ is a weight function. In particular, setting $G = \widehat{U_\tau f}$ leads to the general forecast formula in (1). This procedure can also be described as identifying and averaging, via a data similarity function ρ , moments of the past that can serve as analogs for the present.

Conventional analog forecasting, as first proposed by Lorenz (1969), consists of using only one analog, the “best one,” at any given time. When X is equipped with a distance function D , conventional analog forecasting consists of first identifying the moment in time t^* at which the training data $x(t^*)$ is the nearest neighbor (in data space) to the data $x(s)$ observed at initialization time s : that is,

$$t^* = \operatorname{argmin}_{t \in T} D(x(s), x(t))$$

With the analog identified, the weights are then set to Dirac delta functions $\rho(x(s), \cdot) = \delta_{x(t^*)}$ centered at the analog data point $x(t^*)$, and the forecasting function in (1) becomes

$$F_\tau(t) = f(t^* + \tau).$$

A major drawback to this method is the risk of the resulting forecast F_τ being highly nonsmooth in its argument, as the “best” analog jumps around. Such nonsmooth behavior should be avoided in climate applications and is generally a sign that the forecast is overfitting its training data. In any case, considerations of smoothness aside, it is generally beneficial to incorporate as much information from different analogs as possible.

Kernel analog forecasting takes advantage of the information offered by multiple analogs by having ρ be smooth in its first argument. Since we want P to have the interpretation of being a weighted average, we always require that $(1/N) \sum_{t \in T} \rho(\tilde{x}, x(t)) = 1$ for all $\tilde{x} \in X$. Since enforcing this normalization constraint can make it more difficult to consider all of the possibilities for ρ , it is often convenient to focus on choosing another function $K : X \times X \rightarrow [0, \infty)$, known as the kernel, and letting normalization be taken care of automatically via the formula

$$\rho(x_i, x_j) = \frac{K(x_i, x_j)}{(1/N) \sum_{t \in T} K(x_i, x(t))},$$

for all $x_i, x_j \in X$. The kernel K is also a measure of similarity between data points, except that no normalization

constraint is imposed. One of the most common kernel choices in machine learning is the radial basis function (RBF) $K_\varepsilon(x_i, x_j) = \exp(-\|x_i - x_j\|^2/\varepsilon)$, where the double vertical bar represent the Euclidean distance and ε is a positive bandwidth-controlling parameter. A key property of RBFs is their rapid decay away from $x_i = x_j$ when ε is small; this property allows such kernels to localize on small neighborhoods of datasets with nonlinear geometric structures (in particular, datasets generated by complex dynamical systems with attractors such as the climate system), where the data geometry is approximately linear. As will be described in the next subsection, our choice of kernel is loosely based on RBFs, with several anisotropies introduced.

c. Laplacian pyramid iteration

We have thus far shown, given a training period T on which values of x and $\widehat{U_\tau f}$ are known, how a forecast function F_τ is constructed from a kernel K . For many of the most common forms of kernel, there is at least one tunable parameter whose determination is not trivial. For example, if one determines that RBFs are appropriate kernels, one is still left with the problem of choosing an appropriate bandwidth parameter ε . There is a risk of overfitting if ε is too small and of underfitting if ε is too large. Conventional analog forecasting, and thus overfitting, lies in the extreme case of $\varepsilon \rightarrow 0$, as the weight $\rho(x_i, x_j)$ approaches a point Dirac delta function centered at x_i . Simple averaging, a clear case of underfitting, lies at the opposite extreme of $\varepsilon \rightarrow \infty$, where $F_\tau(t)$ always returns a constant that is equal to the average of the training forecast observables [i.e., $(1/N) \sum_{t \in T} \widehat{U_\tau f}(t)$].

Even if one choice of ε is found for a certain segment of the training data that is neither overfit nor underfit, it can be difficult to find just one choice for which all of the training data are appropriately fit, especially if the data are slowly varying in some parts and rapidly varying in others. When such multiple scales of regularity in the training data arise, it is reasonable to incorporate more than one kernel, with a corresponding multitude of bandwidth scales. This can be done with an iterative generalization of (2), referred to as the Laplacian pyramids extension scheme (Rabin and Coifman 2012; Fernández et al. 2014). Suppose $\varepsilon_1, \varepsilon_2, \dots$ is a sequence of decreasing bandwidth parameters, with corresponding kernels K_{ε_i} and extension operators P_{ε_i} . At each step ℓ in the procedure, we produce a forecast $F_{\tau, \ell}$ and define a residual function $r_\ell : \mathbb{R} \rightarrow \mathbb{R}$, for lead time 0,

$$r_\ell = f - F_{0, \ell},$$

and time shift that function to obtain a residual at lead time τ via

$$r_{\tau,\ell} = U_{\tau} r_{\ell}.$$

This residual is a measure of the difference between the approximation $F_{\tau,\ell}$ and the truth $U_{\tau}f$. The restriction of the residual to just the training period, denoted by $\widehat{r_{\tau,\ell}}$, is a computable function that can be used to improve the overall accuracy of the forecast. The complete recursive procedure is then entirely described by the two formulas:

$$\begin{aligned} F_{\tau,0} &= 0 \quad \text{and} \\ F_{\tau,\ell} &= F_{\tau,\ell-1} + P_{\varepsilon_{\ell}} \widehat{r_{\tau,\ell-1}}. \end{aligned}$$

This scheme can generally be run for as long as necessary before there is no longer an appreciable change in $F_{\tau,\ell}$ from one step to the next. In practice, this means stopping when $r_{\tau,\ell}$ drops below a certain threshold δ , which we take in this paper to be $\delta = 10^{-6}$. The final forecasting function is set to be $F_{\tau} = F_{\tau,L}$, where L is the step at which the procedure is halted.

A common iterative scheme that rapidly moves through multiple scales, and one that we employ in this paper, is that of a decreasing dyadic sequence of bandwidth parameters: that is, letting each parameter be twice as small as the previous one [i.e., $\varepsilon_{\ell} = (1/2)\varepsilon_{\ell-1}$ for $\ell > 1$]. If the algorithm is run to completion, then only the computation time, and not the final results, is sensitive to the initial choice ε_1 , provided that the latter is large enough. One approach is to choose a value of ε_1 that is comparable to the size of a representative of the training data $x(T)$, which lets the algorithm not spend time with bandwidth parameters that are too large to capture any interesting geometric features. In practice, we follow the recommendation in [Fernández et al. \(2014\)](#) of setting ε_1 equal to the median of the pairwise distances between the training data. Note that, with both ε_1 and δ specified, the iterative procedure is automatic, in the sense that there are no free parameters left to be tuned.

d. Choice of observation map

The simplest choice for an observation map x is simply one for which the output at time t is all of the data procured at that time. For example, if $z(t) \in \mathbb{R}^n$ is a vector containing the values of satellite-obtained brightness temperatures at n spatial grid points, then the most immediate candidate for $x(t)$ is simply $z(t)$ itself. It is reasonable, however, to assume that previous values of the data should also figure into predictions about the future. As such, our map of choice is the delay-coordinate map ([Takens 1981](#); [Packard et al. 1980](#); [Broomhead and King 1986](#); [Sauer et al. 1991](#)), defined by

$$x(t) = \{z(t), z(t - \Delta t), \dots, z(t - q(-t)\Delta t)\},$$

where q is a positive integer parameter denoting the number of delays, and Δt is a fixed time interval. The interpretation here is that, at any given time t , we can observe not just a single snapshot of the system, but a succession of snapshots (i.e., a video). The parameter q can be taken as large as we need it to be in order to have the phenomena of interest represented in a single video observation $x(t)$. After embedding, the data $x(t)$ reside in a space $X = \mathbb{R}^N$ of dimension $N = nq$, and in practical applications N can be very high (e.g., in the case of the CLAU T_b data studied in [sections 4](#) and [6](#), N is of order 10^7). However, since both KAF and NLSA are based on localizing kernels, the performance of these methods is sensitive to the intrinsic dimension of the subset of $\mathbb{R}^N$ occupied by $x(t)$, which is typically far smaller than N .

An alternative justification for the use of delay embedding is found in dynamical systems theory. In that framework, the obtained data in X are said to be merely partial observations of a dynamical state in a hidden state space A . It is a classic result ([Packard et al. 1980](#); [Takens 1981](#); [Broomhead and King 1986](#); [Sauer et al. 1991](#)) that, with a sufficiently large window q and weak assumptions on the true system, time-lagged embedding can recover the topology of the system's dynamical attractor. The result is that the system, as represented by a path moving in the space X , rather than Z , is smoother and more Markovian (i.e., predictable). Besides improving Markovianity, time-lagged embedding is known to play an important role in improving the ability of kernel eigenfunctions to recover distinct dynamical time scales ([Berry et al. 2013](#); [Giannakis 2016](#)); we discuss this point further in [section 2f](#).

A detail that is relatively inconsequential to the overall theory, but important in actual implementation, is that appropriate truncations of training data must follow any time-lagged embedding. For example, as will be discussed in [section 3](#), we take one of our training periods T_{long} to span from 1 July 1983 to 30 June 2006, and set q to correspond to 64 days. Since no data are available before 1 July 1983, it is not possible to do full time-lagged embedding for any data obtained before 3 September 1983. If s is the number of samples in the original training period T , then the truncated period has length $s - q + 1$. Instead of always using different notation to differentiate between the original training period and the truncated training period, we sometimes denote both by T , with context determining which meaning is to be understood. Alternatively, we just say that x is undefined for the first q samples.

e. Choice of forecast observable

One of the challenges to forecasting intraseasonal tropical oscillations is the multiplicity of ways to define

them. Nevertheless, one common way to define ISOs is through eigenfunctions of the eigenvalue equation

$$\hat{\mathbf{L}}\phi_k = \lambda_k \phi_k, \quad (3)$$

for a suitable operator $\hat{\mathbf{L}}$ determined from the training data. For instance, in EOF analysis, $\hat{\mathbf{L}}$ is the temporal covariance matrix, and ϕ_k are the principal components. In NLSA, $\hat{\mathbf{L}}$ is a discretized version of a so-called Laplace–Beltrami operator, which is based on a sequence of normalizations of a kernel function $K: X \times X \rightarrow \mathbb{R}$ over a specified (and appropriately truncated) training period T (that is, T is of size $s - q + 1$, to take into account time-lagged embedding). This kernel is of the asymmetric exponential family described in (4) ahead. Letting $K_{ij} = K(x(t_j), x(t_i))$ where t_i, t_j are all of the times available in T , the sequence of normalizations that builds the matrix $\mathbf{L}$, as proposed in the diffusion maps algorithm of Coifman and Lafon (2006), follows

$$\begin{aligned} Q_i &= \sum_{j=1}^s K_{ij}, & \tilde{K}_{ij} &= \frac{K_{ij}}{Q_i Q_j}, \\ D_i &= \sum_{j=1}^s \tilde{K}_{ij}, & P_{ij} &= \frac{\tilde{K}_{ij}}{D_i}, \\ L_{ij} &= \delta_{ij} - P_{ij}, \end{aligned}$$

where δ_{ij} is the Kronecker delta function. Since $\mathbf{L}$ is a matrix of size $s \times s$, where s is typically on the order of tens of thousands of samples in our applications, it is useful to work with a sparse matrix $\hat{\mathbf{L}}$, obtained by retaining only the largest k_{mn} entries per row of $\mathbf{L}$, where k_{mn} is much less than s . Note that the ability to approximate $\mathbf{L}$ by a sparse matrix relies on $K(x_i, x_j)$ decaying sufficiently rapidly to zero away from $x_i = x_j$ [e.g., the RBF kernel and the anisotropic NLSA kernel in (4) ahead both have exponential decay]; this results in a significant reduction of computational cost, which is generally not possible with eigendecomposition techniques based on covariance matrices. In what follows, we generally set k_{mn} so that only about 15% of the entries of $\mathbf{L}$ are retained. It can be shown that the resulting eigenmodes are orthogonal with respect to a weighted inner product $\langle \phi_k, \phi_l \rangle = \sum_{i=1}^s \phi_{ik} \phi_{il} D_i$, where ϕ_{ik} is the i th component of the eigenvector ϕ_k .

Multiple eigenmodes ϕ_k are obtained by solving the eigenvalue problem in (3). In NLSA, the eigenvalues λ_k measure the roughness (Dirichlet energy) of the corresponding eigenfunctions on the manifold sampled by the data, much like the eigenvalues of the Laplacian on a periodic domain are proportional to squared wavenumbers of Fourier modes. Eigenfunctions corresponding to small eigenvalues are therefore smooth functions that can

be robustly extended to previously unseen test data. In what follows, we order all eigenfunctions in order of increasing λ_k . Note that, unlike in EOF analysis, the eigenvalues here are not related to the variance explained by spatiotemporal reconstruction associated with the corresponding eigenfunctions. In practice, however, the small- λ_k eigenfunctions do tend to be the ones explaining large variance.

By investigating some of the features of corresponding spatially reconstructed modes (Székely et al. 2016a), it can be possible to identify a particular mode ϕ_{i^*} to serve as a forecast observable f (e.g., as explained further in section 3, the twelfth eigenmode is identified as being one of the two MJO modes of interest, and thus $i^* = 12$). This mode is then used to define the restrictions onto the training data of the forecast observable $\hat{f}$, and, by consequence, the time-shifted observable $\widehat{U_\tau f}$.

For concreteness, the full, untruncated training period can be represented as $T = \{t_j\}_{j=1}^s$, where s is the total number of samples, and the sampling times are uniformly separated by some interval Δt . The restricted forecast observable function $\hat{f}$ is then defined by

$$\hat{f}(t) = \phi_{i^*}(t), \quad t \in \{t_j\}_{j=q}^s,$$

and the corresponding time-shifted functions $\widehat{U_\tau f}$ become

$$\widehat{U_\tau f}(t) = \hat{f}(t + \tau), \quad t \in \{t_j\}_{j=q}^{s-\tau}.$$

The truncation of the last τ samples in the above expression is due simply to the fact that $\hat{f}(t + \tau)$ is not defined when $t + \tau > s$. However, as mentioned in the final remark of section 2d, it is often useful in an abstract setting to simply let T denote all of the different truncations of the original training period. With this convention, we consider both $\hat{f}$ and $\widehat{U_\tau f}$ as maps from T to Y .

f. Choice of kernel

The NLSA-derived forecast observable f is built from a kernel that, insofar as it assesses local similarity between points, is said to induce a geometry on the data (Berry and Sauer 2016). If we want consistent forecast results, the kernel that is chosen as weights in the weighted ensemble should rely on a similar, if not exactly identical geometry. The simplest way to do so is by using the same kernel form as used in the definition of the forecast observable f ; this choice is also theoretically motivated from the properties of bandlimited observables in reproducing kernel Hilbert spaces (Zhao and Giannakis 2016). For a given parameter ε_ℓ , that kernel is given, as in Giannakis and Majda (2011, 2012, 2013, 2014), Giannakis et al. (2012), Tung et al. (2014), and Székely et al. (2016a,b), by

$$K_{\varepsilon_t}(x_i, x_j) = \exp \left[-\frac{\|x_i - x_j\|^2}{\varepsilon_t \zeta(x_i) \zeta(x_j)} \right], \quad (4)$$

where $\zeta: X \rightarrow \mathbb{R}$ is a bandwidth function, to be specified momentarily, that allows the kernel to modify the size of its influence based on the inputs x_i and x_j . Note that if the denominator in the exponent of the kernel is large, then the inputs x_i and x_j are assigned a high similarity score, whereas this score is small if the denominator is small. This fact is relevant not just to the extraction of NLSA indices, but to the analog weightings as well. In particular, kernels described as above have smaller discriminating power in regions of the phase space where ζ is large and thus will produce indices that do not as aggressively separate modes corresponding to high ζ . As for the effect on weights, analogs $x(t)$ for which $\zeta[x(t)]$ is large are weighted more heavily in the KAF ensemble than they otherwise would be.

The choice of ζ is thus determined by considerations of where in phase space we wish to relax the kernel's discriminating power and which analogs we deem worthy of increased weighting. One intuitive idea is that data corresponding to more energetic activity are relatively more rare and yet more dynamically important and therefore should be treated more generously in both index extraction and analog forecasting steps. One way of measuring activity is through the magnitude of the time tendency of the data, which we refer to as phase speed. In NLSA, if a_i is the hidden state corresponding to the data x_i , we approximate the phase speed through the finite-difference formula $\zeta(x_i) = \|x_i - \pi_X \Phi_{-\Delta t}(a_i)\|$. In practice, since we are dealing with just one realization of the dynamical flow at a time, this expression becomes $\zeta[x(t)] = \|x(t) - x(t - \Delta t)\|$. In this manner, states with large phase velocity are assigned large kernel values. This specific choice of ζ also has a geometrical interpretation as a conformal change (local scaling) in the geometry of the data (Giannakis 2015).

Besides the bandwidth function ζ , a key feature of the NLSA kernel in (4) is that it operates on the time-lagged sequences x_i , as opposed to the individual snapshots z_i . For certain classes of dynamical systems (viz., systems with pure point spectra), it can be shown that, as q increases, the associated kernel eigenfunctions ϕ_k at a given eigenvalue λ_k converge to doubly degenerate periodic pairs with a frequency ω_k and span the same subspace as eigenfunctions of the Koopman operator U_τ of the dynamical system (see section 2a) with eigenvalues $e^{\pm i\omega_k \tau}$ (Giannakis 2016). Thus, in this special case, the eigenfunctions from NLSA are periodic observables (even if the underlying dynamical flow Φ_t is aperiodic), evolving at frequencies intrinsic to the dynamical system. Away from this idealized limit, the eigenfunctions from

NLSA are expected to retain at least some of the desirable properties of Koopman eigenfunctions, including high temporal coherence and physical interpretability, which have been observed in a number of applications (e.g., Giannakis and Majda 2011, 2012, 2013, 2014; Giannakis et al. 2012; Tung et al. 2014; Székely et al. 2016a,b). Note that the effect of time-lagged embedding in the kernel does not, in general, correspond to bandpass filtering at the frequency bandwidth $1/(q\Delta t)$. For instance, in Giannakis et al. (2012), Tung et al. (2014), and Székely et al. (2016a,b), and section 4 ahead, the embedding window $q\Delta t$ has intraseasonal length (64 days), but the NLSA spectrum contains eigenfunctions spanning interannual to diurnal time scales.

g. Out-of-sample extension

The efficacy of the produced forecast F_τ is assessed in a testing period $\bar{T}$ (disjoint from the training period T) with $\bar{s}$ sampling times. One difficulty in doing so is that it requires knowledge of what the “true” values of f , and therefore also those of $U_\tau f$, are over $\bar{T}$. In other words, we must specify $f(\bar{T})$. One way of doing this is to recompute the NLSA eigenmodes over a period that spans both the training and test periods. However, with this choice, the measured forecast errors are due to both intrinsic loss of skill and the fact that the eigenfunctions computed from the union of the training and test data may have differences from those computed from the training data alone (in particular, the MJO and BSISO modes of interest come as nearly degenerate pairs and are therefore arbitrary up to a phase rotation). In addition, recomputing the eigenfunctions for different choices of test data can be unnecessarily expensive. When the testing period is short relative to the training period, it is possible to approximate the full NLSA reconstruction with a so-called out-of-sample extension that is easier to compute.

Out-of-sample extension relies on the fact that we have at our disposal an already carefully constructed function F_τ that can be used for monitoring in addition to forecasting. That is to say, in addition to $F_\tau(t)$ providing forecasts of f at future times $t + \tau$ for $\tau > 0$, it also provides a value for $\tau = 0$. Because $F_\tau(t)$ is smooth in τ , we can say that $F_0(t)$ provides a “forecast” of f at the present time t , or, alternatively, a “nowcast.” As the length of the training period grows, $F_0(t)$ approaches the value $f(t)$, for arbitrary t . We therefore define the “true” signal $\bar{f}$ over $\bar{T}$ to be

$$\bar{f}(t) = F_0(t) \quad \text{for } t \in \bar{T}$$

and say that $\bar{f}$ is the out-of-sample extension of $\hat{f}$. Note that the analog of this approach in the context of EOF analysis is computing EOFs from the training data and

then determining the true values of the corresponding principal components in the testing period by projection of the test data onto those EOFs.

h. Error assessment

Having defined the true signal $\bar{f}$ over a testing period $\bar{T}$, various ways of assessing error can be performed by examining the restriction of the residual error $r_\tau = U_\tau f - F_\tau$ to $\bar{T}$, denoted by $\bar{r}_\tau$. In the case of seasonally dependent phenomena such as the MJO and BSISO, it is more meaningful to perform error assessments over subsets $\tilde{T}$ of the test period where these phenomena are predominantly active. Specifically, in the case of the MJO, we compute skill metrics over the portion of $\bar{T}$ excluding June–August (JJA). For the BSISO, this is done over the portion of $\bar{T}$ excluding December–February (DJF).

The first skill metric we use is the RMSE, defined for a given lead time τ by

$$\text{RMSE}(\tau) = \sqrt{\frac{1}{\tilde{s}} \sum_{t \in \tilde{T}} |\bar{r}_\tau(t)|^2},$$

where $\tilde{s}$ is the number of samples in $\tilde{T}$.

Another valuable and commonly used measure of error is Pearson correlation, also known as pattern correlation,

$$\text{PC}(\tau) = \frac{1}{\tilde{s}} \sum_{t \in \tilde{T}} \frac{[F_\tau(t) - \mu_{F_\tau}][\bar{U}_\tau \bar{f}(t) - \mu_{\bar{U}_\tau \bar{f}}]}{\sigma_{F_\tau} \sigma_{\bar{U}_\tau \bar{f}}},$$

where μ_G and σ_G are means and covariances defined, for any function $G: T \rightarrow \mathbb{R}$, by

$$\mu_G = \frac{1}{\tilde{s}} \sum_{t \in \tilde{T}} G(t) \quad \text{and} \quad \sigma_G^2 = \frac{1}{\tilde{s}} \sum_{t \in \tilde{T}} [G(t) - \mu_G]^2.$$

Because of the recentering and normalization steps in its formula, pattern correlation gives a good measure of how well F_τ tracks $\bar{U}_\tau \bar{f}$ that does not depend on amplitude. Therefore, instead of focusing on whether F_τ and $\bar{U}_\tau \bar{f}$ have the same value, the PC score highlights the degree to which the two signals move together in the same direction at the same time, which is useful in actual forecasting. However, amplitude fidelity can also be important, and thus RMSE is also considered in the following results and discussion.

3. Application to global brightness temperature data

a. NLSA-derived MJO and BSISO modes

Our primary object of study is infrared brightness temperature data recorded under the CLAUS project

over 26 yr, from 1 July 1983 to 30 June 2009, and used in [Giannakis et al. \(2012\)](#), [Tung et al. \(2014\)](#), and [Székely et al. \(2016a,b\)](#). This dataset is often considered to be a reasonable proxy for convective activity in the tropics, with positive T_b anomalies associated with reduced cloudiness and negative anomalies with increased cloudiness. The sampling is confined to the tropical belt between 15°S to 15°N, with a resolution of 1° in both longitude and latitude. Each 2D sample thus consists of $n_{\text{lon}} = 360$ longitudinal and $n_{\text{lat}} = 31$ latitudinal grid points. Each T_b observation $z(t)$ is arranged as a vector, with length $n = n_{\text{lon}} \times n_{\text{lat}} = 11\,160$. These observations are collected at an interval of $\Delta t = 3$ h, for a total sample size of $s_{\text{total}} = 75\,796$ over the 26 yr of the CLAUS record.

To examine the effects of differently sized training datasets, our study in this section is broken into a long training period set T_{long} and a short training period T_{short} . The period T_{long} spans 23 yr from 1 July 1983 to 30 June 2006, thus consisting of $s_{\text{long}} = 67\,208$ samples. The period T_{short} , meanwhile, covers the 9 yr between 1 July 1983 and 30 June 1992, for a sample size of $s_{\text{short}} = 26\,304$. Note that no preprocessing, such as bandpass filtering, seasonal partitioning, or equatorial averaging, is performed.

The testing period $\bar{T}_{\text{long}}$ corresponding to the long training period T_{long} covers the 3 yr from 1 July 2006 to 30 June 2009. The reason for this particular choice of training and testing periods is that we seek to utilize as long a training period as possible while keeping the testing period long enough for forecasts to be verified. The testing period $\bar{T}_{\text{short}}$ associated with the shorter training period, meanwhile, covers the 3 yr between 1 July 1992 and 30 June 1995. The rationale for this second choice of testing period is that it contains the intensive observing period (IOP) of the Tropical Ocean Global Atmosphere Coupled Ocean–Atmosphere Response Experiment (TOGA COARE), which took place between 1 November 1992 and 28 February 1993 ([Webster and Lucas 1992](#)). Our particular choice of T_{short} and $\bar{T}_{\text{short}}$ is thus designed to not only test the robustness of the forecasting algorithm in the face of a shortened training period, but to also benefit from pre-existing research about this well-studied period, during which two large MJO events were observed.

For any given period T consisting of s samples, the observational data are stored in a matrix, denoted by $z(T)$, of dimensions $n \times s$. The time-lagged embedding procedure described in [section 2d](#) is applied to construct a separate matrix $x(T)$, of size nq by $s - q + 1$, where q is the length of embedding window. Following [Giannakis et al. \(2012\)](#), [Tung et al. \(2014\)](#), and [Székely et al. \(2016a,b\)](#), the embedding length is chosen to be long enough to capture key features of intraseasonal

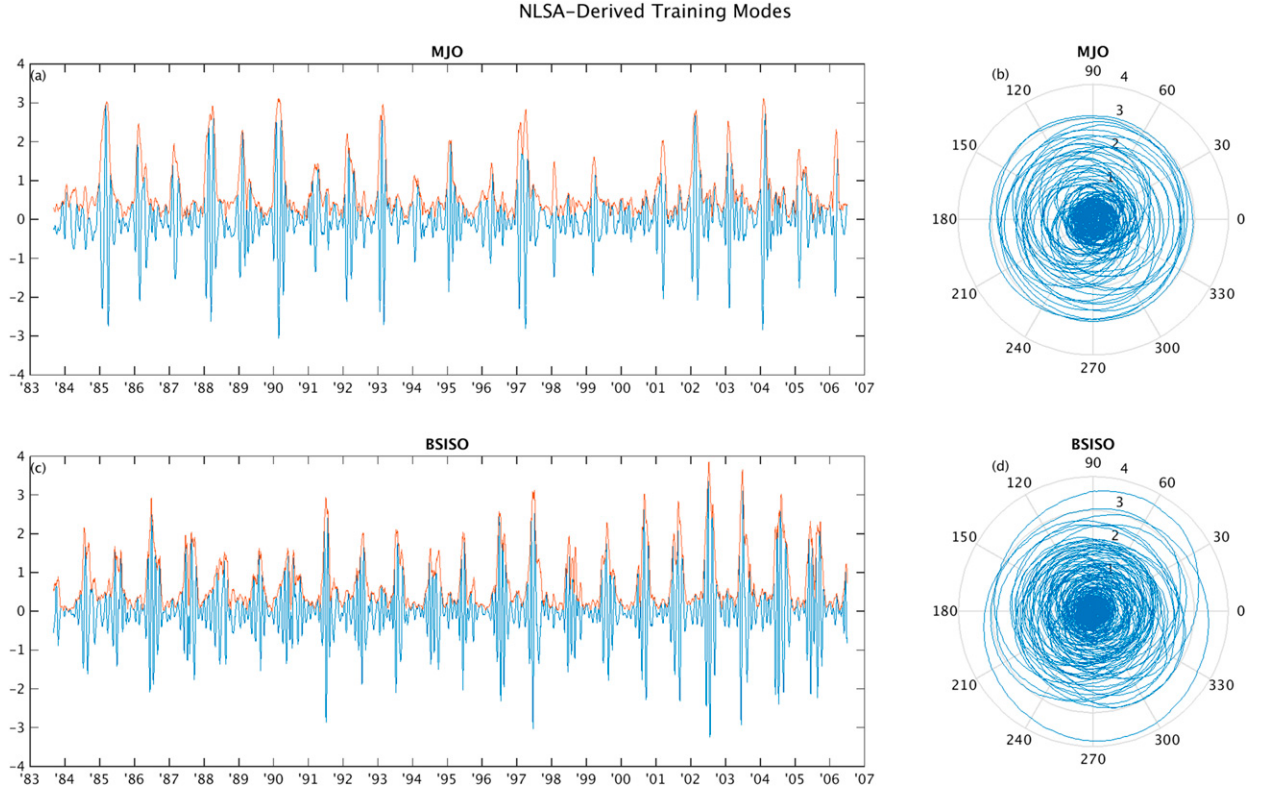


FIG. 1. (a) The time series of one of the MJO modes (blue) and the amplitude of both MJO modes (orange), between 1 Jul 1983 and 30 Jun 2006. (b) Phase plot with the MJO mode shown in (a) on the x axis, and the other mode as the y axis. (c),(d) As in (a) and (b), respectively, but for the BSISO modes.

activity, which occur on less than a 2-month time scale, and short enough compared to the length of the training time series. As such, a lag of 64 days is chosen, which, given the 3-h sampling interval, means $q = 512$. This further means that the embedded vectors $x(t)$ populate a space of dimension $N = nq \approx 2.3 \times 10^7$. As noted in section 2d, NLSA and KAF are well-suited techniques for data analysis and modeling in such high-dimensional spaces, as their performance is governed by the intrinsic dimension of the dataset $x(T)$, which is generally significantly smaller than N .

The forecast observables for T_{long} are extracted from $x(T_{\text{long}})$ via NLSA and stored in time series vectors $f(T_{\text{long}})$ of length $s_{\text{long}} - q + 1$. The bandwidth parameter of the NLSA kernel is set to $\varepsilon = 2$ and the nearest neighbors parameter to $k_{nn} = 5000$ (corresponding to $\sim 10\%$ of the dataset); these are the same parameter values used by Giannakis et al. (2012), Tung et al. (2014), and Székely et al. (2016a,b). Using these parameters, we compute the first 50 eigenmodes, $\phi_i(T_{\text{long}})$, for $i = 1, \dots, 50$, of the discrete Laplace–Beltrami operator $\hat{\mathbf{L}}$. The two MJO modes of interest are the twelfth and fifteenth leading eigenmodes of $\hat{\mathbf{L}}$, while the two BSISO modes of interest are the twenty-first and

twenty-second leading eigenmodes. Given that the two MJO modes are very similar to each other (save for a quarter-phase time shift), and thus produce forecasts of nearly identical quality, we show results for just one of the two MJO modes, setting $f_{\text{MJO}}(T_{\text{long}}) = \phi_{12}(T_{\text{long}})$. This is also the case for the BSISO, for which we set $f_{\text{BSISO}} = \phi_{21}(T_{\text{long}})$. These two representative forecast observables are plotted in Fig. 1. The eigenfunction time series and associated T_b phase composites for these observables are plotted in Figs. 1 and 2.

Selecting which of the NLSA eigenmodes correspond to the MJO or BSISO is done through a one-time qualitative assessment of their associated spatially reconstructed patterns. In particular, the MJO forecast observables were chosen as those corresponding to eastward-propagating wave trains of enhanced and suppressed convection during the boreal winter, initiating over the Indian Ocean, traversing the Maritime Continent and western Pacific warm pool, and eventually decaying in the central Pacific Ocean near the date line. Meanwhile, the BSISO forecast observable was chosen as the one matching boreal summer convective activity initiating in the Indian Ocean and moving north toward the Indian subcontinent. We refer the reader to Székely et al. (2016a,b) for additional

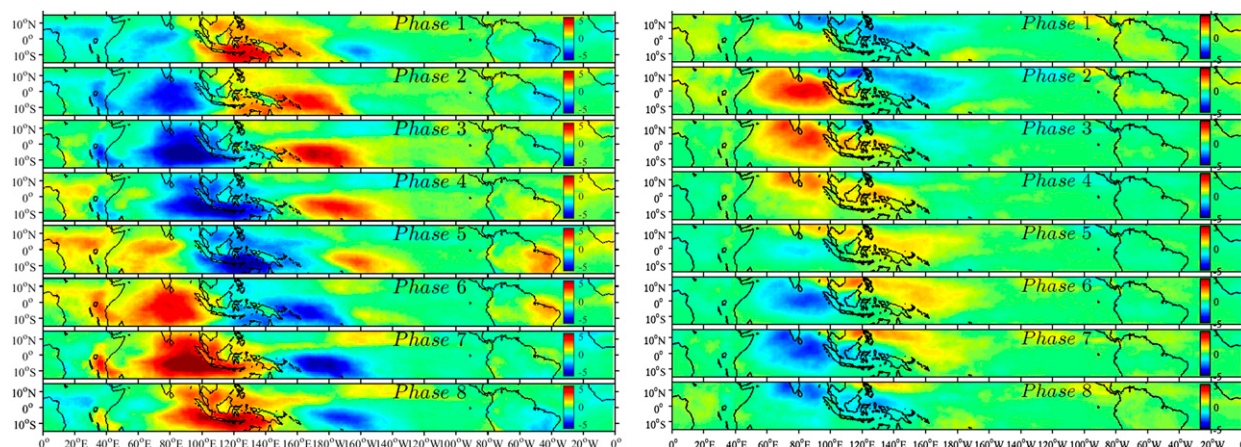


FIG. 2. Phase composites of T_b anomalies (K) for the NLSA-derived (left) MJO and (right) BSISO modes. The composites were created by averaging spatiotemporal T_b reconstructions for these modes over eight temporally sequential phases in the corresponding two-dimensional modal spaces, subject to the requirement that the modal amplitudes exceed one standard deviation of the amplitude time series. Further information on these composites can be found in Székely et al. (2016a,b).

discussions on the properties of these modes, including coarse-grained predictability properties and initiation and termination statistics. Besides the MJO and BSISO modes studied here, the NLSA spectrum recovered from CLAU T_b data contains a multiscale hierarchy of modes (represented by distinct orthogonal eigenfunctions, as described in section 2e), including representations of ENSO, the annual cycle and its harmonics, and the diurnal cycle Székely et al. (2016a). As will be discussed in section 6a, the extracted MJO and BSISO modes from the long training period are qualitatively robust to the embedding length so long as it exceeds ≈ 30 days.

To retain the real-time applicability of kernel analog forecasting, the forecast observables corresponding to the short training period T_{short} are obtained using data restricted to the short training period. That is to say, although it is tempting to simply use the first 9 yr of $f_{\text{MJO}}(T_{\text{long}})$ as $f_{\text{MJO}}(T_{\text{short}})$, it is important to note that each entry of $f_{\text{MJO}}(T_{\text{long}})$ depends on data from the entirety of T_{long} . This erroneous truncation would then correspond to a forecaster living in 1992 using information from the future. We must instead recompute the eigenmodes of the NLSA kernel restricted to just T_{short} . Although several of the resulting eigenmodes are not as clear, the MJO modes are still visible, and, like before, we set $\mathbf{f}_{\text{MJO}}(T_{\text{short}})$ to one of the two modes. The BSISO modes over this shorter period, however, are degraded to the point that no attempt is made to forecast them. The higher sensitivity of the BSISO modes to the training time series length is consistent with the fact that they lie farther down in the NLSA spectrum than the MJO modes. That is, as stated in section 2e, the ordering of the NLSA eigenfunctions is in order of increasing “roughness” on the data manifold (as measured by the

corresponding eigenvalues), and functions with higher roughness are generally expected to be more sensitive to sampling errors. In section 5b, we will see that adequate BSISO modes can be recovered for training time series as short as 17 yr.

b. Comparison with the RMM index

That there is no unambiguously correct measure of intraseasonal variability is part of the rationale for constructing alternative indices based on new techniques. Nevertheless, the RMM index (Wheeler and Hendon 2004) is a commonly accepted and used measure of the MJO. Therefore, for completeness, we include in our analysis correlations between the RMM and our proposed NLSA-derived indices. Figure 3 illustrates what the amplitude of the RMM’s two modes looks like, both before and after performing a 64-day-running-mean smoothing. The correlation of the NLSA MJO amplitude with that of the raw RMM index is small at 0.20 but becomes a more significant 0.46 after the RMM is smoothed. This correlation does not change significantly upon restriction to the DJF period, when the MJO is most active. Unlike the MJO, the year-round BSISO data do not at all correlate with the raw RMM amplitude and, moreover, barely exhibit any increase after smoothing. Upon restriction to the JJA active BSISO period, however, the correlation with the smoothed RMM is also found to be 0.46.

Overall, these results indicate that, while the two sets of modes are by no means equivalent, the NLSA-based MJO and BSISO modes are related to the RMM modes when restricted on the boreal winter and summer periods, respectively, and smoothed to remove high-frequency spectral content. In particular, that NLSA represents the two dominant ISOs through distinct modes with moderately

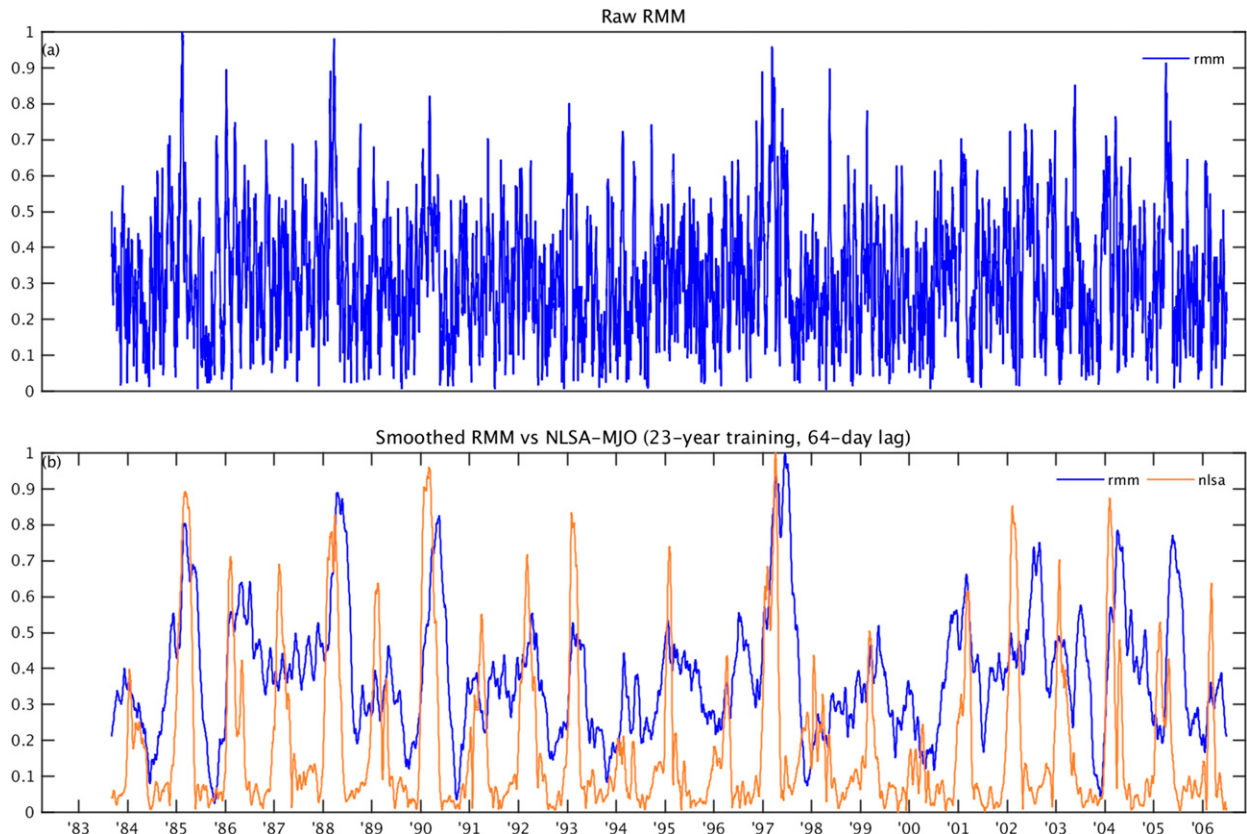


FIG. 3. Visualization of the RMM amplitude, as discussed in section 3b: (a) amplitude of the Wheeler–Hendon RMM modes within the period from 1 Jul 1983 to 30 Jun 2006. (b) The 64-day moving average of the RMM amplitude and the amplitude of the NLSA-derived MJO modes.

narrowband spectra likely plays an important role in the higher predictability of these modes compared to the RMM modes (which represent the two distinct ISOs as a single pair of modes). We will discuss the sensitivity of the correlation results presented in this section on the NLSA embedding window and the RMM smoothing window in section 5a.

4. Hindcast results

In this section, we present the results of applying the KAF method to the NLSA-based MJO and BSISO modes described in section 3. The main result is that the pattern correlations of forecasts remain above 0.6 for 50 days for both the MJO and BSISO when 23 yr of training data are used. Another result is that the RMSE stays below one standard deviation of the historical variability. Other results, such as the relatively worse predictability at the beginning of an ISO event than toward the end and the specifics of individual years, are also presented.

a. MJO 2006–09

Figure 4 shows running forecasts and corresponding RMSE and PC scores for lead times of up to 60 days,

corresponding to the application of KAF in predicting one of the two MJO modes during the 2006–09 testing period, using the 1983–2006 training period. The relatively large amplitude of the monitored MJO signal around January 2007 and January 2008 corresponds to the fact that the winters of 2007 and 2008 contained significant MJO activity, and the relatively small amplitude around January 2009 corresponds to a season of weak MJO in 2009. A key question to ask of any forecasting method is if it can capture these periods of relatively increased and decreased activity. From a strictly visual inspection of Figs. 4a–d, this can be said to be true of KAF with 15- and 30-day lead times and to be false in the case of a 60-day lead time. That the ability to qualitatively discern periods of greater MJO activity is only lost well after 30 days is already an improvement over some of the existing methods discussed in the introduction. One qualitative feature that deteriorates faster than the general ability to discern activity is the ability to detect initial activity. More specifically, the forecast fails to capture the full amplitude of the first spike of the MJO season, occurring in December 2007 and December 2008. The difficulty of

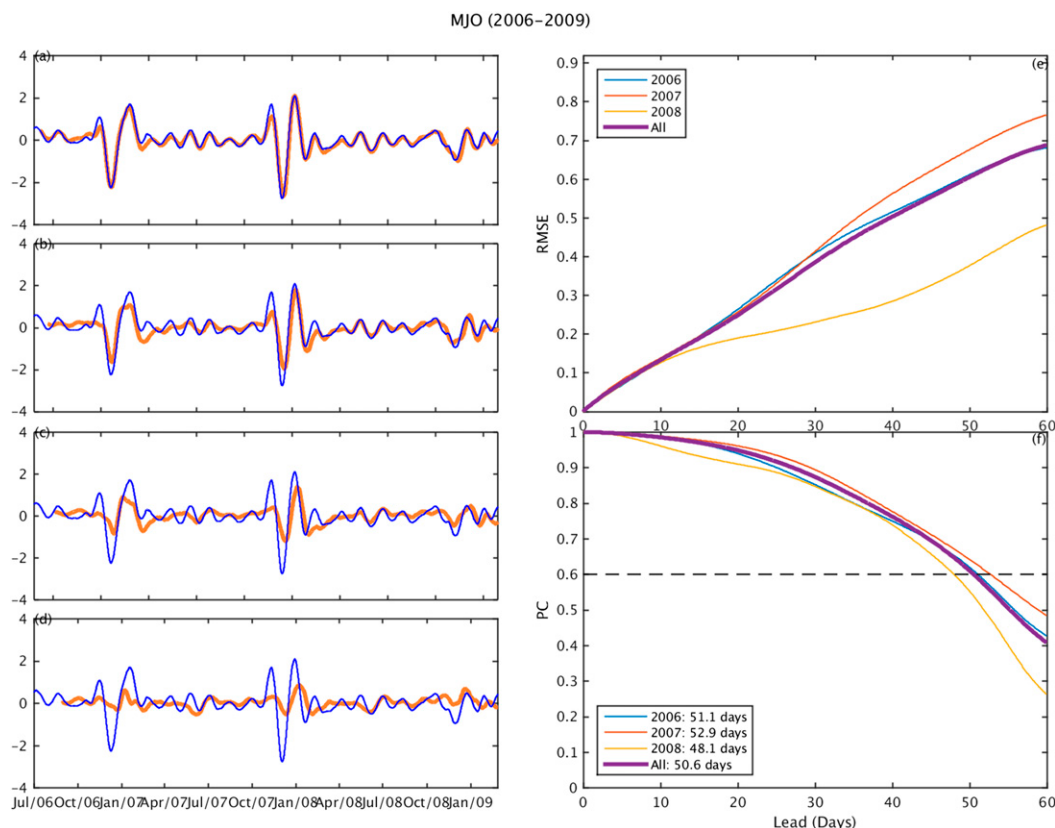


FIG. 4. Kernel analog forecasting of the NLSA-based MJO mode over the 2006–09 testing period using the 1983–2006 training period, as discussed in section 4a. (a)–(d) Running forecasts (orange) with lead times of 15, 30, 45, and 60 days, respectively, along with the true signal (blue). (e) RMSE and (f) 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 JJA, as discussed in section 2h.

predicting initiations, however, is a challenge that is not unique to this method.

A quantitative analysis of these results is obtained through RMSE and PC evaluation. A typically used threshold for separating skillful and unskillful forecasts is a PC score of 0.6. As such, the 0.97 and 0.86 scores for the 15- and 30-day lead forecasts in Figs. 4e and 4f reflect the qualitatively good nature of those forecasts, whereas the 0.41 pattern correlation of the 60-day lead reflects the natural decrease in forecast skill at long lead times.

The RMSE and PC of forecasts with lead times ranging from 0 to 60 days are also displayed. The decrease in pattern correlation score for the entire testing period is very slight up to 20 days. The decrease is slightly more modest between 20 and 50 days of lead time. The skill dips below 0.6, and is thus said to become unskillful, when the lead time exceeds 50 days.

The RMSE and PC plots also show scores over three shorter time periods (July 2006–July 2007, July 2007–June 2008, and July 2008–June 2009), which we refer to

by the year in which the period begins. Although the scores of the 2006 and 2007 periods are similar to those over the entire testing period, the 2008 scores have notable differences. First, the RMSE of the 2008 period diverges from that of the other periods at around $\tau = 10$ days of lead time and stays significantly below the others. Second, the PC of the 2008 period dips faster than that of the other periods after about 40 days of lead time. It may seem paradoxical at first how a period can simultaneously have better RMSE and a worse PC. However, both aspects are explained by the fact that the 2008 period contains less activity than the others; that is, it contains few large peaks, if any. Thus, the relatively small RMSE of the 2008 period is because there is only modest deviation from the mean.

b. BSISO 2006–09

Figure 5 shows the results of KAF applied to predicting one of the BSISO modes during the 2006–09 period, using the same 1983–2006 training data as before. The significant

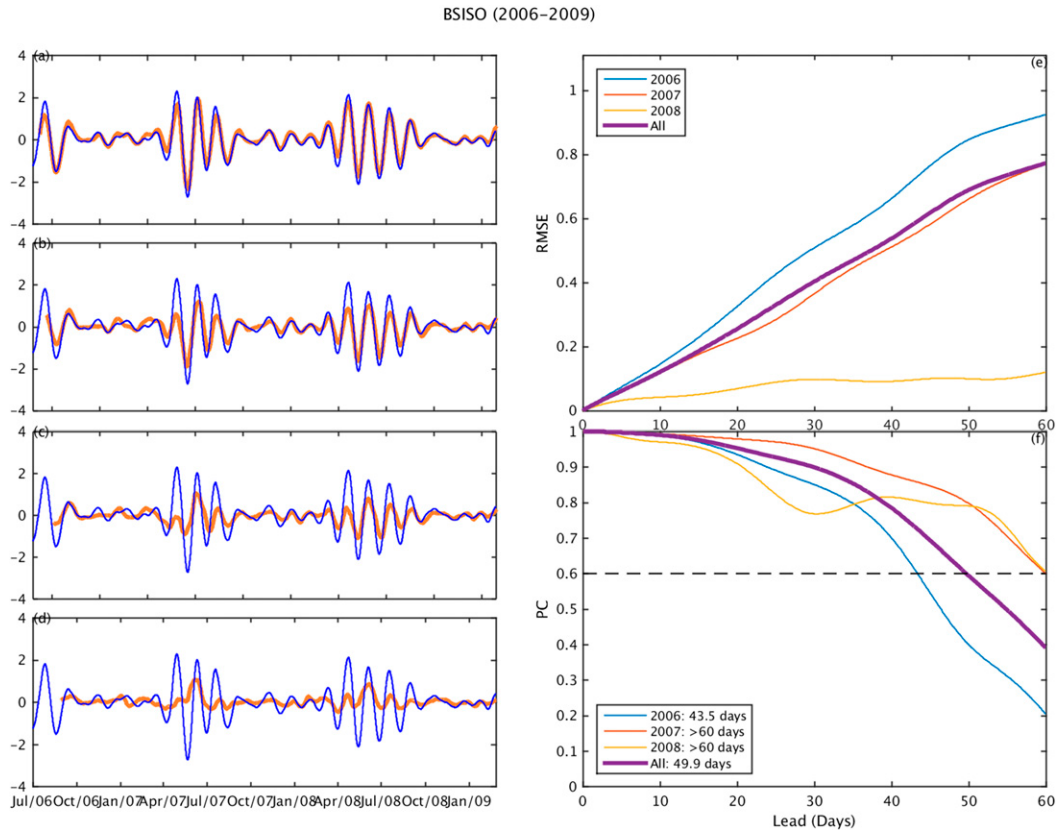


FIG. 5. As in Fig. 4, but for the NLSA-based BSISO mode, and discussed in section 4b. The RMSE and PC skill scores are calculated by excluding DJF, as discussed in section 2h.

BSISO events during this time occurred in the summers of 2007 and 2008, as shown in the plots, and their durations are longer than those of MJO events. As in the previous two cases, the 15- and 30-day-lead-time forecasts are qualitatively good, while the 60-day-lead-time forecast is not. Although the skillful PC score extends slightly farther than for the MJO, not falling below 0.6 until about 50 days, there is more variance, with the 2006 PC score dipping below 0.6 at 45 days, and the 2007 and 2008 scores doing so after 50 days. This may be explained by the seeming property of KAF forecasts to be markedly better when locked onto a regular, oscillatory event. Therefore, the fact that the KAF method produces a slightly better PC score for the BSISO mode is probably due to the longer and more sustained BSISO events.

c. MJO 1992–95

Figure 6 shows the results of applying KAF to predict values of one of the MJO modes during the 1992–95 testing period, using the 1983–92 training period. As stated in section 3, the reason this testing period was chosen is that it contains the well-documented large MJO events that occurred during the TOGA COARE

IOP (Yanai et al. 2000). The plots show that the true signal captures the succession of these two events, as well as a similarly large MJO event in the winter of 1995 and a couple of smaller ones in the winter of 1994. The goals of applying KAF with these testing and training periods are twofold: to determine the effects on the KAF of 1) a shorter training period and 2) a testing period with well-documented MJO events. Note that the shortened training period impacts the KAF skill in two distinct ways: namely, through poorer-quality MJO indices in the NLSA step and fewer available analogs in the forecast step. We will return to this point in section 6.

Qualitatively speaking, many of the results in Fig. 2 are similar to those of Fig. 1: the KAF forecasts perform reasonably well with lead times of 15 and 30 days, but not so well with a lead time of 60 days. That being said, the initial detection of MJO events is more difficult because of the smaller training sample size. Quantitatively, although the overall RMSE scales similarly as in the previous case, there is more variance, with a low RMSE for 1993 and high RMSE for 1995. There is also more variance in the PC scores, as the 1994 forecast remains well above a score of 0.5 beyond 50 days. Most significantly, however, is

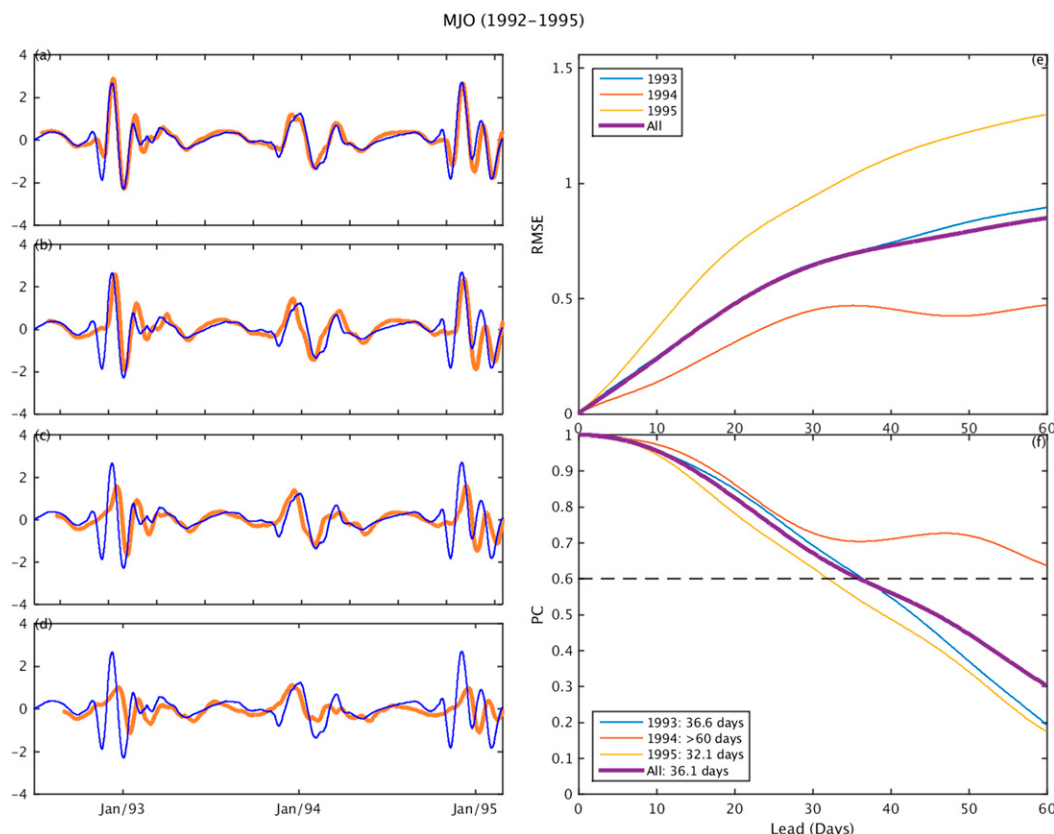


FIG. 6. As in Fig. 4, but for the short 1983–92 training period and the 1992–95 testing period. Discussed in section 4c.

that the overall PC score dips below 0.6 after about 37 days, which is 13 days earlier than in the previous case.

5. Sensitivity analysis

The main factors affecting the robustness of the hindcast results presented in section 4 are the choice of NLSA and KAF parameters, the sampling frequency, and the length of the training and test intervals. The NLSA parameter values, sampling frequency, and training interval influence the properties of the extracted ISO modes (i.e., the truth signal), the KAF parameters affect the predictive skill of the forecast model, and the length of the test interval affects the robustness of the skill scores computed for the given choice of NLSA parameters, training interval, and KAF parameters. In this section, we present assessments of the sensitivity of our hindcast results on these factors, focusing on the influence of the embedding window size (section 5a) and length of the training and test time series (section 5b). To manage the computational cost of this study, we reduce the frequency of the time sampling of the raw CLAUS data from eight times a day to four times a day; this downsampling cuts the overall computational cost of the pairwise kernel

evaluations (which scale quadratically with the number of samples) by a factor of 4. Aside from a moderate decrease in forecast skill (by approximately 10 days of $PC > 0.6$ lead time), the resulting MJO and BSISO modes are qualitatively very similar to those presented in section 3.

a. Sensitivity to lagged embedding window length

The parameters of the combined NLSA–KAF algorithm described in section 2 are the number of embedding lags q (section 2d), the kernel bandwidth ε and nearest neighbor truncation k_{nn} used in the training stage (section 2f), and the bandwidth ε_t and termination threshold δ used in the prediction stage (section 2c). While all of these parameters will generally affect forecast skill, q plays a distinguished role, as it controls the time-scale separation in the recovered eigenfunctions (see section 2f), influencing their physical significance and potential predictability.

Figure 7 displays the truth signals and RMSE and PC scores for the MJO, evaluated using the T_{long} and $\bar{T}_{\text{long}}$ training and testing periods (at the 6-h sampling interval), respectively, and embedding window sizes Δt of 48, 64, 80, and 96 days. All other NLSA and KAF parameters are kept fixed to the same values as in section 3.

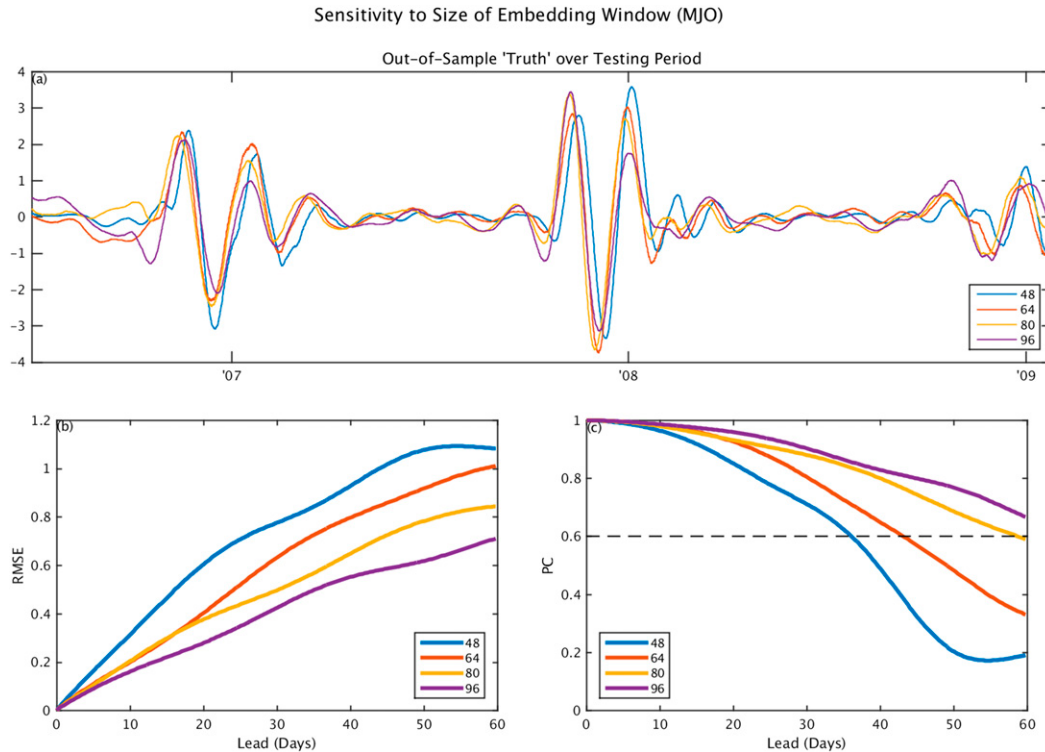


FIG. 7. (a) Out-of-sample extensions of MJO signals to the testing period from 1 Jul 2006 to 30 Jun 2009, with sizes of embedding window varying from 48 to 96 days. (b) RMSE and (c) PC for each experiment (with JJA excluded, as per section 2h).

As shown in Fig. 7a, while the MJO modes from all four cases are qualitatively robust, increasing q leads to an increase of temporal coherence; that is, as q increases, the MJO modes acquire the structure of amplitude-modulated oscillators with a distinct intraseasonal carrier frequency and a low-frequency modulating envelope. Correspondingly, there is a monotonic increase of forecast skill as measured by both the RMSE (Fig. 7b) and PC (Fig. 7c) scores, which appears to saturate for the larger q values. In particular, the PC = 0.6 crossing times for the 48-, 64-, 80-, and 96-day windows are approximately 35, 42, 56, and 58 days, respectively. This increase of skill is consistent with the theoretically expected property of the NLSA eigenfunctions to become increasingly temporally coherent (and hence more predictable), with increasing embedding window size. At large q values, this increase of predictability can come at the expense of introducing additional modes that are related to the phenomenon of interest. Here, we observe that, besides the MJO pair shown in Fig. 7, additional MJO-like modes are present in the NLSA spectra for the 80- and 96-day embedding windows. It is possible that, for these embedding window sizes, the “fundamental” MJO modes represent events that undergo at least one full cycle of the oscillation (and such events are more predictable once

they initiate), and the “secondary” MJO modes are harmonics required to capture weak or stalling events. Our nominal choice of a 64-day embedding window is a compromise between a sufficiently long window so that the NLSA eigenfunctions separate the MJO from other modes of tropical variability (in particular, BSISO), and a sufficiently short window so that secondary MJO-like modes are not present in the top part of the NLSA spectrum. Similar behavior also occurs for the BSISO (shown in Fig. 8).

The effect of changing the size of the embedding window is further analyzed in Fig. 9 by computing correlations of the NLSA-derived MJO and BSISO amplitudes with the RMM amplitude for the NLSA embedding windows examined above and different values of (backward) running-average smoothing of the RMM amplitude. For both MJO and BSISO, the RMM smoothing window for maximum correlation is an increasing function of the NLSA embedding window, but that relationship is not proportional and appears to saturate at the larger (80 and 96 day) Δt values examined. In the case of the MJO (Fig. 9a) the maximum correlation between NLSA and RMM is 0.51 and occurs for $\Delta t = 80$ days and a ~ 65 -day RMM smoothing window. As stated in section 3b, the correlation between NLSA–BSISO and RMM

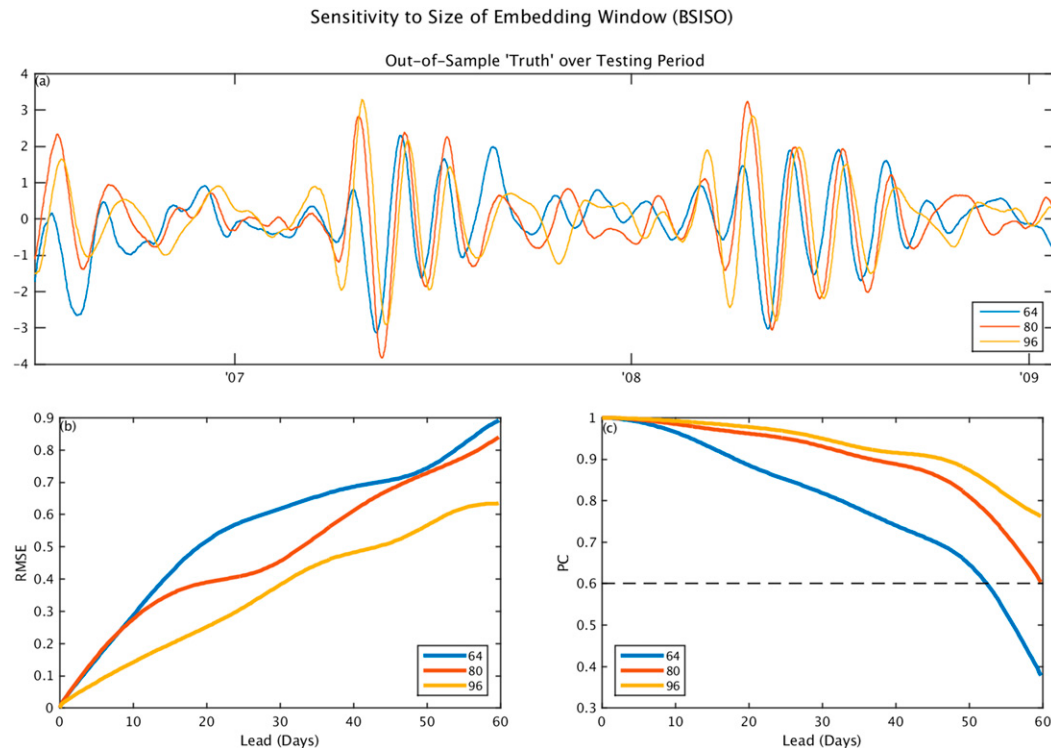


FIG. 8. As in Fig. 7, but for the BSISO mode (and thus excluding DJF from the calculation of the RMSE and PC scores, as per section 2).

(Fig. 9b) is significantly higher when conditioned on JJA. In particular, the highest correlation is 0.53 and occurs for $\Delta t = 96$ days and an ~ 80 -day RMM smoothing window.

b. Sensitivity to training size

As with any statistical method, the size of training dataset is a critical ingredient of KAF, as it influences the quality of the extracted eigenfunctions (the truth signal) and the availability of adequate analogs for prediction given previously unseen initial data. The size of the test dataset is also important, as it influences the robustness of skill scores. In particular, besides being important for accurately assessing the performance of the method in a hindcast setting, the availability of accurate skill scores is also important in operational forecast settings, where the parameters of the method would be tuned in a validation stage (analogous to the hindcasts performed here) prior to its deployment in actual forecasts.

In this section, we examine the robustness of the forecast skill results presented in section 3 by comparing PC and RMSE scores from multiple hindcast experiments of the MJO and BSISO with different sizes of training and test data. In all cases, we work with 6-hourly sampled data, a 64-day embedding window, and the

same ε , ε_ℓ , and δ parameters as in section 3. Moreover, the start of the training period is fixed at 1 July 1983 for all experiments, and the end of the training period is 30 June of various years from 1992 to 2006. We vary the number of nearest neighbors k_m so that it is roughly equal to 15% of the training data in each experiment. In addition, we set the test period to either 1 July 2006 to 30 June 2009, or to the largest test interval available in each case (i.e., the interval starting immediately after the end of the training period and ending on 30 June 2009). Note that we do not individually tune the NLSA and KAF parameters in each experiment in order to optimize skill, and this likely increases the spread of the computed scores but reduces the risk of overestimating skill. Also, in the case of BSISO we do not show hindcast results for training periods shorter than 1983–2000, as we find that for those periods the BSISO modes are significantly degraded.

The RMSE and PC scores for the MJO and BSISO hindcasts from these experiments are shown in Figs. 10 and 11, respectively. There it is evident that the scores have noticeable spread (particularly at large leads); they are largely consistent with those in Figs. 4 and 5, respectively. In particular, even for the shortest training period examined in each case, both MJO and BSISO

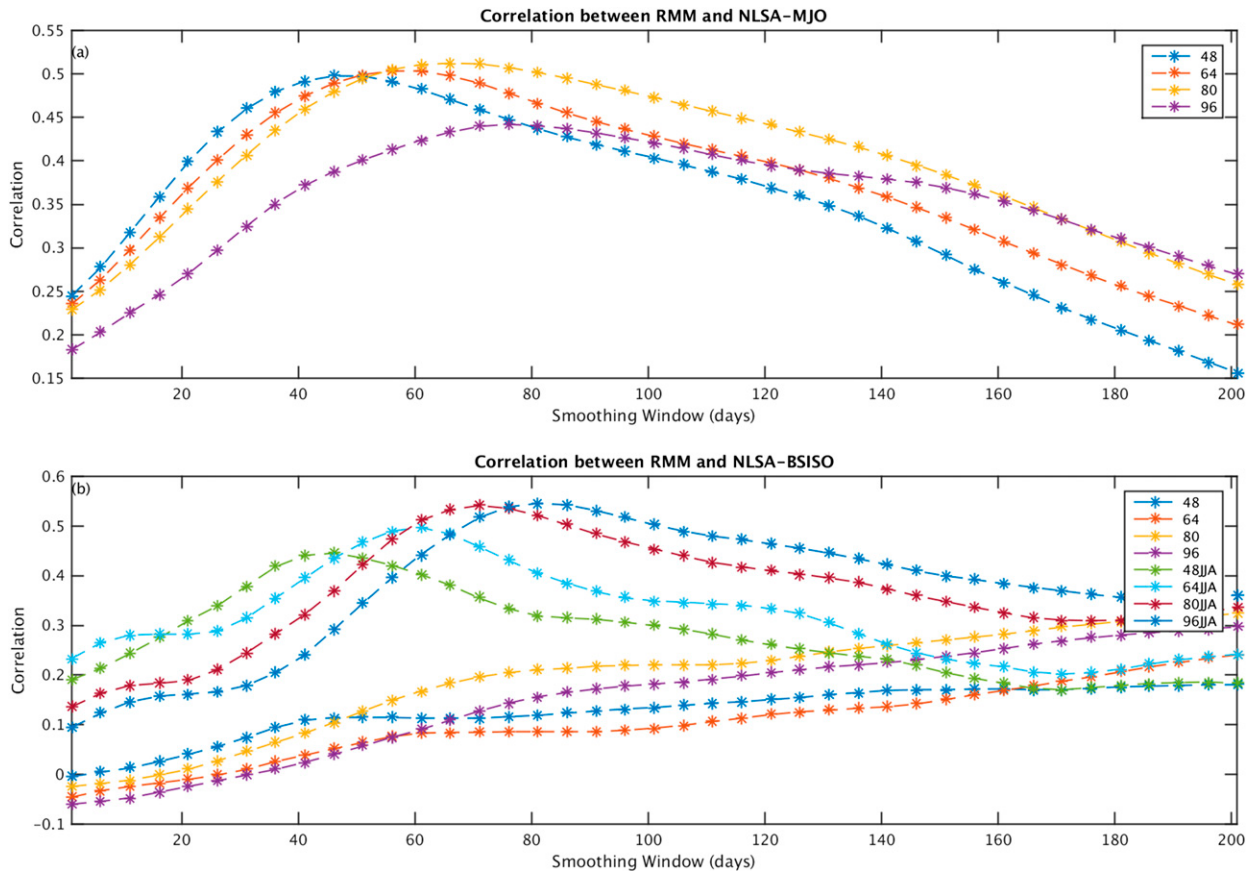


FIG. 9. Correlations between NLSA-derived modes and RMM, for lengths of RMM averaging. Discussed in both [sections 3b](#) and [5a](#).

have PC scores greater than 0.6 out to ~ 40 -day leads. Moreover, the skill scores computed for 2006–09 and the longest possible test periods are mutually consistent. This suggests that in an operational setting it should be possible to tune the parameters of the method using a modest, ~ 3 -yr, validation period and utilizing the rest of the available samples to obtain high-quality eigenfunctions and analogs. Note that the RMSE and PC scores in [Figs. 10](#) and [11](#) are not monotonic functions of the training or test period size. This is likely due to both lack of optimality of NLSA/KAF parameters and variance of the skill scores (particularly at large leads).

[Figure 12](#) shows more precisely the increase in training data dependency with respect to forecast time by plotting ratio of the standard deviation of the skill scores at given lead times, across all available experiments, to the mean amount of skill deterioration by that time. Roughly speaking, for each of the plots in [Figs. 10](#) and [11](#), we are plotting the ratio of the vertical spread (as measured by the standard deviation) to the mean increase in RMSE (or mean decrease in PC). This metric provides a way of approximating how much of the change in skill is due to choice of training data. Overall,

[Fig. 12](#) shows that this ratio remains bounded between 0% and 23%.

6. Discussion

Although it is tempting to directly compare the PC scores of KAF to those of other methods, it is important to note that it is often not the case that the same MJO and BSISO definitions are used in different methods. While KAF could be used to predict other commonly defined ISO indices (e.g., the RMM index), investing too much in making such comparisons risks missing the important point that all indices are representations of physical phenomena that depend on a choice of data analysis technique and that much of the appeal of KAF is that it incorporates the same class of kernel operators for both ISO definition and prediction in a unified scheme. Instead, a true comparison with other methods would require assessing how each method fares in predicting physical observables of interest (e.g., precipitation on intraseasonal time scales) conditioned on the predicted values of the indices. Such comparisons are outside the scope of this work, but we believe that the

Sensitivity to Size of Training Data (MJO)

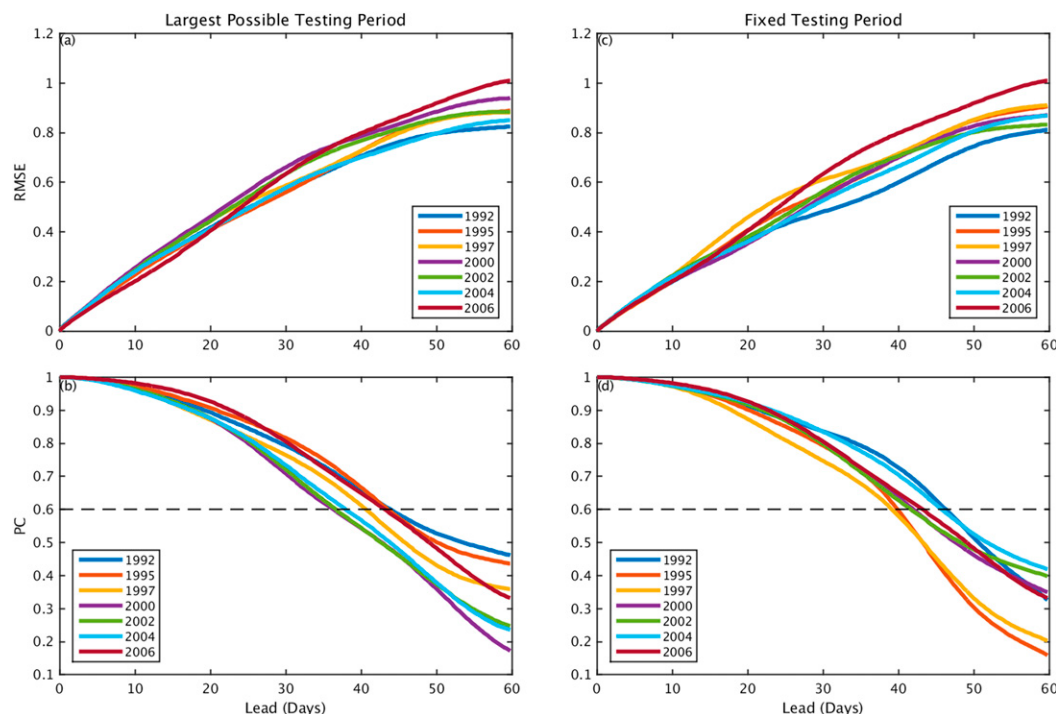


FIG. 10. RMSE and PC of forecasts created by training data spanning the range from 1 Jul 1983 to 30 Jun of the year specified by the legend, excluding DJF as per [section 2h](#). (a) RMSE and (c) PC over the testing range from 1 Jul of the specified year up to 30 Jun 2009; (b) RMSE and (d) PC over a fixed range from 1 Jul 2006 to 30 Jun 2009. See [section 5b](#).

high predictability and coherent spatiotemporal structure of the NLSA-based ISO modes are encouraging properties for future predictability studies of physical observables.

Despite the difficulty of making objective comparisons to other methods, it is nevertheless valuable to place our results in context with existing ISO forecasting techniques. For instance, when efforts to forecast the MJO first began in the 1990s, global climate models (GCMs) were frequently unable to achieve more than 6 days of predictability ([Chen and Alpert 1990](#); [Lau and Chang 1992](#); [Slingo et al. 1996](#); [Jones et al. 2000](#); [Hendon et al. 2000](#)). Much of the difficulty of these early GCMs stemmed from their inability to sufficiently represent organized convection, which eventually led to a shift in research focus to empirical methods that are not affected by model error in GCMs. That said, the switch in focus to empirical methods did not yield immediate benefits, as many of the earliest attempts, such as applying principal oscillation pattern (POP) analysis to 200-hPa equatorial velocity potentials ([von Storch and Jinsong 1990](#)), still did not attain predictability beyond 1 week.

Multiweek MJO predictability with empirical methods was eventually attained through improvements to both

ISO definition and forecasting methods. For instance, multiple-field EOF analysis that includes outgoing long-wave radiation data, as first recommended by [Kousky and Kayano \(1993\)](#), was performed by [Waliser et al. \(1999\)](#) to obtain MJO predictability for up to 15–20 days. Time-lagged embedding, meanwhile, was used in a singular spectrum analysis of similar data by [Mo \(2001\)](#) to consistently obtain 20-day predictability. That KAF uses elements of these techniques, in particular the use of cloudiness data (in this case T_b) and time-lagged embedding, can help explain why we obtained limits of predictability that exceed these previous results. More recently, methods based on physics-constrained, low-order, nonlinear stochastic models ([Chen and Majda 2015b](#); [Kondrashov et al. 2013](#)) have demonstrated 30–40-day skill for the RMM index, though these results are not for real-time forecasts. Such low-order stochastic models have been used to establish 40–50-day internal predictability of the NLSA-based MJO [Chen et al. \(2014\)](#) and BSISO [Chen and Majda \(2015a\)](#) indices, respectively, over the 1983–2006 training interval used in this work. The KAF results presented here show that these predictability limits can also be attained without the use of such stochastic models.

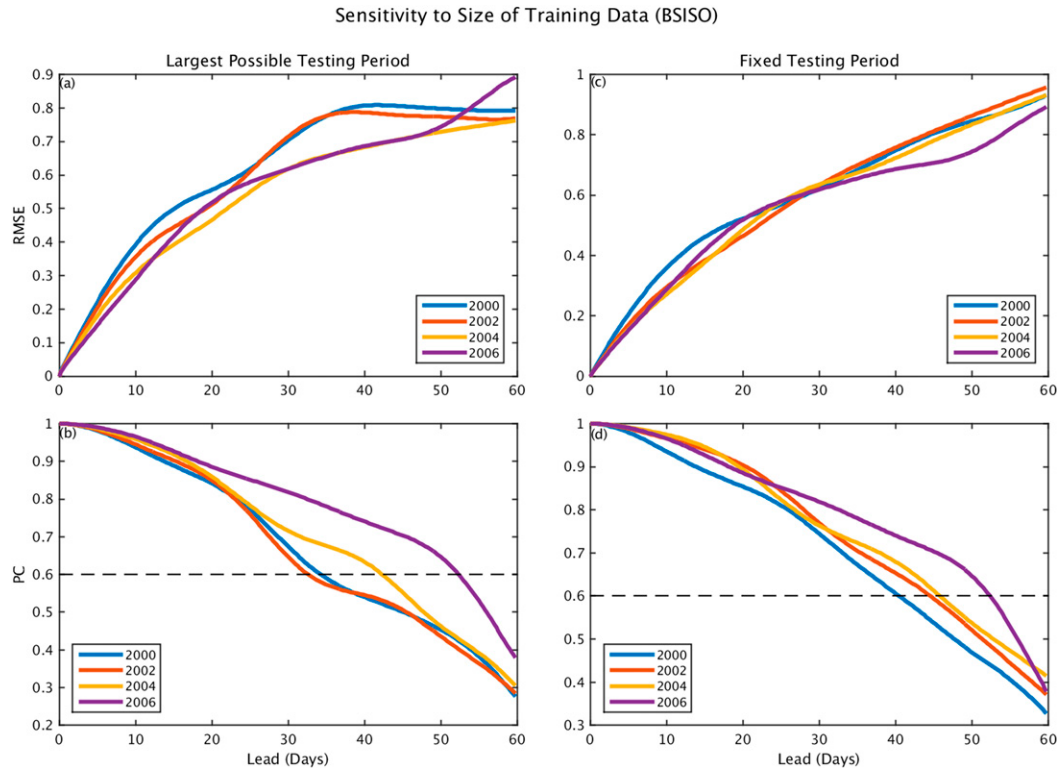


FIG. 11. As in Fig. 10, but for the BSISO (and thus excluding DJF from the calculation of the RMSE and PC scores, as per section 2). See section 5b.

As with any empirical method, two important factors affecting the skill of KAF are the length of the training time series and its relevance to the future behavior of the system. As discussed in sections 4c and 5b, the length of the training time series affects prediction skill in terms of both the quality of the extracted ISO indices and the availability of analogs matching the current initial data. Those results illustrate that decreasing the length of the training dataset generally leads to a degradation of the quality of the NLSA ISO modes, in the sense that the modes of interest become mixed with modes that are unrelated to ISOs, with the effect more pronounced for the BSISO than for the MJO. For the MJO, 35-day predictability is still within reach when the training size is reduced to 9 yr, and for the BSISO, 30-day predictability is available when the training size decreases to 17 yr. While these experiments may appear somewhat artificial (since there is no reason why a forecaster would not use all of the available CLAUS data in practice), they nevertheless illustrate some of the long-term impacts of forced climate variability (as well as low-frequency natural variability) on future analog ISO forecasts. In particular, GCM simulations suggest that climate change on decadal time scales will have statistically significant impacts on the characteristics of ISOs

(Subramanian et al. 2014), and such changes would limit the effective time span of training data available for extracting faithful ISO indices. Similarly, the effective number of analogs would be limited to the latter portions of the training data commensurate with the characteristic time scale of climate change. For tropical variability, that time scale is expected to be in the interdecadal range (e.g., Deser et al. 2012), suggesting that the useful length of available training data is comparable (and possibly longer) than the 23-yr training interval employed in sections 4a and 4b, which was sufficient for skillful MJO and BSISO forecasts. Thus, KAF methods should be useful for ISO forecasts even in the presence of climate change.

7. Conclusions

In this paper, we have demonstrated that qualitative features of tropical ISOs can be forecasted, in an empirical and nonparametric manner, on a scale of 5–7 weeks with appropriate kernel algorithms for ISO index definition and analog forecasting. In particular, using kernels developed in the context of NLSA algorithms, it is possible to obtain indices from unprocessed CLAUS T_b data that are fair representations of the MJO and BSISO, all the while being

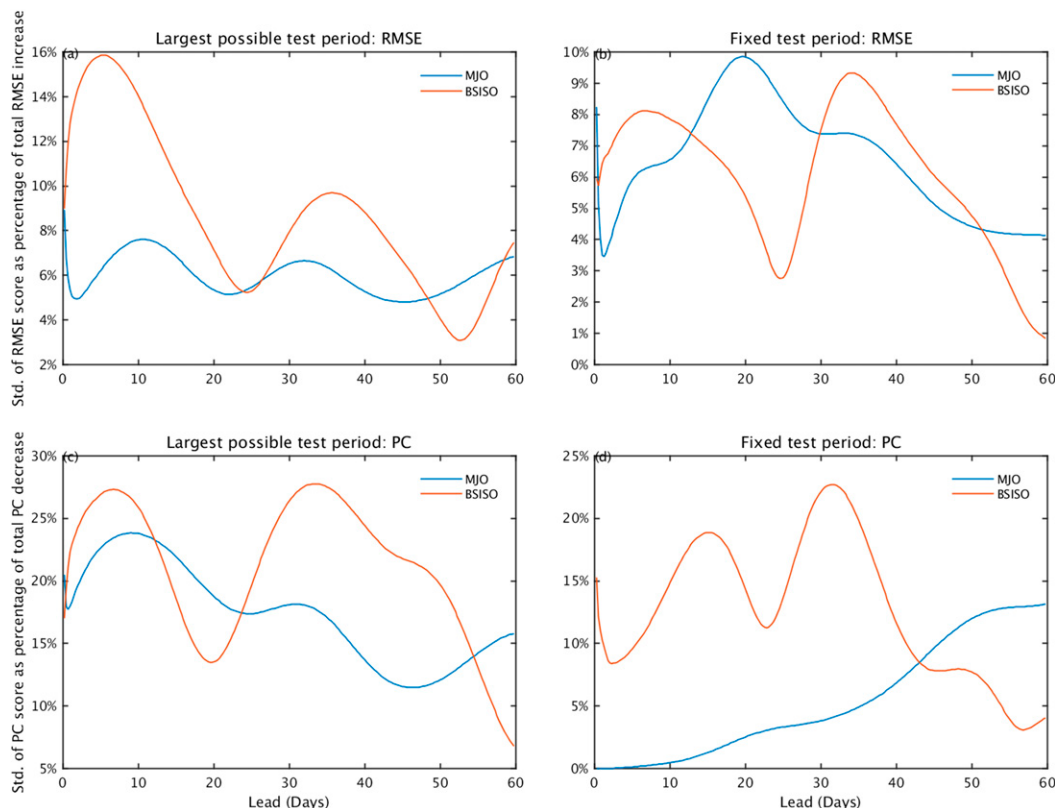


FIG. 12. For both the MJO and BSISO, ratio of the standard deviation of skill across all available training data to the mean amount of (a),(b) RMSE gained and (c),(d) PC lost by a given lead time. See [section 5b](#).

easier to predict than alternate definitions based on classical linear approaches (e.g., the RMM index). Specifically, the PC of forecasts produced via KAF remains above 0.6 for 50 days for both the MJO and BSISO when 23 yr of training data are used.

The overall robustness of the KAF method applied to MJO and BSISO forecasting should continue to be investigated. Varying both the type of kernels used and the size and type of training data is important. One class of kernels to be tried is the so-called cone kernel family ([Giannakis 2015](#)), which takes into account not only the speed at which data varies, but also the direction in which it changes. Relaxing or tightening certain regularity conditions is another potentially interesting approach, as well as incorporating additional predictor variables (e.g., circulation) in a multivariate kernel analysis. Equally importantly, KAF should be assessed in forecasts of physical variables, such as intraseasonal precipitation.

Combining aspects of KAF with other methods has the potential to extend overall tropical ISO predictability even further than what has already been shown. Several numerical models, such as the European Centre for Medium-Range Weather Forecasts model ([Vitart](#)

[2014](#)), a 10-petaflop “K” supercomputer ([Miyakawa et al. 2014](#)), and a coupled GFDL model ([Xiang et al. 2015](#)) have recently obtained MJO predictability of up to 27 days with the RMM as the baseline definition of the MJO. It should be explored whether using NLSA indices in these models would lead to greater predictability. Furthermore, an ensemble of numerical simulations and empirical KAF forecasting can be combined to create an optimized method that produces forecasts more accurately when initialized with novel conditions and yet more quickly when given familiar conditions.

Acknowledgments. The authors gratefully acknowledge the financial support given by the Earth System Science Organization, Ministry of Earth Sciences, Government of India (Grant/Project MM/SERP/CNRS/2013/INT-10/002) to conduct this research under the Monsoon Mission. D. Giannakis, E. Székely, and Z. Zhao acknowledge support from ONR Grant N00014-14-0150 and ONR MURI Grant 25-74200-F7112. D. Giannakis and Z. Zhao also acknowledge support from NSF Grant DMS-1521775. The authors thank Nan Chen and Andrew Majda for stimulating discussions on low-order modeling of intraseasonal oscillations.

REFERENCES

- Belkin, M., and P. Niyogi, 2003: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.*, **15**, 1373–1396, doi:[10.1162/089976603321780317](https://doi.org/10.1162/089976603321780317).
- Berry, T., and T. Sauer, 2016: Local kernels and the geometric structure of data. *Appl. Comput. Harmonic Anal.*, **40**, 439–469, doi:[10.1016/j.acha.2015.03.002](https://doi.org/10.1016/j.acha.2015.03.002).
- , R. Cressman, Z. Greguric Ferencek, and T. Sauer, 2013: Time-scale separation from diffusion-mapped delay coordinates. *SIAM J. Appl. Dyn. Syst.*, **12**, 618–649, doi:[10.1137/12088183X](https://doi.org/10.1137/12088183X).
- Broomhead, D. S., and G. P. King, 1986: Extracting qualitative dynamics from experimental data. *Physica D*, **20**, 217–236, doi:[10.1016/0167-2789\(86\)90031-X](https://doi.org/10.1016/0167-2789(86)90031-X).
- Budisić, M., R. Mohr, and I. Mezić, 2012: Applied Koopmanism. *Chaos*, **22**, 047510, doi:[10.1063/1.4772195](https://doi.org/10.1063/1.4772195).
- Chen, N., and A. J. Majda, 2015a: Predicting the cloud patterns for the boreal summer intraseasonal oscillation through a low-order stochastic model. *Math. Climate Wea. Forecasting*, **1**, 1–20, doi:[10.1515/mcwf-2015-0001](https://doi.org/10.1515/mcwf-2015-0001).
- , and —, 2015b: Predicting the real-time multivariate Madden–Julian oscillation index through a low-order nonlinear stochastic model. *Mon. Wea. Rev.*, **143**, 2148–2169, doi:[10.1175/MWR-D-14-00378.1](https://doi.org/10.1175/MWR-D-14-00378.1).
- , —, and D. Giannakis, 2014: Predicting the cloud patterns of the Madden–Julian oscillation through a low-order nonlinear stochastic model. *Geophys. Res. Lett.*, **41**, 5612–5619, doi:[10.1002/2014GL060876](https://doi.org/10.1002/2014GL060876).
- Chen, T.-C., and J. C. Alpert, 1990: Systematic errors in the annual and intraseasonal variations of the planetary-scale divergent circulation in NMC medium-range forecasts. *Mon. Wea. Rev.*, **118**, 2607–2623, doi:[10.1175/1520-0493\(1990\)118<2607:SEITAA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1990)118<2607:SEITAA>2.0.CO;2).
- Coifman, R. R., and S. Lafon, 2006: Diffusion maps. *Appl. Comput. Harmonic Anal.*, **21**, 5–30, doi:[10.1016/j.acha.2006.04.006](https://doi.org/10.1016/j.acha.2006.04.006).
- Comeau, D., Z. Zhao, D. Giannakis, and A. J. Majda, 2017: Data-driven prediction strategies for low-frequency patterns of North Pacific climate variability. *Climate Dyn.*, doi:[10.1007/s00382-016-3177-5](https://doi.org/10.1007/s00382-016-3177-5), in press.
- Deser, C., A. Phillips, and V. Bourdette, 2012: Uncertainty in climate change projections: The role of internal variability. *Climate Dyn.*, **38**, 527–546, doi:[10.1007/s00382-010-0977-x](https://doi.org/10.1007/s00382-010-0977-x).
- Fernández, A., N. Rabin, D. Fishelov, and J. R. Dorronsoro, 2014: Auto-adaptive Laplacian pyramids for high-dimensional data analysis. arXiv.org, 11 pp. [Available online at <https://arxiv.org/abs/1311.6594>.]
- Giannakis, D., 2015: Dynamics-adapted cone kernels. *SIAM J. Appl. Dyn. Syst.*, **14**, 556–608, doi:[10.1137/140954544](https://doi.org/10.1137/140954544).
- , 2016: Data-driven spectral decomposition and forecasting of ergodic dynamical systems. arXiv.org, 55 pp. [Available online at <https://arxiv.org/abs/1507.02338>.]
- , and A. J. Majda, 2011: Time series reconstruction via machine learning: Revealing decadal variability and intermittency in the North Pacific sector of a coupled climate model. *Proc. Conf. on Intelligent Data Understanding 2011*, Mountain View, CA, NASA, 107–117.
- , and —, 2012: Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proc. Natl. Acad. Sci. USA*, **109**, 2222–2227, doi:[10.1073/pnas.1118984109](https://doi.org/10.1073/pnas.1118984109).
- , and —, 2013: Nonlinear Laplacian spectral analysis: Capturing intermittent and low-frequency spatiotemporal patterns in high-dimensional data. *Stat. Anal. Data Min.*, **6**, 180–194, doi:[10.1002/sam.11171](https://doi.org/10.1002/sam.11171).
- , and —, 2014: Data-driven methods for dynamical systems: Quantifying predictability and extracting spatiotemporal patterns. *Mathematical and Computational Modeling: With Applications in Engineering and the Natural and Social Sciences, Engineering, and the Arts*, R. Melnik, Ed., Wiley, 135–191, doi:[10.1002/9781118853887.ch7](https://doi.org/10.1002/9781118853887.ch7).
- , W.-w. Tung, and A. J. Majda, 2012: Hierarchical structure of the Madden–Julian oscillation in infrared brightness temperature revealed through nonlinear Laplacian spectral analysis. *2012 Conference on Intelligent Data Understanding (CIDU 2012)*, K. Das, N. V. Chawla, and A. N. Srivastava, Eds., IEEE, 55–62, doi:[10.1109/CIDU.2012.6382201](https://doi.org/10.1109/CIDU.2012.6382201).
- Goswami, B. N., 2011: South Asian monsoon. *Intraseasonal Variability in the Atmosphere–Ocean Climate System*, W. K. M. Lau and D. E. Waliser, Eds., Springer, 19–61.
- Hendon, H. H., B. Liebmann, M. Newman, J. D. Glick, and J. Schemm, 2000: Medium-range forecast errors associated with active episodes of the Madden–Julian oscillation. *Mon. Wea. Rev.*, **128**, 69–86, doi:[10.1175/1520-0493\(2000\)128<0069:MRFEAW>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<0069:MRFEAW>2.0.CO;2).
- Hodges, K., D. Chappell, G. Robinson, and G. Yang, 2000: An improved algorithm for generating global window brightness temperatures from multiple satellite infrared imagery. *J. Atmos. Oceanic Technol.*, **17**, 1296–1312, doi:[10.1175/1520-0426\(2000\)017<1296:AIAFGG>2.0.CO;2](https://doi.org/10.1175/1520-0426(2000)017<1296:AIAFGG>2.0.CO;2).
- Jones, C., D. Waliser, J.-K. Schemm, and W. Lau, 2000: Prediction skill of the Madden and Julian oscillation in dynamical extended range forecasts. *Climate Dyn.*, **16**, 273–289, doi:[10.1007/s003820050327](https://doi.org/10.1007/s003820050327).
- Kikuchi, K., B. Wang, and Y. Kajikawa, 2012: Bimodal representation of the tropical intraseasonal oscillation. *Climate Dyn.*, **38**, 1989–2000, doi:[10.1007/s00382-011-1159-1](https://doi.org/10.1007/s00382-011-1159-1).
- Kiladis, G. N., J. Dias, K. H. Straub, M. C. Wheeler, S. N. Tulich, K. Kikuchi, K. M. Weickmann, and M. J. Ventrice, 2014: A comparison of OLR and circulation-based indices for tracking the MJO. *Mon. Wea. Rev.*, **142**, 1697–1715, doi:[10.1175/MWR-D-13-00301.1](https://doi.org/10.1175/MWR-D-13-00301.1).
- Kondrashov, D., M. D. Chekroun, A. W. Robertson, and M. Ghil, 2013: Low-order stochastic model and “past-noise forecasting” of the Madden–Julian oscillation. *Geophys. Res. Lett.*, **40**, 5305–5310, doi:[10.1002/grl.50991](https://doi.org/10.1002/grl.50991).
- Kousky, V. E., and M. T. Kayano, 1993: Real-time monitoring of intraseasonal oscillations. *Proc. 18th Annual Climate Diagnostics Workshop*, Boulder, CO, NOAA, 1–5.
- Lau, K., and F. Chang, 1992: Tropical intraseasonal oscillation and its prediction by the NMC operational model. *J. Climate*, **5**, 1365–1378, doi:[10.1175/1520-0442\(1992\)005<1365:TIOAIP>2.0.CO;2](https://doi.org/10.1175/1520-0442(1992)005<1365:TIOAIP>2.0.CO;2).
- Lee, J.-Y., B. Wang, M. C. Wheeler, X. Fu, D. E. Waliser, and I.-S. Kang, 2013: Real-time multivariate indices for the boreal summer intraseasonal oscillation over the Asian summer monsoon region. *Climate Dyn.*, **40**, 493–509, doi:[10.1007/s00382-012-1544-4](https://doi.org/10.1007/s00382-012-1544-4).
- Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26**, 636–646, doi:[10.1175/1520-0469\(1969\)26<636:APABN>2.0.CO;2](https://doi.org/10.1175/1520-0469(1969)26<636:APABN>2.0.CO;2).
- Madden, R. A., and P. R. Julian, 1972: Description of global-scale circulation cells in the tropics with a 40–50 day period. *J. Atmos. Sci.*, **29**, 1109–1123, doi:[10.1175/1520-0469\(1972\)029<1109:DOGCC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1972)029<1109:DOGCC>2.0.CO;2).
- Mezić, I., 2005: Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dyn.*, **41**, 309–325, doi:[10.1007/s1071-005-2824-x](https://doi.org/10.1007/s1071-005-2824-x).
- Miyakawa, T., and Coauthors, 2014: Madden–Julian oscillation prediction skill of a new-generation global model demon-

- strated using a supercomputer. *Nat. Commun.*, **5**, 3769, doi:[10.1038/ncomms4769](https://doi.org/10.1038/ncomms4769).
- Mo, K. C., 2001: Adaptive filtering and prediction of intraseasonal oscillations. *Mon. Wea. Rev.*, **129**, 802–817, doi:[10.1175/1520-0493\(2001\)129<0802:AFAPOI>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0802:AFAPOI>2.0.CO;2).
- Neena, J. M., J. Y. Lee, D. Waliser, B. Wang, and X. Jiang, 2014: Predictability of the Madden–Julian oscillation in the Intraseasonal Variability Hindcast Experiment (ISVHE). *J. Climate*, **27**, 4531–4543, doi:[10.1175/JCLI-D-13-00624.1](https://doi.org/10.1175/JCLI-D-13-00624.1).
- Packard, N. H., and Coauthors, 1980: Geometry from a time series. *Phys. Rev. Lett.*, **45**, 712–716, doi:[10.1103/PhysRevLett.45.712](https://doi.org/10.1103/PhysRevLett.45.712).
- Rabin, N., and R. R. Coifman, 2012: Heterogeneous datasets representation and learning using diffusion maps and Laplacian pyramids. *Proceedings of the 2012 SIAM International Conference on Data Mining*, J. Ghosh et al., Eds., Proceedings, Society for Industrial and Applied Mathematics, 189–199, doi:[10.1137/1.9781611972825.17](https://doi.org/10.1137/1.9781611972825.17).
- Sauer, T., J. A. Yorke, and M. Casdagli, 1991: Embedology. *J. Stat. Phys.*, **65**, 579–616, doi:[10.1007/BF01053745](https://doi.org/10.1007/BF01053745).
- Slingo, J., and Coauthors, 1996: Intraseasonal oscillations in 15 atmospheric general circulation models: Results from an AMIP diagnostic subproject. *Climate Dyn.*, **12**, 325–357, doi:[10.1007/BF00231106](https://doi.org/10.1007/BF00231106).
- Subramanian, A., M. Jochum, A. J. Miller, R. Neale, H. Seo, D. Waliser, and R. Murtugudde, 2014: The MJO and global warming: A study in CCSM4. *Climate Dyn.*, **42**, 2019–2031, doi:[10.1007/s00382-013-1846-1](https://doi.org/10.1007/s00382-013-1846-1).
- Székely, E., D. Giannakis, and A. J. Majda, 2016a: Extraction and predictability of coherent intraseasonal signals in infrared brightness temperature data. *Climate Dyn.*, **46**, 1473–1502, doi:[10.1007/s00382-015-2658-2](https://doi.org/10.1007/s00382-015-2658-2).
- , —, and —, 2016b: Initiation and termination of intraseasonal oscillations in nonlinear Laplacian spectral analysis indices. *Math. Climate Wea. Forecasting*, **2**, 1–25, doi:[10.1515/mcwf-2016-0001](https://doi.org/10.1515/mcwf-2016-0001).
- Takens, F., 1981: Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence*, D. Rand and L.-S. Young, Ed., Lecture Notes in Mathematics, Vol. 898, Springer, 366–381, doi:[10.1007/bfb0091924](https://doi.org/10.1007/bfb0091924).
- Tung, W.-w., D. Giannakis, and A. J. Majda, 2014: Symmetric and antisymmetric signals in the Madden–Julian Oscillation. Part I: Basic modes in infrared brightness temperature. *J. Atmos. Sci.*, **71**, 3302–3326, doi:[10.1175/JAS-D-13-0122.1](https://doi.org/10.1175/JAS-D-13-0122.1).
- Vitart, F., 2014: Evolution of ECMWF sub-seasonal forecast skill scores. *Quart. J. Roy. Meteor. Soc.*, **140**, 1889–1899, doi:[10.1002/qj.2256](https://doi.org/10.1002/qj.2256).
- von Storch, H., and X. Jinsong, 1990: Principal oscillation pattern analysis of the 30- to 60-day oscillation in the tropical troposphere. Part I: Definition of an index and its prediction. *Climate Dyn.*, **4**, 175–190, doi:[10.1007/BF00209520](https://doi.org/10.1007/BF00209520).
- Waliser, D. E., 2011: Predictability and forecasting. *Intraseasonal Variability in the Atmosphere–Ocean Climate System*, 2nd ed. W. K. M. Lau and D. E. Waliser, Eds., Springer, 433–468, doi:[10.1007/978-3-642-13914-7_12](https://doi.org/10.1007/978-3-642-13914-7_12).
- , C. Jones, J.-K. E. Schemm, and N. E. Graham, 1999: A statistical extended-range tropical forecast model based on the slow evolution of the Madden–Julian oscillation. *J. Climate*, **12**, 1918–1939, doi:[10.1175/1520-0442\(1999\)012<1918:ASERTF>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<1918:ASERTF>2.0.CO;2).
- Wang, B., F. Huang, Z. Wu, J. Yang, X. Fu, and K. Kikuchi, 2009: Multi-scale climate variability of the South China Sea monsoon: A review. *Dyn. Atmos. Oceans*, **47**, 15–37, doi:[10.1016/j.dynatmoce.2008.09.004](https://doi.org/10.1016/j.dynatmoce.2008.09.004).
- Webster, P. J., and R. Lucas, 1992: TOGA COARE: The Coupled Ocean–Atmosphere Response Experiment. *Bull. Amer. Meteor. Soc.*, **73**, 1377–1416, doi:[10.1175/1520-0477\(1992\)073<1377:TCTCOR>2.0.CO;2](https://doi.org/10.1175/1520-0477(1992)073<1377:TCTCOR>2.0.CO;2).
- Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917–1932, doi:[10.1175/1520-0493\(2004\)132<1917:AARMMI>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2).
- Xavier, P. K., and B. N. Goswami, 2007: An analog method for real-time forecasting of summer monsoon subseasonal variability. *Mon. Wea. Rev.*, **135**, 4149–4160, doi:[10.1175/2007MWR1854.1](https://doi.org/10.1175/2007MWR1854.1).
- Xiang, B., M. Zhao, X. Jiang, S. J. Lin, T. Li, X. Fu, and G. Vecchi, 2015: The 3–4-week MJO prediction skill in a GFDL coupled model. *J. Climate*, **28**, 5351–5364, doi:[10.1175/JCLI-D-15-0102.1](https://doi.org/10.1175/JCLI-D-15-0102.1).
- Yanai, M., B. Chen, and W.-w. Tung, 2000: The Madden–Julian oscillation observed during the TOGA COARE IOP: Global view. *J. Atmos. Sci.*, **57**, 2374–2396, doi:[10.1175/1520-0469\(2000\)057<2374:TMJOOD>2.0.CO;2](https://doi.org/10.1175/1520-0469(2000)057<2374:TMJOOD>2.0.CO;2).
- Zhang, C., 2005: Madden–Julian oscillation. *Rev. Geophys.*, **43**, RG2003, doi:[10.1029/2004RG000158](https://doi.org/10.1029/2004RG000158).
- , J. Gottschalck, E. D. Maloney, M. W. Moncrieff, F. Vitart, D. E. Waliser, B. Wang, and M. C. Wheeler, 2013: Cracking the MJO nut. *Geophys. Res. Lett.*, **40**, 1223–1230, doi:[10.1002/grl.50244](https://doi.org/10.1002/grl.50244).
- Zhao, Z., and D. Giannakis, 2016: Analog forecasting with dynamics-adapted kernels. *Nonlinearity*, **29**, 2888–2939, doi:[10.1088/0951-7715/29/9/2888](https://doi.org/10.1088/0951-7715/29/9/2888).