Testing methods of pattern extraction for climate data using synthetic modes

Testing Methods of Pattern Extraction for Climate Data Using Synthetic Modes

D. James Fulton[a] and Gabriele C. Hegerl[a]

[a] *School of GeoSciences, University of Edinburgh, Edinburgh, United Kingdom*

ABSTRACT: In this paper we develop a method to quantify the accuracy of different pattern extraction techniques for the additive space–time modes often assumed to be present in climate data. It has previously been shown that the standard technique of principal component analysis (PCA; also known as empirical orthogonal functions) may extract patterns that are not physically meaningful. Here we analyze two modern pattern extraction methods, namely dynamical mode decomposition (DMD) and slow feature analysis (SFA), in comparison with PCA. We develop a Monte Carlo method to generate synthetic additive modes that mimic the properties of climate modes described in the literature. The datasets composed of these generated modes do not satisfy the assumptions of any pattern extraction method presented. We find that both alternative methods significantly outperform PCA in extracting local and global modes in the synthetic data. These techniques had a higher mean accuracy across modes in 60 out of 60 mixed synthetic climates, with SFA slightly outperforming DMD. We show that in the majority of simple cases PCA extracts modes that are not significantly better than a random guess. Finally, when applied to real climate data these alternative techniques extract a more coherent and less noisy global warming signal, as well as an El Niño signal with a clearer spectral peak in the time series, and more a physically plausible spatial pattern.

KEYWORDS: Algorithms; Data mining; Empirical orthogonal functions; Pattern detection; Principal components analysis; Statistical techniques; Climate variability; Data science; Dimensionality reduction; Model interpretation and visualization; Pattern recognition

1. Introduction

Much climate research relies on extracting space and time signatures of large-scale climate processes such as El Niño–Southern Oscillation (ENSO). The spatial patterns and time series extracted from data are used to direct research into the physical mechanisms (e.g., Wanner et al. 2001) associated with these signals, and thus to better understand physical processes that generate climate variability and teleconnections. The signal time series are often used to search for other phenomena driven by these large-scale patterns, such as heatwaves, droughts, and other temporary shifts in local climate (Weisheimer et al. 2017), and are also relevant for future forecasts (Wu et al. 2009; Johnson et al. 2014). Our understanding of the physics and impacts of this emergent behavior is therefore strongly impacted by the quality and purity of the modes we extract from data.

One of the common tools used to extract modes from climate data is principal component analysis (PCA), also known as empirical orthogonal functions (EOFs) (von Storch and Zwiers 2001). The spatial vectors, known as principal components (PCs), returned by this method are often interpreted as if they represent individual climate processes, such as ENSO. ENSO (Timmermann et al. 2018) is a well-established physical process and the PCs are often used merely as a visualization of its variability. However, the debate around different flavors of ENSO is based on PCA at least in part (Ashok et al. 2007). Other physical modes with less understood mechanisms, such

as the Pacific decadal oscillation (PDO) (Newman et al. 2016), rely heavily on PCA for their definition, specification of spatial pattern, and time index, and as a starting point in the search for their mechanism. However, PCA is merely designed for compressing data and capturing as much variance as possible in the first *n* modes. Therefore, interpreting these modes individually or using them as evidence to support a climate mechanism can be misleading. Dommenget and Latif (2002) showed that PCA is prone to finding dipoles where none exist and Monahan et al. (2009) showed that the PCs do not correspond to individual dynamical modes, do not represent individual kinematic degrees of freedom of the system, and are not statistically independent of each other.

It is often assumed that the first few modes returned by PCA, with highest variance, should be physically interpretable, and should match some underlying signal in the data. However, even the first PC cannot capture a true underlying mode, in examples where true modes exist, if these modes are nonorthogonal to each other (proof in appendix A).

Many other methods of extracting modes from climate data have been proposed and, as we will discuss, are theoretically more consistent with a physical interpretation of what a climate mode might be. These newer methods have been developed both inside and outside the climate community, and we review some of these in section 3. However, there is no ground truth for the space–time structure of climate modes. This is true both for observational data and for data from climate simulations. While climate simulations provide complete information about the climate state, there is still no ground truth for climate modes as these are emergent properties of the system.

Therefore, we previously have not been able to quantitatively evaluate the accuracy of any method of mode extraction.

Some efforts have been made to devise statistical tests to guard against spurious modes, but these efforts have not been directed toward accessing how well these patterns match underlying known signals in the data. They have instead tried to stop researchers from choosing too many modes given the statistical uncertainty (North et al. 1982), or to stop them identifying modes that may be attributable to red noise fluctuations (Dommenget 2007).

In this paper we propose a Monte Carlo method to generate physically plausible synthetic climate modes. We assume that these modes are additive, and mix them with generated background noise to create synthetic climate-like datasets. With these datasets we can perform pseudoexperiments to test the accuracy of each method of mode extraction, and quantify how well it is able to separate known modes from each other, given our assumptions. The datasets we create are motivated to be an idealized simple climate with spatial and temporal autocorrelation. They do not strictly satisfy the assumptions of any mode discovery method commonly used. Therefore this will allow us to explore systematic errors associated with each mode extraction technique on datasets that are motivated to be more realistic than those assumed by the method's definition.

2. Generating Monte Carlo climate data

a. What is a "mode" of climate variability?

A standard definition of a mode in climate data has never been established to our knowledge. In physics, modes are defined as eigenstates of the time evolution operator of an oscillator system. These eigenstates evolve in time predictably and independently from each other, and are derived analytically rather than from data. The full state of the system can be expressed as a superposition of its eigenstates. We cannot apply the same definition to climate, as the time evolution operator is either an entire global circulation model or the real climate system itself, and if any eigenstates of this operator existed they would be intractable. It is also highly likely that the physical quantities of any eigenstate of this operator would be complex combinations of climate variables, not all of them well observed.

In this paper we take inspiration from the definition of physical modes to inform our definition and generation of climate modes. We assume they are additive, such that a superposition of these space–time modes $\mathbf{D}^{(i)}(t)$ make up the observed climate state $\mathbf{X}(t)$ at any time, that is, $\mathbf{X}(t) = \sum_i \mathbf{D}^{(i)}(t)$. An example of this would be the Pacific sea surface temperatures being a sum of ENSO and PDO projections. In this simple case the projections $\mathbf{D}^{(i)}(t)$ are just the spatial patterns of ENSO and PDO multiplied by their time indices. We also expect that modes should be approximately independent, so that the phase of one mode does not strongly imply that of others. This prevents the negative and positive phases of a phenomenon from being split into two different modes, and also from being split geographically into multiple modes covering different regions. We also expect that modes should show time coherence in their evolution. These

assumptions are reasonable for modes caused by slow physical mechanisms and are consistent with known climate modes (e.g., Christensen et al. 2013).

Note that these are modes of climate data, and not of the physical climate. Whether some physical effect induces a mode signal in observations depends on the sampling frequency and the length of observations. At low sampling frequency weather is approximately white in time [see discussion in Hasselmann (1976)], but at high sampling frequency weather events could be modes under our definition. If the sequence is long enough then glacial cycles would also show up in our signal as modes. At intermediate sampling frequency even the seasonal cycle would be a mode, but not a particularly interesting one. However, here we focus on modes inspired by the literature on the time scale of several months to tens of years.

That these modes are features of the coupled physics–measurement system, not the physical system alone, is true in other climate research where sample frequency is defined, a geographical region is chosen, high-frequency components of the signal are filtered out, or the global warming trend is removed. It is only after we extract these modes from data that we can think about any associated physical mechanisms. This is true in both observations and simulations. Since these signals direct our research we should care about the magnitude of biases in their extraction.

Some climate modes may affect each other via teleconnections as explored in Runge et al. (2015), hence their dynamics may not be entirely independent. Due to the red nature of the climate system, it is likely that no set of modes could be defined that are truly independent from each other. Even ENSO is strongly tied to the seasonal cycle.

Based on this definition of climate data modes we now generate synthetic data to test mode extraction methods. We note that it already appears unlikely that PCA will be the optimal tool to estimate these modes. We define modes to have time coherence and PCA does not consider any of the time information available, and instead maximizes variance. The same PCs are calculated when the time series are shuffled as when not.

b. Monte Carlo mode generation

The spatial patterns of a synthetically generated mode should be consistent with being on a spherical surface. To generate these patterns we take ideas from Gaussian processes (Rasmussen and Williams 2006). We randomly draw the spatial pattern of each mode $\mathbf{v}^{(i)}$ from a multivariate normal distribution $\mathbf{v}^{(i)} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, where we use the following function to create the covariance matrix

$$\boldsymbol{\Sigma}(\mathbf{x}_1, \mathbf{x}_2) = \exp\left[-\frac{\Theta(\mathbf{x}_1, \mathbf{x}_2)^2}{2\Delta\theta^2}\right]. \qquad (1)$$

Here $\mathbf{x}_1$ and $\mathbf{x}_2$ are two points on the surface of a sphere, and the covariance matrix is defined between a regular spaced latitude–longitude grid of points. The term $\Theta(\mathbf{x}_1, \mathbf{x}_2)$ is the central angle between the two points, and $\Delta\theta$ is the characteristic angular range of correlations. The first column of Fig. 1 shows examples of these sampled spatial patterns.
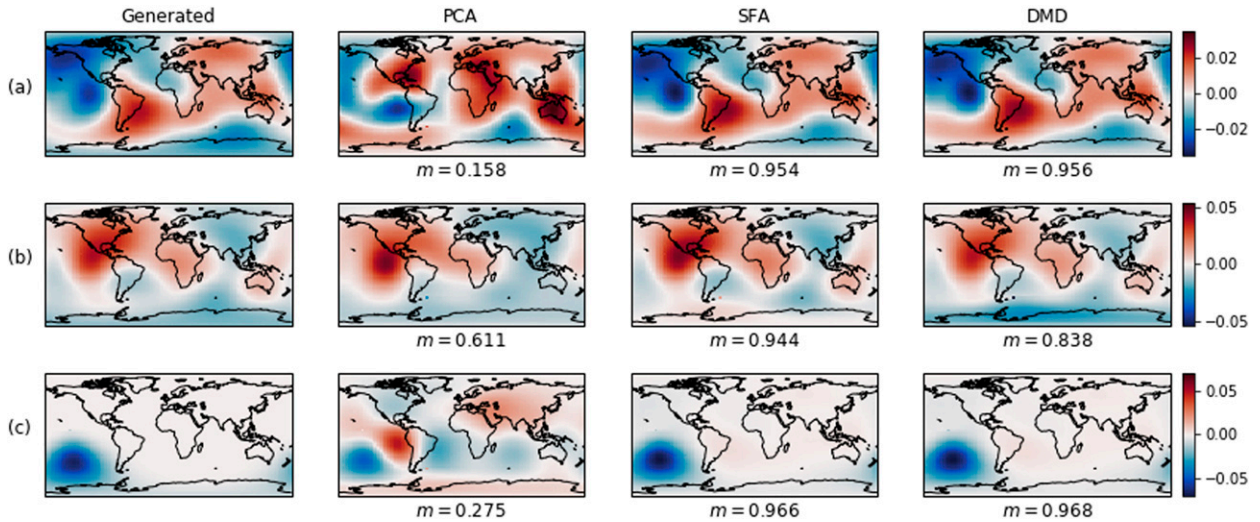
FIG. 1. (a)–(c) A comparison between three synthetic generated modes and their reconstructions using three latent variable methods. The left column shows three of the eight modes generated for this dataset. The PCA, SFA, and DMD columns show the mode discovered by each method that best matches the generated mode in the same row. The projection correlations for each match are stated below each subplot. These scores represent how similar the modes are. Note that the coastlines are fictitious but included to act as an eye guide and for demonstration purposes.

Each sampled vector $\mathbf{v}^{(i)}$ stores the numerical values of a singular mode pattern for all latitude–longitude grid points. Due to the covariance matrix chosen, elements in the vector that are adjacent on the globe's surface are sampled with high covariance, and so are likely to have similar values. While those far apart are more likely to have very different values in $\mathbf{v}^{(i)}$. Equation (1) shows that this covariance decays exponentially with distance, and this function is known in the Gaussian processes literature as the *covariance kernel function*.

Numerically, the method to create one of these vectors is to decompose the covariance matrix $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$. We use Cholesky decomposition in particular. Then simply use $\mathbf{v}^{(i)} = \mathbf{n}^{(i)}\mathbf{L}^{\mathrm{T}}$, where $\mathbf{n}^{(i)}$ is a vector of values with the same shape $\mathbf{v}^{(i)}$, where each value is randomly sampled from a univariate normal distribution.

These sampled modes are *spatially dense* maps, meaning that most points on the globe have nonzero elements in the vector $\mathbf{v}^{(i)}$ (see first column of Figs. 1a and 1b for examples). By slightly modifying our sampling from the normal distribution we can also generate patterns which are *spatially sparse*, and are confined to a smaller area of the globe as in Fig. 1c. We exclude exact detail of this sampling from our discussion but make all source code available (see acknowledgments). These spatially dense and sparse maps mimic climate modes that are either global scale with teleconnections, or more localized respectively, similar to the different modes presented in Dommenget and Latif (2002). These modes have space–time separable, linear projections onto the observations. Their projection is $\mathbf{D}^{(i)}(t) = z^{(i)}(t)\mathbf{v}^{(i)}$ where $z^{(i)}(t)$ is the mode's time series. We will refer to these as *linear modes* because the projections are linear with $z^{(i)}(t)$.

Seasonal autoregressive moving average (SARMA) time series models are often used to model climate indices (von Storch and Zwiers 2001) and so we use these to generate feasible synthetic time indices $z^{(i)}(t)$ for our modes. The autoregressive and moving average parameters used were sampled over the range of parameters of SARMA models fit to observational monthly indices of ENSO, PDO, DMI (dipole mode index), AMO (Atlantic multidecadal oscillation), and AO (Arctic Oscillation) (NOAA/OAR/ESRL PSL 2019).

This method of construction of linear modes with SARMA time series produces synthetic modes that are strongly motivated by signals found in climate data. As we assume modes are additive, we create a pseudodataset by summing a number of modes $\mathbf{X}(t) = \sum_i \mathbf{D}^{(i)}(t)$. These modes are not constrained to be orthogonal to each other, which is as expected in real climate data.

We can generate *nonlinear modes* (modes that are not space–time separable) by extending the covariance kernel function [Eq. (1)] to add a new phase dimension. This modification in Eq. (2) means that the spatial pattern of modes is not constant like the linear modes, and the mode projections are not linear with $z^{(i)}(t)$.

We are motivated to do this by considering ENSO, which is nonlinear—La Niña events have slightly different spatial patterns from El Niño events. Although the mode discovery methods we use will only search for linear modes, we wish to know how much nonlinear modes confound them and whether they make good linear estimates of nonlinear modes.

When we sample vectors from the new covariance kernel function [Eq. (2)], we generate 3D space-phase modes instead of 2D spatial patterns. These 3D modes are on a regular latitude–longitude–phase grid, where the phase values of the grid are chosen to be in the range [0,1] and the latitude–longitude values cover the global surface. Some examples of these modes are shown in the source code repository in gif format (see acknowledgments).

$$\boldsymbol{\Sigma}_{\mathrm{NL}}(\mathbf{x}_1, s_1, \mathbf{x}_2, s_2) = \boldsymbol{\Sigma}(\mathbf{x}_1, \mathbf{x}_2)\exp[-\mathcal{S}(s_2, s_1)^2]. \quad (2)$$

In this new nonlinear covariance kernel Σ_{NL}, we must define a function for *phase distance* $\mathcal{S}(s_2, s_1)$. This function tells us how similar the spatial patterns at phase s_1 and s_2 will be for a single mode (i.e., how similar a synthetic set of La Niña and El Niño spatial patterns would be). We choose $\mathcal{S}(s_2, s_1) = (1/\sqrt{2}\Delta s)(s_1 - s_2)$ so that the covariance between a mode's spatial patterns at different phases decreases exponentially with phase difference. Here Δs is the characteristic phase range of correlations.

Once the nonlinear modes have been sampled we must generate a phase series $s^{(i)}(t)$ rather than a time index series $z^{(i)}(t)$. The projection of these nonlinear modes onto observations is $\mathbf{D}^{(i)}(t) = \mathbf{v}^{(i)}[s(t)]$. To generate the phase series we first generate a SARMA time series $\zeta^{(i)}(t)$ like before and then apply a sigmoid function

$$s^{(i)}(t) = (1 + e^{-a\zeta^{(i)}(t)})^{-1} \tag{3}$$

to ensure that $s(t) \in [0, 1] \forall t$. Each space-phase vector $\mathbf{v}^{(i)}(s)$ was randomly sampled for a discrete grid of s coordinates, so we use cubic interpolation to map to continuous values of s for the projections.

Further we can generate nonlinear modes with cyclic projections. This means the mode projection grows and shrinks along different pathways. This mode type is motivated to be even more similar to ENSO, as the pattern of transitioning from a La Niña to an El Niño state is not the exact reverse of transitioning from El Niño to La Niña. To accomplish this we use a cyclic phase-distance function $\mathcal{S}(s_2, s_1) = (\sqrt{2}/\Delta s)\sin[\pi(s_2 - s_1)]$. Here the projections for $s = 0$ and $s = 1$ are identical as $\mathcal{S}(1, 0) = 0$. The projection moves on a nonlinear closed path. In this case we set $s(t)$ to a randomly generated monotonically increasing time series to give the modes a defined direction. We will refer to these as *cyclic nonlinear modes*.

Finally we can generate *moving wave modes*, like the Madden–Julian oscillation (Zhang 2005). We approximate these by creating a linear spatial vector $\mathbf{v}^{(i)}$ and simply rotate it around the surface of the sphere.

We generate red noise to add to this synthetic dataset that has shorter range spatial correlations than our modes, and time series with a smaller amount of time autocorrelation. We do this by using a covariance kernel function like Eq. (1) but with smaller $\Delta\theta$ to give a covariance matrix Σ_{noise}. Instead of sampling from $\mathcal{N}(0, \Sigma_{noise})$ as before, we decompose Σ_{noise} into a full set of basis vectors. If there are N grid points then we will have N vectors. We generate a moving average (MA) time series for each basis vector, and sum the projections of all of these to make our noise component. This gives us our desired red noise as a background, which has space and time correlations. It has equal variance in all directions, and so is "spherical." The modes will generally stand out against this spherical background as they have more variance in a single direction.

The modes we have created here are consistent with what we discussed in section 2a and are, by construction, additive with independent dynamics and time series. This is often assumed to be approximately the case in climate mode research. There remains a question about whether any additive mode signals actually exist in real climate data. We do not intend to answer

this here, but we note that essentially all climate mode research quietly assumes that this is true.

The spatial structure and time sequences of the modes are constructed to be similar to climate modes proposed in the literature. However, we have added complexity by also creating nonlinear modes.

These generated modes can now be used to test additive mode extraction techniques against true known additive modes. They can help us assess how likely we are to be led astray by different mode extraction techniques in our search for physical mechanisms.

3. Review of latent variable methods for pattern extraction

Extracting patterns from high-dimensional data is a common problem in various research fields. Therefore a range of methods have been devised for this task. These methods are known as latent variable methods (LVMs), of which PCA is one example, and many of them have similar mathematical forms. Some of these LVMs have been extended to extract nonlinear modes, using nonlinear maps between time series and projections, that is, $\mathbf{D}^{(i)}(t) = \mathbf{v}^{(i)}[s(t)]$. Here we limit ourselves to use only the linear versions, even though some of our generated modes, and the real climate modes, are nonlinear. This was chosen because the focus of this study is how these LVMs can reliably direct our research in climate physics, and linear maps are easier to interpret than nonlinear ones.

Most of the LVMs appropriate to climate data fall into one of three classes. In this section, we review these methods and select one form to take forward from each class. Here, we present a general explanation of the methods and then provide their mathematical definition.

a. Variance maximizing methods

As mentioned, the most common technique for analyzing high-dimensional data in climate research is PCA. This is simply the eigendecomposition of the covariance matrix and is known in other communities as principal orthogonal decomposition or empirical orthogonal functions, and shares much mathematical detail with singular value decomposition. This method produces vectors of spatial patterns which maximize the variance captured in each successive component. Autoencoder neural networks (Kramer 1991) are also a member of this class, but the linear version of these networks return modes equivalent to PCA's modes. We will be using PCA only.

b. Slowness maximizing methods

There are several techniques that have been proposed, which aim to capture slowly varying aspects of a system. The underlying assumption of these methods is the slowness principle (Goodfellow et al. 2016), which states that the most important aspects of a system are those that change slowly. In monthly means of climate data over tens of years, this means we focus on indices that may represent global warming, slow ocean circulation, or other long-term coherent patterns, as opposed to the chaotic variation of short-term

weather. A selection of these is detailed below, but note that many more variants exist.

- Slow feature analysis (SFA) (Wiskott and Sejnowski 2002) is the simplest of these methods and finds the slowest varying components of the system. This is achieved by finding spatial maps that minimize the mean square differential of the associated time series.
- Optimally persistent patterns (OPP) (DelSole 2001) are a well-established method in climate research. This method finds spatial maps that maximize the autocorrelation of the time series across multiple steps behind and ahead.
- Low-frequency component analysis (LFCA) (Wills et al. 2018) is a recently proposed method that aims to find patterns that maximize the ratio of low-frequency variance to total variance of the time series.

We have chosen to use SFA as a representative model from this group because it has a simple optimization objective, which is highlighted in section 3a, which we believe is more intuitive and interpretable. It also does not require additional parameter values to be chosen. In contrast, LFCA requires definition of a cutoff between low and high frequency. OPP also has an extra free parameter, which is the limit of time steps ahead or behind for which the correlations are considered (as shown in section 3a). Choosing SFA is also an opportunity to explore this technique, which has not been widely used in climate research despite being commonly employed in other fields.

c. Time dynamics methods

Many proposed time dynamics methods share an identical simplest form. This is to model the system dynamics as $\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t - \Delta t) + \boldsymbol{\eta}(t)$, where $\mathbf{A}$ is a real or complex matrix that defines the dynamics and $\boldsymbol{\eta}(t)$ is a random noise vector sampled at each time step. This form is shared by dynamic mode decomposition (DMD) (Schmid 2010; Kutz et al. 2016a), principal oscillation patterns (Hasselmann 1988), discrete time linear dynamical systems (Dahleh et al. 2011), and first-order vector autoregressive methods (Hamilton 1994). Through eigendecomposition of $\mathbf{A}$ we retrieve modes that represent independent dynamic modes in the data vectors $\mathbf{x}(t)$. These modes are predictable in their dynamics and the eigenvalues dictate whether they grow, decay, and oscillate. The idea behind this method is that modes that evolve coherently are more likely to be physically meaningful. Although all of these methods are mathematically equivalent, we follow DMD as this has had more recent developments (Kutz et al. 2016b) and can be efficiently applied to high-dimensional data (Tu et al. 2014).

It is important to note that the slowness maximizing and time dynamics methods produce spatial patterns that are not constrained to be orthogonal in the observable variables. This is a desirable feature as the modes we create, and any real climate modes, are not constrained in this way either.

We also note that the forward model of DMD, that is, $\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t - \Delta t) + \boldsymbol{\eta}(t)$, is similar to the forward model used to create our synthetic data. Our forward model would be

identical if we limited ourselves to generating only linear modes, if we used AR (autoregressive) time series rather than SARMA time series, and if the noise we add was white rather than red in time. Additionally, if the generated modes are orthogonal then DMD and SFA calculate the exact same modes. However, the modes we generate in this paper are not orthogonal.

From this theoretical basis we should expect SFA and DMD to better capture the generated linear modes. But since some of their assumptions are not met by our data generation, we also should not expect them to be perfect, and we do not know quantitatively how each will perform.

d. Latent variable methods as eigendecomposition problems

All of the methods described above can be expressed as eigendecomposition problems. In the following we list the matrices that when decomposed to a set of eigenvectors and eigenvalues represent the spatial patterns and key values for each method.

SFA, OPP, and LFCA are performed on a whitened rotation (Kessy et al. 2018) of the original data, that is, on $\boldsymbol{\Sigma}_{\mathbf{x}}^{-1/2}\mathbf{x}(t)$ instead of $\mathbf{x}(t)$. Here we also use the notation $\boldsymbol{\Sigma}_{\mathbf{x}}^{-1/2} = \mathbf{U}\boldsymbol{\Lambda}^{-1/2}\mathbf{U}^{\mathrm{T}}$ where the matrix $\mathbf{U}$ is the column PCs, and the diagonal matrix $\boldsymbol{\Lambda}$ is the PCA eigenvalues of $\mathbf{x}(t)$. Therefore the matrix $\boldsymbol{\Sigma}_{\mathbf{x}}^{-1/2}$ performs a transform that would remove all correlation between variables in the data matrix $\mathbf{x}(t)$ and set the variance of each variable to one. The spatial eigenvectors of SFA, OPP, and LFCA must be rotated back into the observation space after computation by left-multiplying by the matrix $\boldsymbol{\Sigma}_{\mathbf{x}}^{1/2} = \mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^{\mathrm{T}}$.

We use the notation $\langle \dots \rangle$ to indicate the time average of the quantity inside the brackets.

PCA: The highest eigenvalues represent spatial patterns that capture most variance:

$$\langle \mathbf{x}(t)\mathbf{x}(t)^{\mathrm{T}} \rangle. \tag{4}$$

SFA: The lowest eigenvalues represent spatial patterns with time series that change the slowest (Δt is the sampling time step):

$$\langle \boldsymbol{\Sigma}_{\mathbf{x}}^{-1/2}[\mathbf{x}(t) - \mathbf{x}(t - \Delta t)][\mathbf{x}(t) - \mathbf{x}(t - \Delta t)]^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1/2\mathrm{T}} \rangle. \tag{5}$$

OPP: The highest eigenvalues represent spatial patterns with time series that have the highest average correlation across multiple steps ahead and behind.

$$\sum_{n=-k}^{k} \langle \boldsymbol{\Sigma}_{\mathbf{x}}^{-1/2}\mathbf{x}(t)\mathbf{x}(t - n\Delta t)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1/2\mathrm{T}} \rangle. \tag{6}$$

LFCA: The highest eigenvalues represent spatial patterns with time series that have the highest proportion of variance in low frequencies. Here $\mathcal{L}()$ is a Lanczos low-pass filter:

$$\langle \boldsymbol{\Sigma}_{\mathbf{x}}^{-1/2}\mathcal{L}[\mathbf{x}(t)]\mathcal{L}[\mathbf{x}(t)]^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1/2\mathrm{T}} \rangle. \tag{7}$$

DMD: Large eigenvalues represent spatial patterns with time series that have high one-step-ahead autocorrelation.

Here $\mathbf{x}^{-1}(t - \Delta t)$ is the pseudoinverse of the matrix $\mathbf{x}(t - \Delta t)$ $\forall\ t$:

$$\langle \mathbf{x}(t)\mathbf{x}^{-1}(t - \Delta t)\rangle. \qquad (8)$$

4. Defining mode similarity

We need a way to quantify how well the discovered modes match to the generated modes. To do this we compare the projections of the discovered modes $\tilde{\mathbf{D}}^{(j)}(t)$ to the known synthetic patterns $\mathbf{D}^{(i)}(t)$. In the case of complex discovered mode projections we take the real part of $\tilde{\mathbf{D}}^{(j)}(t)$. To compare them we subtract the time mean at each location from the projections, unroll them in space and time into 1D vectors, and then calculate the cosine similarity between these. We will refer to this as the *projection correlation* and denote its value between generated mode i and discovered mode j as m_{ij}. More explicitly,

$$m_{ij} = \frac{\langle [\mathbf{D}^{(i)}(t) - \langle \mathbf{D}^{(i)}(t)\rangle] \cdot [\tilde{\mathbf{D}}^{(j)}(t) - \langle \tilde{\mathbf{D}}^{(j)}(t)\rangle]\rangle}{\sigma^{(i)}\tilde{\sigma}^{(j)}}, \qquad (9)$$

where

$$\sigma^{(i)} = [\langle \mathbf{D}^{(i)}(t) \cdot \mathbf{D}^{(i)}(t)\rangle - \langle \mathbf{D}^{(i)}(t)\rangle \cdot \langle \mathbf{D}^{(i)}(t)\rangle]^{1/2} \qquad (10)$$

and $\tilde{\sigma}^{(j)}$ is similar.

If the discovered and generated modes are both linear, then the projection correlation is equal to the Pearson correlation coefficient between their time series multiplied by the dot product between their normalized spatial vectors (detailed in appendix B).

This makes the metric of choice easily interpretable in simple cases and but also generalized for more complex, nonlinear cases. This metric gives a maximum value of unity when the projections are perfect and a minimum value of zero when the projections are orthogonal.

The order of the estimated LVM modes will not match the order of input modes, and we always have more LVM estimated modes than input modes. Therefore we need to match them up by creating a set of one-to-one index matches $\{i_k \rightarrow j_k\}$ between the generated modes and the LVM modes. We do this in a "greedy" fashion by first calculating the projection correlation values between all mode pairs m_{ij}, then select the index pair i and j that maximize m_{ij} and add this to our set of matches. Next we find the index pair that maximizes m_{ij} subject to the constraint that neither of the index values have been selected previously, and repeat this step until each generated mode has a matched LVM mode. Using these matches we then define a mean projection correlation for the dataset for that LVM:

$$M = \frac{1}{n}\sum_{k=1}^{n} m_{i_k j_k}, \qquad (11)$$

where n is the number of generated modes.

5. Pseudoclimate experiments

To assess systematic biases in the chosen LVMs we propose a series of mode recovery experiments in which the pseudoclimate comprises different combinations of modes and noise.

Note that each generated pseudoclimate dataset has 7008 spatial points and 14 400 time measurements unless stated otherwise. These values were chosen to mimic monthly average data from a long climate simulation experiment.

The experiments we performed were as follows.

1) Precision. This is the simplest case for a pseudoclimate and gives us an estimate of the natural spread in the metric scores. We create 30 datasets each with eight modes. Each mode has the same variance, and is linear and spatially dense. A small amount of noise, which has variance equal to the variance of one mode, is added to the dataset.

2) Spatial sparsity. It is important that our LVMs can estimate both global and local patterns, as this may direct our search to understand the physics. These generated datasets are identical to experiment 1 except that four of the generated modes are spatially sparse and the other four spatially dense.

3) Noise sensitivity. Our climate signals may contain high amounts of noise. We would like methods that are still able to find modes among the noise. These generated datasets are identical to experiment 1 except that we add noise with variance 0.5–32 times that of any single mode.

4) Record length sensitivity. Real observational data are quite short in length while our climate model runs can be many hundreds of years. We would like to know if any of the LVMs are more negatively affected by limited length of the data. In this experiment we generate modes as in experiment 1 but with time scales equating to 20–2000 years of monthly data.

5a) Nonlinear confounding. This experiment is designed to explore the effect of having nonlinear modes mixed among linear modes. We would like our linear LVMs to approximately capture nonlinear modes and not let the residuals of these modes leak into the others. We create 30 pseudoclimate datasets each with eight modes. Half of the datasets have one nonlinear mode, and the other datasets have two. The remaining modes in each dataset are linear dense modes and all modes have the same variance and a low-level noise as in experiment 1.

5b) Nonlinear cyclic confounding. As experiment 5a, but using the nonlinear cyclic modes.

5c) Wave confounding. As in experiment 5a, but using moving wave modes.

6a) Dominant mode. Having one dominant mode such as ENSO may make it more difficult to identify other modes. We generate 32 datasets each with eight linear dense modes, seven of which have the same variance and one of which has a variance 2–32 times greater. Noise with variance equal to one nondominant modes was also included.

6b) Dominant nonlinear mode. We perform the same experiment as 6a except that the dominant mode is nonlinear. This puts any method we use at particular risk of mixing the residual of the linear fit to the dominant nonlinear mode into the other modes.

7) Mixed. This is the most generalized of all the experiments and is designed so that the mean and spread of the projection correlation scores from this experiment can be

used to approximate the general ability of each LVM for climate-like data that consist of additive modes.

In this experiment we created 60 pseudoclimates that are randomly generated from all the mode types mentioned, with randomly drawn variances for each and with randomly drawn amounts of noise.

The number of modes in each generated pseudoclimate is randomly drawn from the distribution $P(n) \propto \text{Poisson}(n; \lambda = 8)$ $H(n - 1)$, where $H()$ is the Heaviside step function (this gave 5–12 modes per dataset).

Each drawn mode has a 45% chance of being linear, 45% nonlinear, and 10% wavelike. Each linear mode has a 50% chance of being dense or sparse. Each nonlinear mode has a 35% chance of being cyclic.

The variance for each mode is drawn from the distribution $P(\sigma^2) = \text{gamma}(\sigma^2; \alpha = 3, \beta = 5)$ (this gave variances of 0.99–58.7). The variance of the noise in each dataset was chosen to be the median variance of the modes multiplied by a factor drawn from $P(\sigma^2) \propto \text{gamma}(f; \alpha = 1, \beta = 1)H(f - 0.5)$ (this gave noise variances of 5.5–71.2).

Note that the choices for these distribution functions are necessarily arbitrary. We simply aimed to avoid creating very simple datasets, with few modes or very low levels of noise compared to the modes. We also desired a diverse set of generated pseudoclimates.

In each of the experiments there are eight generated modes (excluding experiment 7, which had a randomly drawn quantity n). Before doing greedy matching we only keep the first 12 LVM modes ($n + 2$ for experiment 7) as in practice we are unlikely to look through the sample of 7008 modes that fully describe the datasets. We also believe the first modes are more likely to be meaningful as they are more variant, slower changing, or more predictable. This also reduces the probability that a good match to the mode was found just by chance.

To provide a baseline method for correlations between random modes, we sample a new random set of dense linear vectors $\tilde{\mathbf{v}}^{(i)}$ from the same spatial covariance matrix as used to generate the input modes. Then we project the full dataset onto these column vectors as $\tilde{\mathbf{D}}^{(i)}(t) = \tilde{\mathbf{v}}^{(i)}\tilde{\mathbf{v}}^{(i)\mathrm{T}}\mathbf{X}(t)$. Therefore the modes of the baseline method are essentially random guesses given the covariance matrix.

6. Synthetic mode extraction results

From experiment 1 we find that the projection correlations between the generated modes and those discovered by each LVM have large spreads (Fig. 2). This shows that no matter the method we apply, we cannot assume that there is a precise match between any particular discovered mode and any underlying signal. However, the modes calculated by SFA and DMD had substantially better matches than those calculated by PCA and are likely to capture any additive mode more accurately if it exists.

In this experiment, the distribution of the PCA and baseline projection correlations overlap substantially. Only 27% of the PCA projection correlation values were larger than the 95th percentile of values from the baseline method (i.e., only 27% of
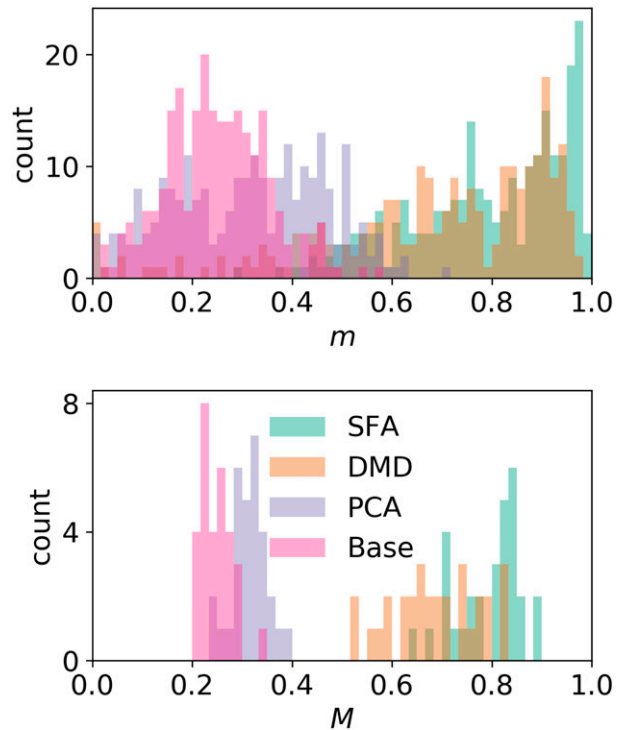

FIG. 2. Histogram of projection correlation scores from experiment 1, achieved by our three candidate mode extraction methods and a baseline. (top) The m are all of the greedily matched scores across 30 pseudoclimates of 8 modes each. (bottom) The M are the mean of the 8 greedily matched projection correlation for each of the 30 pseudoclimates.

the PCA modes had projections that were significantly better than a random guess). For SFA and DMD this value was 98% and 87% respectively.

A subset of generated modes and the matching modes from PCA, SFA, and DMD are shown in Fig. 1. These are taken from a typical dataset from experiment 2. We have plotted coastlines as an eye guide and to draw comparisons to the real application of these methods.

Figure 1a shows a mode where the match from PCA is inaccurate. PCA has failed to capture the positive area over South America and the negative areas in the Pacific. It has also exaggerated the positive areas over Australia and the Middle East. This retrieved map is qualitatively different from the generated one, and if it were to occur while analyzing observations or climate simulations then we might be led to research and propose physical mechanisms that are incorrect.

Figure 1b shows the mode of the dataset in which PCA achieved its highest projection correlation. This projection correlation is also in the top 3% of PCA projection correlations across all datasets in experiment 2. This represents a best case scenario for PCA, and even here it narrows the highly positive region over North America and misses the positive area over Australia. If this were a real dataset, this could lead us to believe that this signal connects to a localized equatorial ocean process when it does not. In both of these cases SFA and DMD
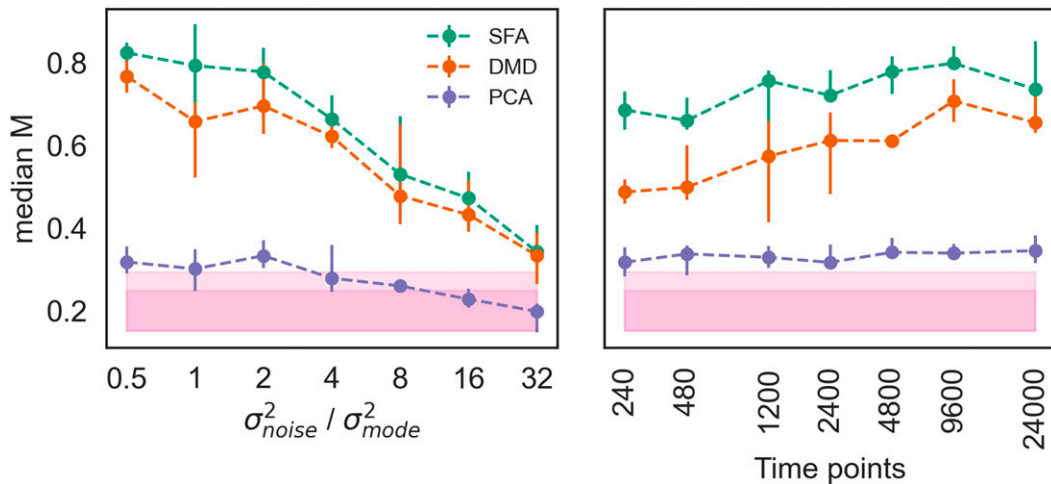
FIG. 3. The median value and uncertainty of the dataset mean projection correlation M, achieved by our three candidate LVMs for various background noise levels and number of time points, from (left) experiment 3 and (right) experiment 4. Shaded regions represent the 50% and 95% values of the baseline method from experiment 1.

capture the mode more accurately both quantitatively and qualitatively.

In Fig. 1c, PCA has suggested a dipole for a monopolar signal. It has also incorrectly linked other areas of the globe to this mode. This pole multiplication and spatial signal leaking happened for every sparse mode in all experiments, and is similar to the simplified example used in Dommenget and Latif (2002). In that study the authors discuss the DMI in relation to PCA's dipole creating bias, and question whether this phenomenon really is dipolar. We note that each of the 30 datasets in experiment 2 in this paper is broadly equivalent to one of the examples presented in Dommenget and Latif (2002). Our results here suggest that SFA and DMD would be better choices for the examples posed in their paper. The range of scores for spatially sparse and dense modes were very similar, and similar distortion of modes happens for both kinds. However the sparse modes, due to their simplicity, better highlight the qualitative significance of the distortion.

This spatial signal leaking is directly related to PCA's optimization objective. Each PC is optimized to capture as much variability as possible, so if two localized (spatially sparse) generated modes are nonorthogonal, then PCA will capture as much variance from both of them as possible in a single PC. This causes the PC to have a mix of the spatial structures of the two generated modes, and so it will not be localized.

In experiment 2 even VARIMAX rotation (von Storch and Zwiers 2001), a common variant of PCA that targets localized variability, would perform poorly on the given mixture of global and localized modes. The optimization objective of VARIMAX PCA encourages it to make most elements in $\tilde{\mathbf{v}}^{(i)}$ approximately zero, and allow the rest to have high value. This forces the estimated modes to be spatially localized and therefore would distort truly global modes. SFA and DMD have optimization objectives which do not lead them to favor either global or local modes as PCA and VARIMAX PCA do.

Figure 3 shows that all of the methods were sensitive to the level of background noise and the number of observations. This is expected as both of these parameters affect the signal-to-noise ratio.

Over the range of background noise tested, 0.5–32 times the variance of a single mode, we found that all methods had the same proportional noise sensitivity. The median value of M decreased by approximately the same fraction for a given increased noise level for all methods. Therefore SFA and DMD performed better than PCA for all noise levels tested.

Over all observation lengths tested we found only a slight sensitivity, with the equivalent observations to 2000 years of monthly data having a median value of M only 7%–30% greater than 20 years.

In experiments 5a–5c we found that nonlinear modes were estimated less well by any single LVM mode than the linear ones. Table 1 shows that regular nonlinear modes were reconstructed better than the cyclic nonlinear modes. These in turn were reconstructed better than moving wave modes, indicating that reliable detection of the latter will be a challenge with linear mode extraction methods.

Experiments 6a and 6b showed that the dominant modes were better reconstructed than the nondominant modes, and the higher their variance the better reconstructed they were. This is shown in Fig. 4 and is essentially a signal-to-noise boost for SFA and DMD and is a direct result of the PCA algorithm as discussed in appendix A.

The performance of PCA on the dominant mode approaches that of DMD and SFA if the dominant mode has a variance of about 8 or more times that the other modes. However, PCA still underperforms on the remaining modes.

Dominant modes such as ENSO have been extensively studied so any lesser modes would be of particular interest. Even in cases where we believe we have fully removed these dominant modes from our data, which may be impossible due

TABLE 1. The average of the matched projection correlation scores m_{ij}, for each mode type. The number stated is the median with the uncertainty in the last digit in parentheses. The subscript and superscript values are the difference between the median and the 15.9 and 84.1 percentile values, respectively (i.e., one standard deviation), included to show the spread and overlap in values. Some experiments were not carried out for the baseline method as these experiments were not relevant; these entries are left blank.

Experiment	Mode type	SFA	DMD	PCA	Base
1. Precision	Linear	$0.82(2)^{+0.14}_{-0.21}$	$0.73(2)^{+0.18}_{-0.26}$	$0.32(1)^{+0.15}_{-0.18}$	$0.246(5)^{+0.096}_{-0.094}$
2. Spatial sparsity	Linear	$0.84(2)^{+0.12}_{-0.18}$	$0.76(1)^{+0.18}_{-0.21}$	$0.33(1)^{+0.13}_{-0.18}$	$0.242(9)^{+0.101}_{-0.098}$
	Linear sparse	$0.84(2)^{+0.11}_{-0.17}$	$0.74(1)^{+0.15}_{-0.19}$	$0.35(2)^{+0.16}_{-0.21}$	$0.25(1)^{+0.12}_{-0.11}$
5a. Nonlinear confounding	Linear	$0.906(8)^{+0.058}_{-0.227}$	$0.84(1)^{+0.12}_{-0.30}$	$0.36(1)^{+0.13}_{-0.22}$	—
	Nonlinear	$0.831(1)^{+0.053}_{-0.062}$	$0.830(9)^{+0.045}_{-0.064}$	$0.35(2)^{+0.11}_{-0.16}$	—
5b. Nonlinear cyclic confounding	Linear	$0.78(2)^{+0.17}_{-0.24}$	$0.77(3)^{+0.16}_{-0.28}$	$0.37(2)^{+0.12}_{-0.23}$	—
	Nonlinear cyclic	$0.48(2)^{+0.19}_{-0.14}$	$0.55(2)^{+0.14}_{-0.18}$	$0.297(9)^{+0.053}_{-0.074}$	—
5c. Wave confounding	Linear	$0.78(2)^{+0.15}_{-0.25}$	$0.74(2)^{+0.20}_{-0.31}$	$0.36(1)^{+0.15}_{-0.21}$	—
	Moving wave	$0.39(3)^{+0.11}_{-0.1}$	$0.38(4)^{+0.18}_{-0.17}$	$0.29(2)^{+0.09}_{-0.06}$	—
6a. Dominant mode	Linear	$0.89(1)^{+0.07}_{-0.26}$	$0.87(2)^{+0.05}_{-0.39}$	$0.348(9)^{+0.161}_{-0.210}$	—
	Linear dominant	$0.994(2)^{+0.004}_{-0.014}$	$0.988(3)^{+0.009}_{-0.012}$	$0.95(2)^{+0.04}_{-0.29}$	—
6b. Dominant nonlinear mode	Linear	$0.89(2)^{+0.06}_{-0.24}$	$0.789(7)^{+0.142}_{-0.195}$	$0.30(1)^{+0.13}_{-0.17}$	—
	Nonlinear dominant	$0.87(1)^{+0.06}_{-0.03}$	$0.87(1)^{+0.06}_{-0.03}$	$0.87(2)^{+0.07}_{-0.22}$	—
7. Mixed	Linear	$0.85(1)^{+0.09}_{-0.29}$	$0.84(2)^{+0.10}_{-0.42}$	$0.44(2)^{+0.28}_{-0.27}$	$0.20(2)^{+0.14}_{-0.09}$
	Linear sparse	$0.74(4)^{+0.20}_{-0.34}$	$0.68(8)^{+0.25}_{-0.39}$	$0.30(2)^{+0.32}_{-0.20}$	$0.20(1)^{+0.15}_{-0.09}$
	Nonlinear	$0.57(2)^{+0.20}_{-0.28}$	$0.54(2)^{+0.22}_{-0.31}$	$0.33(2)^{+0.17}_{-0.18}$	$0.21(1)^{+0.11}_{-0.12}$
	Nonlinear cyclic	$0.43(2)^{+0.18}_{-0.25}$	$0.30(2)^{+0.23}_{-0.18}$	$0.23(2)^{+0.16}_{-0.1}$	$0.17(1)^{+0.09}_{-0.07}$
	Moving wave	$0.27(4)^{+0.18}_{-0.15}$	$0.21(2)^{+0.16}_{-0.12}$	$0.22(1)^{+0.12}_{-0.11}$	$0.109(9)^{+0.062}_{-0.059}$

to our additive assumption of mode imprints (discussed in appendix C), SFA and DMD would still be better choices.

Finally in experiment 7 we found that SFA and DMD significantly outperformed PCA on every mode type across the generated pseudoclimates. The exception to this is moving wave modes, which were not estimated well by one linear mode from any method.

Out of 480 modes SFA, DMD, and PCA extracted more similar modes to the generated modes than a random mode baseline in 78%, 67%, and 44% of modes respectively when taking account of different thresholds for different mode types. SFA, DMD, and PCA recovered the most accurate reconstruction of 326, 94, and 60 modes respectively and had the highest mean projection correlation M in 53, 7, and 0 datasets. A full comparison of mode projection correlation scores is shown in Fig. 5.

Our results so far have shown us that SFA and DMD are significantly better at estimating additive modes in a range of scenarios. Using the results of these experiments we should expect SFA and DMD to perform better than PCA in a range of conditions where we have few expectations about the composition of the modes that exist.

7. Application to climate data

SFA and DMD recover known additive modes in our synthetic data more accurately than PCA. However, there is no robust or principled way to test whether any LVM mode matches a true synthetic mode, without already knowing the ground truth. We have shown that LVMs risk mixing these modes, and in real climate data there is no guarantee that additive ground truth modes exist, and no objective way to distinguish which method's extracted modes are most realistic.

Some methods such as testing against a Hasselmann-type red noise model (Hasselmann 1976) have been proposed, and applied to PCA (Dommenget 2007), and could be easily extended for use on the DMD and SFA. However, this only tests whether the discovered modes could be explained by a red noise model.

We do not propose any method of guarding against discovering mixed up modes. However we do note that from our Monte Carlo experiments that SFA and DMD are more likely than PCA to correctly separate any independent additive modes if they do exist. So if any tool is to be used to find additive modes or explore a reduced dimension version of climate data, as is often done in this research area, SFA and DMD would be significantly better choices than PCA.

In Fig. 6 we have applied our three LVMs to monthly surface temperature anomalies from HadCRUT4 observational medians (Morice et al. 2012), from a single CESM2 historically forced run, and from a single CESM2 preindustrial control run (Danabasoglu 2019) to further demonstrate these methods on real data. We subtracted the seasonal cycle from each dataset and applied root-cosine-latitude weighting prior to analysis. To use the observational data, which have missing values, we apply probabilistic PCA (Tipping and Bishop 1999) and truncate to the first 36 dimensions before applying the different LVMs.

After we apply the LVMs to each dataset we perform, as done in the synthetic case, greedy matching between modes discovered by the different methods by calculating the projection correlations between modes. Figure 6 shows the set of modes that appeared the most like global warming, and the most like a broad ENSO pattern. These patterns are nonorthogonal to
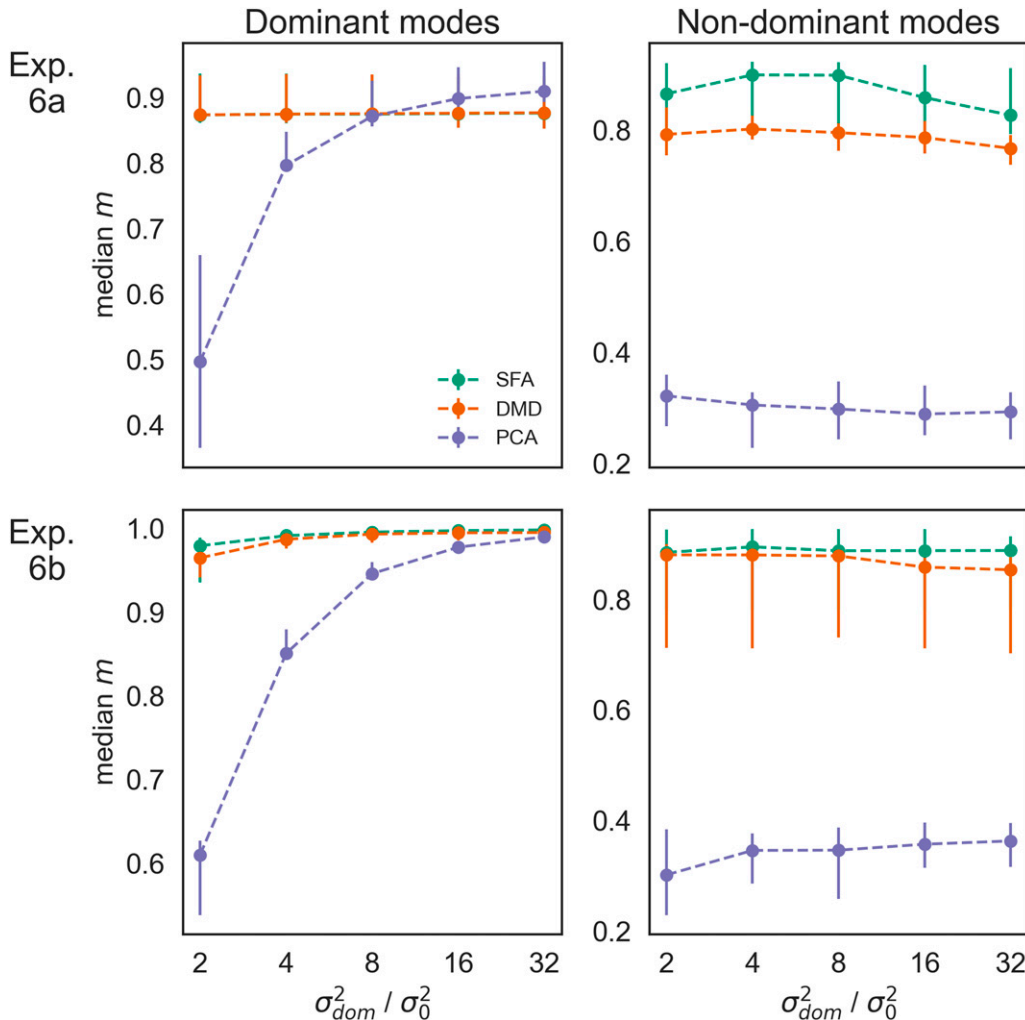
FIG. 4. The median value and uncertainty of the projection correlation scores m achieved by the three LVMs for the dominant modes and nondominant modes. The x axis shows the variance ratio between the dominant and the nondominant modes. Results are shown from (top) experiment 6a with all linear modes and (bottom) experiment 6b with nonlinear dominant modes and linear nondominant modes.

each other and evolve on different time scales. ENSO is spatially more localized, but may have teleconnections to other parts of the globe, is nonlinear, and should have a clear spectral peak. Global warming is spatially global and more linear than ENSO (see pattern scaling discussion; Collins et al. 2013), and both are reasonably well understood. Therefore we can consider differences in the LVM modes with little uncertainty of what the patterns should look like, as would be the case if we were considering combinations of modes such as ENSO and the PDO.

Usually PCA is applied only to the Pacific region to focus on the spatial pattern of ENSO. However, we apply these LVMs globally to find ENSO. This is a useful qualitative test of the methods, as we may wish to find global ENSO teleconnections.

Usually global warming is not considered a mode and is removed before modes are extracted. However we could consider it an emission-driven mode, or a mode of the anthroposphere. There is also some complexity in removing the global warming

trend in a way suitable for these mode extraction techniques (an equivalent problem of correctly removing the seasonal cycle is discussed in appendix C).

The first row of the figure shows the HadCRUT4 derived modes that appear most like global warming. These matched modes happen to be the slowest, most predictable, and most variant as defined by our three respective LVMs. The SFA and DMD time series have a much greater proportion of their variance at lower frequencies, and therefore the trend is much easier to see without any temporal averaging. SFA has reduced the weighting over landmasses compared to the other two methods. This is to be expected as the temperature variance over land has much more power at higher frequencies. This spatial pattern therefore gives the optimally slow time signal.

The third row shows the most ENSO-like modes recovered from the observations. We note that the DMD and particularly the SFA modes are spatially simpler than the PCA mode.
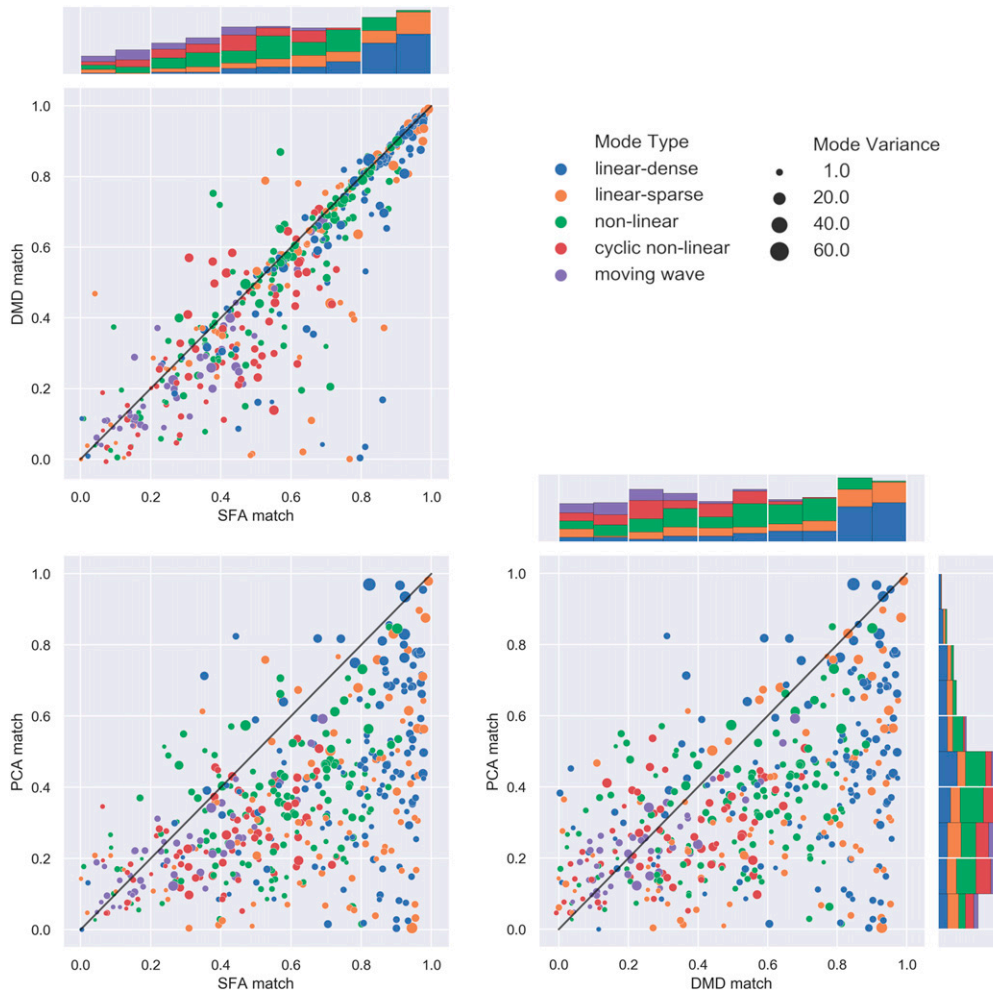
FIG. 5. The projection correlation scores between the input modes and those discovered by the three LVMs. These results are from experiment 7. The black diagonal shows where the scores are even between the different LVMs. In the plot margins are the distributions for the three different LVMs and are broken down by mode type. In the distributions the mode types are stacked from bottom to top in order: linear-dense, linear sparse, nonlinear, cyclic nonlinear, and moving wave.

The ENSO-like modes discovered by SFA, DMD, and PCA were respectively the second slowest, second most predictable, and seventh most variant comprising only 4% of global variance. PCA suggests that the positive phase of this mode is associated with warmer European temperatures and warmer temperatures in parts of Asia. The power spectrum of the normalized time series of each mode shows that SFA and DMD have a higher proportion of their total power at lower frequencies, and they more clearly peak in the 3–7-yr range, which is consistent with ENSO. Considering the two CESM2 datasets the SFA and DMD ENSO-like modes have a more consistent spatial structure than the equivalent PCA modes.

Analysis of the historically forced run (CESM2 Hist) is unable to separate the ENSO-like signal from global warming, and mixes in some of ENSO's low-frequency signal in all methods. The spatial pattern associated with this global warming mode has a strong component in the ENSO region, and the time series has enhanced variability in the 3–7-yr band, as well as some apparent multidecadal variability. The discovered ENSO-like mode in this dataset is spatially narrower than those of the other two datasets, and has less power at a period of 5 years or greater than the other datasets. This is because the longer-term ENSO variability is already accounted for.

This result is an interesting contrast to the observational data, where ENSO and global warming were separated well. This mixing happened consistently in all runs tested from the CESM2 historically forced ensemble and so was not specific to this one dataset. This could simply be because the CESM2 data do not contain observational noise as in HadCRUT4, and so this mixed signal is slower than the two separated signals. It may also suggest differences in the relative strength and variability of ENSO and global warming between model and observations, which we do not investigate further here.
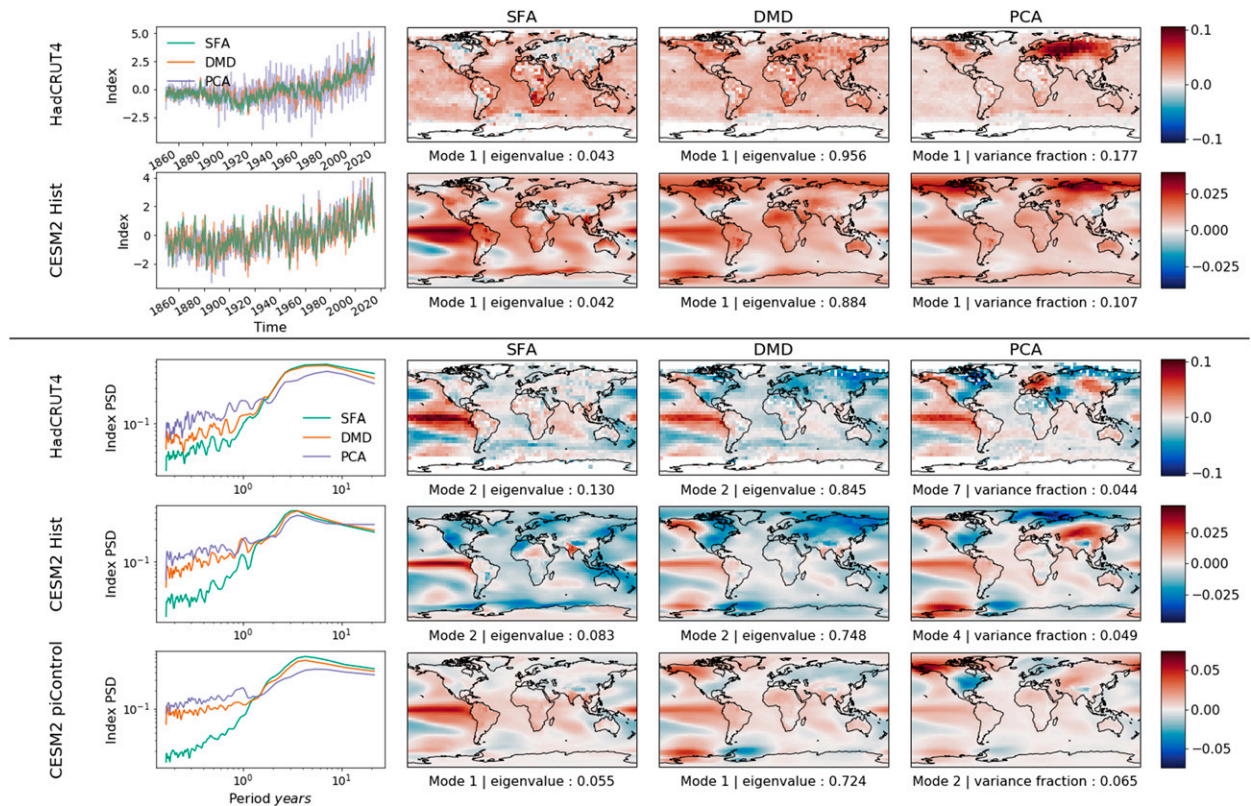
FIG. 6. The modes discovered by three different latent variable methods, SFA, DMD, and PCA, computed on HadCRUT4 observational data, a single CESM2 historically forced run, and a single CESM2 preindustrial control run. Above the line are the most global warming–like modes discovered by each method along with their time series. Below the line are the most ENSO-like patterns and their normalized power spectral density. Each row shows the spatial patterns for the dataset labeled at the left of the row. The time series or power spectral density of the normalized index time series for the three modes is shown on the left, and the color bar for the entire row is on the right. Below each spatial pattern is the mode number ordered by the normal LVM criteria (see section 3d) and the eigenvalue for each mode (or in the case of PCA the fraction of total variance).

We also carried out the same analysis on annual as opposed to monthly means and found very similar results (not shown).

Figure 7 shows the different LVMs applied to datasets derived from an ensemble of CESM2 historically forced simulations. The global warming mode found by SFA in the ensemble mean data is remarkably similar to the global warming mode extracted from the observational data in Fig. 6. This may present an opportunity to use SFA alone to remove the global warming signal from observations instead of relying on subtracting a climate model ensemble mean.

In Fig. 7 the *detrended* data are the *single run* data with the ensemble mean subtracted. The SFA ENSO-like mode from Fig. 7 detrended is very similar to that from the preindustrial (piControl) run of Fig. 6. This similarity can also be seen in the DMD modes, but not the PCA modes.

The fact that DMD and SFA find very similar modes in the detrended historical and preindustrial control runs while PCA did not is another reason to prefer these alternate techniques. We should expect these datasets to have the same modes, unless historical forcing influences strongly influence climate variability. However, both methods deal differently with the

variance difference between land and ocean, something not tested in synthetic cases.

8. Conclusions

In this study we have developed a Monte Carlo (MC) method to generate additive space-time modes. Methods of mode extraction are often developed analytically with highly idealized assumptions of the form of modes and the background noise. Using our MC generated datasets allows us to quantitatively assess how accurately these mode extraction methods perform when their assumptions are not strictly met. This is important as the extracted modes direct our research into the physical mechanisms that imprint these mode signals.

We have shown that two alternate mode extraction methods, SFA and DMD, vastly outperform the commonly used method PCA (also known as empirical orthogonal functions) in separating multiple additive synthetic modes. Both SFA and DMD methods target modes that are temporally autocorrelated, which is the case for many modes of climate variability in data. Even the North Atlantic Oscillation, which has fairly
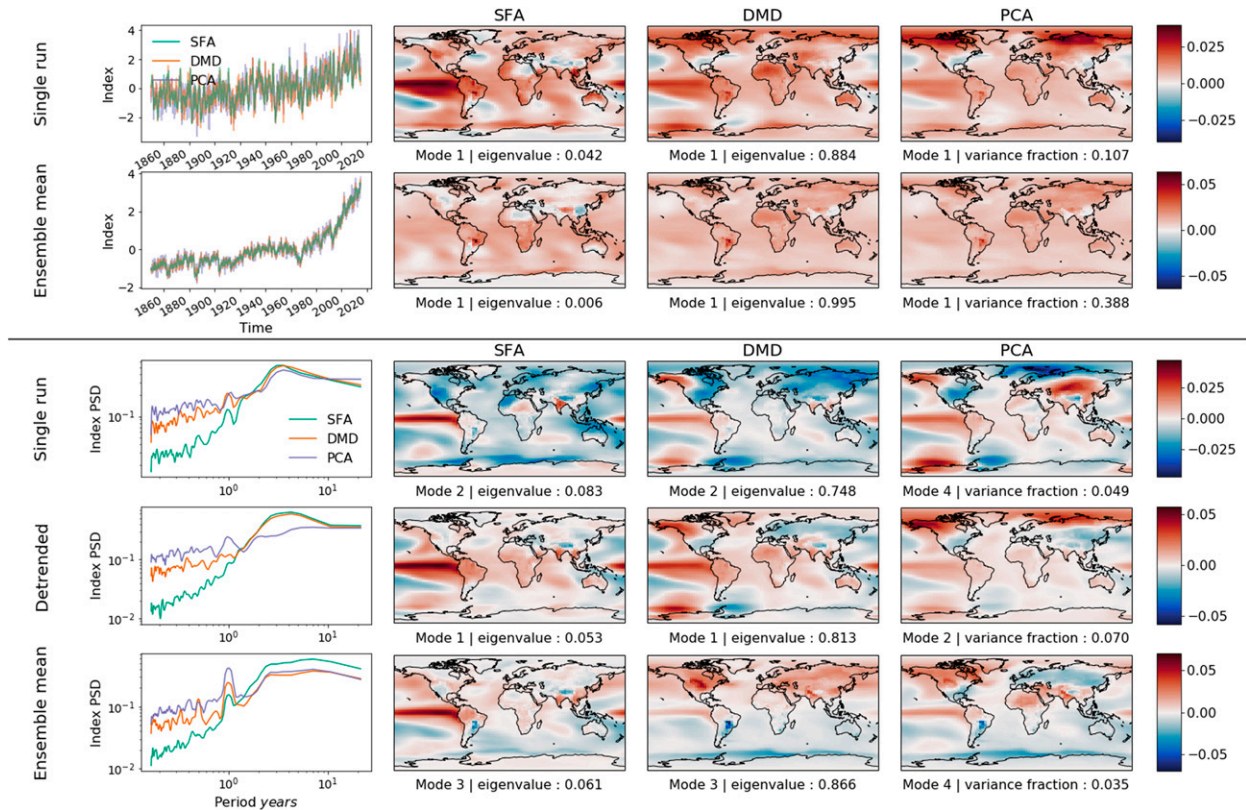
FIG. 7. The modes discovered by three different latent variable methods, SFA, DMD, and PCA, on an ensemble of CESM2 historically forced simulations. The datasets used are a single run that is identical to CESM2 Hist in Fig. 6, the ensemble mean of 10 historically forced run members, and the single run with this ensemble mean subtracted, which we refer to as *detrended*. All further detail is the same as in Fig. 6. See appendix C for more discussion on power spectra 1-yr peaks despite seasonal detrending.

white time series, exhibits substantial decadal variability (Iles and Hegerl 2017).

This performance gap may have been predicted from definitions of the mode extraction methods, but importantly we have quantified this gap under various scenarios, and presented an example of how this could misdirect research efforts.

These two techniques, and particularly SFA, significantly outperform PCA whether the synthetic modes are linear, or nonlinear, or a mixture. This was true whether the generated climate included one or more modes with higher variance and also where the signal-to-noise ratio is small, as might be expected in climate observations and simulations.

We find that a particularly problematic property of PCA is its inability to discover localized modes, always deforming them into global modes. In all synthetic datasets it extracted false dipoles instead of monopoles, which may lead to serious misinterpretation of extracted modes. More generally we found that PCA tends to mix independent spatial regions into single modes.

Overall, this evidence based on pseudoclimate data may suggest that SFA and DMD are better suited to analyzing real climate data from observations and simulations, assuming that climate modes could be reasonably approximated as additive.

When applied to observational data we found that SFA and DMD produced less noisy time indices for a global warming signal. The global warming pattern from observational data matched well with that from an ensemble mean of historically forced simulations. This suggests that SFA and/or DMD could be used to remove the global warming signal from a dataset without the need of an external reference such as a model ensemble.

ENSO-like modes discovered by SFA and DMD also appear to be more realistic than those found by PCA. Their time indices had a higher proportion of variance in the 3–7-yr period band and their spatial patterns are much more localized to the equatorial pacific, as would be expected from an ENSO signal. The ENSO-like PCA modes had anomalous projections onto spurious parts of the globe which were not consistent between datasets, whereas SFA and DMD were much more self-consistent across datasets. This suggests that comparison between modes derived from different data sources would be more meaningful using SFA and DMD.

APPENDIX A

PCA and Nonorthogonal Modes

We mention in main the text that PCA (EOFs) cannot recover any true mode signal if these modes are nonorthogonal. This is contrary to the seemingly common assumption that the first couple of principal components, as ordered by decreasing variance, are reliable while the later ones may be wrong due to being constrained to be orthogonal to the earlier ones. We show here analytically, that if the true underlying modes are nonorthogonal then even the first principal component (PC) will be incorrect.

Suppose we have a set of k true modes of climate variability $\{\mathbf{v}^{(i)} | i \in \mathbb{N}, i \le k\}$. These modes might be the signal imprint on surface temperature anomalies, precipitation, or pressure anomalies etc. For simplicity we will suppose that these modes are space–time separable (linear), meaning that each mode can be expressed as a column vector, and we will assume that they are not necessarily orthogonal but are normalized. We will define a basis matrix of all of the linear modes $\mathbf{V}$, where

$$\mathbf{V} = \begin{bmatrix} | & | & & | \\ \mathbf{v}^{(1)} & \mathbf{v}^{(2)} & \cdots & \mathbf{v}^{(k)} \\ | & | & & | \end{bmatrix}. \tag{A1}$$

Each of these modes has a corresponding time series $\mathbf{z}^{(i)}$ that states how much of that mode is expressed at any time. These time series, $\{\mathbf{z}^{(i)} | i \in N, i \le k\}$, can be thought of as climate indices associated with each mode. We will define that they have zero mean and are not correlated. These indices can also be stacked into a matrix of column vectors.

$$\mathbf{Z} = \begin{bmatrix} | & | & & | \\ \mathbf{z}^{(1)} & \mathbf{z}^{(2)} & \cdots & \mathbf{z}^{(k)} \\ | & | & & | \end{bmatrix}. \tag{A2}$$

The full dataset matrix $\mathbf{X}$ that represents observations of the state of the climate, taken at regular time intervals and at defined locations, is therefore

$$\mathbf{X} = \mathbf{V}\mathbf{Z}^{\mathrm{T}} \tag{A3}$$

and has a shape of $[N, T]$, where T is the number of time snapshots and N is the number of measurements taken at each time. From our definitions of $\mathbf{v}^{(i)}$ and $\mathbf{z}^{(i)}$, the data matrix $\mathbf{X}$ has a zero mean in the time dimension and so can be used to represent measurement anomalies. Here we assume that measurements are composed from only the superposition of linear modes with no noise, either red or white.

To show that PCA will not recover any of the true nonorthogonal modes, we consider performing PCA by eigendecomposition of the covariance matrix $\boldsymbol{\Sigma} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$. Any column vector $\mathbf{u}$ that is an eigenvector of $\boldsymbol{\Sigma}$ has the property that

$$\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{u} - \lambda\mathbf{u} = 0. \tag{A4}$$

We use the substitution that $\mathbf{X}\mathbf{X}^{\mathrm{T}} = \mathbf{V}\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\mathbf{V}^{\mathrm{T}}$ and note that since the index time series are uncorrelated, $\mathbf{z}^{(i)\mathrm{T}}\mathbf{z}^{(j)} = \delta_{ij}\sigma_i^2$. In this case $\boldsymbol{\Gamma} = \mathbf{Z}^{\mathrm{T}}\mathbf{Z}$ is the diagonal covariance matrix where $\boldsymbol{\Gamma} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)$. Following this,

$$\mathbf{V}\boldsymbol{\Gamma}\mathbf{V}^{\mathrm{T}}\mathbf{u} - \lambda\mathbf{u} = 0. \tag{A5}$$

And since $\boldsymbol{\Gamma}$ is diagonal,

$$\left[\sum_{i=1}^{k} \sigma_i^2 (\mathbf{v}^{(i)\mathrm{T}}\mathbf{u})\mathbf{v}^{(i)} \right] - \lambda\mathbf{u} = 0. \tag{A6}$$

Using the equation above we can see that the eigenvector $\mathbf{u}$ can only be an element of the original mode set $\mathbf{u} = \mathbf{v}^{(n)}$, if at least one of the following conditions is true:

1) $\sigma_i^2 = \delta_{in}\lambda = \delta_{in}\sigma_i^2$ (i.e., there is only one mode).
2) $\mathbf{v}^{(i)\mathrm{T}}\mathbf{v}^{(n)} = \delta_{in}$ (i.e., the mode $\mathbf{u} = \mathbf{v}^{(n)}$ is orthogonal to every other mode).
3) $\sum_{i=1, i \ne n}^{k} \sigma_i^2 [\mathbf{v}^{(i)\mathrm{T}}\mathbf{v}^{(n)}]\mathbf{v}^{(i)} = (\lambda - \sigma_n^2)\mathbf{v}^{(n)}$ (i.e., extreme coincidence). The mode $\mathbf{u} = \mathbf{v}^{(n)}$ is an extremely specific and unlikely linear combination of the other modes.

Therefore in practice PCA cannot find even a single correct mode if the modes are nonorthogonal. This is true whether we consider the first PC with the most variance or the later ones. PCA is a greedy algorithm and so as much variance as possible from all of the input modes is captured in each subsequent PC, and so the first PC is not the true mode with the most variance.

There is no physical reason for climate modes to be orthogonal. It is also extremely unlikely for climate modes to be arranged "just so" as required by condition 3. Therefore if our climate is composed of multiple modes of variability such as ENSO, NAO, PDO, AMO, DMI, etc., then PCA cannot extract the "real" modes for any of these and will return PCs which are impure mixtures of these modes as well as the added complication of noise. We note that PDO is defined by the leading PC of SST anomalies the North Pacific basin and so is likely mischaracterized.

A limited case where the first PC could approximately match an input mode is where the variance in one true mode is much greater than the variance of the other modes, $\sigma_n^2 \gg \sigma_i^2 \,\forall\, i \ne n$. However, in this case all other modes that are not completely orthogonal to the first mode will be distorted by it, absorbing some small part of its variance to the point where they may be corrupted and spatially unrecognizable. This is in fact the

reason why when applying PCA we need to remove the seasonal component. This is by far the strongest mode of climate variability and if we do not remove it we have this stated problem.

It is easy to follow this reasoning to modes further down the variance hierarchy. The next most dominant signal of any climate data, if global warming is removed, is ENSO. So we should expect the ENSO signal to dominate other modes thus projecting parts of ENSO into them. This of course must be a major concern for teleconnection studies as spurious lead–lag relations might be the product of this variability leakage.

The same reasoning as presented above can be applied to extended EOFs to say that they cannot return the correct spatiotemporal modes if the system contains multiple spatiotemporal modes which are nonorthogonal. However since the modes are extended into the time dimension, it is more likely that any modes will be approximately orthogonal as they will likely have different dominant frequencies.

APPENDIX B

Equivalence to the Projection Correlation

In the text we state that when the discovered and generated modes are both linear, then the projection correlation is equal to the Pearson correlation coefficient between their time series multiplied by the dot product between their normalized spatial vectors.

Assuming the modes are linear so $\mathbf{D}^{(i)}(t) = z^{(i)}(t)\mathbf{v}^{(i)}$, then Eq. (9) becomes

$$
\begin{aligned}
m_{ij} &= \frac{\langle [z^{(i)}(t)\mathbf{v}^{(i)} - \langle z^{(i)}(t)\mathbf{v}^{(i)}\rangle] \cdot [\tilde{z}^{(j)}(t)\tilde{\mathbf{v}}^{(j)} - \langle \tilde{z}^{(j)}(t)\tilde{\mathbf{v}}^{(j)}\rangle]\rangle}{\boldsymbol{\sigma}^{(i)}\tilde{\boldsymbol{\sigma}}^{(j)}} \\
&= \frac{\langle [z^{(i)}(t) - \langle z^{(i)}(t)\rangle]\mathbf{v}^{(i)} \cdot [\tilde{z}^{(j)}(t) - \langle \tilde{z}^{(j)}(t)\rangle]\rangle \tilde{\mathbf{v}}^{(j)}}{\boldsymbol{\sigma}^{(i)}\tilde{\boldsymbol{\sigma}}^{(j)}} \\
&= \frac{\langle [z^{(i)}(t) - \langle z^{(i)}(t)\rangle][\tilde{z}^{(j)}(t) - \langle \tilde{z}^{(j)}(t)\rangle]\rangle}{\boldsymbol{\sigma}^{(i)}\tilde{\boldsymbol{\sigma}}^{(j)}}\mathbf{v}^{(i)} \cdot \tilde{\mathbf{v}}^{(j)}.
\end{aligned} \tag{B1}
$$

If we make the same substitutions into Eq. (10), and use the fact that the spatial vectors are normalized, then it becomes

$$
\begin{aligned}
\sigma^{(i)} &= [\langle z^{(i)}(t)\mathbf{v}^{(i)} \cdot z^{(i)}(t)\mathbf{v}^{(i)}\rangle - \langle z^{(i)}(t)\mathbf{v}^{(i)}\rangle \cdot \langle z^{(i)}(t)\mathbf{v}^{(i)}\rangle]^{1/2} \\
&= [\langle z^{(i)}(t)z^{(i)}(t)\rangle - \langle z^{(i)}(t)\rangle\langle z^{(i)}(t)\rangle]^{1/2}.
\end{aligned} \tag{B2}
$$

Once Eq. (B2) is substituted into Eq. (B1) it becomes obvious that the term on the left of Eq. (B1) is indeed the Pearson correlation coefficient between the modes' time series and the term on the right is the dot product between their normalized spatial vectors.

APPENDIX C

The Difficulty of Removing a Nonlinear Yearly Cycle

In the main text we claim that it may not be possible to fully remove the yearly cycle or other dominant modes from our climate signal if they are nonadditive. This is what we observe in all power spectra in Figs. 6 and 7 and is particularly noticeable in the final row of 7. Despite deseasonalizing the data as usual where we subtract from each month J the mean value of all months J, this peak remains.

Lets assume we have some climate data $\mathbf{x}(t)$ that are made up of some mode projection $\mathbf{D}_s(t)$ and some other part $\mathbf{x}'(t)$, which is the sum of the other modes and noise, so that

$$
\mathbf{x}(t) = \mathbf{x}'(t) + \mathbf{D}_s(t). \tag{C1}
$$

Here $\mathbf{D}_s(t)$ is seasonal such that $\mathbf{D}_s(t + s) = \mathbf{D}_s(t)$ for some seasonal time length s.

If $\mathbf{D}_s(t)$ were the seasonal cycle we would remove this using the usual method to get the perfectly deseasonalized data $\mathbf{x}'(t)$. However, if the seasonality is not additive as in

$$
\mathbf{x}(t) = \mathbf{x}'(t) + \mathcal{A}_s(t)\mathbf{x}'(t) + \mathbf{D}_s(t), \tag{C2}
$$

where $\mathcal{A}_s(t) = \mathcal{A}_s(t + s)$ is some seasonal scalar function, then the seasonality cannot be removed as normal. In the climate system this could be a higher summer or winter variability and $\mathcal{A}_s(t)$ could even be a matrix operator to account for spatially heterogeneous variability effects. If we applied the normal method of deseasonalizing where we subtract the monthly means we would get $\mathbf{x}'(t) + \mathcal{A}_s(t)\mathbf{x}'(t)$.

Since all of our LVM methods as well as calculating the power spectrum involve taking some form of the inner product of the data with itself, this will maintain some seasonal information as $\langle \mathbf{x}'(t)\mathbf{x}'(t)^{\mathrm{T}}\rangle$ is replaced by

$$
\langle [1 + \mathcal{A}_s(t)]\mathbf{x}'(t)\mathbf{x}'(t)^{\mathrm{T}}[1 + \mathcal{A}_s(t)^{\mathrm{T}}]\rangle. \tag{C3}
$$

If $\mathcal{A}_s(t)$ is simply a scalar factor as opposed to a matrix then this is $\langle [1 + 2\mathcal{A}_s(t) + \mathcal{A}_s(t)^2]\mathbf{x}'(t)\mathbf{x}'(t)^{\mathrm{T}}\rangle$. This explains the one year peak observed in the normalized power spectra in Figs. 6 and 7 despite the data being deseasonalized in the usual way.

This kind of seasonality may affect our calculated modes as the factor $1 + 2\mathcal{A}_s(t) + \mathcal{A}_s(t)^2$ weights anomalies by time of year.

REFERENCES

Ashok, K., S. K. Behera, S. A. Rao, H. Weng, and T. Yamagata, 2007: El Niño Modoki and its possible teleconnection. *J. Geophys. Res. Oceans*, **112**, C11007, https://doi.org/10.1029/2006JC003798.

Christensen, J. H., and Coauthors, 2013: Climate phenomena and their relevance for future regional climate change. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al. Eds., Cambridge University Press, 1217–1308.

Collins, M., and Coauthors, 2013: Long-term climate change: Projections, commitments and irreversibility. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 1029–1136.

Dahleh, M., M. A. Dahleh, and G. Verghese, 2011: 6.241j: Lectures on dynamic systems and control. Massachusetts Institute of Technology: MIT OpenCourseWare, 354 pp., https://viterbi-web.usc.edu/~mihailo/courses/ee585/f17/mit-notes//mit-notes.pdf.

Danabasoglu, G., 2019: NCAR CESM2 model output prepared for CMIP6 CMIP. Earth System Grid Federation, accessed 2021, https://doi.org/10.22033/ESGF/CMIP6.2185.

DelSole, T., 2001: Optimally persistent patterns in time-varying fields. *J. Atmos. Sci.*, **58**, 1341–1356, https://doi.org/10.1175/1520-0469(2001)058<1341:OPPITV>2.0.CO;2.

Demo, N., M. Tezzele, and G. Rozza, 2018: PyDMD: Python dynamic mode decomposition. *J. Open Source Softw.*, **3**, 530, https://doi.org/10.21105/joss.00530.

Dommenget, D., 2007: Evaluating EOF modes against a stochastic null hypothesis. *Climate Dyn.*, **28**, 517–531, https://doi.org/10.1007/s00382-006-0195-8.

——, and M. Latif, 2002: A cautionary note on the interpretation of EOFs. *J. Climate*, **15**, 216–225, https://doi.org/10.1175/1520-0442(2002)015<0216:ACNOTI>2.0.CO;2.

Goodfellow, I., Y. Bengio, and A. Courville, 2016: *Deep Learning*. MIT Press, 800 pp., http://www.deeplearningbook.org.

Hamilton, J. D., 1994: *Time Series Analysis*. Princeton University Press, 816 pp.

Hasselmann, K., 1976: Stochastic climate models Part I. Theory. *Tellus*, **28**, 473–485, https://doi.org/10.1111/j.2153-3490.1976.tb00696.x.

——, 1988: PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.*, **93**, 11 015–11 021, https://doi.org/10.1029/JD093iD09p11015.

Hoyer, S., and J. Hamman, 2017: xarray: N-D labeled arrays and datasets in Python. *J. Open Res. Software*, **5**, 10, https://doi.org/10.5334/jors.148.

Iles, C., and G. Hegerl, 2017: Role of the North Atlantic Oscillation in decadal temperature trends. *Environ. Res. Lett.*, **12**, 114010, https://doi.org/10.1088/1748-9326/aa9152.

Johnson, N. C., D. C. Collins, S. B. Feldstein, M. L. L'Heureux, and E. E. Riddle, 2014: Skillful wintertime North American temperature forecasts out to 4 weeks based on the state of ENSO and the MJO. *Wea. Forecasting*, **29**, 23–38, https://doi.org/10.1175/WAF-D-13-00102.1.

Kessy, A., A. Lewin, and K. Strimmer, 2018: Optimal whitening and decorrelation. *Amer. Stat.*, **72**, 309–314, https://doi.org/10.1080/00031305.2016.1277159.

Kramer, M. A., 1991: Nonlinear principal component analysis using autoassociative neural networks. *AIChE J.*, **37**, 233–243, https://doi.org/10.1002/aic.690370209.

Kutz, J. N., S. L. Brunton, B. W. Brunton, and J. L. Proctor, 2016a: Dynamic mode decomposition: An introduction. *Dynamic Mode Decomposition*, SIAM, 1–24, https://doi.org/10.1137/1.9781611974508.ch1.

——, X. Fu, and S. L. Brunton, 2016b: Multiresolution dynamic mode decomposition. *SIAM J. Appl. Dyn. Syst.*, **15**, 713–735, https://doi.org/10.1137/15M1023543.

Monahan, A. H., J. C. Fyfe, M. H. Ambaum, D. B. Stephenson, and G. R. North, 2009: Empirical orthogonal functions: The medium is the message. *J. Climate*, **22**, 6501–6514, https://doi.org/10.1175/2009JCLI3062.1.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophys. Res.*, **117**, D08101, https://doi.org/10.1029/2011JD017187.

Newman, M., and Coauthors, 2016: The Pacific decadal oscillation, revisited. *J. Climate*, **29**, 4399–4427, https://doi.org/10.1175/JCLI-D-15-0508.1.

NOAA/OAR/ESRL PSL, 2019: Climate timeseries. Accessed 8 July 2019, https://psl.noaa.gov/gcos_wgsp/Timeseries/index.html.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.*, **110**, 699–706, https://doi.org/10.1175/1520-0493(1982)110<0699:SEITEO>2.0.CO;2.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830.

Rasmussen, C. E., and C. K. Williams, 2006: *Gaussian Processes for Machine Learning*, Vol. 2. MIT Press, 248 pp.

Runge, J., and Coauthors, 2015: Identifying causal gateways and mediators in complex spatio-temporal systems. *Nat. Commun.*, **6**, 8502, https://doi.org/10.1038/ncomms9502.

Schmid, P. J., 2010: Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.*, **656**, 5–28, https://doi.org/10.1017/S0022112010001217.

Timmermann, A., and Coauthors, 2018: El Niño–Southern Oscillation complexity. *Nature*, **559**, 535–545, https://doi.org/10.1038/s41586-018-0252-6.

Tipping, M. E., and C. M. Bishop, 1999: Probabilistic principal component analysis. *J. Roy. Stat. Soc.*, **61B**, 611–622, https://doi.org/10.1111/1467-9868.00196.

Tu, J. H., C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, 2014: On dynamic mode decomposition: Theory and applications. *J. Comput. Dyn.* **1**, 391–421, https://doi.org/10.3934/jcd.2014.1.391.

von Storch, H., and F. W. Zwiers, 2001: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.

Wanner, H., S. Brönnimann, C. Casty, D. Gyalistras, J. Luterbacher, C. Schmutz, D. B. Stephenson, and E. Xoplaki, 2001: North Atlantic Oscillation—Concepts and studies. *Surv. Geophys.*, **22**, 321–381, https://doi.org/10.1023/A:1014217317898.

Weisheimer, A., N. Schaller, C. O'Reilly, D. A. MacLeod, and T. Palmer, 2017: Atmospheric seasonal forecasts of the twentieth century: Multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution. *Quart. J. Roy. Meteor. Soc.*, **143**, 917–926, https://doi.org/10.1002/qj.2976.

Wills, R. C., T. Schneider, J. M. Wallace, D. S. Battisti, and D. L. Hartmann, 2018: Disentangling global warming, multidecadal variability, and El Niño in Pacific temperatures. *Geophys. Res. Lett.*, **45**, 2487–2496, https://doi.org/10.1002/2017GL076327.

Wiskott, L., and T. J. Sejnowski, 2002: Slow feature analysis: Unsupervised learning of invariances. *Neural Comput.*, **14**, 715–770, https://doi.org/10.1162/089976602317318938.

Wu, Z., B. Wang, J. Li, and F.-F. Jin, 2009: An empirical seasonal prediction model of the East Asian summer monsoon using ENSO and NAO. *J. Geophys. Res.*, **114**, D18120, https://doi.org/10.1029/2009JD011733.

Zhang, C., 2005: Madden-Julian oscillation. *Rev. Geophys.*, **43**, RG2003, https://doi.org/10.1029/2004RG000158.

Zito, T., N. Wilbert, L. Wiskott, and P. Berkes, 2009: Modular toolkit for data processing (MDP): A Python data processing framework. *Front. Neuroinform.*, **2**, 8, https://doi.org/10.3389/neuro.11.008.2008.