

# Comparing low-frequency and intermittent variability in comprehensive climate models through nonlinear Laplacian spectral analysis

Dimitrios Giannakis<sup>1</sup> and Andrew J. Majda<sup>1</sup>

Received 9 March 2012; revised 23 April 2012; accepted 26 April 2012; published 30 May 2012.

[1] Nonlinear Laplacian spectral analysis (NLSA) is a recently developed technique for spatiotemporal analysis of high-dimensional data, which represents temporal patterns via natural orthonormal basis functions on the nonlinear data manifold. Through such basis functions, determined efficiently via graph-theoretic algorithms, NLSA captures intermittency, rare events, and other nonlinear dynamical features which are not accessible through linear approaches (e.g., singular spectrum analysis (SSA)). Here, we apply NLSA to study North Pacific SST monthly data from the CCSM3 and ECHAM5/MPI-OM models. Without performing spatial coarse graining (i.e., operating in ambient-space dimensions up to  $1.6 \times 10^5$  after lagged embedding), or seasonal-cycle subtraction, the method reveals families of periodic, low-frequency, and intermittent spatiotemporal modes. The intermittent modes, which describe variability in the Western and Eastern boundary currents, as well as variability in the subtropical gyre with year-to-year reemergence, are not captured by SSA, yet are likely to have high significance in a predictive context and utility in cross-model comparisons. **Citation:** Giannakis, D., and A. J. Majda (2012), Comparing low-frequency and intermittent variability in comprehensive climate models through nonlinear Laplacian spectral analysis, *Geophys. Res. Lett.*, *39*, L10710, doi:10.1029/2012GL051575.

## 1. Introduction

[2] The ability to detect physically-meaningful spatiotemporal patterns of variability in the atmosphere–ocean climate system is crucial for enhancing our understanding and prediction of a host of phenomena, including regime shifts in oceanic circulation [Mantua and Hare, 2002], tropical-extratropical interactions on intraseasonal [Vitart and Jung, 2010] to decadal [Newman et al., 2003] time-scales, and other applications of wide interest. Such phenomena are governed by nonlinear dynamical laws, and influenced strongly by external factors (e.g., solar forcing), meaning that the dynamically-important features of these systems are not necessarily those carrying a high energy content (explained

variance). For instance, *Crommelin and Majda* [2004] demonstrate that the leading empirical orthogonal functions (EOFs) cannot reproduce chaotic regime transitions between zonal and blocked states in a reduced model of the atmosphere; *Overland et al.* [2010] allude to the importance of intermittent events of 1–2 yr duration in Northern Hemisphere ocean variability. The role of low-energy, short-time events has also been highlighted in the fluid dynamics literature [Aubry et al., 1993]. In general, classical linear methods, such as singular spectrum analysis (SSA) [Vautard and Ghil, 1989; Aubry et al., 1991; Ghil et al., 2002] and its variants, are likely to have poor skill in extracting dynamically-important, low-energy modes from large-scale numerical or observational climate datasets.

[3] A variety of nonlinear principal components analysis (NLPCA) algorithms [e.g., Hsieh, 2009], including kernel- [Lima et al., 2009] and neural-network-based algorithms [Monahan, 2000; Hsieh, 2007], have been proposed as ways of addressing the above shortcomings of SSA in the geosciences. However, even though the nonlinear mapping functions employed by these methods greatly increase the flexibility to describe nonlinear aspects of the data, the structure of those functions is generally *ad hoc* and susceptible to overfitting the data [Christiansen, 2005]. Moreover, nonlinear mapping functions preclude a straightforward determination of spatial EOFs corresponding to the principal components (PCs) [Lima et al., 2009], limiting the utility of this class of algorithms in deriving bases for reduced dynamical modeling. Another important issue is that the computational cost of NLPCA tends to scale poorly with the ambient space dimension. Indeed, a number of applications in the geophysics literature either deal explicitly with low-dimensional systems [Monahan, 2000], or preprocess the data through subtraction of the seasonal cycle and/or projection onto linear PCs [Hsieh, 2007], thus removing several of their important nonlinear features.

[4] Recently, a method called nonlinear Laplacian spectral analysis (NLSA) was developed [Giannakis and Majda, 2012, also Nonlinear Laplacian spectral analysis: Capturing intermittent and low-frequency spatiotemporal patterns in high-dimensional data, arXiv:1202.6103v1, 2012], which addresses the aforementioned shortcomings of SSA and NLPCA by combining aspects of both approaches. Similarly to SSA, NLSA produces a spatiotemporal decomposition of the data through spectral analysis of linear maps (denoted here by  $\mathcal{A}^l$ ), which are defined objectively from the way the data is presented in ambient space. Thus, there is no need to specify feature maps, and the correspondence between EOFs and PCs is unambiguous and straightforward. However, the  $\mathcal{A}^l$  maps differ crucially from SSA in that they are tailored to

<sup>1</sup>Center for Atmosphere Ocean Science and Courant Institute of Mathematical Sciences, New York University, New York, New York, USA.

Corresponding author: D. Giannakis, Courant Institute of Mathematical Sciences, New York University, 251 Mercer St., New York, NY 10012, USA. (dimitris@cims.nyu.edu)

Copyright 2012 by the American Geophysical Union.  
0094-8276/12/2012GL051575

the nonlinear geometry of the data manifold,  $M$ . Specifically, the key principle underlying NLSA is that for efficient analysis of high-dimensional complex data, temporal patterns (analogous to PCs) should belong in natural low-dimensional spaces of functions on  $M$ , spanned by the leading-few eigenfunctions of graph-theoretic Laplace-Beltrami operators. These eigenfunctions are evaluated efficiently in high ambient-space dimensions using sparse algorithms developed in machine learning [Belkin and Niyogi, 2003; Coifman and Lafon, 2006], thus providing natural orthonormal bases to extract spatiotemporal patterns through singular value decomposition (SVD) of  $\mathcal{A}^l$ . The parameter  $l$  controls the scale (“resolution”) on the data manifold resolved by the temporal modes, and can be set via spectral entropy criteria to prevent overfitting (Giannakis and Majda, arXiv:1202.6103v1, 2012).

[5] Here, we apply the NLSA methodology in a comparative study of the spatiotemporal variability of sea surface temperature (SST) in monthly-averaged, North Pacific data from three control experiments used in the Intergovernmental Panel on Climate Change assessment report 4 (IPCC AR4). Applying no spatial coarse-graining or seasonal-cycle subtraction, we identify a number of physically-distinct spatiotemporal processes, allowing comparison of the SST variability in these models at a significantly more refined level than possible through classical SSA. In particular, besides the familiar periodic and decadal modes of SST variability [Mantua and Hare, 2002; Bond et al., 2003; Di Lorenzo et al., 2008], NLSA reveals a family of modes with strongly intermittent behavior, describing variability in the Eastern and Western boundary currents, as well as mid-basin variability with year-to-year reemergence [Alexander et al., 1999]. The bursting-like behavior of these modes, a hallmark of strongly-nonlinear dynamics, means that they carry little variance, and therefore are not captured by SSA.

## 2. Summary of NLSA Algorithms

[6] Consider a timeseries  $\mathbf{x}_t$  of a  $d$ -dimensional climatic variable sampled uniformly with time step  $\delta t$  for  $s$  samples. NLSA produces a decomposition of  $\mathbf{x}_t$  into  $l$  spatiotemporal patterns,  $\tilde{\mathbf{x}}_t^1, \dots, \tilde{\mathbf{x}}_t^l$ , taking into account the fact that the underlying trajectory of the dynamical system lies on a nonlinear manifold  $M$  in phase space. Below, we outline the key elements of the procedure, which is developed in detail in Giannakis and Majda [2012, arXiv:1202.6103v1, 2012].

[7] *Time-lagged embedding.* In typical climate applications (including the North Pacific systems studied here), the data are highly non-Markovian, generally because the phase space in which the dynamics operate is not fully accessible and/or the influence of external factors is strong. In such cases, it is possible to recover some of the phase-space information lost in the observation process, and alleviate non-dynamical effects caused by external factors, by taking histories of observations  $\mathbf{X}_t = (\mathbf{x}_t, \mathbf{x}_{t-\delta t}, \dots, \mathbf{x}_{t-(q-1)\delta t})$  over a window  $\Delta t = (q-1)\delta t$ , and treating the timeseries  $\mathbf{X}_t$  in the so-called embedding space (of dimension  $n = qd$ ) as input data [Broomhead and King, 1986]. This step can be applied in either of the SSA and NLSA algorithms, and will be particularly important in the analysis of the SST datasets

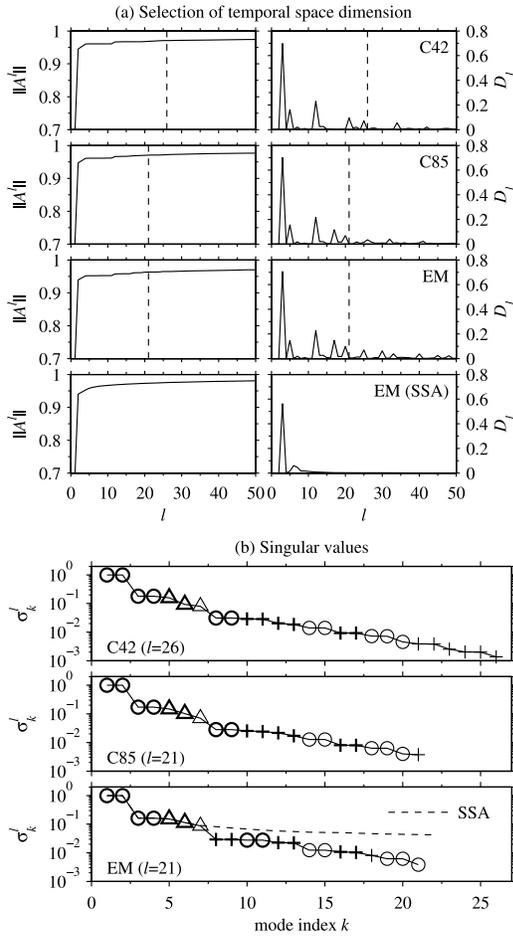
in Section 4, which are strongly influenced by annually varying solar forcing.

[8] *Natural spaces of temporal patterns.* In NLSA, regularity of temporal patterns on the nonlinear data manifold  $M$  is regarded as an essential ingredient of efficient spatiotemporal decompositions of high-dimensional complex data. This requirement is enforced by replacing the temporal space  $\mathbb{R}^s$  of classical SSA by a family of  $l$ -dimensional spaces  $V_l$  (with  $l \ll s$ ) spanned by natural orthonormal functions on the discretely-sampled data manifold. The basis functions in question are the leading  $l$  eigenfunctions,  $\phi_0, \dots, \phi_{l-1}$ , of a graph-theoretic Laplace-Beltrami operator, evaluated by means of algorithms developed in data mining and machine learning [e.g., Belkin and Niyogi, 2003; Coifman and Lafon, 2006]. Giannakis and Majda [2012] modified the conventional version of these algorithms to make use of adaptive Gaussian kernel widths determined through the local phase-space velocity,  $\epsilon_i = \|\mathbf{X}_{t_i} - \mathbf{X}_{t_i-\delta t}\|$ , alleviating the need to tune  $\epsilon_i$  by hand. The basis functions  $\phi_i = (\phi_{i1}, \dots, \phi_{is})$  constructed in this manner are analogous to smooth orthonormal scalar functions on differentiable manifolds; e.g.,  $\phi_i$  would sample the spherical harmonics if  $M$  were a two-dimensional spherical surface with standard measure. In particular, they are orthonormal with respect to a weighted inner product,  $\langle \phi_i, \phi_j \rangle = \sum_{k=1}^s \mu_k \phi_{ki} \phi_{kj}$ , where the weights  $\mu_i$  can be thought of as the volume (Riemannian measure) occupied by sample  $X_{t_i}$  in  $M$ . That  $\mu_i$  is large for states visited infrequently by the system enhances the skill of NLSA in capturing intermittency and rare events. In summary, each element of the temporal space  $V_l = \text{span}\{\phi_0, \dots, \phi_{l-1}\}$  can be represented by an  $l$ -dimensional column vector  $\mathbf{f} = (f_1, \dots, f_l)^T$ , with corresponding temporal pattern  $f(t_i) = \sum_{j=1}^l f_j \phi_{i,j-1}$ .

[9] *Singular value decomposition.* Associated with each  $V_l$  is a linear map  $\mathcal{A}^l$  providing the link between temporal and spatial patterns (analogous to EOFs) through a weighted sum of the data. In the eigenfunction basis,  $\mathcal{A}^l$  is represented by an  $n \times l$  matrix  $A^l$  with elements  $A_{ij}^l = \sum_{k=1}^s \mu_k X_{ik} \phi_{k,j-1}$ , where  $X_{ik}$  denotes the  $i$ -th component (e.g., gridpoint value) of datapoint  $\mathbf{X}_k$  in embedding space. Thus, the outcome of acting on  $\mathbf{f} \in V_l$  with  $\mathcal{A}^l$  is the spatial pattern  $\mathbf{y} = (y_1, \dots, y_n)^T$  with components  $y_i = \sum_{j=1}^l A_{ij}^l f_j$ . The spatial and temporal patterns associated with  $\mathcal{A}^l$  are determined through the SVD of  $A^l$ , viz.

$$A_{ij}^l = \sum_{k=1}^r u_{ik} \sigma_k^l v_{jk}, \quad r = \min\{l, n\}, \quad (1)$$

where  $u_{ik}$  and  $v_{jk}$  are elements of  $n \times n$  and  $l \times l$  orthogonal matrices, respectively, and  $\sigma_k^l \geq 0$  are singular values (ordered in order of decreasing  $\sigma_k^l$ ). Each  $\mathbf{u}_k = (u_{1k}, \dots, u_{nk})$  is a spatial pattern in  $\mathbb{R}^n$ . Moreover, the entries  $(v_{1k}, \dots, v_{lk})$  are the expansion coefficients of the corresponding temporal pattern in the  $\{\phi_i\}$  basis for  $V_l$ , associated with the temporal process  $v_k(t_i) = \sum_{j=1}^l \phi_{i,j-1} v_{jk}$ . Therefore, the decomposition of the signal  $\mathbf{X}_t$  in terms of the  $l$  modes of  $\mathcal{A}^l$  is  $\mathbf{X}_t^l = \sum_{k=1}^l \tilde{\mathbf{X}}_t^k$ , with  $\tilde{\mathbf{X}}_t^k = \mathbf{u}_k \sigma_k^l v_k(t)$ . Note that because  $l \ll s$ , the computational cost of the SVD step in equation (1) is significantly lower than SSA, which involves the full  $n \times s$  data matrix. If lagged embedding has been performed,



**Figure 1.** (a) Operator norm  $\|\mathcal{A}^l\|$  and spectral entropy  $D_l$  versus temporal space dimension  $l$  for the C42, C85, and EM datasets. Dashed vertical lines indicate the selected  $l$  values. (b) Singular values  $\sigma_k^l$  from NLSA (normalized so that  $\sigma_1^l = 1$ ), showing the periodic (O), low-frequency ( $\Delta$ ), and intermittent (+) modes. Boldface markers indicate the modes displayed in Figures 2 and 3 and Animations S1–S3.

the spatiotemporal patterns  $\tilde{\mathbf{X}}_t^k$  are mapped back to  $d$ -dimensional physical space via a uniform-weight projection [Horenko, 2008] to obtain  $\tilde{\mathbf{x}}_t^k$ .

[10] *Selection criteria for the temporal space dimension.* Heuristically, the parameter  $l$  controls the lengthscale on the data manifold resolved by the eigenfunctions spanning  $V_l$ . Working at large  $l$  is desirable, since the lengthscale on  $M$  resolved by the eigenfunctions spanning  $V_l$  generally becomes smaller as  $l$  grows, but the sampling error in the graph-theoretic eigenfunctions  $\phi_l$  increases with  $l$  for a fixed number of samples  $s$ . In other words,  $\phi_l$  will generally depend more strongly on  $s$  for large  $l$ , resulting in an overfit of the discrete data manifold. A useful way of establishing a tradeoff between improved resolution and risk of overfitting is to monitor a relative spectral entropy  $D_l$ , measuring changes in the energy distribution among the modes of  $\mathcal{A}^l$  as  $l$  grows (Giannakis and Majda, arXiv:1202.6103v1, 2012). This measure is given by the formula  $D_l = \sum_{i=1}^l p_i^{l+1} \log(p_i^{l+1}/\pi_i^{l+1})$ , with  $p_i^l = \sigma_i^{l2}/(\sum_{i=1}^l \sigma_i^{l2})$ ,  $\pi_i^l = \hat{\sigma}_i^{l2}/(\sum_{i=1}^l \hat{\sigma}_i^{l2})$ , and

$(\hat{\sigma}_1^l, \dots, \hat{\sigma}_{l-1}^l, \hat{\sigma}_l^l) = (\sigma_1^{l-1}, \dots, \sigma_{l-1}^{l-1}, \sigma_l^{l-1})$ . The appearance of qualitatively new features in the spectrum of  $\mathcal{A}^l$  is accompanied by spikes in  $D_l$  (see Figure 1a), and therefore a reasonable truncation level is the minimum  $l$  beyond which  $D_l$  settles to small values.

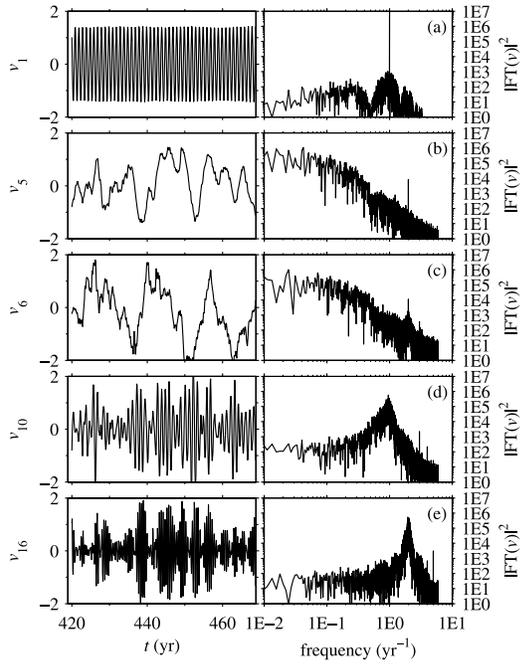
### 3. Datasets

[11] We study monthly-averaged North Pacific SST data from three long ( $\geq 500$  yr) equilibrated control experiments used in IPCC AR4. Two of the datasets are taken from integrations of the Community Climate System Model version 3 (CCSM3) with 1990 greenhouse forcings [Bryan *et al.*, 2005; Collins *et al.*, 2006], and sampled on the model's native ocean grid ( $1^\circ$  nominal horizontal resolution) in the region  $20^\circ\text{N}$ – $65^\circ\text{N}$  and  $120^\circ\text{E}$ – $110^\circ\text{W}$ . These datasets, designated C42 and C85, differ in their temporal extent and atmosphere resolution. C42 spans 900 yr of model output using a T42 atmosphere, whereas C85 covers a 500 yr interval, but employs a finer, T85 atmosphere grid. Our third dataset, which we refer to as EM, consists of monthly-averaged SST output with pre-industrial greenhouse forcing from the coupled atmosphere (ECHAM) and ocean (MPI-OM) models developed by the Max Planck Institute for Meteorology (MPI) [Roeckner *et al.*, 2006]. The EM dataset covers the same spatial domain as C42 and C85 for an interval of 506 years, but the publicly available SST data have been regridded on a T63 atmosphere grid ( $1.5^\circ$  nominal resolution). Compared to the depth-averaged, upper-300 m ocean temperature data in Branstator and Teng [2010], Teng and Branstator [2010], and Giannakis and Majda [2012, arXiv:1202.6103v1, 2012], the datasets studied here are of significantly higher dimension, and are influenced more strongly by atmospheric noise. The number of North Pacific gridpoints is  $d = 6671$  (C42 and C85), and 1204 (EM), amounting to embedding space dimension  $n = qd = 160, 104$  and  $28,896$ , respectively, for the  $\Delta t = 2$  yr lag window ( $q = 24$ ) employed in Section 4. A summary of the datasets is provided in Table S1 in the auxiliary material.<sup>1</sup>

### 4. Results and Discussion

[12] We process all three datasets with the NLSA algorithm using an embedding window  $\Delta t = 2$  yr, which is equal to the  $\Delta t$  value in our earlier work on depth-averaged upper ocean data [Giannakis and Majda, 2012, arXiv:1202.6103v1, 2012]. We also consider results from SSA (again with  $\Delta t = 2$  yr), but due to the high computational cost associated with the SVD of the  $n \times s$  data matrix, we have performed this analysis only on the lower-resolution EM dataset. Physically, our choice of embedding window was motivated by the fact that the annually-varying solar forcing is one of the strongest sources of non-Markovianity in the data. As discussed in more detail in Giannakis and Majda (arXiv:1202.6103v1, 2012), a sufficiently-long  $\Delta t$  (here, spanning at least two solar-forcing periods) is crucial to separate the seasonal cycle from decadal variability in either of the NLSA- or SSA-based spatiotemporal patterns, but  $\Delta t$  is not directly related to the longest frequency captured by those modes. We have assessed the robustness of our results through a sequence of NLSA runs on

<sup>1</sup>Auxiliary materials are available in the HTML. doi:10.1029/2012GL051575.



**Figure 2.** Representative temporal patterns  $v_k(t)$  from the C85 dataset evaluated through NLSA with temporal space dimension  $l=21$ . (a) Annual mode. (b) PDO mode. (c) NPGO mode. (d) Leading intermittent mode (boundary currents and subtropical gyre). (e) Leading intermittent mode with semiannual base frequency (Kuroshio-Oyashio transition area).

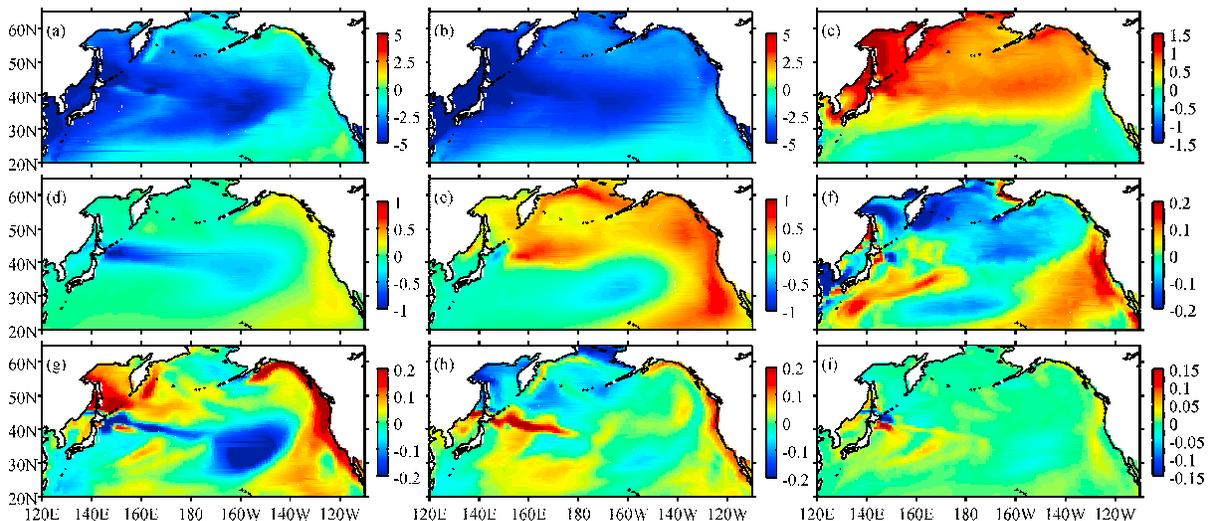
the C42 dataset coarse-grained on the T42 atmospheric grid with representative embedding windows in the interval 2–10 yr. The resulting spatiotemporal patterns depended weakly on  $\Delta t$  in that interval, and were also in good qualitative agreement with the high-resolution results reported here.

[13] To select suitable values for the temporal-space dimension, we examine the  $l$  dependence of the spectral

entropy measure  $D_l$ , shown in Figure 1a. For all three datasets processed through NLSA,  $D_l$  exhibits a series of spikes for  $l \approx 25$ , and is small for larger  $l$  values. These spikes correspond to qualitatively-new types of modes entering the spectrum of  $\mathcal{A}^l$ , representing periodic, decadal, and intermittent SST variability (as described below). In contrast, the  $D_l$  measure for SSA applied to the EM dataset, with  $l$  corresponding here to the number of modes used in truncated PC/EOF expansions, becomes negligible at significantly smaller values ( $l \approx 7$ ). Indeed, the higher SSA modes do not reveal qualitatively new patterns of spatiotemporal variability (at least up to mode number 100, which we have checked). This situation is also reflected in the weak decay of the singular values in Figure 1b. As a result of independent interest, Figure 1a also shows the  $l$ -dependence of the operator norm,  $\|\mathcal{A}^l\| = (\sum_i \sigma_i^2)^{1/2}$ , which, being dominated by a small number of highly energetic periodic and decadal modes, reaches a plateau around  $l = 7$ . Evidently, a norm-based criterion would have significantly less discriminating power for selecting  $l$  than spectral entropy.

[14] On the basis of the  $D_l$  results above, hereafter we set  $l = 26, 21$ , and  $21$  for the C42, C85, and EM datasets, respectively; the corresponding singular values  $\sigma_i^l$  from equation (1) are shown in Figure 1b. In all three cases, the spatiotemporal patterns identified through NLSA fall into three families of periodic, low-frequency (decadal) and intermittent modes, summarized in Table S2, and illustrated for the high-resolution atmosphere dataset C85 in Figures 2 and 3. The dynamic evolution of spatiotemporal patterns from all three datasets is presented in Animations S1–S3.

[15] *Periodic modes.* The periodic modes come in doubly-degenerate pairs, and have the structure of sinusoidal waves (Figure 2a) with phase difference  $\pi/2$  and frequency equal to integer multiples of  $1 \text{ yr}^{-1}$  down to the Nyquist limit of  $1/6 \text{ yr}^{-1}$ . The leading four modes in this family represent the annual and semiannual variability of SST, and carry more energy than any of the low-frequency or intermittent modes



**Figure 3.** Snapshots of spatiotemporal patterns  $\tilde{x}_t^k$  of the SST (annual mean subtracted at each gridpoint) evaluated through NLSA for February of year 451 of the C85 dataset. (a) Raw data. Composites evaluated using modes (b) 1 and 2 (annual); (c) 3 and 4 (semiannual); (d) 5 (PDO); (e) 6 (NPGO); (f) 8 and 9 (4-month periodic); (g) 10 and 11 (intermittent, boundary currents and subtropical gyre); (h) 12 and 13 (intermittent, Kuroshio extension and Alaskan Gyre); (i) 16 and 17 (intermittent with 1/2-yr base frequency). The dynamical evolution in Animation S1 is much more revealing.

in the datasets studied here (cf. the depth-averaged data in *Giannakis and Majda* [2012, arXiv:1202.6103v1, 2012]). In the spatial domain (Figures 3b and 3c), the annual and semiannual modes have larger amplitude in the western part of the basin and relatively weak zonal variability. The higher-frequency periodic modes (e.g., Figure 3f) have lower amplitude and stronger zonal variability. Note that the annual mode is not equivalent to monthly climatology; this is because the other modes in the spectrum have non-zero month-conditional means. Note also that annual and semiannual modes are present in the SSA spectrum (see Animation S3), but these modes are somewhat mixed with a low-frequency modulating signal. We found no other periodic modes in the SSA spectrum with  $l \leq 100$ .

[16] *Low-frequency modes.* These modes are characterized by high spectral power over interannual to interdecadal timescales, and strongly suppressed power over annual or shorter time scales (Figures 2b and 2c). The leading mode in this family (Figure 3d) features a basin-scale horseshoe-like temperature anomaly pattern, developing along the Kuroshio extension, together with an anomaly of the opposite sign along the west coast of North America. This is the familiar PDO pattern, determined through EOF analysis of seasonally-detrended data [*Mantua and Hare*, 2002, and references therein]. The second low-frequency SST mode has been described as a North Pacific Gyre Oscillation (NPGO) by *Di Lorenzo et al.* [2008] (through EOF analysis in the North-eastern Pacific), and is known to explain aspects of Pacific decadal variability which are not represented by the PDO [*Bond et al.*, 2003]. The spatial structure of the NPGO differs somewhat across the three datasets, but the its main features (viz., SST anomaly of the same sign in the Eastern boundary currents and the Alaskan gyre; see Figure 3e) are represented consistently.

[17] *Intermittent modes.* As illustrated in Figures 2d and 2e, the key feature of modes of this family is temporal intermittency, arising out of oscillations at integer multiples of  $1 \text{ yr}^{-1}$ , which are modulated by relatively sharp envelopes with a temporal extent in the 2–7 yr regime. The resulting Fourier spectrum is dominated by a peak centered at the base frequency, exhibiting some skewness towards lower frequencies. The most prominent features of the leading two intermittent modes (Figure 3g) are anomalies of the same sign in the Eastern boundary currents (Alaska and California currents) and the Oyashio current, accompanied by anomalies of the opposite sign in the Kuroshio extension and the interior of the subtropical gyre. As one can verify from timeseries of basin-integrated temperature anomaly, extrema of the subtropical gyre oscillation in these modes tend to coincide with the February extremum of the annual SST cycle. Because these modes carry significant power on interannual timescales (Figure 2d), this results in a reemergence of SST anomalies between consecutive late winters without persisting through the intervening summer, as discussed by *Alexander et al.* [1999, 2006]. The intermittent modes further down the spectrum exhibit more localized spatial structures, such as variability in the Kuroshio extension coupled to the Alaskan gyre (Figure 3h). Moreover, the leading intermittent modes with semiannual base frequency (Figure 2e) are concentrated in the Kuroshio-Oyashio transition area (Figure 3i). In general, the C42 and EM datasets with coarser, less noisy atmospheres feature stronger intermittent modes than C85, but in all cases the modes can

be grouped into those involving (1) the boundary currents and subtropical gyre, (2) the Kuroshio extension and Alaskan gyre, and (3) semiannual variability in the Kuroshio-Oyashio transition area; see Animations S1–S3.

[18] That the intermittent modes carry less energy than the periodic or low-frequency ones does not mean their dynamical significance is unimportant. In fact, an adequate representation of intermittent and transitory states can be crucial for generating the right dynamics of complex systems through reduced modeling [*Aubry et al.*, 1991; *Crommelin and Majda*, 2004; *Giannakis and Majda*, 2012]. Being able to capture this intrinsically nonlinear behavior constitutes a major strength of NLSA algorithms, which also allows for cross-model comparisons at a significantly more refined level than possible through linear SSA. The consistency of the identified modes across all three datasets studied here is remarkable given the differences in model structure, greenhouse forcings, and atmosphere resolutions used to generate the data. In future work we plan to use these modes to build predictive models.

[19] **Acknowledgments.** This work was supported by NSF grant DMS-0456713 and ONR DRI grants N25-74200-F6607 and N00014-10-1-0554.

[20] The Editor thanks Chris Wikle and an anonymous reviewer for assisting with the evaluation of this paper.

## References

- Alexander, M., et al. (2006), Extratropical atmosphere–ocean variability in CCSM3, *J. Clim.*, *19*, 2496–2525, doi:10.1175/JCLI3743.1.
- Alexander, M. A., C. Deser, and M. S. Timlin (1999), The reemergence of SST anomalies in the North Pacific Ocean, *J. Clim.*, *12*, 2419–2433, doi:10.1175/1520-0442(1999)012<2419:trosai>2.0.co;2.
- Aubry, N., R. Guyonnet, and R. Lima (1991), Spatiotemporal analysis of complex signals: Theory and applications, *J. Stat. Phys.*, *64*, 683–739, doi:10.1007/bf01048312.
- Aubry, N., W.-Y. Lian, and E. S. Titi (1993), Preserving symmetries in the proper orthogonal decomposition, *SIAM J. Sci. Comput.*, *14*, 483–505, doi:10.1137/0914030.
- Belkin, M., and P. Niyogi (2003), Laplacian eigenmaps for dimensionality reduction and data representation, *Neural Comput.*, *15*, 1373–1396, doi:10.1162/089976603321780317.
- Bond, N. A., J. E. Overland, M. Spillane, and P. Stabeno (2003), Recent shifts in the state of the North Pacific, *Geophys. Res. Lett.*, *30*(23), 2183, doi:10.1029/2003GL018597.
- Branstator, G., and H. Teng (2010), Two limits of initial-value decadal predictability in a CGCM, *J. Clim.*, *23*, 6292–6311, doi:10.1175/2010JCLI3678.1.
- Broomhead, D. S., and G. P. King (1986), Extracting qualitative dynamics from experimental data, *Physica D*, *20*(2–3), 217–236, doi:10.1016/0167-2789(86)90031-x.
- Bryan, F. O., et al. (2005), Response of the North Atlantic thermohaline circulation and ventilation to increasing carbon dioxide in CCSM3, *J. Clim.*, *19*, 2382–2397, doi:10.1175/JCLI3757.1.
- Christiansen, B. (2005), The shortcomings of nonlinear component analysis in identifying circulation regimes, *J. Clim.*, *18*, 4814–4823, doi:10.1175/JCLI3569.1.
- Coifman, R. R., and S. Lafon (2006), Diffusion maps, *Appl. Comput. Harmonic Anal.*, *21*, 5–30, doi:10.1016/j.acha.2006.04.006.
- Collins, W. D., et al. (2006), The Community Climate System Model version 3 (CCSM3), *J. Clim.*, *19*, 2122–2143, doi:10.1175/JCLI3761.1.
- Crommelin, D. T., and A. J. Majda (2004), Strategies for model reduction: Comparing different optimal bases, *J. Atmos. Sci.*, *61*, 2206–2217, doi:10.1175/1520-0469(2004)061<2206:sfmrcd>2.0.co;2.
- Di Lorenzo, E., et al. (2008), North Pacific Gyre Oscillation links ocean climate and ecosystem change, *Geophys. Res. Lett.*, *35*, L08607, doi:10.1029/2007GL032838.
- Ghil, M., et al. (2002), Advanced spectral methods for climatic time series, *Rev. Geophys.*, *40*(1), 1003, doi:10.1029/2000RG000092.
- Giannakis, D., and A. J. Majda (2012), Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability, *Proc. Natl. Acad. Sci. U. S. A.*, *109*(7), 2222–2227, doi:10.1073/pnas.1118984109.

- Horenko, I. (2008), On simultaneous data-based dimension reduction and hidden phase identification, *J. Atmos. Sci.*, *65*, 1941–1954, doi:10.1175/2007JAS2587.1.
- Hsieh, W. W. (2007), Nonlinear principle component analysis of noisy data, *Neural Networks*, *20*, 434–443, doi:10.1016/j.neunet.2007.04.018.
- Hsieh, W. W. (2009), *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels*, Cambridge Univ. Press, Cambridge, U. K.
- Lima, C. H. R., U. Lall, T. Jebara, and A. G. Barnston (2009), Statistical prediction of ENSO from subsurface sea temperature using a nonlinear dimensionality reduction, *J. Clim.*, *22*, 4501–4519, doi:10.1175/2009jcli2524.1.
- Mantua, N. J., and S. R. Hare (2002), The Pacific Decadal Oscillation, *J. Oceanogr.*, *58*, 35–44, doi:10.1023/A:1015820616384.
- Monahan, A. H. (2000), Nonlinear principal component analysis by neural networks: Theory and application to the Lorenz system, *J. Clim.*, *13*, 821–835, doi:10.1175/1520-0442(2000)013<0821:npcabn>2.0.co;2.
- Newman, M., G. P. Compo, and M. A. Alexander (2003), ENSO-forced variability of the Pacific Decadal Oscillation, *J. Clim.*, *16*, 3853–3857, doi:10.1175/1520-0442(2003)016<3853:evotpd>2.0.co;2.
- Overland, J. E., et al. (2010), Climate controls on marine ecosystems and fish populations, *J. Mar. Sys.*, *79*, 305–315, doi:10.1016/j.jmarsys.2008.12.009.
- Roeckner, E., M. Lautenschlager, and M. Esch (2006), IPCC-AR4 MPI-ECHAM5 T63L31 MPI-OM\_GR1.5 L40 PlCntrl (pre-industrial control experiment): Atmosphere monthly mean values MPImet/MaD Germany, [http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=EH5-T63L31\\_OM-GR1.5L40\\_PRE\\_6H](http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=EH5-T63L31_OM-GR1.5L40_PRE_6H), World Data Cent. for Clim., Hamburg, Germany
- Teng, H., and G. Branstator (2010), Initial-value predictability of prominent modes of North Pacific subsurface temperature in a CGCM, *Clim. Dyn.*, *36*, 1813–1834, doi:10.1007/s00382-010-0749-7.
- Vautard, R., and M. Ghil (1989), Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, *Physica D*, *35*, 395–424, doi:10.1016/0167-2789(89)90077-8.
- Vitart, F., and T. Jung (2010), Impact of the Northern Hemisphere extratropics on the skill in predicting the Madden Julian Oscillation, *Geophys. Res. Lett.*, *37*, L23805, doi:10.1029/2010GL045465.