

OPEN

Extended-range statistical ENSO prediction through operator-theoretic techniques for nonlinear dynamics

Xinyang Wang ¹, Joanna Slawinska² & Dimitrios Giannakis ^{1*}

Forecasting the El Niño-Southern Oscillation (ENSO) has been a subject of vigorous research due to the important role of the phenomenon in climate dynamics and its worldwide socioeconomic impacts. Over the past decades, numerous models for ENSO prediction have been developed, among which statistical models approximating ENSO evolution by linear dynamics have received significant attention owing to their simplicity and comparable forecast skill to first-principles models at short lead times. Yet, due to highly nonlinear and chaotic dynamics (particularly during ENSO initiation), such models have limited skill for longer-term forecasts beyond half a year. To resolve this limitation, here we employ a new nonparametric statistical approach based on analog forecasting, called kernel analog forecasting (KAF), which avoids assumptions on the underlying dynamics through the use of nonlinear kernel methods for machine learning and dimension reduction of high-dimensional datasets. Through a rigorous connection with Koopman operator theory for dynamical systems, KAF yields statistically optimal predictions of future ENSO states as conditional expectations, given noisy and potentially incomplete data at forecast initialization. Here, using industrial-era Indo-Pacific sea surface temperature (SST) as training data, the method is shown to successfully predict the Niño 3.4 index in a 1998–2017 verification period out to a 10-month lead, which corresponds to an increase of 3–8 months (depending on the decade) over a benchmark linear inverse model (LIM), while significantly improving upon the ENSO predictability “spring barrier”. In particular, KAF successfully predicts the historic 2015/16 El Niño at initialization times as early as June 2015, which is comparable to the skill of current dynamical models. An analysis of a 1300-yr control integration of a comprehensive climate model (CCSM4) further demonstrates that the enhanced predictability afforded by KAF holds over potentially much longer leads, extending to 24 months versus 18 months in the benchmark LIM. Probabilistic forecasts for the occurrence of El Niño/La Niña events are also performed and assessed via information-theoretic metrics, showing an improvement of skill over LIM approaches, thus opening an avenue for environmental risk assessment relevant in a variety of contexts.

Previous studies on improving the skill of conventional LIMs^{1,2} have highlighted the importance of nonlinearity in ENSO dynamics, such as surface-subsurface interactions and surface winds³, which are usually underestimated in linear dynamics approximations. Adequate representation of such processes within statistical ENSO models is important to attain optimal long-term forecast skill, influencing two major components of model design, namely (i) the construction of predictor vectors to extract pertinent information about the state of the climate system; and (ii) the assumed evolution dynamics employed to make predictions beyond the training period. For instance, LIMs use the leading principal components from empirical orthogonal function (EOF) analysis as predictors, and approximate their tendencies by a Markov prediction model similar to a linear regression. A linear structure is imposed in two aspects here; that is, the linear predictors obtained by EOF analysis and the linear dynamical model treating these predictors as state vectors. This suggests that two corresponding types of improvement of conventional LIMs can be sought. Indeed, a number of studies^{3,4} have shown that inclusion of LIM residuals at the

¹Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, New York, USA. ²Department of Physics, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, USA. *email: dimitris@cims.nyu.edu

present and recent past states can implicitly capture subsurface forcing and its nonlinear interactions with SST, consequently contributing to more skillful ENSO predictions. Meanwhile, replacing EOF analysis by nonlinear dimension reduction techniques⁵ has also led to improved performance by LIMs under certain circumstances, especially for long-term ENSO forecasts.

Other statistical approaches for ENSO prediction^{6–9} have included methods based on Lorenz’s analog forecasting technique¹⁰ or neural networks (NNs)¹¹. The core idea of analog methods^{6–8} is to employ observational or free-running (non-initialized) model data as libraries of states, through which forecasts can be made by identifying states in the library that closely resemble the observed data at forecast initialization, and following the evolution of the quantity of interest on these so-called analogs to yield a prediction. Advantages of analog forecasting include avoiding the need for an expensive, and potentially unstable, initialization system, as well as reducing structural model error if natural analogs are employed. For phenomena such as ENSO exhibiting a reasonably small number of effective degrees of freedom, the performance of analog techniques has been shown to be comparable to, and sometimes exceed, the skill of initialized forecasts by coupled general circulation models (CGCMs)^{7,8}. Meanwhile, NN models build a representation of the predictand as a nonlinear function of the predictor variables by optimization within a high-dimensional parametric class of functions. A recent study⁹ utilizing convolutional NNs trained on GCM output and reanalysis data has demonstrated forecast skill for the 3-month averaged Niño 3.4 index extending out to 17 months.

The KAF approach^{12,13} employed in this work can be viewed as a generalization of conventional analog forecasting, utilizing nonlinear-kernel algorithms for statistical learning¹⁴ and operator-theoretic ergodic theory^{15,16} to optimally capture the evolution of a response variable (here, an ENSO index) under partially observed nonlinear dynamics. In particular, unlike conventional approaches, the output of KAF is not given by a local average over analog states (or linear combinations of states in the case of constructed analogs⁶), but is rather determined by a projection operation onto a function space of observables of high regularity, learned from high-dimensional training data. This hypothesis space has as a basis a set of eigenvectors of a judiciously constructed kernel matrix, which depends nonlinearly on the input data, thus potentially capturing richer patterns of variability than the linear principal components from EOF analysis. It should be noted, in particular, that the eigenvectors employed in KAF are not expressible as linear projections of the data onto corresponding EOFs. By virtue of this structure, the forecast function from KAF can be shown to converge in an asymptotic limit of large data to the conditional expectation of the response function, acted upon by the Koopman evolution operator of the dynamical system over the desired lead time, conditioned on the input (predictor) data at forecast initialization¹³. As is well known from statistics, the conditional expectation is optimal in the sense of minimal expected square error (L^2 error) among all forecasts utilizing the same predictors.

Here, we build KAF models using lagged histories of Indo-Pacific SST fields as input data. Specifically, we employ the class of kernels introduced in the context of nonlinear Laplacian spectral analysis (NLSA)^{17,18} algorithms, whose eigenfunctions provably capture modes of high dynamical coherence^{19,20}, including representations of ENSO, ENSO combination modes, and Pacific decadal variability^{21–23}. This capability is realized without ad hoc prefiltering of the input data through the use of delay-coordinate maps²⁴ and a nonlinear (normalized Gaussian) kernel function²⁵ measuring similarity between delay-embedded sequences of SST fields. A key advantage of the methodology employed here over LIMs or other parametric models is its nonparametric nature, in the sense that KAF does not impose explicit assumptions on the dynamical model structure (e.g., linearity), thus avoiding systematic model errors that oftentimes preclude parametric models from attaining useful skill at long lead times. Moreover, by employing delays, KAF enhances the information content of the predictor data beyond the information present in individual SST snapshots, allowing it to capture important contributors to ENSO predictability, such as upper-ocean heat content and surface winds^{3,26}.

A distinguishing aspect of the approach presented here over recent analog^{7,8} and NN⁹ methods utilizing large, GCM-derived, multivariate datasets for training is that our observational forecasts are trained only on industrial-era observational/reanalysis SST fields. Indeed, due to the large number of parameters involved, NN methods cannot be adequately trained from industrial-era data alone⁹, and are trained instead using thousands of years of model output. In contrast, KAF is an intrinsically nonparametric method, involving only a handful of meta-parameters (such as embedding window and hypothesis space dimension), and thus able to yield robust ENSO forecasts from a comparatively modest training dataset spanning ~120 years. It should also be noted that KAF exhibits theoretical convergence and optimality guarantees¹³ which, to our knowledge, are not available for other comparable statistical ENSO forecasting methodologies.

Results

We present prediction results obtained via KAF and LIMs in a suite of prediction experiments for (i) the Niño 3.4 index in industrial-era HadISST data²⁷; (ii) the Niño 3.4 index in a 1300-yr control integration of CCSM4²⁸, a CGCM known to exhibit fairly realistic ENSO dynamics and Pacific decadal variability²⁹; and (iii) the probability of occurrence of El Niño/La Niña events in both HadISST and CCSM4. All experiments use SST fields from the respective datasets in a fixed Indo-Pacific domain as training data. The Niño 3.4 prediction skill is assessed using root-mean-square error (RMSE) and pattern correlation (PC) scores. We employ PC = 0.6 and 0.5 as representative thresholds separating useful from non-useful predictions. The probabilistic El Niño/La Niña experiments are assessed using the Brier skill score (BSS)^{30,31}, as well as information-theoretic relative entropy metrics^{32–34}. For consistency with an operational forecasting environment, all experiments employ disjoint, temporally consecutive portions of the available data for training and verification, respectively. In the CCSM4 experiments, Niño 3.4 indices are computed relative to monthly climatology of the training data; in HadISST we use anomalies relative to the 1981–2010 monthly climatology as per current convention in many modeling centers³⁵.

Figure 1 displays KAF and LIM prediction results for the Niño 3.4 index in the HadISST dataset over a 1998–2017 verification period, using 1870–1997 for training. In Fig. 1(a,b) it is clearly evident that even though there is

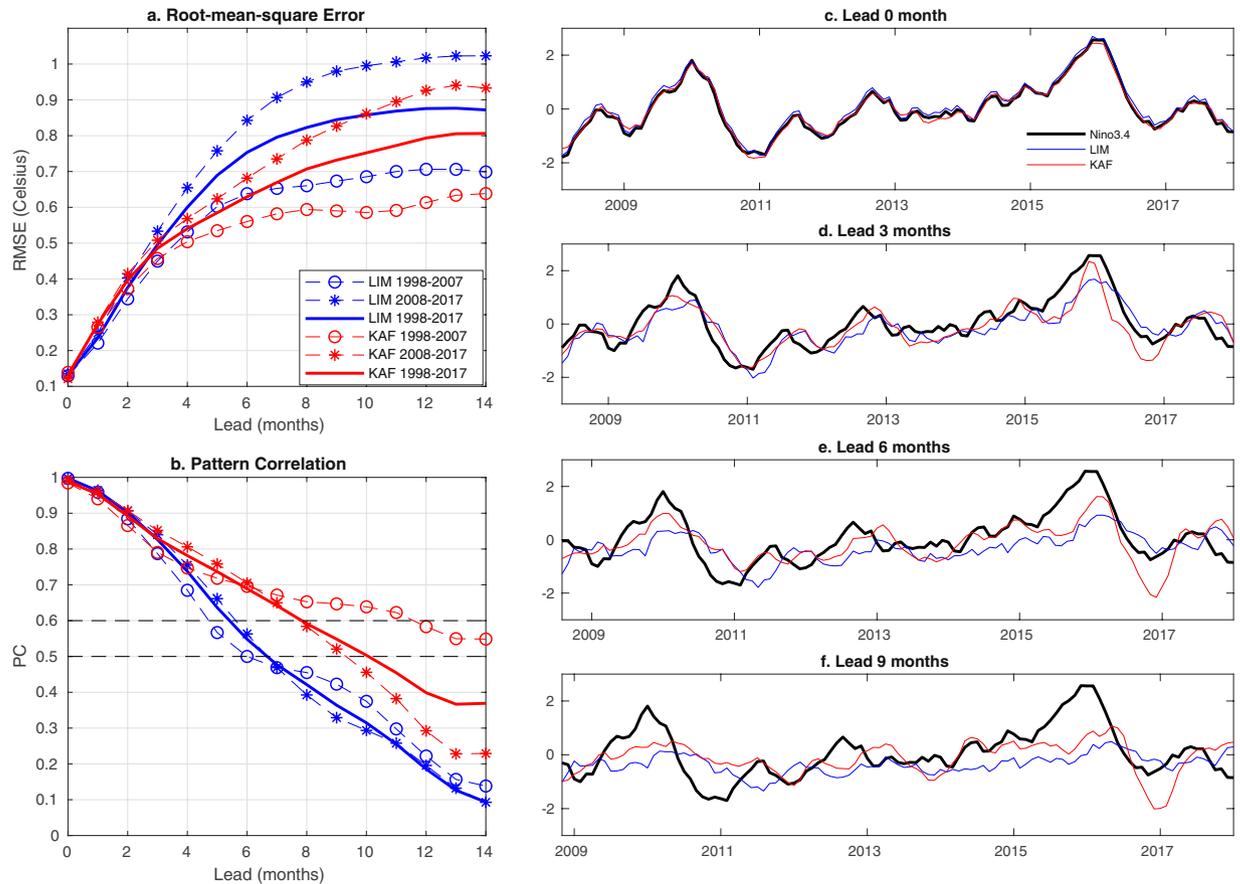


Figure 1. Forecasts of the Niño 3.4 index in industrial-era HadISST data during 1998–2017 obtained via KAF (red) and LIM (blue). **(a,b)** RMSE and PC scores, respectively, as a function of lead time, computed for 1998–2007 (○ markers), 2008–2017 (* markers), and the full 1998–2017 verification period (solid lines). The 0.6 and 0.5 PC levels are highlighted in **(b)** for reference. **(c–f)** Running forecasts for representative lead times in the range 0–9 months. Black lines show the true signal at the corresponding verification times.

significant decadal variability of skill, KAF consistently outperforms the LIM in terms of both RMSE and PC scores, with an apparent increase of 4–7 months of useful forecast horizon. In particular, KAF attains the longest useful skill in the first decade of the verification period, 1998–2007, despite an initially faster decrease of skill over 0–6 months. The KAF PC score for 1998–2007 is seen to cross the 0.6 threshold at approximately 12 months, and remains above 0.5 at least out to 14 months, compared to 5 and 6 months for LIM, respectively. In 2008–2017, the PC skill of KAF exhibits a more uniform rate of decay, crossing the $PC = 0.5$ threshold at a lead time of 9 months, compared to 6–7 months for LIM. When measured over the full, 20-year verification period, KAF provides 3 to 4 months of additional useful skill, as well as a markedly slower rate of skill reduction beyond 4 months. As a comparison with state-of-art dynamical forecasts, the skill scores in Fig. 1 are generally comparable, and in some cases exceed, recently reported scores from the North American Multimodel Ensemble³⁶ (though it should be kept in mind that the verification periods of that study and ours are different).

In separate calculations, we have verified that incorporating delays in the kernel is a significant contributor to the skill of KAF, particularly at lead times greater than 2 months. This is consistent with previous work, which has shown that delay embedding increases the capability of kernel algorithms to extract dynamically intrinsic coherent patterns in the spectrum of the Koopman operator^{19,20}, including ENSO and its linkages with seasonal and decadal variability of the climate system^{21–23}. Given that our main focus here is on the extended-range regime, Fig. 1 depicts KAF results for a single embedding window of 12 months. Operationally, however, one would employ different embedding windows at different leads to optimize performance.

Next, in order to assess the potential skill of KAF and LIM approaches in an environment with more data available for training and verification (contributing to more robust model construction and skill assessment, respectively) than HadISST, we examine Niño 3.4 prediction results for the CCSM4 dataset, using the first 1100 years for training and the last 200 years for verification. Figure 2 demonstrates that while both KAF and LIM exhibit significantly higher skill compared to the HadISST results in Fig. 1(a,b), the performance of KAF is particularly strong, with PC scores computed over the 200-year verification period crossing the 0.6 and 0.5 levels at 17- and 24-month leads, respectively. The latter values correspond to an increase of predictability over HadISST by 9 and 14 months, measured with respect to the same PC thresholds. In comparison, the $PC = 0.6$ and 0.5 predictability horizons for LIM are 11 and 18 months, respectively, corresponding to an increase of 6 and 11

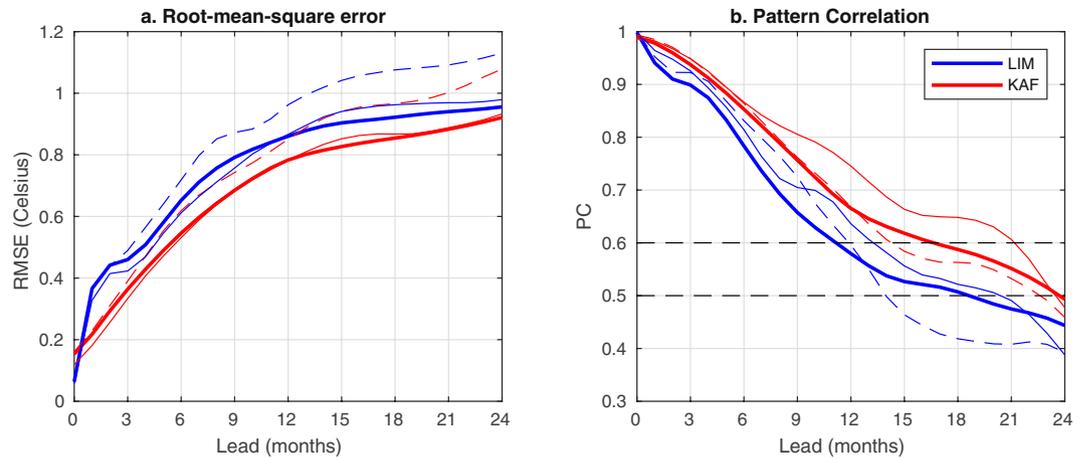


Figure 2. As in Fig. 1(a,b), but for a 200-year verification period corresponding to CCSM4 simulation years 1100–1300. Solid and dashed lines correspond to scores computed from the fourth and last decade of the verification period, respectively. Thick lines show scores computed from the full 200-year verification period.

months over HadISST. As illustrated in Fig. 2, the RMSE and PC scores from individual decades in the CCSM4 verification period exhibit significant variability, and similarly to HadISST, KAF consistently outperforms LIM on individual verification decades. In general, KAF appears to benefit from the larger CCSM4 dataset more than LIM, which is consistent with the former method better leveraging the information content of large data sets due to its nonparametric nature and theoretical ability to consistently approximate an optimal prediction function (the conditional expectation). In contrast, LIM’s performance is limited to some extent by structural model error due to linear dynamics, and this type of error may not be possible to eliminate even with arbitrarily large training datasets.

Another important consideration in ENSO prediction is the seasonal dependence of skill, exhibiting the so-called “spring barrier”³⁷. Figure 3 shows the month-to-month distribution of the KAF and LIM Niño 3.4 PC scores, computed for 6- and 9-month leads in HadISST and 0–24-month leads in CCSM4 using the same training and verification datasets as in Figs. 1 and 2, respectively. These plots feature characteristic spring barrier curves, with the highest predictability occurring in late winter to early spring and a clear drop of skill in summer^{3,4}. The diminished summer predictability is thought to be caused by the low amplitude of SST anomalies developing then, making ENSO dynamics more sensitive to high-frequency atmospheric noise³. As we have shown in previous work²², the class of kernels employed here for forecasting is adept at capturing the effects of atmospheric noise and its nonlinear impact on ENSO statistics, including positive SST anomaly skewness underlying El Niño/La Niña asymmetry. The method has also been found highly effective in capturing the phase locking of ENSO to the seasonal cycle through a hierarchy of combination modes^{21,23,38}. Correspondingly, while both methods exhibit a reduction of skill during summer, KAF fares significantly better than the benchmark LIMs in both of the HadISST and CCSM4 datasets. More specifically, in the HadISST results in Fig. 3(a,b), the LIM’s PC score drops rapidly to small, 0–0.3 values between June and September, while KAF maintains scores of about 0.4 or greater over the same period. In particular, the June–September PC scores for KAF at 6-month leads hover between 0.5 and 0.6, which are values that one could consider as “beating” the spring barrier (albeit marginally). Similarly, in CCSM4 (Fig. 3(d)), the PC = 0.6 contour for KAF decreases less appreciably and abruptly over May–August than in the case of the LIM, indicating that KAF is better at maintaining skill during the transition from spring to summer. Since strong El Niño events are often triggered by westerly wind bursts in spring, advecting warm surface water to the east³, the better performance of KAF in summer suggests that it is more capable of capturing the triggering mechanism²², aiding its ability in predicting the onset of ENSO. Indeed, this is consistent with the running forecast time series for HadISST in Fig. 1(c,d), in which KAF often yields more accurate forecasts at the beginning of El Niño events, e.g., during 2009/10 and 2015/16.

As a final deterministic hindcast experiment, we consider prediction of the 2015/16 El Niño event with KAF. This event has received considerable attention by researchers and stakeholders across the world³⁵, in part due to it being the first major El Niño since the 1997/98 event, providing an important benchmark to assess the advances in modeling and observing systems in the nearly two intervening decades. Figure 4 shows KAF forecast trajectories of 3-month running means of the Niño 3.4 index in HadISST for the period May 2015 to May 2017, initialized in May, June, and July of 2015 (i.e., 7, 6, and 5 months prior to the December 2015 peak of the event). In this experiment, we use Indo-Pacific SST data from HadISST over the period January 1870 to December 2007 for training. Moreover, we perform prediction of 3-month-averaged Niño 3.4 indices to highlight seasonal evolution³⁵. Prediction results for the instantaneous (monthly) Niño 3.4 index are broadly consistent with the results in Fig. 4.

We find that the KAF hindcast initialized in July 2015 accurately predicts the 3-month averaged Niño 3.4 evolution through February 2016, including the 2015/16 El Niño peak, yielding a moderate over-prediction over the ensuing months which nevertheless correlates well with the true signal. The trajectory initialized in June 2015 also predicts a strong ($\geq 2^\circ\text{C}$) El Niño event peaking in December 2015, but underestimates the true 2.5°C

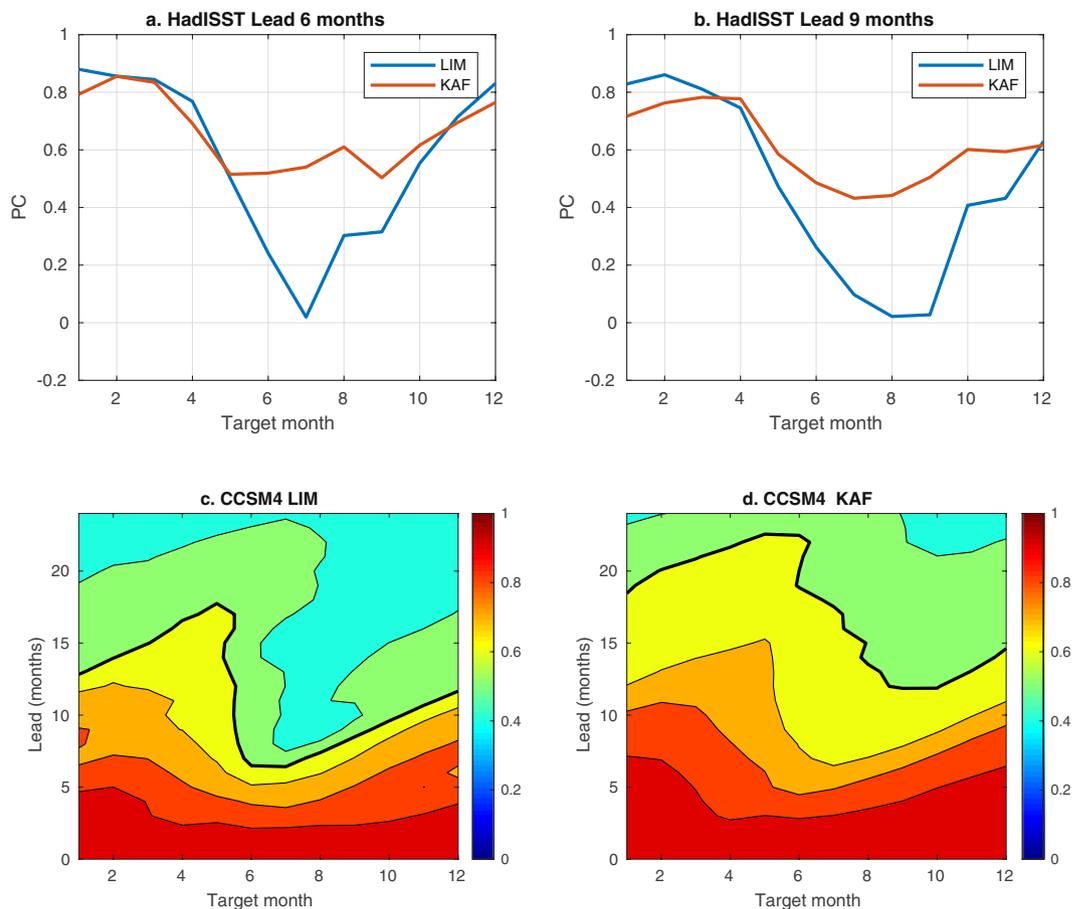


Figure 3. (a,b) Seasonal dependence of Niño 3.4 PC scores for KAF (red lines) and the benchmark LIM at 6- and 9-month leads in industrial-era HadISST data during 1998–2017. (c,d) LIM and KAF PC scores for 0–24-month leads in CCSM4 during simulation years 1100–1300. Contours are drawn at the PC = 0.9, 0.8, ..., 0.5 levels. Thick black lines highlight PC = 0.6 contours for reference.

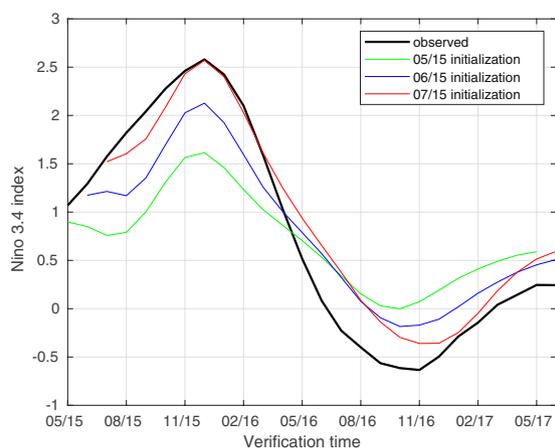


Figure 4. KAF hindcasts of the 3-month-averaged Niño 3.4 index in HadISST during the 2015/16 El Niño period. Hindcast trajectories are initialized in the May, June, and July preceding the December 2015 peak of the event.

anomaly peak by approximately 0.4 °C. Meanwhile, the trajectory initialized in May 2015 predicts a 1.5 °C positive anomaly in the ensuing December, which falls short of the true event by 1 °C but still qualifies as an El Niño event by conventional definitions. Overall, this level of performance compares favorably with many statistical model forecasts of the 2015/16 El Niño (which typically did not yield a possibility of a ≥ 2 °C event until August 2015³⁵).

In fact, the forecast trajectories in Fig. 4 have comparable, and sometimes higher, skill than forecasts from many operational weather prediction systems, despite KAF utilizing far-lower numerical/observational resources and temporal resolution.

We now turn to the skill of probabilistic El Niño/La Niña forecasts. A distinguishing aspect of probabilistic prediction over point forecasts (e.g., the results in Figs. 1–3) is that it provides more direct information about uncertainty, and as a result more actionable information to decision makers. Probabilistic prediction with either first-principles or statistical models, or ensembles of models, is typically conducted by binning collections of point forecasts, e.g., realized by random draws from distributions of initial conditions and/or model parameters. Here, we employ an alternative approach, which to our knowledge has not been previously pursued, whereby KAF and LIM approaches are used to predict the probability of occurrence of El Niño or La Niña events directly, without generating ensembles of trajectories. Our approach is based on the fact that predicting conditional probability is equivalent to predicting the conditional expectation of a characteristic function indexing the event of interest¹³. As a result, this task can be carried out using the same KAF and LIM approaches described above, replacing the Niño 3.4 index by the appropriate characteristic function as the response variable.

Here, we construct characteristic functions for El Niño and La Niña events in the HadISST and CCSM4 data using the standard criterion requiring that at least five overlapping seasonal (3-month running averaged) Niño 3.4 indices to be greater (smaller) than 0.5 (−0.5) °C, respectively³⁵. In this context, natural skill metrics are provided by the BSS^{30,31}, as well as relative entropy from information theory^{32,33} (see Methods). For binary outcomes, such as occurrence vs. non-occurrence of El Niño or La Niña events, the BSS is equivalent to a climatology-normalized RMSE score for the corresponding characteristic function. As information-theoretic metrics, we employ two relative entropy functionals³⁴, denoted \mathcal{D} and \mathcal{E} , which measure respectively the precision relative to climatology and ignorance (lack of accuracy) relative to the truth in a probabilistic forecast. For a binary outcome, \mathcal{D} attains a maximal value \mathcal{D}_* , corresponding to a maximally precise binary forecast relative to climatology. On the other hand, \mathcal{E} is unbounded, but has a natural scale \mathcal{E}_* corresponding to the ignorance of a probabilistic forecast based on random draws from the climatology; this makes $\mathcal{E} = \mathcal{E}_*$ a natural threshold indicating loss of skill. Overall, a skillful binary probabilistic forecast should simultaneously have $BSS \ll 1$, $\mathcal{D} \simeq \mathcal{D}_*$, and $\mathcal{E} \ll \mathcal{E}_*$. Note that we only report values of these scores in the CCSM4 experiments, as we found that the HadISST verification is not sufficiently long for statistically robust computation of relative-entropy scores.

The results of probabilistic El Niño/La Niña prediction for CCSM4 and HadISST are shown in Figs. 5 and 6, respectively. As is evident in Fig. 5(a,b), in the context of CCSM4, KAF performs markedly better than LIM in terms of the BSS and \mathcal{E} metric for all examined lead times, and for both El Niño and La Niña events. The \mathcal{D} scores in Fig. 5(c) start from $\simeq 0.3$ at forecast initialization (zero lead) for both methods, and while the KAF scores exhibit a monotonic decrease with lead time, in the case of LIM they exhibit an oscillatory behavior, hovering around $\simeq 0.25$ values. The latter behavior, in conjunction with a steady decrease of BSS and \mathcal{E} seen in Fig. 5(a,b), is a manifestation of the fact that, as the lead time grows, LIM produces biases, likely due to dynamical model error. Note, in particular, that a forecast of simultaneously high ignorance (i.e., large error) and high precision, as the LIM forecast at late times, must necessarily be statistically biased as it underestimates uncertainty. In contrast, the KAF-derived results exhibit a simultaneous increase of ignorance and decrease of precision, as expected for an unbiased forecast under complex dynamics with intrinsic predictability limits. Turning to the HadISST results, Fig. 6 again demonstrates that KAF performs noticeably better than LIM for both El Niño and La Niña prediction. Note that the apparent “false positives” in the KAF-based La Niña results around 2017 are not necessarily unphysical. In particular, there are weak La Niña events documented during 2016–2017 and 2017–2018, which are excluded from the true characteristic function for La Niñas, but may exhibit residuals in the approximate characteristic function output by KAF.

In conclusion, this work has demonstrated the efficacy of KAF in statistical ENSO prediction, with a robust improvement in useful prediction horizon in observational data by 3–7 months over LIM approaches, and three appealing characteristics—namely, more skillful forecasts at long lead times with slower reduction of skill, reduced spring predictability barrier, and improved prediction of event onset. Moreover, in the setting of model data with larger sample sizes, the enhanced performance of KAF becomes more pronounced, with skill extending out to 24 month leads. Aside from the higher skill in predicting ENSO indices, the method is also considerably more skillful in probabilistic El Niño/La Niña prediction. We attribute these improvements to the nonparametric nature of KAF, which can consistently approximate optimal prediction functions via conditional expectation in the presence of nonlinear dynamics through a rigorous connection with Koopman operator theory. A combination of this approach with delay-coordinate maps further aids its capability to extract dynamically coherent predictor variables (through kernel eigenfunctions), including seasonal variability associated with ENSO combination modes, representations of higher-order ENSO statistics due to atmospheric noise, and Pacific decadal variability. Even though nonparametric ENSO prediction methods are not particularly common, this study shows that methods such as KAF have high potential for skillful ENSO forecasts at long lead times, and can be naturally expected to be advantageous in forecasting other geophysical phenomena and their impacts, particularly in situations where the underlying dynamics is unknown or partially understood.

Methods

Datasets. The observational data used in this study consists of monthly averaged SST fields from the Hadley Centre Sea Ice and Sea Surface Temperature (HadISST) dataset^{27,39}, sampled on a $1^\circ \times 1^\circ$ latitude–longitude grid, and spanning the industrial era, 1870–2017. The modeled SST data are from a 1300-year, pre-industrial control integration of the Community Climate System Model version 4 (CCSM4)^{28,40}, monthly averaged, and sampled on the model’s native ocean grid of approximately 1° nominal resolution. All experiments use SST on the Indo-Pacific longitude-latitude box $28^\circ\text{E}–70^\circ\text{W}$, $30^\circ\text{S}–20^\circ\text{N}$ as input predictor data, and the corresponding Niño 3.4

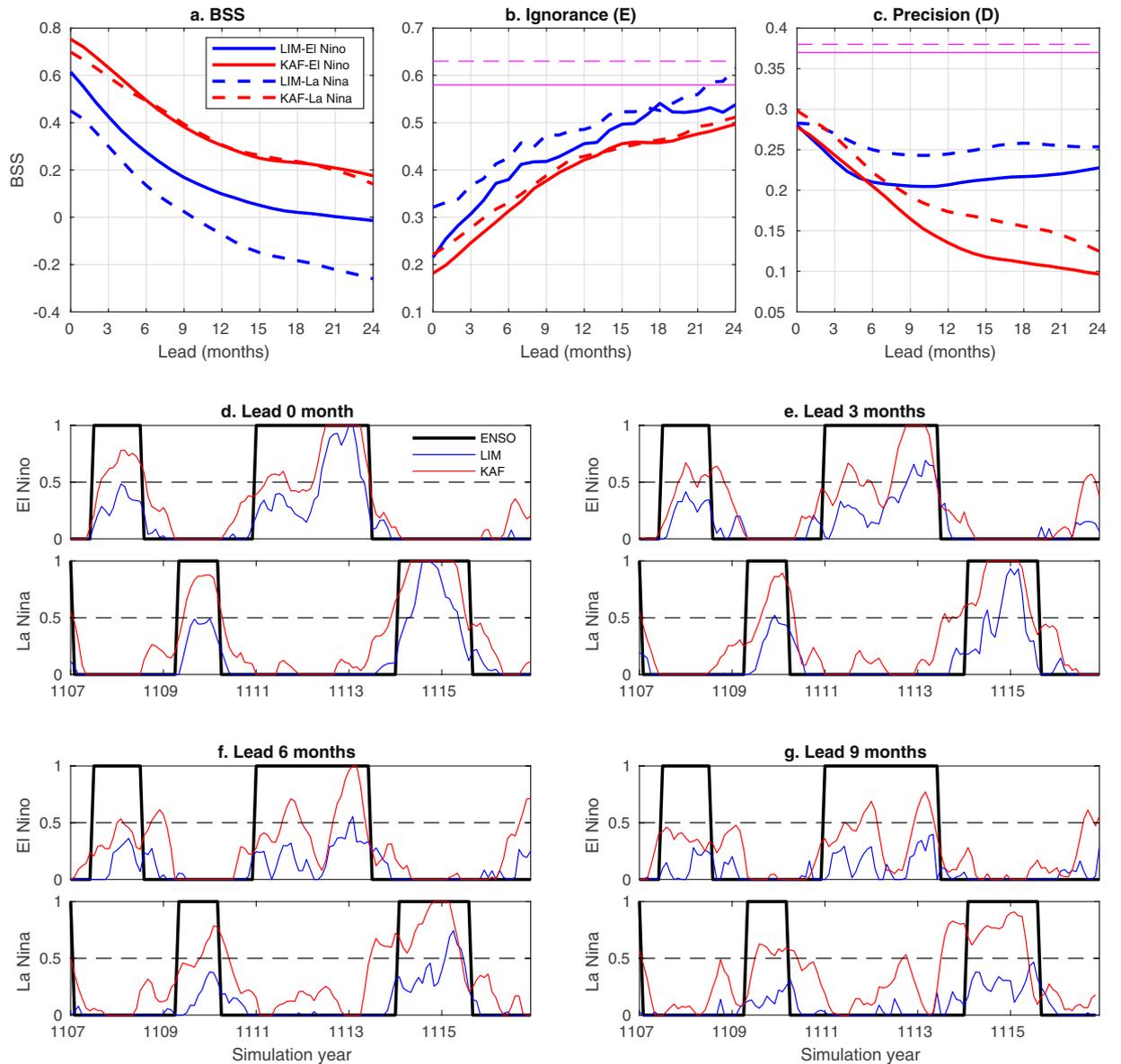


Figure 5. Probabilistic El Niño (solid lines) and La Niña (dashed lines) forecasts in CCSM4 data during simulation years 1100–1300, using KAF (red lines) and LIMs (blue lines). (a–c) BSS and information-theoretic ignorance (\mathcal{E}) and precision (\mathcal{D}) metrics, respectively, as a function of lead time. Magenta lines in (b,c) indicate the entropy thresholds \mathcal{E}_* and \mathcal{D}_* , respectively. (d–g) Running forecasts of the characteristic function for El Niño/La Niña events, representing conditional probability, for representative lead times in the range 0–9 months. Black lines show the true signal at the corresponding lead times.

indices as target (predictand) variables. The latter are defined as SST anomalies relative to a monthly climatology (to be specified below) computed from each dataset, spatially averaged over the region 5°N – 5°S , 170° – 120°W . The number p of spatial gridpoints in the HadISST and CCSM4 Indo-Pacific domains is 11,315 and 31,984, respectively.

Each dataset is divided into disjoint, temporally consecutive, training and verification periods. In the HadISST results in Figs. 1, 3 and 6, these periods are January 1870 to December 1997 and January 1998 to December 2017, respectively. The 2015/16 El Niño forecasts in Fig. 4 use January 1870 to December 2014 for training and May 2015 to May 2017 for verification. In the CCSM4 results in Figs. 2 and 5, the training and verification periods correspond to simulation years 1–1100 and 1101–1300, respectively. In separate calculations, we have removed portions of the training data to perform parameter tuning via hold-out validation. Note that we use hold-out validation as opposed to cross-validation in order to reduce the risk of overfitting the training data and estimating artificial skill in the verification phase. In the HadISST experiments, SST anomalies are computed by subtracting monthly means of the period 1981–2010. In CCSM4, anomalies are computed relative to the monthly means of the 900 yr training periods. See Eq. (8) ahead for an explicit formula relating SST data vectors to Niño 3.4 indices.

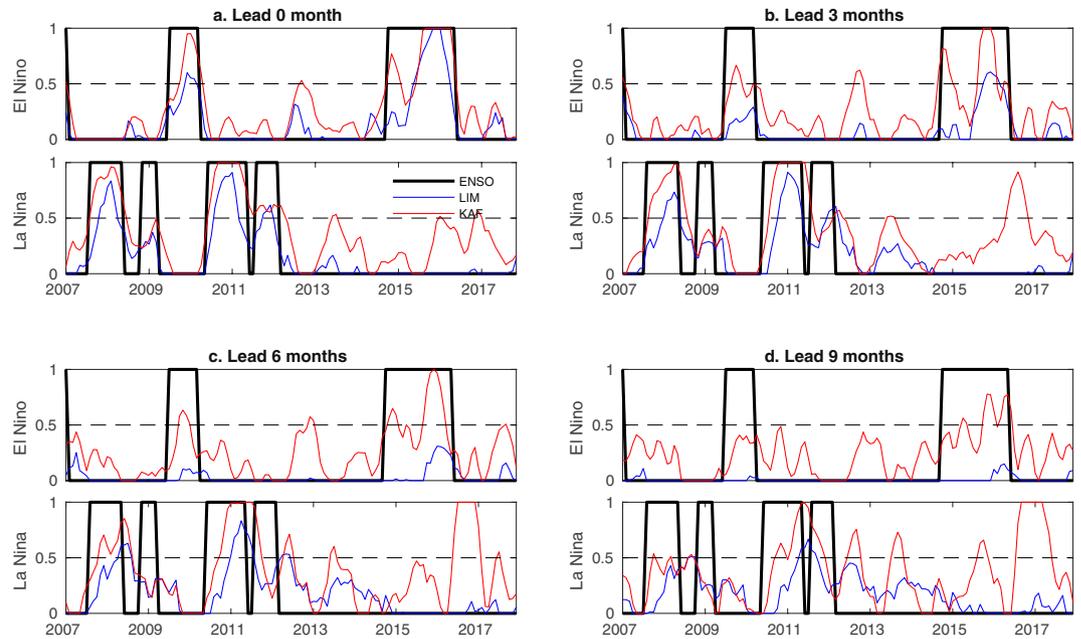


Figure 6. As in Fig. 5(d–g), but for probabilistic El Niño/La Niña forecasts in industrial-era HadISST data during 2007–2017.

Kernel analog forecasting. Kernel analog forecasting (KAF)^{12,13} is a kernel-based nonparametric forecasting technique for partially observed, nonlinear dynamical systems. Specifically, it addresses the problem of predicting the value of a time series $y(t + \tau) \in \mathbb{R}$, where t and τ are the forecast initialization and lead times, respectively, given a vector $x(t) \in \mathbb{R}^m$ of predictor variables observed at forecast initialization, under the assumption that $y(t + \tau)$ and $x(t)$ are generated by an underlying dynamical system. To make a prediction of $y(t + \tau)$ at lead time $\tau = q\Delta t$, where q is a non-negative integer and $\Delta t > 0$ a fixed interval, it is assumed that a time-ordered dataset consisting of pairs (x_n, y_{n+q}) , with

$$x_n = x(t_n), \quad y_n = y(t_{n+q}), \quad t_n = n\Delta t, \quad (1)$$

and $n \in \{0, \dots, N - 1\}$, is available for training. In the examples studied here, $y(t)$ is the Niño 3.4 index, $x(t)$ is a lagged sequence of Indo-Pacific SST snapshots (to be defined in Eq. (4) below), Δt is equal to 1 month, and N is the number of samples in the training data.

For every such lead time τ , KAF constructs a forecast function $f_\tau: \mathbb{R}^m \rightarrow \mathbb{R}$, such that $f_\tau(x(t))$ approximates $y(t + \tau)$. To that end, it treats $x(t)$ and $y(t)$ as the values of observables (functions of the state) along a trajectory of an abstract dynamical system (here, the Earth's climate system), operating on a hidden state space Ω . On Ω , the dynamics is characterized by a flow map $\Phi^t: \Omega \rightarrow \Omega$, such that $\Phi^t(\omega)$ corresponds to the state reached after dynamical evolution by time t , starting from a state $\omega \in \Omega$. Moreover, there exist functions $X: \Omega \rightarrow \mathbb{R}^m$ and $Y: \Omega \rightarrow \mathbb{R}$ such that, along a dynamical trajectory initialized at an arbitrary state $\omega \in \Omega$, we have $x(t) = X(\Phi^t(\omega))$ and $y(t) = Y(\Phi^t(\omega))$. That is, in this picture, $x(t)$ and $y(t)$ are realizations of random variables, X and Y , referred to as predictor and response variables, respectively, defined on the state space Ω .

It is a remarkable fact, first realized in the seminal work of Koopman⁴¹ and von Neumann⁴² in the 1930s, that the action of a general nonlinear dynamical system on observables such as X and Y can be described in terms of intrinsically linear operators, called Koopman operators. In particular, taking $\mathcal{F} = \{f: \Omega \rightarrow \mathbb{R}\}$ to be the vector space of all real-valued observables on Ω (note that Y lies in \mathcal{F}), the Koopman operator U_τ at time τ is defined as the linear operator mapping $f \in \mathcal{F}$ to $g = U_\tau f \in \mathcal{F}$, such that $g(\omega) = f(\Phi^\tau(\omega))$. From this perspective, the problem of forecasting $y(t + \tau)$ at lead time τ becomes a problem of approximating the observable $Y_\tau = U_\tau Y$; for, if Y_τ were known, one could compute $Y_\tau(\Phi^t(\omega)) = y(t + \tau)$, where $\omega \in \Omega$ is the state initializing the observed dynamical trajectory. It should be kept in mind that while \mathcal{F} is by construction a linear space, on which U_τ acts linearly without approximation, the elements of \mathcal{F} may not (and in general, will not) be linear functions satisfying a relationship such as $f(\omega + c\omega') = f(\omega) + cf(\omega')$ for $\omega, \omega' \in \Omega$ and $c \in \mathbb{R}$. In fact, in many applications Ω is not a linear set (e.g., it could be a nonlinear manifold), in which case no elements of \mathcal{F} are linear functions.

In the setting of forecasting with initial data determined through X , a practically useful approximation of $U_\tau Y$ must invariably be through a function $F_\tau \in \mathcal{F}$ that can be evaluated given values of X alone. That is, we must have $F_\tau(\Phi^t(\omega)) = f_\tau(x(t))$, where $f_\tau: \mathbb{R}^m \rightarrow \mathbb{R}$ is a real-valued function on data space, referred to above as the forecast function, and $x(t) = X(\Phi^t(\omega))$. In real-world applications, including the ENSO forecasting problem studied here, the observed data $x(t)$ are generally not sufficiently rich to uniquely infer the underlying dynamical state on Ω

(i.e., X is a non-invertible map). In that case, any forecast function f_τ will generally exhibit a form of irreducible error. The goal then becomes to construct F_τ optimally given the training data (x_n, y_{n+q}) so as to incur minimal error with respect to a suitable metric. Effectively, this is a learning problem for a function in an infinite-dimensional function space, which KAF addresses using kernel methods from statistical learning theory¹⁴.

Following the basic tenets of statistical learning theory, and in particular kernel principal component regression, KAF searches for an optimal f_τ in a finite-dimensional hypothesis space \mathcal{H}_L , of dimension L , which is a subspace of a reproducing kernel Hilbert space (RKHS), \mathcal{H} , of real-valued functions on data space \mathbb{R}^m . For our purposes, the key property that the RKHS structure of \mathcal{H} allows is that a set of orthonormal basis functions $\psi_0, \dots, \psi_{L-1}$ for \mathcal{H}_L can be constructed from the observed data x_0, \dots, x_{N-1} , such that $\psi_l(x)$ can be evaluated at any point $x \in X$, not necessarily lying in the set of training points x_n . In particular, $\psi_l(x(t))$ can be evaluated for the data $x(t)$ observed at forecast initialization. As with every RKHS, \mathcal{H} is uniquely characterized by a symmetric, positive-definite kernel function $k: \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$, which can be intuitively thought of as a measure of similarity, or correlation, between data points. As a simple example, EOF analysis is based on a covariance kernel, $k(x, x') = x^\top x'$, also known as a linear kernel, but note that the kernel k used in our experiments is a Markov-normalized, nonlinear Gaussian kernel (whose construction will be described below).

Associated with any kernel function k and the training data x_0, \dots, x_{N-1} is an $N \times N$ symmetric, positive-semidefinite kernel matrix \mathbf{K} with elements $K_{ij} = k(x_{i-1}, x_{j-1})$, and a corresponding orthonormal basis of \mathbb{R}^N consisting of eigenvectors ϕ_1, \dots, ϕ_N , satisfying

$$\mathbf{K}\phi_l = \lambda_l\phi_l, \quad \lambda_l \geq 0, \quad \phi_l = (\phi_{0l}, \dots, \phi_{N-1,l})^\top.$$

By convention, we order the eigenvalues λ_l in decreasing order. Assuming then that λ_L is strictly positive, the basis functions $\psi_l: \mathbb{R}^m \rightarrow \mathbb{R}$ of the hypothesis space \mathcal{H}_L are given by

$$\psi_l(x) = \frac{1}{N\lambda_l^{1/2}} \sum_{n=0}^{N-1} k(x, x_n)\phi_{nl}.$$

Given these basis functions, the KAF forecast function f_τ at lead time $\tau = q\Delta t$ is expressed as the linear combination

$$f_\tau = \sum_{l=1}^L \frac{1}{\lambda_l^{1/2}} c_l(\tau)\psi_l, \quad c_l(\tau) \in \mathbb{R},$$

where the expansion coefficients $c_l(\tau)$ are determined by regression of the time-shifted response values $y(t_n + \tau) = y_{n+q}$ against the eigenvectors ϕ_l , viz.

$$c_l(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} \phi_{nl} y_{n+q}.$$

One can verify that with this choice of expansion coefficients $c_l(\tau)$, f_τ minimizes an empirical risk functional $\mathcal{E}(f_\tau) = \sum_{n=0}^N |f_\tau(x_n) - y_{n+q}|^2 / N$ over functions in the hypothesis space \mathcal{H}_L .

A key aspect of the KAF forecast function is its asymptotic behavior as the number of training data N and the dimension of the hypothesis space L grow. In particular, it can be shown¹³ that if the dynamics on Ω has an ergodic invariant probability measure (i.e., there is a well-defined notion of climatology, which can be sampled from long dynamical trajectories), and the eigenvalues of the kernel matrix \mathbf{K} are all strictly positive for every n , then under mild regularity assumptions on X, Y , and the kernel k (related to continuity and finite variance), F_τ converges in a joint limit of $N \rightarrow \infty$ followed by $L \rightarrow \infty$ to the conditional expectation $\mathbb{E}(U_\tau Y | X)$ of the response observable $U_\tau Y$ evolved under the Koopman operator for the desired lead time, conditioned on the data available at forecast initialization through X . In particular, it is a consequence of standard results from probability theory and statistics that $\mathbb{E}(U_\tau Y | X)$ is optimal in the sense of minimal RMSE among all finite-variance approximations of $U_\tau Y$ that depend on the values of X alone. Note that no linearity assumption on the dynamics was made in order to obtain this result. It should also be noted that while, as with many statistical forecasting techniques, stating analogous asymptotic convergence or optimality results in the absence of measure-preserving, ergodic dynamics is a challenging problem, the KAF formulation described above remains well-posed in quite general settings, including the long-term climate change trends present in the observational SST data studied in this work.

Choice of predictors and kernel. Our choice of predictor function X and kernel k is guided by two main criteria: (i) X should contain relevant information to ENSO evolution beyond the information present in individual SST snapshots. (ii) k should induce rich hypothesis spaces \mathcal{H}_L ; in particular, the number of positive eigenvalues λ_l (which controls the maximal dimension of \mathcal{H}_L) should grow without bound as the size N of the training dataset grows. First, note that the covariance kernel employed in EOF analysis is not suitable from the point of view of the latter criterion, since in that case the number of positive eigenvalues is bounded above by the dimension of the data space m (which also bounds the number of linearly independent, linear functions of the input data). This means that one cannot increase the hypothesis space dimension L to beyond m , even if plentiful data

is available, i.e., $N \gg m$. In response, following earlier work^{17,18,43}, to construct k we start from a kernel of the form $\tilde{k}(x, x') = h(d(x, x'))$, where $h: \mathbb{R} \rightarrow \mathbb{R}$ is a nonlinear shape function, set here to a Gaussian,

$$h(u) = e^{-u^2/\epsilon},$$

with $\epsilon > 0$, and $d: \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$ a distance-like function on data space. We set d to an anisotropic modification of the Euclidean distance, shown to improve skill in capturing slow dynamical timescales⁴³. Specifically, choosing the dimension m of the predictor space to be an even number, and (for reasons that will become clear below) partitioning the predictor vectors into two $(m/2)$ -dimensional components, i.e., $x = (v_0, v_{-1})$ with v_0 and v_{-1} column vectors in $\mathbb{R}^{m/2}$, we define

$$d^2(x, x') = \frac{\|\Delta x\|^2}{\|\xi\| \|\xi'\|} \sqrt{(1 - \zeta \cos^2 \theta)(1 - \zeta \cos^2 \theta')}. \quad (2)$$

Here, ζ is a parameter lying in the interval $[0, 1)$, $\|\cdot\|$ is the standard (Euclidean) 2-norm, and $\Delta x, \xi, \xi'$ are vectors in $\mathbb{R}^{m/2}$ given by

$$\Delta x = v'_0 - v_0, \quad \xi = v_0 - v_{-1}, \quad \xi' = v'_0 - v_{-1}.$$

Moreover, θ and θ' are the angles between Δx and ξ and ξ' , respectively, so that

$$\cos \theta = \frac{\Delta x^\top \xi}{\|\Delta x\| \|\xi\|}, \quad \cos \theta' = \frac{\Delta x^\top \xi'}{\|\Delta x\| \|\xi'\|}.$$

The kernel \tilde{k} is then Markov-normalized to obtain k using the normalization procedure introduced in the diffusion maps algorithm²⁵. This involves the following steps:

$$k(x, x') = \frac{\hat{k}(x, x')}{\sum_{n=0}^{N-1} \hat{k}(x, x_n)}, \quad \hat{k}(x, x') = \frac{\tilde{k}(x, x')}{\sum_{n=0}^{N-1} \tilde{k}(x', x_n)},$$

where x_n are the predictor values from the training dataset. With this construction, all eigenvalues of K are positive if the bandwidth parameter ϵ is small-enough.

What remains is to specify the predictor function X . For that, we follow the popular approach employed, among many techniques, in singular spectrum analysis (SSA)⁴⁴, extended EOF analysis, and NLSA^{17,18}, which involves augmenting the dimension of data space using the method of delay-coordinate maps²⁴. Specifically, using $S: \Omega \rightarrow \mathbb{R}^p$ to denote the function on state space such that

$$s = S(\omega) \quad (3)$$

is equal to the values of the SST field sampled at the p gridpoints on the Indo-Pacific domain corresponding to climate state ω , we set

$$\tilde{X}(\omega) = (S(\omega), S(\Phi^{-\Delta t}(\omega)), \dots, S(\Phi^{-(Q-1)\Delta t}(\omega))) \in \mathbb{R}^{m/2},$$

where Q is the number of delays, and $m = 2pQ$. It is well known from dynamical systems theory²⁴ that for sufficiently large Q , $\tilde{X}(\omega)$ generically becomes in one-to-one correspondence with ω , meaning that as Q grows, $\tilde{X}(\omega)$ becomes a more informative predictor than $S(\omega)$. We then build our final predictor function $X: \Omega \rightarrow \mathbb{R}^m$ by concatenating \tilde{X} with its one-timestep lag, viz.

$$X(\omega) = (\tilde{X}(\omega), \tilde{X}(\Phi^{-\Delta t}(\omega))). \quad (4)$$

In this manner, the quantities ξ and ξ' in the anisotropic distance in Eq. (2) measure the local time-tendencies of the predictor data, and due to the presence of the $\cos \theta$ and $\cos \theta'$ terms, $k(x, x')$ takes preferentially larger values on pairs of predictors where the mutual displacement Δx is aligned with the local time-tendencies. Together, the delay-embedding and anisotropic distance contribute to the ability of $k(x, x')$ to identify states that evolve in a similar manner (i.e., are good analogs of one-another) over longer periods of time than kernels operating on individual snapshots.

Elsewhere^{19,20}, it was shown that as Q increases, the eigenvectors of the corresponding kernel matrix K , and thus the hypothesis spaces \mathcal{H}_L , become increasingly adept at capturing intrinsic dynamical timescales associated with the spectrum of the Koopman operator of the dynamical system. In particular, it has been shown^{21–23} that for interannual embedding windows, $Q \gtrsim 24$, the leading eigenvectors of K become highly efficient at capturing coherent modes of variability associated with ENSO evolution, as well as its linkages with the seasonal cycle and decadal variability. Due to this property, the kernels employed in this work are expected to be useful for ENSO prediction since the associated eigenspaces can capture nonlinear functions of the input data, and within those eigenspaces, meaningful representations of ENSO dynamics are possible using a modest number of eigenfunctions.

Linear inverse models. Under the classical LIM ansatz¹, the dynamics of ENSO can be well modeled as linear system driven by stochastic external forcing represented as temporally Gaussian white noise. The method used here to conduct LIM experiments follows closely the procedure described by Penland and Sardeshmukh². Specifically, the dynamics is governed by a stochastic differential equation

$$\dot{\psi}(t) = B\psi(t) + \dot{W}(t), \tag{5}$$

where $\psi(t) \in \mathbb{R}^L$ is the LIM state vector at time t , B is an $L \times L$ matrix representing a stable dynamical operator, and $W(t)$ is a Gaussian white noise process.

Let $s(t)$ be an SST vector from Eq. (3), observed at forecast initialization time t , and $s'(t) = s(t) - \bar{s}(t)$ be the corresponding anomaly vector relative to the monthly climatology $\bar{s}(t)$ at time t (computed as described above for the HadISST and CCSM4 datasets). Then, the corresponding LIM state vector $\psi(t)$ is given by projection onto the leading L EOFs $e_l \in \mathbb{R}^p$, computed for the $p \times p$ spatial covariance matrix C based on the anomalies $s'(t_n)$ in the training data, i.e.,

$$Ce_l = \lambda_l e_l, \quad C = SS^T, \\ \psi(t) = (\psi_1(t), \dots, \psi_L(t))^T, \quad \psi_l(t) = e_l^T s'(t) / \lambda_l^{1/2}, \tag{6}$$

where S is the $p \times N$ data matrix whose n -th column is equal to $s'(t_{n-1})$.

The RMSE-optimal estimate for $\psi(t + \tau)$ at an arbitrary lead time τ under the assumed dynamics in Eq. (5) is then given by the solution of the deterministic part with initial data $\psi(t)$, viz.

$$\hat{\psi}(t + \tau) = G(\tau)\psi(t),$$

where $G(\tau) = \exp(B\tau)$. Given this estimate, the predicted Niño 3.4 index at time $t + \tau$ is given by

$$\hat{y}(t + \tau) = \sum_{l=1}^L z_l \hat{\psi}_l(t + \tau), \tag{7}$$

where $\hat{\psi}_l(t + \tau)$ is the l -th component of $\hat{\psi}(t + \tau)$, and z_l is the regression coefficient of the Niño 3.4 index $y(t_n)$ against the l -th SST principal component time series $\psi_l(t_n)$ in the training phase,

$$z_l = \frac{1}{N} \sum_{n=0}^{N-1} \psi_l(t_n) y(t_n).$$

Note that because the Niño 3.4 index $y(t)$ is a linear function of the Indo-Pacific SST anomalies $s'(t)$, forecasts via Eq. (7) are equivalent to (but numerically cheaper than) first computing LIM forecasts $\hat{s}'(t + \tau)$ of the full Indo-Pacific SST anomalies $s'(t + \tau)$, and then deriving from these forecasts a predicted Niño 3.4 index. Specifically, let J denote the set of indices j of the components $s'_j(t)$ of $s'(t)$ lying within the Niño 3.4 region, and $|J|$ the number of elements of J . Then, we have

$$y(t_n) = \frac{1}{|J|} \sum_{j \in J} s'_j(t_n), \tag{8}$$

so that

$$z_l = \frac{1}{N} \sum_{n=0}^{N-1} \psi_l(t_n) \frac{1}{|J|} \sum_{j \in J} s'_j(t_n) = \frac{1}{|J|} \sum_{j \in J} \frac{1}{N} \sum_{n=0}^{N-1} s'_j(t_n) \psi_l(t_n) \\ = \frac{\lambda_l^{1/2}}{|J|} \sum_{j \in J} e_{jl},$$

where e_{jl} is the j -th component of EOF e_l . Noticing that the j -th component of the LIM forecast $\hat{s}(t + \tau)$ is given by

$$\hat{s}_j(t + \tau) = \sum_{l=1}^L \lambda_l^{1/2} e_{jl} \hat{\psi}_l(t + \tau),$$

it then follows that

$$\hat{y}(t + \tau) = \sum_{l=1}^L \frac{\lambda_l^{1/2}}{|J|} \sum_{j \in J} e_{jl} \hat{\psi}_l(t + \tau) = \frac{1}{|J|} \sum_{j \in J} \hat{s}'_j(t + \tau), \tag{9}$$

from which we deduce that predicting using Eq. (7) is equivalent to Eq. (9). For completeness, we note that had one implemented KAF using the covariance kernel, the eigenvectors ϕ_l of the $N \times N$ kernel matrix $K = S^T S$ corresponding to nonzero eigenvalues would be principal components, given by linear projections of the data onto the corresponding EOF from Eq. (6); that is,

$$\phi_l = \mathbf{e}_l^\top \mathbf{S} / \lambda_l^{1/2}. \tag{10}$$

We emphasize that Eq. (10) is special to eigenvectors associated with covariance kernels, and in particular does not hold for the nonlinear Gaussian kernels employed in this work.

Returning to the LIM implementation, in practice, $\mathbf{G}(\tau_0)$ is first estimated at some time $\tau_0 = q_0 \Delta t$, $q_0 \in \mathbb{N}$, from the training SST data $\mathbf{s}(t_n)$ sampled at the times t_n from (1),

$$\mathbf{G}(\tau_0) = \left(\frac{1}{N} \sum_{n=0}^{N-1} \psi(t_n + \tau_0) \psi(t_n)^\top \right) \left(\frac{1}{N} \sum_{n=0}^{N-1} \psi(t_n) \psi(t_n)^\top \right)^{-1}, \tag{11}$$

and then $\mathbf{G}(\tau)$ is computed at the desired lead time τ by

$$\mathbf{G}(\tau) = (\mathbf{G}(\tau_0))^{\tau/\tau_0}. \tag{12}$$

All LIM-based forecasts reported in this paper were obtained via Eq. (7) with $\mathbf{G}(\tau)$ from Eq. (12). Note that under analogous ergodicity assumptions to those employed in KAF, the empirical time averages in Eq. (11) converge to climatological ensemble averages. It should also be noted that, among other approximations, the model structure in Eq. (5) assumes that the dynamics is seasonally independent.

Validation. The main tunable parameters of the KAF method employed here are the number of delays Q , the Gaussian kernel bandwidth ϵ , and the hypothesis space dimension L . Here, we use throughout the values $Q = 12$, $\epsilon = 1$, $\zeta = 0.99$, and $L = 400$. As stated above, these values were determined by hold-out validation, i.e., by varying the parameters seeking optimal prediction skill in validation datasets. Note that this search was not particularly exhaustive, as we found fairly mild dependence of forecast skill under modest parameter changes around our nominal values.

The LIMs in this work have two tunable parameters, namely τ_0 in Eq. (12) and the number L of principal components employed. We set $\tau_0 = 2$ months and $L = 20$, using the same hold-out validation procedure as in KAF.

Probabilistic forecasting. Our approach for probabilistic El Niño/La Niña forecasting is based on the standard result from probability theory that the conditional probability of a certain event to occur is equal to the conditional expectation of its associated characteristic function. Specifically, let, as above, $Y: \Omega \rightarrow \mathbb{R}$ be the real-valued function on state space Ω such that $Y(\omega)$ is equal to the Niño 3.4 index corresponding to climate state $\omega \in \Omega$. Let also $\bar{Y}: \Omega \rightarrow \mathbb{R}$ be the 3-month running-averaged Niño 3.4 index, i.e.,

$$\bar{Y}(\omega) = \frac{1}{3} \sum_{q=-1}^1 Y(\Phi^q \Delta t(\omega)).$$

Here, we follow a standard definition for El Niño events³⁵, which declares ω to be an El Niño state if $\bar{Y}(\omega) > 0.5$ °C for a period of five consecutive months about ω . This leads to the characteristic function $\chi_+ : \Omega \rightarrow \mathbb{R}$, such that $\chi_+(\omega) = 1$ if there exists a set J of five consecutive integers in the range $[-4, 4]$, such that $\bar{Y}(\Phi^j \Delta t(\omega)) > 0.5$ for all $j \in J$, and $\chi_+(\omega) = 0$ otherwise. Similarly, we define a characteristic function χ_- for La Niña events, requiring that $\bar{Y}(\omega) < -0.5$ °C for a five-month period about ω . With these definitions, the conditional probabilities $P_{\pm, \tau}(x(t))$ for El Niño/La Niña, respectively, to occur at lead time τ , given the predictor vector $x(t)$ at forecast initialization time t , is equal to the conditional expectation $\mathbb{E}(U_\tau \chi_\pm | X = x(t))$ of χ_\pm acted upon by the time- τ Koopman operator U_τ . In particular, because the values of χ_\pm are available to us over the training period, we can estimate $P_{\pm, \tau}(x(t))$ using the KAF and LIM methodologies analogously to the Niño 3.4 predictions described above. All probabilistic forecast results reported in this paper were obtained in this manner.

Forecast skill quantification. Let $\tilde{\omega}_n = \Phi^{n \Delta t}(\tilde{\omega}_0)$, with $\tilde{\omega}_0 \in \Omega$ and $n \in \{0, 1, \dots, \tilde{N} - 1\}$, be the climate states in Ω over a verification period consisting of \tilde{N} samples, and $\tilde{x}_n = X(\tilde{\omega}_n)$ and $\tilde{y}_{n+q} = U_\tau Y(\tilde{\omega}_n) = Y(\tilde{\omega}_{n+q})$ be the corresponding values of the predictor and response (Niño 3.4) functions at lead time $\tau = q \Delta t$. We assess the forecast skill of the prediction function f_τ for $U_\tau Y$ using the root-mean-square error (RMSE) and Pearson correlation (PC; also known as pattern correlation) scores, defined as

$$\text{RMSE}(\tau) = \sqrt{\frac{1}{\tilde{N}} \sum_{n=0}^{\tilde{N}-1} (f_\tau(\tilde{x}_n) - \tilde{y}_{n+q})^2}, \text{ PC}(\tau) = \frac{1}{\tilde{N}} \sum_{n=0}^{\tilde{N}-1} \frac{(f_\tau(\tilde{x}_n) - \hat{\mu}_\tau)(\tilde{y}_{n+q} - \mu_\tau)}{\hat{\sigma}_\tau \sigma_\tau},$$

respectively. Here, μ_τ ($\hat{\mu}_\tau$) and σ_τ ($\hat{\sigma}_\tau$) are the empirical means and standard deviations of y_{n+q} ($f_\tau(\tilde{x}_n)$), respectively.

In the case of the probabilistic El Niño/La Niña forecasts, we additionally employ the BSS and relative-entropy scores, which are known to provide natural metrics for assessing the skill of statistical forecasts³¹⁻³⁴. In what follows, we outline the construction of these scores in the special case of binary response functions, such as the characteristic functions χ_\pm for El Niño/La Niña events, taking values in the set $\{0, 1\}$. First, note that every probability measure on $\{0, 1\}$ can be characterized by a single real number $\pi \in [0, 1]$ such that the probability to

obtain 0 and 1 is given by π and $1 - \pi$, respectively. Given two such measures characterized by $\pi, \rho \in [0, 1]$, we define the quadratic distance function

$$D(\pi, \rho) = (\pi - \rho)^2.$$

Moreover, under the condition that $\rho = 0$ only if $\pi = 0$, we define the relative entropy (Kullback-Leibler divergence)

$$D_{\text{KL}}(\pi \parallel \rho) = \pi \log_2 \left(\frac{\pi}{\rho} \right) + (1 - \pi) \log_2 \left(\frac{1 - \pi}{1 - \rho} \right),$$

where, by convention, we set $\pi \log_2(\pi/\rho)$ or $(1 - \pi) \log_2[(1 - \pi)/(1 - \rho)]$ equal to zero whenever ρ or $1 - \rho$ is equal to zero, respectively. This quantity has the properties of being non-negative, and vanishing if and only if $\pi = \rho$. Thus, D_{KL} can be thought of as a distance-like function on probability measures, though note that it is non-symmetric (i.e., in general, $D_{\text{KL}}(\pi \parallel \rho) \neq D_{\text{KL}}(\rho \parallel \pi)$), and does not obey the triangle inequality. Intuitively, $D_{\text{KL}}(\pi \parallel \rho)$ can be interpreted as measuring the precision (additional information content) of the probability measure characterized by π relative to that characterized by ρ , or, equivalently, the ignorance (lack of information) of ρ relative to π .

In order to assess the skill of probabilistic El Niño forecasts, for each state $\tilde{\omega}_n$ in the verification dataset, we consider three probability measures on $\{0, 1\}$, characterized by (i) the predicted probability $P_{+, \tau}(\tilde{x}_n)$ determined by KAF or LIM; (ii) the climatological probability \bar{P}_+ , which is equal to the fraction of states $\tilde{\omega}_n$ in the verification dataset corresponding to El Niño events (i.e., those states for which $\chi_+(\tilde{\omega}_n) = 1$); and (iii) the true probability equal to the value $\chi_+(\tilde{\omega}_n)$ of the characteristic function indexing the event. Using these probability measures, for each lead time $\tau = q \Delta t, q \in \mathbb{N}_0$, we define the instantaneous scores

$$\begin{aligned} \mathcal{B}(\tau; \tilde{\omega}_n) &= D(\chi_+(\tilde{\omega}_{n+q}), P_{+, \tau}(\tilde{x}_n)), & \bar{\mathcal{B}}(\tau; \tilde{\omega}_n) &= D(\chi_+(\tilde{\omega}_{n+q}), \bar{P}_+), \\ \mathcal{D}(\tau; \tilde{x}_n) &= D_{\text{KL}}(P_{+, \tau}(\tilde{x}_n) \parallel \bar{P}_+), & \mathcal{E}(\tau; \tilde{\omega}_n) &= D_{\text{KL}}(\chi_+(\tilde{\omega}_{n+q}) \parallel P_{+, \tau}(\tilde{x}_n)). \end{aligned}$$

Among these, $\mathcal{B}(\tau; \tilde{\omega}_n)$ and $\bar{\mathcal{B}}(\tau; \tilde{\omega}_n)$ measure the quadratic error of the probabilistic binary forecasts by $P_{+, \tau}(\tilde{x}_n)$ and the climatological distribution \bar{P}_+ , respectively, relative to the true probability $\chi_+(\tilde{\omega}_{n+q})$. In particular, $\mathcal{B}(\tau; \tilde{\omega}_n)$ vanishes if and only if $P_{+, \tau}(\tilde{x}_n)$ is equal to 1 whenever $\chi_+(\tilde{\omega}_{n+q})$ is equal to 1 (i.e., if whenever the forecast model predicts an El Niño event, the event actually occurs). Moreover, based on the interpretation of relative entropy stated above, $\mathcal{D}(\tau; \tilde{x}_n)$ measures the additional precision in the forecast distribution relative to climatology, and $\mathcal{E}(\tau; \tilde{\omega}_n)$ the ignorance of the forecast distribution relative to the truth. In our assessment of a forecasting framework such as KAF and LIM we consider time-averaged, aggregate scores over the verification dataset, viz.

$$\begin{aligned} \mathcal{B}(\tau) &= \frac{1}{\tilde{N}} \sum_{n=0}^{\tilde{N}-1} \mathcal{B}(\tau; \tilde{\omega}_n), & \bar{\mathcal{B}}(\tau) &= \frac{1}{\tilde{N}} \sum_{n=0}^{\tilde{N}-1} \bar{\mathcal{B}}(\tau; \tilde{\omega}_n), \\ \mathcal{D}(\tau) &= \frac{1}{\tilde{N}} \sum_{n=0}^{\tilde{N}-1} \mathcal{D}(\tau; \tilde{x}_n), & \mathcal{E}(\tau) &= \frac{1}{\tilde{N}} \sum_{n=0}^{\tilde{N}-1} \mathcal{E}(\tau; \tilde{\omega}_n). \end{aligned}$$

Note that $\mathcal{B}(\tau)$ is equal to the mean square difference between the characteristic function χ_+ and the forecast function P_+ evaluated on the verification dataset. Similarly, $\bar{\mathcal{B}}(\tau)$ is equal to the mean square difference between χ_+ and the constant function equal to the climatological probability \bar{P}_+ . As is customary, we normalize $\mathcal{B}(\tau)$ by the climatological error $\bar{\mathcal{B}}(\tau)$, and subtract the result from 1, leading to the Brier skill score³¹

$$\text{BSS}(\tau) = 1 - \frac{\mathcal{B}(\tau)}{\bar{\mathcal{B}}(\tau)}.$$

Note that in the expression above we have tacitly assumed that the climatological forecast has nonzero error, i.e., $\bar{\mathcal{B}}(\tau) \neq 0$, which holds true apart from trivial cases. With these definitions, a skillful model for probabilistic El Niño prediction should have small values of BSS(τ), large values of $\mathcal{D}(\tau)$, and small values of $\mathcal{E}(\tau)$.

Let now $P_{+,*}$ be defined such that it is equal to 1 if $\bar{P}_+ \leq 0.5$, and 0 if $\bar{P}_+ > 0.5$. Note that the binary probability distribution represented by $P_{+,*}$ places all probability mass to the outcome in $\{0, 1\}$ that is least probable with respect to the climatological distribution, which in the present context corresponds to an El Niño event. One can verify that the precision score $\mathcal{D}(\tau)$ is bounded above by the relative entropy $\mathcal{D}_* = D_{\text{KL}}(P_{+,*} \parallel \bar{P}_+)$. The latter quantity provides a natural scale for $\mathcal{D}(\tau)$ corresponding to the precision score of a probabilistic forecast that is maximally precise relative to climatology, in the sense of predicting with probability 1 the climatologically least likely outcome. To define a natural scale for $\mathcal{E}(\tau)$, we consider the time-averaged relative entropy $\mathcal{E}_* = \sum_{n=0}^{\tilde{N}-1} D_{\text{KL}}(\chi_+(\tilde{\omega}_n) \parallel \bar{P}_+) / \tilde{N}$, i.e., the average ignorance of the climatological distribution relative to the truth. This quantity sets a natural threshold for useful probabilistic El Niño forecasts, in the sense that such forecasts should have $\mathcal{E}(\tau) < \mathcal{E}_*$. The BSS and relative entropy scores and thresholds for probabilistic La Niña prediction are derived analogously to their El Niño counterparts using the characteristic function χ_- .

Data availability

The datasets analyzed during the current study are available in the Earth System Grid repository, <https://www.earthsystemgrid.org/dataset/ucar.cgd.cesm4.joc.b40.1850.track1.1deg.006.html>, and the Haley Centre Ice and Sea Surface Temperature repository, <http://www.metoffice.gov.uk/hadobs/hadisst/data/download.html>. The datasets generated during the current study are available from the corresponding author on reasonable request.

Received: 21 June 2019; Accepted: 20 January 2020;

Published online: 14 February 2020

References

1. Penland, C. & Magorian, T. Prediction of Niño 3 sea surface temperatures using linear inverse modeling. *J. Climate* **6**, 1067–1076 (1993).
2. Penland, C. & Sardeshmukh, P. D. The optimal growth of tropical sea surface temperature anomalies. *J. Climate* **8**, 1999–2024 (1995).
3. Chapman, D., Cane, M. A., Henderson, N., Lee, D. E. & Chen, C. A vector autoregressive ENSO prediction model. *J. Climate* **28**, 8511–8520 (2015).
4. Kondrashov, D., Kravtsov, S., Robertson, A. W. & Ghil, M. A hierarchy of data-based ENSO models. *J. Climate* **18**, 4425–4444 (2005).
5. Lima, C. H., Lall, U., Jebara, T. & Barnston, A. G. Statistical prediction of ENSO from subsurface sea temperature using a nonlinear dimensionality reduction. *J. Climate* **22**, 4501–4519 (2009).
6. Van den Dool, H. *Empirical Methods in Short-Term Climate Prediction*. (Oxford University Press, Oxford, 2006).
7. Ding, H., Newman, M., Alexander, M. A. & Wittenberg, A. T. Skillful climate forecasts of the tropical Indo-Pacific Ocean using model-analogs. *J. Climate* **31**, 5437–5459 (2018).
8. Ding, H., Newman, M., Alexander, M. A. & Wittenberg, A. T. Diagnosing secular variations in retrospective ENSO seasonal forecast skill using CMIP5 model-analogs. *Geophys. Res. Lett.* **46**, 1721–1730 (2019).
9. Ham, Y.-G., Kim, J.-H. & Luo, J.-J. Deep learning for multi-year ENSO forecasts. *Nature* **573**, 568–572 (2019).
10. Lorenz, E. N. Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.* **26**, 636–646 (1969).
11. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
12. Zhao, Z. & Giannakis, D. Analog forecasting with dynamics-adapted kernels. *Nonlinearity* **29**, 2888 (2016).
13. Alexander, R. & Giannakis, D. Operator-theoretic framework for forecasting nonlinear time series with kernel analog techniques. *Phys. D*, <https://arxiv.org/abs/1906.00464>, In minor revision (2019).
14. Cucker, F. & Smale, S. On the mathematical foundations of learning. *Bull. Amer. Math. Soc.* **39**, 1–49 (2001).
15. Eisner, T., Farkas, B., Haase, M. & Nagel, R. *Operator Theoretic Aspects of Ergodic Theory*, vol. 272 of *Graduate Texts in Mathematics* (Springer, 2015).
16. Budisić, M., Mohr, R. & Mezić, I. Applied Koopmanism. *Chaos* **22**, 047510 (2012).
17. Giannakis, D. & Majda, A. J. Comparing low-frequency and intermittent variability in comprehensive climate models through nonlinear Laplacian spectral analysis. *Geophys. Res. Lett.* **39**, L10710 (2012).
18. Giannakis, D. & Majda, A. J. Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proc. Natl. Acad. Sci.* **109**, 2222–2227 (2012).
19. Giannakis, D. Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Appl. Comput. Harmon. Anal.* **47**, 338–396 (2019).
20. Das, S. & Giannakis, D. Delay-coordinate maps and the spectra of Koopman operators. *J. Stat. Phys.* **175**, 1107–1145 (2019).
21. Slawinska, J. & Giannakis, D. Indo-Pacific variability on seasonal to multidecadal time scales. Part I: Intrinsic SST modes in models and observations. *J. Climate* **30**, 5265–5294 (2017).
22. Giannakis, D. & Slawinska, J. Indo-Pacific variability on seasonal to multidecadal time scales. Part II: Multiscale atmosphere-ocean linkages. *J. Climate* **31**, 693–725 (2018).
23. Wang, X., Giannakis, D. & Slawinska, J. Antarctic circumpolar waves and their seasonality: Intrinsic traveling modes and ENSO teleconnections. *Int. J. Climatol.* **39**, 1026–1040 (2019).
24. Sauer, T., Yorke, J. A. & Casdagli, M. Embedology. *J. Stat. Phys.* **65**, 579–616 (1991).
25. Coifman, R. R. & Lafon, S. Diffusion maps. *Appl. Comput. Harmon. Anal.* **21**, 5–30 (2006).
26. Kleeman, R., Moore, A. W. & Neville, R. S. Assimilation of subsurface thermal data into a simple ocean model for the initialization of an intermediate tropical coupled ocean–atmosphere forecast model. *J. Climate* **123**, 3103–3113 (1995).
27. Rayner, N. A. *et al.* Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108** (2003).
28. Gent, P. R. *et al.* The Community Climate System Model version 4. *J. Climate* **24**, 4973–4991 (2011).
29. Deser, C. *et al.* ENSO and Pacific decadal variability in the Community Climate System Model Version 4. *J. Climate* **25**, 2622–2651 (2012).
30. Brier, G. W. Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.* **78**, 1–3 (1950).
31. Weigel, A. P., Liniger, M. A. & Appenzeller, C. The discrete Brier and ranked probability skill scores. *Mon. Wea. Rev.* **135**, 118–124 (2007).
32. Kleeman, R. Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.* **59**, 2057–2072 (2002).
33. DelSole, T. & Tippett, M. K. Predictability: Recent insights from information theory. *Rev. Geophys.* **45**, RG4002 (2007).
34. Giannakis, D., Majda, A. J. & Horenko, I. Information theory, model error, and predictive skill of stochastic models for complex nonlinear systems. *Phys. D* **241**, 1735–1752 (2012).
35. L’Heureux, M. L. *et al.* Observing and predicting the 2015/16 El Niño. *Bull. Amer. Meteorol. Soc.* **98**, 1363–1382 (2017).
36. Barnston, A. G., Tippett, M. K., Ranganathan, M. & L’Heureux, M. L. Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. *Climate Dyn.* **53**, 7215–7234 (2019).
37. Barnston, A. G. & Ropelewski, C. F. Prediction of ENSO episodes using canonical correlation analysis. *J. Climate* **5**, 1316–1345 (1991).
38. Stuecker, M. F., Timmermann, A., Jin, F.-F., McGregor, S. & Ren, H.-L. A combination mode of the annual cycle and the El Niño/Southern Oscillation. *Nat. Geosci.* **6**, 540–544 (2013).
39. HadISST. Hadley Centre Sea Ice and Sea Surface Temperature (HadISST1) data, <http://www.metoffice.gov.uk/hadobs/hadisst/data/download.html>, Accessed March 2019 (2013).
40. CCSM. Community Climate System Model Version 4 (CCSM4) data, <https://www.earthsystemgrid.org/dataset/ucar.cgd.cesm4.joc.b40.1850.track1.1deg.006.html>, Accessed March 2019 (2010).
41. Koopman, B. O. Hamiltonian systems and transformation in Hilbert space. *Proc. Natl. Acad. Sci.* **17**, 315–318 (1931).
42. Koopman, B. O. & von Neumann, J. Dynamical systems of continuous spectra. *Proc. Natl. Acad. Sci.* **18**, 255–263 (1931).
43. Giannakis, D. Dynamics-adapted cone kernels. *SIAM J. Appl. Dyn. Sys.* **14**, 556–608 (2015).
44. Ghil, M. *et al.* Advanced spectral methods for climatic time series. *Rev. Geophys.* **40** (2002).

Acknowledgements

D.G. received support by NSF grants DMS-1521775, 1551489, 1842538, and ONR YIP grant N00014-16-1-2649. X.W. was supported as a PhD student under the first NSF grant. J.S. was supported as a postdoctoral fellow under the second and third NSF grants.

Author contributions

X.W., J.S. and D.G. designed research. X.W. and J.S. performed numerical experiments. X.W., J.S. and D.G. analyzed the results. X.W. and D.G. wrote the main manuscript. All authors reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to D.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020