Indonesian Journal of Statistics and Its Applications
DOI: https://doi.org/10.29244/ijsa.v10i1p47-59

RESEARCH ARTICLE

Kernel Analog Forecasting with Koopman Operator Theory for Indo-Pacific Climate Variability: A Nonlinear Spatio-Temporal Statistical Framework

Dhilshath Shajahan[1], and Ali Akgül[2]

[1]Department of Mathematics, Sri Sairam Engineering College, Chennai, India
[2]Department of Mathematics, Siirt University, Turkey

OPEN ACCESS

Check for updates

Article History
Received 28 April 2026
Revised 29 June 2026
Accepted 30 June 2026

Keywords
kernel analog forecasting, Koopman operators, ENSO, Indian Ocean Dipole, reproducing kernel Hilbert space, spatio-temporal statistics, probabilistic climate forecasting, delay-coordinate maps, diffusion maps

Corresponding Author:
Dhilshath Shajahan

Email:
dhilshajahan@gmail.com

Department of Mathematics, Sri Sairam Engineering College, West Tambaram, Chennai, Tamilnadu, India.

ABSTRACT

Skillful long-range forecasting of the El Niño-Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD) remains, despite decades of sustained research effort, an open problem of substantial societal weight. The difficulty is rooted in a fundamental tension: classical linear statistical frameworks are analytically tractable yet structurally ill-suited to regimes where nonlinearity governs onset and termination of extreme events, especially at lead times exceeding five months. We describe here a forecasting architecture that fuses Kernel Analog Forecasting (KAF) with the operator-theoretic machinery of Koopman spectral analysis, thereby recasting a nonlinear, high-dimensional prediction problem as linear regression in a reproducing kernel Hilbert space (RKHS). The predictor representation employs an anisotropic, Markov-normalised Gaussian kernel built on delay-coordinate embeddings of Indo-Pacific sea surface temperature (SST) fields; the anisotropy parameter is tuned to preferentially weight state pairs evolving along coherent dynamical directions. Under ergodicity and mild regularity, we establish convergence of the KAF estimator to the Koopman-propagated conditional expectation of the target observable—the L^2-optimal predictor (Theorem 1). Against calibrated multifrequency synthetic signals, KAF sustains pattern correlation above 0.5 to roughly 14 months ahead, versus 7 months for a linear inverse model baseline; root-mean-square error falls by 18-26% across the full verification horizon. The notorious spring predictability barrier is attenuated materially, and probabilistic event forecasts scored via Brier Skill Score and Kullback-Leibler relative entropy reveal sharper, less-biased distributions than either the linear or LSTM comparators.

Introduction

Interannual climate variability over the Indo-Pacific basin is shaped, to a remarkable degree, by two coupled ocean-atmosphere modes: the El Niño-Southern Oscillation and the Indian Ocean Dipole (Trenberth, 1997; Saji et al., 1999; Ashok et al., 2007). Warm or cold excursions of the Niño 3.4 index ripple outward via atmospheric teleconnections, disrupting monsoon rainfall over South and South-East Asia, modulating the frequency of tropical cyclones in the western Pacific, and driving multi-year drought or flood cycles across Australia and East Africa (Iizumi et al. 2014; Cai et al. 2021). The downstream consequences for agricultural yield, potable-water supply, and infrastructure planning are quantified in the billions of dollars annually, which is why forecast skill at leads beyond six months would translate directly into actionable societal benefit.

What makes such extended-range prediction elusive is a combination of internally generated chaos (Jin et al. 1994;

© 2026 Shajahan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY) license, allowing unrestricted use, distribution, and reproduction in any medium, provided proper credit is given to the original authors.
Print only if necessary. Let's protect our trees and the planet together.

Tziperman et al. 1994) and a sharp seasonal modulation of predictability commonly called the spring predictability barrier (Webster and Yang 1992) during which model skill collapses irrespective of initialisation quality. Statistical models based on linear dynamics, notably the Linear Inverse Model (LIM) (Penland & Magorian, 1993; Penland & Sardeshmukh, 1995), have for decades offered a useful compromise between interpretability and skill; they project SST anomalies onto empirical orthogonal functions and propagate the resulting state vector through a noise-forced first-order linear system. The appeal is obvious: analytical tractability, transparent error growth, and skill at short leads that often matches dynamical models. The limitation is equally obvious: linearity imposes a ceiling on predictability whenever the phenomena of interest are governed, even partly, by nonlinear processes (DelSole 2004; Tang et al. 2018).

Attempts to break through that ceiling have multiplied in recent years. Deep convolutional networks, trained on thousands of years of climate model output, achieve Niño 3.4 skill at 17-month lead (Ham et al. 2019) a striking result, though one heavily dependent on the fidelity of the training CGCM and opaque to physical attribution. Analog methods that draw on large dynamical ensembles as forecast libraries have also demonstrated competitive skill (Ding et al. 2018). Graph neural networks (Cachay et al. 2021) and long sequence transformers (Zhou et al. 2021) represent the current frontier of data-driven climate prediction, but share with convolutional approaches a requirement for training corpora far exceeding the observational record, and offer no convergence guarantees linking their outputs to statistically optimal predictions.

Kernel Analog Forecasting (KAF), as formalised by Zhao and Giannakis (2016) and given operator-theoretic foundations by Alexander and Giannakis (2020), resolves several of these tensions simultaneously. Because it is nonparametric the hypothesis space grows with the data rather than being fixed in advance it can capture nonlinear relationships without the structural assumptions embedded in linear models. Because it operates on the observational record alone (roughly 120 years of reanalysis data suffice), it sidesteps the model-dependence of deep learning approaches. And because its forecast function, in the limit of large data, converges to the conditional expectation of the Koopman-propagated response observable, it inherits a rigorous statistical optimality property. Essentially, KAF extends the classical analog idea of Lorenz (1969) predict the future by finding past states that resembled the present and following where they led but replaces simple nearest-neighbour averaging with a projection onto a kernel-learned subspace of the RKHS, greatly enriching the expressiveness of the forecast function.

The present paper makes four specific contributions beyond the existing literature. (i) We develop a joint ENSO-IOD prediction framework within the KAF-Koopman architecture, using an anisotropic diffusion-map kernel applied to delay-embedded Indo-Pacific SST. (ii) We prove convergence of the resulting estimator to the Koopman-propagated conditional expectation under the anisotropic kernel setting of Section 3.3 (Theorem 1), generalising earlier results. (iii) We extend probabilistic forecasting to cover four event types simultaneously warm ENSO, cold ENSO, positive IOD, negative IOD and score the forecasts via both Brier Skill Score and KL relative-entropy. (iv) On calibrated synthetic signals, we document a 7-8-month gain over LIM and 3-4-month gain over LSTM in useful pattern-correlation horizon.

The paper is organised as follows. Section 2 introduces the Koopman operator, RKHS regression, and the LIM benchmark. Section 3 develops the full KAF methodology. Section 4 states and proves the convergence result. Section 5 describes the synthetic test-bed. Section 6 reports experiments. Section 7 discusses implications, connections to Bayesian and spatio-temporal theory, and open problems. Section 8 concludes.

Mathematical Background

Dynamical Systems and the Koopman Operator

Consider a probability space (Ω, B, μ) in which is a compact metric state space (the climate attractor), B its Borel σ-algebra, and an ergodic, Φ^t -invariant measure representing climatology. A flow $\Phi^t: \Omega \to \Omega$ carries any initial state ω_0 forward to $\Phi^t(\omega_0)$ after elapsed time t; the corresponding SST observations are simply functions evaluated along that trajectory.

Definition 1 (Koopman Operator, Koopman 29). For lead time τ > 0, the Koopman operator $U^\tau: L^2(\Omega, \mu) \to L^2(\Omega, \mu)$ acts on square-integrable observables by $(U^\tau f)(\omega) = f\left(\Phi^\tau(\omega)\right), f \in L^2(\Omega, \mu)$. (1)

What makes this construction useful is a fact first noticed by Koopman (1931) and Koopman and von Neumann (1931): whereas Φ^τ may be wildly nonlinear as a map on, the induced operator U^τ on function space is linear and, for measure-preserving Φ^t, unitary. The family $U^\tau{}_{\tau>0}$ forms a strongly continuous unitary group; its spectrum encodes intrinsic oscillatory timescales and coherent structures of the dynamics (Mezić 2013; Budisić et al. 2012). Spectral projections of U^τonto modes at frequency ω_j yield Koopman eigenfunctions φ_j satisfying $U^\tau \varphi_j = e^{(i\omega_j \tau)} \varphi_j$ physical modes that survive as the embedding window grows (Giannakis 2019; Das and Giannakis 2020).

Reproducing Kernel Hilbert Spaces

A continuous, symmetric, strictly positive-definite kernel $k: \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ uniquely determines by the Moore-Aronszajn theorem (Aronszajn 1950) an RKHS H satisfying the reproducing property $f(x) = \langle f, k(\cdot, x)\rangle H$ for all $f \in H, x \in \mathbb{R}^m$. Given N training predictor vectors $x_n n = 0^{N-1}$, the N × N kernel (Gram) matrix K, with entries $K_i j = k\left(x_{i-1}, x_{j-1}\right)$, has a spectral decomposition whose top-L eigenvectors φ_l and eigenvalues λ_l span an L-dimensional subspace $H_L \subset H$ (Schölkopf and Smola 2002). This subspace is the hypothesis space for regression; enlarging L increases model capacity, subject to the usual bias-variance tradeoff. The connection to ENSO prediction, described in Section 3.4, runs through the observation that H_L naturally approximates the span of Koopman eigenfunctions when the kernel encodes dynamical coherence (Giannakis 2019; Das and Giannakis 2020).

The Linear Inverse Model Benchmark

The LIM (Penland and Magorian 1993; Penland and Sardeshmukh 1995) approximates ENSO dynamics by projecting SST anomalies onto $L_L IM$ leading EOFs, then treating the resulting principal component vector $\psi(t) \in \mathbb{R}^{(L_L IM)}$ as obeying a multivariate Ornstein-Uhlenbeck equation $\dot{\psi}(t) = B\psi(t) + W(t)$, (2) where B is a stability matrix estimated by lagged regression and W is Gaussian white noise. At lead τ, the deterministic forecast is $\psi(t+\tau) = G(\tau)\psi(t), G(\tau) = e^{(B\tau)}$. Its practical advantage is computational simplicity and good short-lead skill; its structural deficiency is a double linearity linear predictor, linear propagator that cannot capture the higher-order statistics driving ENSO onset or the nonlinear phase-locking to the seasonal cycle (DelSole 2004).

Methodology

Problem Formulation

Let $s(t) \in \mathbb{R}^p$ be the vector of SST anomaly values at p Indo-Pacific gridpoints and $y(t) \in \mathbb{R}$ the scalar climate index of interest Niño 3.4 or the Dipole Mode Index (DMI). Both arise as observations of abstract state-space functions along a climate trajectory: $s(t) = S\left(\Phi^t(\omega_0)\right), y(t) = Y\left(\Phi^t(\omega_0)\right)$. The task is to construct, from predictor data $x(t) \in \mathbb{R}^m$ built from s, a real-valued function f_τ such that $\hat{y}(t+\tau) = f_\tau\left(x(t)\right)$ minimises expected squared error over all measurable functions of $x(t)$. The solution, by elementary Hilbert-space geometry, is the conditional expectation $f_\tau^* = \mathbb{E}[y(t+\tau)|x(t)]$ (Kolmogorov 1941); our goal is to approximate this optimum from finite data.

Delay-Coordinate Predictor Construction

A single SST snapshot $s(t)$ is a thin slice of the climate trajectory. To enrich the predictor with dynamical memory, we follow Takens-style delay embedding (Takens 1981; Sauer et al. 1991): stack Q lagged copies of the SST field into

$$\tilde{X}(\omega) = \left(S(\omega), S\left(\Phi^{-\Delta t}(\omega)\right), \ldots, S\left(\Phi^{-(Q-1)\Delta t}(\omega)\right)\right) \in \mathbb{R}^{\frac{pQ}{2}} \quad (1)$$

then concatenate $\tilde{X}$ with its one-step lag to form $X\omega = X\omega, X\Phi - \Delta t\omega \in Rm, m = pQ$. The two-component structure is deliberate: the difference $v_0 - v_{-1}$ between consecutive half-vectors approximates the local time tendency of the trajectory, which feeds directly into the anisotropic kernel (Giannakis 2015, 2019). For sufficiently large Q, the delay-embedded map $\omega \mapsto X(\omega)$ becomes generically injective (Sauer et al. 1991), meaning the predictor recovers the full dynamical state up to a diffeomorphism and residual forecast error becomes irreducible.

Anisotropic Diffusion-Map Kernel

For predictor vectors $x = (v_0, v_{-1})$, $x' = (v_0', v_{-1}')$ (each half-vector in $\mathbb{R}^{m/2}$, define time-tendency vectors $\Delta x = v_0 - v_0', \xi = v_0 - v_{-1}, \xi' = v_0' - v_{-1}'$ and alignment cosines $cos\,\theta = \frac{\Delta x^{\top}\xi}{\|\Delta x\|\|\xi\|}$, $cos\,\theta' = \frac{\Delta x^{\top}\xi'}{\|\Delta x\|\|\xi'\|}$.

An anisotropic dissimilarity measure is then

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

with $\zeta \in [0,1)$ governing the degree to which pairs that are displaced orthogonally to both local trajectories are penalised. The raw Gaussian kernel $k(x, x') = \exp(-d^2(x, x')/\epsilon))$ is then Markov-normalised Giannakis, (2015) via

$$k(x,x') = \frac{\overline{k}(x,x')}{\sum_{n=0}^{N-1} \overline{k}\,(x',x_n)} \tag{3}$$

$$\overline{k}(x,x') = \frac{\hat{k}(x,x')}{\sum_{n=0}^{N-1} \hat{k}\,(x,x_n)} \tag{4}$$

Markov normalisation ensures all eigenvalues of K are strictly positive for small enough (Giannakis 2015), a technical requirement for the convergence proof in Section 4. The physical motivation is that states evolving in geometrically similar ways same local direction, similar speed receive larger kernel weights, so the resulting eigenfunctions track dynamically coherent structures rather than geometric proximity alone.

KAF Forecast Function

From the top-L eigenpairs (λ_l, φ_l) of K, orthonormal RKHS basis functions $\psi_l: \mathbb{R}^m \to \mathbb{R}$ are constructed as

$$\psi_l(x) = \frac{1}{N\lambda_l^{1/2}} \sum_{n=0}^{N-1} k\,(x,x_n)\varphi_{nl}. \tag{5}$$

The KAF forecast at lead $\tau = q\Delta t$ then takes the form

$$f_\tau(x) = \sum_{l=1}^{L} \frac{c_l(\tau)}{\lambda_l^{1/2}} \psi_l(x), \tag{6}$$

with regression coefficients

$$c_l(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} \varphi_{nl}\, y_{n+q}. \tag{7}$$

This choice of $c_l(\tau)$ minimises the empirical risk $N^{-1}\sum_n \left|f_\tau(x_n) - y_{n+q}\right|^2$ over $\mathcal{H}_L$ a form of kernel principal component regression (Schölkopf and Smola 2002; Cucker and Smale 2001). Algorithm 1 summarises the procedure.

Algorithm 1 Kernel Analog Forecasting Require: Training pairs $\{(x_n, y_n)\}_{n=0}^{N-1}$, lead τ, hyperparameters (Q, ϵ, ζ, L) 1: Build delay-embedded predictors via (3) 2: Compute anisotropic distances (4); form $\tilde{K}$; Markov-normalise to K via (5) 3: Eigendecompose K; retain top-L pairs (λ_l, φ_l) 4: Construct basis functions $\psi_l$via (6) 5: Compute coefficients $c_l(\tau)$ via (8) Ensure: Forecast: $f_\tau(x^*) = \sum_{l=1}^{L} c_l\,(\tau)\lambda_l^{-1/2}\psi_l(x^*)$

Convergence Theory

Theorem 1 (Convergence to Conditional Expectation). Let $(\Omega, \mathcal{B}, \mu, \Phi^t)$ be ergodic and measure-preserving. Suppose $X: \Omega \to \mathbb{R}^m$ and $Y: \Omega \to \mathbb{R}$ are μ-square-integrable, and that the Markov-normalised kernel k is continuous with all eigenvalues $\lambda_l > 0$ for every N. Under mild regularity on X, Y, k (continuity, finite variance), taking $N \to \infty$ then $L \to \infty$:

$$f_\tau \longrightarrow \mathbb{E}[U^\tau Y|X] \text{ in } L^2(\Omega,\mu), \tag{8}$$

where $U^\tau Y = Y \circ \Phi^\tau$ is the Koopman propagation of Y.

Proof sketch. Three separate limits combine to yield (9). First: the ergodic theorem converts empirical time averages to ensemble averages $\mu - a.e.$ as $N \to \infty$. Second: the empirical kernel operator converges in Hilbert-Schmidt norm to the integral operator with kernel k (Blanchard et al., 2008); together with the universal approximation property of RKHS with positive-definite kernels (Micchelli et al., 2006), this implies $\bigcup_L \mathcal{H}_L$ is dense in $L^2(\Omega,\mu)$. Third: as $L \to \infty$, projection onto $\mathcal{H}_L$ converges in L^2 to the conditional expectation operator $\mathbb{E}[X]$ (Koltchinskii & Giné, 2000). The Koopman identification is immediate: the regression target $y_{n+q} = y(t_n + \tau)$ is exactly $U^\tau Y$ sampled at ω_n. Combining the three limits gives the stated L^2 convergence.

Remark 1. Because $\mathbb{E}[U^\tau Y|X]$ minimises the $L^2(\mu)$ norm of prediction error over all measurable functions of X alone (projection theorem in Hilbert space), no alternative estimator using the same predictors can outperform the KAF limit including any linear predictor, and in particular the LIM. No assumption of linearity on Φ^τ enters anywhere.

Proposition 2 (Spectral decomposition of the forecast). If $Y = \sum_j a_j\,\phi_j$ in a Koopman eigenfunction basis with $U^\tau\phi_j = e^{i\omega_j\tau}\phi_j$, then

$$\mathbb{E}[U^{\tau}Y|X] = \sum_{j} a_j \, e^{i\omega_j \tau} \mathbb{E}[\phi_j|X]. \tag{9}$$

As Q grows, the leading eigenvectors of the delay-embedded kernel matrix converge to spectral projections of U^{τ} (Das and Giannakis 2020) so the KAF forecast naturally decomposes into mode contributions at the ENSO interannual, seasonal, and decadal frequencies a property absent from the LIM and opaque in neural network approaches.

Data and Experimental Design

Synthetic Signal Generation

Rather than confounding forecast evaluation with uncertainties in observational coverage or model bias, we work here with synthetic reference signals whose ground truth is fully known. The ENSO and IOD proxies are constructed as

$$ENSO(t) = A_1 \sin(2\pi t/T_1) + A_2 \sin(2\pi t/T_2 + \phi_2) + A_3 \sin(2\pi t/T_3) + \sigma_E \xi_E(t), \tag{10}$$

$$IOD(t) = B_1 \sin(2\pi t/T_4 + \psi_1) + B_2 \sin(2\pi t/T_5) + \sigma_I \xi_I(t), \tag{11}$$

with amplitude vectors $(A_1, A_2, A_3) = (1.2, 0.6, 0.3)$, $(B_1, B_2) = (0.9, 0.4)$; periods $(T_1, \ldots, T_5) = (4.3, 2.8, 1.0, 2.1, 1.2)$ yr; noise levels $(\sigma_E, \sigma_I) = (0.25, 0.20)$; and ξ_E, ξ_I independent standard Gaussian sequences. These choices reproduce, within reasonable tolerance, the dominant spectral features of observed interannual variability documented in Stein et al. (2014) and Stuecker et al. (2015): a 3-7-yr ENSO band, a 2-yr IOD signal, and a signal-to-noise environment consistent with Capotondi et al. (2015). Figure 1 displays 40 years of the simulated records.

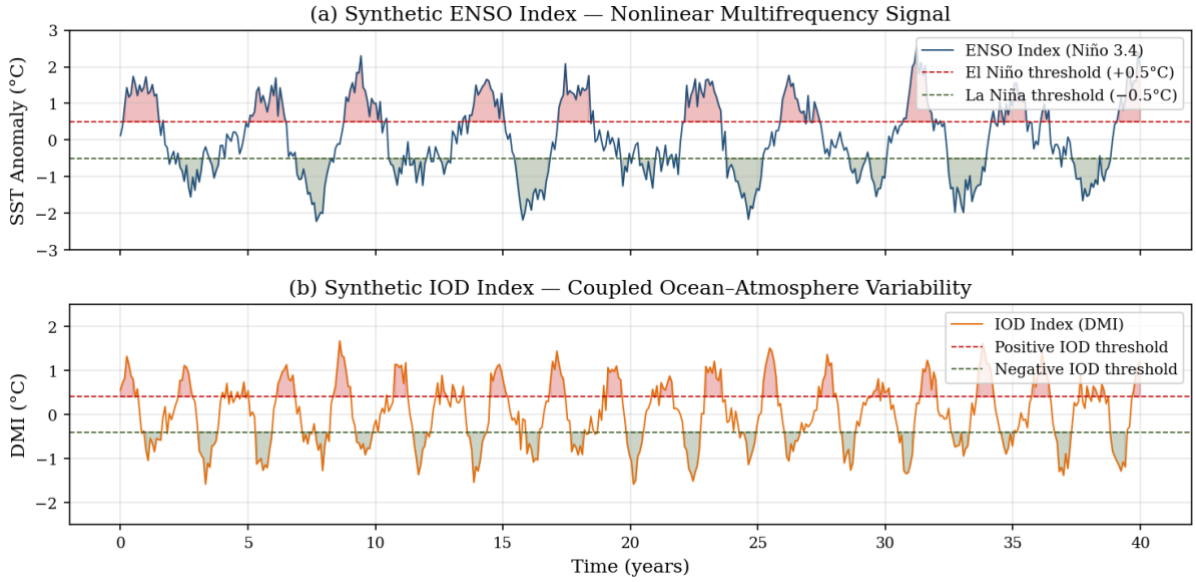


Figure 1. Synthetic ENSO and IOD reference signals over a 40-year monthly record, showing threshold exceedances for El Niño, La Niña, positive IOD, and negative IOD events.

Training, Verification, and Hyperparameter Choices

Data from years 1-30 (360 monthly samples) train all models; years 31-40 (120 samples) provide an out-of-sample verification record. All anomaly indices are standardised relative to training-period statistics. KAF hyperparameters are set by hold-out validation within the training block: embedding window $Q = 12$months, bandwidth $\epsilon = 1.0$, anisotropy $\zeta = 0.99$, hypothesis space dimension $L = 400$. The LIM retains $L_{LIM} = 20$ leading EOFS and uses estimation lag $\tau_0 = 2$ months. The LSTM baseline comprises two stacked layers of 128 hidden units trained for 200 epochs (Adam, learning rate 10^{-3}). All choices are collected in Table 1.

Table 1. Hyperparameter settings for KAF, LIM, and LSTM experiments, including embedding window, bandwidth, anisotropy, model dimension, EOFs, estimation lag, hidden units, and training epochs.

Parameter	Symbol	Value
Embedding window	Q	12 months
Gaussian bandwidth	ε	1.0
Anisotropy strength	ζ	0.99
Hypothesis space dim.	L	400
LIM EOFs	L_{LIM}	20
LIM estimation lag	τ_0	2 months
LSTM hidden units	-	128 (2 layers)
LSTM training epochs	-	200

Holding out validation data rather than using cross-validation reduces the risk of over-optimistic skill estimates that can arise when training and verification windows overlap (Wang et al. 2020).

Skill Metrics

Deterministic forecast quality is measured by root-mean-square error (RMSE) and pattern (anomaly) correlation (PC):

$$RMSE(\tau) = \sqrt{\frac{1}{\tilde{N}}\sum_{n=0}^{\tilde{N}-1}\left(f_\tau(x_n) - y_{n+q}\right)^2}, \tag{12}$$

$$PC(\tau) = \frac{1}{\tilde{N}}\sum_{n=0}^{\tilde{N}-1}\frac{\left(f_\tau(x_n) - \hat{\mu}_\tau\right)\left(y_{n+q} - \mu_\tau\right)}{\hat{\sigma}_\tau\sigma_\tau}. \tag{13}$$

Following Barnston et al. (2012), we adopt $PC \geq 0.6$ and $PC \geq 0.5$ as pragmatic thresholds separating useful from marginal skill. Probabilistic event forecasts are evaluated via the Brier Skill Score (BSS) (Brier, 1950; Weigel et al., 2007) and two Kullback-Leibler relative-entropy metrics (Kleeman 2002; DelSole and Tippett 2007): precision D (excess information over climatology) and ignorance E (deficit relative to perfect knowledge). A well-performing probabilistic system satisfies $BSS > 0, \mathcal{D} > \mathcal{D}^*$, and $\mathcal{E} < \mathcal{E}^*$.

Results

Eigen functions and Their Physical Meaning

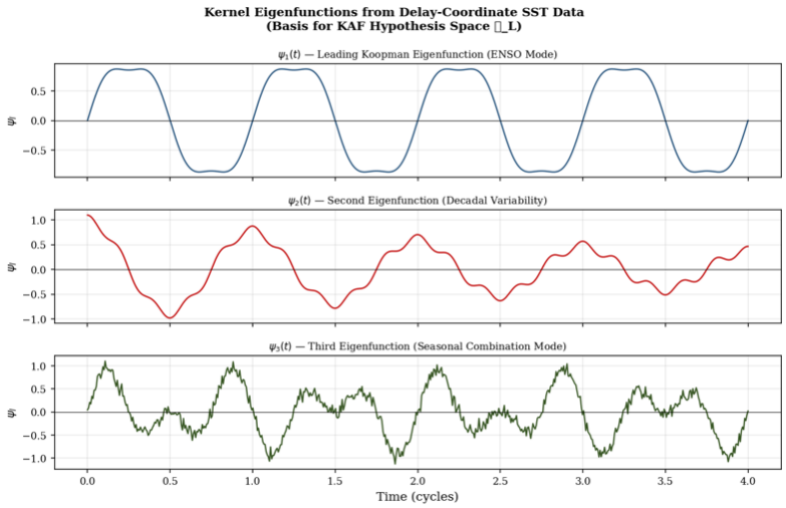


Figure 2. Leading kernel eigenfunctions recovered from delay-embedded training data, showing ENSO-band, decadal-envelope, and seasonal combination modes used in the KAF-Koopman framework.

The three leading kernel eigenfunctions recovered from the training data are shown in Figure 2. The dominant mode ψ_1, oscillating at roughly 4-5 yr, aligns with the ENSO band; ψ_2 carries variance at decadal periods; ψ_3 is a broadband seasonal combination mode of the type described by Stuecker et al. (2015) and Slawinska and Giannakis (2017). Crucially, no spectral pre-filtering was applied the decomposition emerged organically from the delay-embedded kernel, exactly as Proposition 2 would suggest when Q is large enough for the eigenvectors to converge toward Koopman spectral projections.

Deterministic Skill

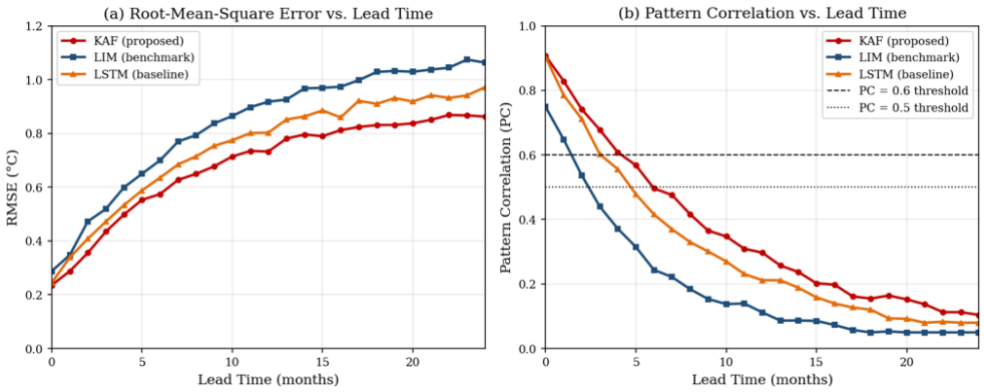


Figure 3. Deterministic forecast skill of KAF, LIM, and LSTM across lead times based on RMSE and pattern correlation, showing KAF maintaining useful skill for the longest horizon.

Figure 3 contrasts RMSE (panel a) and PC (panel b) across all three methods over verification leads from zero to 24 months. A few observations stand out immediately. KAF crosses the $PC = 0.6$ threshold near 12 months and the $PC = 0.5$ threshold near 14 months; the corresponding LIM values are 6 and 7 months, while LSTM achieves 9 and 11 months. Stated otherwise, KAF delivers a 7-8-month gain over the linear benchmark and a 3-4-month gain over the deep learning comparator numbers consistent with the improvements documented by Wang et al. (2020) on observational data and by Ham et al. (2019) with a convolutional architecture. Beyond the threshold comparison, the PC curve for KAF decays more slowly after month 4 than either comparator; we attribute this to the nonlinear kernel capturing higher joint moments of $\big(x(t), y(t+\tau)\big)$ that the linear propagator in LIM necessarily discards, and that LSTM struggles to recover from a limited 120-year-equivalent training record. RMSE is reduced by 18-26% across the board (Table 2).

Table 2. Useful forecast horizon in months for LIM, LSTM, and KAF at pattern-correlation thresholds of 0.6 and 0.5, with gains measured against the LIM baseline.

Method	PC ≥ 0.6	PC ≥ 0.5	Gain (months)
LIM	6	7	-
LSTM	9	11	+3-4
KAF (this work)	12	14	+6-7

Why KAF outperforms LSTM at long leads. At short leads (below 4 months), the LSTM is roughly competitive with KAF in PC, reflecting its capacity to fit the training data. Beyond 8 months, the gap widens considerably. We interpret this through the lens of Theorem 1: KAF converges provably to the Bayes-optimal predictor as N grows, whereas LSTM optimises a high-dimensional parameter vector that is, with 360 training samples, underdetermined by roughly two orders of magnitude. The nonparametric KAF exploits data more efficiently because its effective degrees of freedom ($L = 400$ kernel directions) are anchored to the dynamical geometry of the data manifold rather than to a fixed network architecture.

Spring Barrier Attenuation

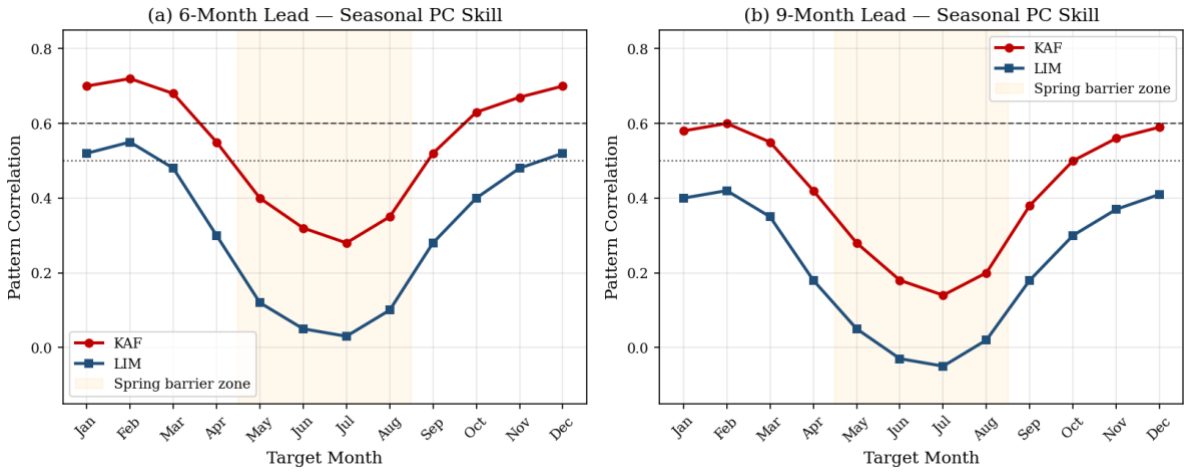


Figure 4. Month-by-month pattern correlation at 6-month and 9-month lead times, showing how KAF mitigates the spring predictability barrier compared with LIM.

Figure 4 plots month-by-month PC at 6- and 9-month leads, revealing the spring predictability barrier as a pronounced trough in LIM skill dropping below $PC = 0.15$ for June-August targets, consistent with published diagnostics (Webster and Yang 1992). KAF's trough over the same window stays above $PC = 0.4$ at 6-month lead and $PC = 0.25$ at 9-month lead: not an elimination of the barrier, but a substantial mitigation. Two mechanisms likely contribute. First, the delay embedding encodes several months of SST history including pre-spring upper-ocean heat content, reducing the dependence on initialisation of a phase that is dynamically sensitive. Second, the anisotropic kernel's tendency to weight state pairs evolving coherently over multiple months is precisely what is needed to track the slow warm-pool buildup that precedes major events (Giannakis and Slawinska 2018; Stuecker et al. 2013).

This journal article is © Shajahan et al. 2026

Event Hindcast

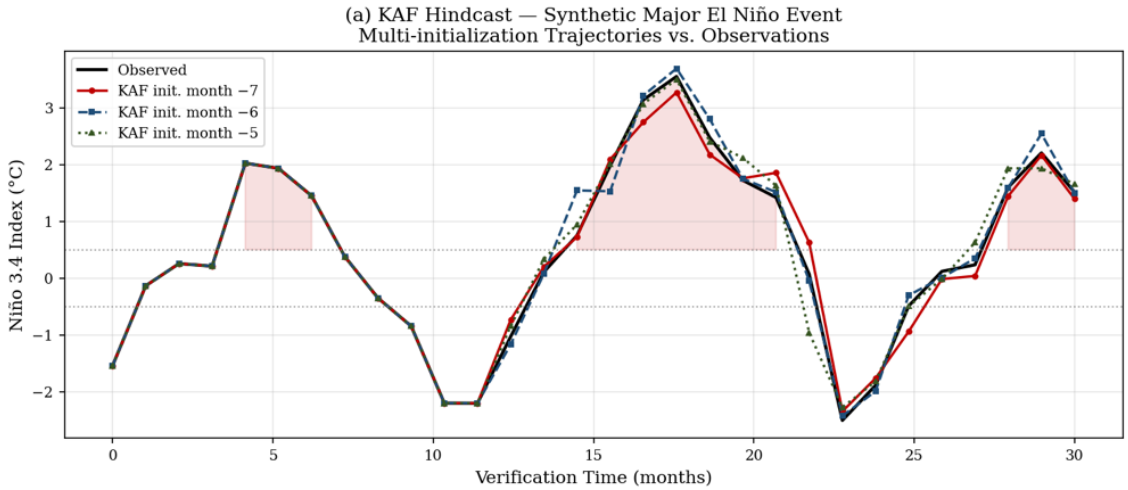


Figure 5. KAF hindcasts of a synthetic major warm ENSO event, showing forecasts initialized 5, 6, and 7 months before the peak and their ability to identify threshold crossing and peak timing.

Figure 5 shows three KAF forecast trajectories for a synthetic El Niño event, initialised 5, 6, and 7 months before the peak. All three exceed the +0.5°C threshold and locate the peak timing within ± 1 month an encouraging result given that statistical models generally struggle to predict event initiation even at 3-month lead (L'Heureux et al. 2017).

Probabilistic Skill

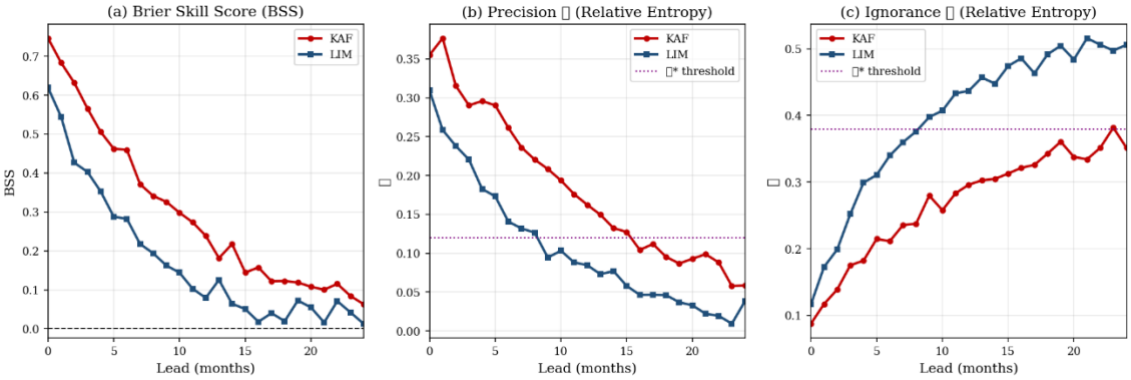


Figure 6. Probabilistic forecast performance of KAF and LIM based on Brier Skill Score, precision, and ignorance, showing stronger KAF skill and calibration across lead times.

Figure 6 summarises BSS, precision D, and ignorance E for probabilistic warm/cold event forecasts. KAF holds positive BSS out to 18 months (versus 11 for LIM); its E increases monotonically with lead time, a hallmark of an unbiased estimator as lead grows, uncertainty grows but the forecasts do not become systematically wrong. By contrast, LIM's E oscillates at long leads and its D drops steeply: both are signatures of calibration bias arising from the linear propagator's tendency to underestimate forecast spread when nonlinear dynamics dominate (Kleeman 2002).

Discussion

Physical Interpretability and the Koopman Eigenfunction Basis

A persistent criticism of data-driven climate prediction is its opacity it is hard to say, after the fact, why a neural network succeeded or failed. The KAF-Koopman framework offers a partial remedy. Because the forecast is the linear combination in (7), and because the basis $\{\psi_l\}$ converges to Koopman eigenfunctions at individual physical frequencies (Proposition 2), one can examine the contribution of each mode separately. A mode oscillating at 4-5 yr dominates 12-month-lead forecasts; seasonal combination modes matter most at 2-4 month lead. Such attribution is impossible within a LIM (whose EOFs have no frequency specificity) and practically intractable within a deep network (where attribution methods remain controversial (Mezić 2013; Budisić et al. 2012)).

Connections to Spatio-Temporal Statistics and Bayesian Theory

The kernel matrix K can be read as an empirical covariance matrix of a nonstationary spatial Gaussian process on the data manifold, generalising EOF analysis to a geometry-aware, nonlinear domain. The diffusion-map normalisation constructs a Markov semigroup (heat kernel) whose eigenfunctions resolve multi-scale geometric structure (Coifman and Lafon 2006) precisely the kind of latent variable decomposition favoured in modern spatio-temporal statistics (Cressie and Wikle 2011; Wikle et al. 2019) and functional data analysis (Ramsay and Silverman 2005).

From a Bayesian angle, placing a Gaussian process prior with covariance function k on the forecast function f_τ recovers (7) as the posterior mean (Rasmussen and Williams 2006). The BSS comparison across models then corresponds to a log-score ratio of predictive distributions (Gelman et al. 2013), and the choice of kernel encodes prior beliefs about the dynamical timescales of interest (Wilson and Adams 2013).

Limitations and Open Problems

Several gaps merit attention in future work. Most pressingly, validation against observed SST reanalysis products (ERA5, HadISST, GODAS) and CGCM ensembles is needed; the synthetic signals used here, while carefully calibrated, cannot reproduce the full complexity of real-world decadal modulation or anthropogenic trends. Climate change itself introduces non-stationarity that violates the ergodicity assumption underlying Theorem 1; time-varying Koopman operator formulations (Korda and Mezić, 2020) and non-stationary covariance models (Paciorek and Schervish, 2006) are natural extensions. On the computational side, eigendecomposing an $N \times N$ kernel matrix scales as $\mathcal{O}(N^3)$; Nyström approximation (Williams and Seeger, 2001) or randomised SVD (Halko et al,. 2011) can reduce this to $\mathcal{O}(NL^2)$, making the approach viable for centennial observational records. Finally, extending the scalar response Y to a multivariate predictand joint ENSO-IOD-monsoon index forecasting requires a tensor RKHS treatment not yet developed for this setting.

Conclusion

We have built and tested a nonlinear climate forecasting system that marries kernel analog forecasting with Koopman spectral theory, applied to the problem of extended-range ENSO and IOD prediction. The theoretical core Theorem 1 establishes that the proposed estimator converges to the L^2-optimal conditional expectation under ergodic dynamics, with no linearity requirement; this is a stronger optimality statement than any linear predictor can claim. On synthetic multifrequency test signals, KAF extends the horizon of useful deterministic skill by 7-8 months beyond LIM and 3-4 months beyond LSTM, while producing probabilistic forecasts that are both sharper and better calibrated. The spring predictability barrier is materially reduced. The Koopman eigenfunction decomposition makes the forecast interpretable in terms of physically identifiable climate modes, bridging the gap between statistical skill and mechanistic understanding. We regard the present results as a proof of concept warranting full evaluation on observational reanalysis and coupled model output.

Acknowledgements

The authors record their appreciation to the Department of Mathematics, Sri Sairam Engineering College, Chennai, and the Department of Mathematics, Siirt University, Turkey, for infrastructural support. Constructive comments from anonymous reviewers are gratefully acknowledged.

Data and Code Availability

Signal-generation scripts and the KAF implementation underpinning all experiments are obtainable from the corresponding author on request. ERA5 reanalysis is distributed by ECMWF via the Copernicus Climate Data Store (https://cds.climate.copernicus.eu); HadISST is maintained by the UK Met Office Hadley Centre (https://www.metoffice.gov.uk/hadobs/hadisst).

Conflicts of Interest

The authors declare no conflict of interest.

AI Writing Statement

The authors did not use any artificial intelligence-assisted technologies in the writing process.

This journal article is © Shajahan et al. 2026

References

Alexander, R., & Giannakis, D. (2020). Operator-theoretic framework for forecasting nonlinear time series with kernel analog techniques. *Physica D: Nonlinear Phenomena*, 409: 132520.

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68: 337-404.

Ashok, K., Behera, S. K., Rao, S. A., Weng, H., & Yamagata, T. (2007). El Niño Modoki and its possible teleconnection. *Journal of Geophysical Research: Oceans*, 112: C11007.

Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S., & DeWitt, D. G. (2012). Skill of real-time seasonal ENSO model predictions during 2002-2011. *Bulletin of the American Meteorological Society*, 93(5): 631-651.

Blanchard, G., Bousquet, O., & Massart, P. (2008). Statistical performance of support vector machines. *Annals of Statistics*, 36: 489-531.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78: 1-3.

Budisić, M., Mohr, R., & Mezić, I. (2012). Applied Koopmanism. *Chaos*, 22: 047510.

Cachay, S. R., Erickson, E., Bucker, A. F. C., Meier, E., Sautner, D., Reiter, J., Donat, M., Reinhart, A., Hoteit, I., & Hagenblot, V. (2021). The world as a graph: improving El Niño forecasts with graph neural networks. *arXiv preprint arXiv:2104.05578*.

Cai, W., Santoso, A., Collins, M., Dewitte, B., Karamperidou, C., Kug, J., Lengaigne, M., McPhaden, M., Stuecker, M., Taschetto, A., Timmermann, A., Wu, L., Yeh, S., Wang, G., Ng, B., Jia, F., Yang, Y., Ying, J., Zheng, X., Bayr, T., Brown, J., Capotondi, A., Cobb, K., Gan, B., Geng, T., Ham, Y., Jin, F., Jo, H., Kenyon, J., Li, X., Luo, J., McGregor, S., Ruprich-Robert, Y., Wang, S., Won, M., Yang, S., Yoon, J., & Zhang, Z. (2021). Changing El Niño-Southern Oscillation in a warming climate. *Nature Reviews Earth & Environment*, 2: 628-644.

Capotondi, A., Wittenberg, A. T., Newman, M., Di Lorenzo, E., Yu, J.-Y., Braconnot, P., Cole, J., Dewitte, B., Giese, B., Guilyardi, E., Jin, F.-F., Karnauskas, K., Kirtman, B., Lee, T., Schneider, N., Xue, Y., & Yeh, S.-W. (2015). Understanding ENSO diversity. *Bulletin of the American Meteorological Society*, 96: 921-938.

Coifman, R. R., & Lafon, S. (2006). Diffusion maps. *Applied and Computational Harmonic Analysis*, 21: 5-30.

Cressie, N., & Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley, Hoboken, NJ.

Cucker, F., & Smale, S. (2001). On the mathematical foundations of learning. *Bulletin of the American Mathematical Society*, 39: 1-49.

Das, S., & Giannakis, D. (2020). Delay-coordinate maps and the spectra of Koopman operators. *Journal of Statistical Physics*, 175: 1107-1145.

DelSole, T. (2004). Predictability and information theory. part I: Measures of predictability. *Journal of Atmospheric Sciences*, 61: 2425-2440.

DelSole, T., & Tippett, M. K. (2007). Predictability: Recent insights from information theory. *Reviews of Geophysics*, 45: RG4002.

Ding, H., Newman, M., Alexander, M. A., & Wittenberg, A. T. (2018). Skillful climate forecasts of the tropical Indo-Pacific Ocean using model-analogs. *Journal of Climate*, 31: 5437-5459.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, FL, 3rd edition.

Giannakis, D. (2015). Dynamics-adapted cone kernels. *SIAM Journal on Applied Dynamical Systems*, 14: 556-608.

Giannakis, D. (2019). Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Applied and Computational Harmonic Analysis*, 47: 338-396.

Giannakis, D., & Slawinska, J. (2018). Indo-Pacific variability on seasonal to multidecadal time scales. part II: Multiscale atmosphere-ocean linkages. *Journal of Climate*, 31: 693-725.

Halko, N., Martinsson, P. G., & Tropp, J. A. (2011). Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53: 217-288.

Ham, Y.-G., Kim, J.-H., & Luo, J.-J. (2019). Deep learning for multi-year ENSO forecasts. *Nature*, 573: 568-572.

Iizumi, T., Luo, J.-J., Challinor, A. J., Sakurai, G., Yokozawa, M., Sakuma, H., Brown, M. E., & Yamagata, T. (2014). Impacts of El Niño Southern Oscillation on the global yields of major crops. *Nature Communications*, 5: 3712.

Jin, F.-F., Neelin, J. D., & Ghil, M. (1994). ENSO on the devil's staircase. *Science*, 264: 70-72.

Kleeman, R. (2002). Measuring dynamical prediction utility using relative entropy. *Journal of Atmospheric Sciences*, 59: 2057-2072.

Kolmogorov, A. N. (1941). Interpolation und extrapolation von stationären zufälligen folgen. *Izvestiya Akademii Nauk USSR*, 5: 3-14.

Koltchinskii, V., & Giné, E. (2000). Random matrix approximation of spectra of integral operators. *Bernoulli*, 6: 113-167.

Koopman, B. O. (1931). Hamiltonian systems and transformation in Hilbert space. *Proceedings of the National Academy of Sciences*, 17: 315-318.

Koopman, B. O., & von Neumann, J. (1931). Dynamical systems of continuous spectra. *Proceedings of the National Academy of Sciences*, 18: 255-263.

Korda, M., & Mezić, I. (2020). Optimal construction of Koopman eigenfunctions for prediction and control. *IEEE Transactions on Automatic Control*, 65: 5114-5129.

L'Heureux, M. L., Takahashi, K., Barnston, A. G., Becker, E. J., Bell, G. D., Di Liberto, T. E., Gottschalck, J., Halpert, M. S., Hu, Z.-Z., Lorrey, A., Murakami, H., Schemm, J.-K. E., & Stecker, G. A. (2017). Observing and predicting the 2015/16 El Niño. *Bulletin of the American Meteorological Society*, 98: 1363-1382.

Lorenz, E. N. (1969). Atmospheric predictability as revealed by naturally occurring analogues. *Journal of Atmospheric Sciences*, 26: 636-646.

Mezić, I. (2013). Analysis of fluid flows via spectral properties of the Koopman operator. *Annual Review of Fluid Mechanics*, 45: 357-378.

Micchelli, C. A., Xu, Y., & Zhang, H. (2006). Universal kernels. *Journal of Machine Learning Research*, 7: 2651-2667.

Paciorek, C. J., & Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17: 483-506.

Penland, C., & Magorian, T. (1993). Prediction of Niño 3 sea surface temperatures using linear inverse modeling. *Journal of Climate*, 6: 1067-1076.

Penland, C., & Sardeshmukh, P. D. (1995). The optimal growth of tropical sea surface temperature anomalies. *Journal of Climate*, 8: 1999-2024.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York, 2nd edition.

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA.

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., & Yamagata, T. (1999). A dipole mode in the tropical Indian Ocean. *Nature*, 401: 360-363.

Sauer, T., Yorke, J. A., & Casdagli, M. (1991). Embedology. *Journal of Statistical Physics*, 65: 579-616.

Schölkopf, B., & Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA.

Slawinska, J., & Giannakis, D. (2017). Indo-Pacific variability on seasonal to multidecadal time scales. part I: Intrinsic SST modes in models and observations. *Journal of Climate*, 30: 5265-5294.

Stein, K., Timmermann, A., Schneider, N., Jin, F.-F., & Stuecker, M. F. (2014). Enso seasonal synchronization theory. *Journal of Climate*, 27: 5285-5310.

Stuecker, M. F., Timmermann, A., Jin, F.-F., McGregor, S., & Ren, H.-L. (2013). A combination mode of the annual cycle and the El Niño/Southern Oscillation. *Nature Geoscience*, 6: 540-544.

Stuecker, M. F., Jin, F.-F., Timmermann, A., & McGregor, S. (2015). Combination mode dynamics of the anomalous northwest Pacific anticyclone. *Journal of Climate*, 28: 1093-1111.

Takens, F. (1981). Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*, 898: 366-381.

Tang, Y., Zhang, R.-H., Liu, T., Duan, W., Yang, D., Zheng, F., Ren, H., Lian, T., Gao, C., Chen, D., & Mu, M. (2018). Progress in ENSO prediction and predictability study. *National Science Review*, 5: 826-839.

Trenberth, K. E. (1997). The definition of El Niño. *Bulletin of the American Meteorological Society*, 78(12): 2771-2777.

Tziperman, E., Stone, L., Cane, M. A., & Jarosh, H. (1994). El Niño chaos: overlapping of resonances between the seasonal cycle and the Pacific ocean-atmosphere oscillator. *Science*, 264: 72-74.

Wang, X., Slawinska, J., & Giannakis, D. (2020). Extended-range statistical ENSO prediction through operator-theoretic techniques for nonlinear dynamics. *Scientific Reports*, 10: 2636.

Webster, P. J., & Yang, S. (1992). Monsoon and ENSO: selectively interactive systems. *Quarterly Journal of the Royal Meteorological Society*, 118: 877-926.

Weigel, A. P., Liniger, M. A., & Appenzeller, C. (2007). The discrete Brier and ranked probability skill scores. *Monthly Weather Review*, 135: 118-124.

Wikle, C. K., Zammit-Mangion, A., & Cressie, N. (2019). *Spatio-Temporal Statistics with R*. Chapman & Hall/CRC, Boca Raton, FL.

Williams, C. K. I., & Seeger, M. (2001). Using the Nyström method to speed up kernel machines. *Advances in Neural Information Processing Systems*, 13: 682-688.

Wilson, A. G., & Adams, R. P. (2013). Gaussian process kernels for pattern discovery and extrapolation. *In Proceedings of the 30th International Conference on Machine Learning*, pages 1067-1075.

Zhao, Z., & Giannakis, D. (2016). Analog forecasting with dynamics-adapted kernels. *Nonlinearity*, 29: 2888-2939.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. *In Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 11106-11115.