

Moment-based approach to cyclostationary nonlinear inverse modeling under persistent noise forcing

Justin Lien ^{*}*Mathematical Institute, Tohoku University, Sendai, Japan*Hiroyasu Ando *Advanced Institute for Materials Research, Tohoku University, Sendai, Japan
and Mathematical Institute, Tohoku University, Sendai, Japan*Yong-Yub Kim *Geophysical Institute, University of Bergen, Bergen, Norway;
Bjerknes Centre for Climate Research, Bergen, Norway;
IBS Center for Climate Physics, Busan, Republic of Korea;
and Pusan National University, Busan, Republic of Korea*Ingo Richter [†]*Application Laboratory, Research Institute for Value-Added-Information Generation,
Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan*

(Received 19 December 2025; accepted 6 April 2026; published 18 May 2026)

This study extends the moment-based nonlinear inverse modeling (nLIM) framework from the stationary to the cyclostationary (CS) regime, introducing CS-Colored-nLIM. The proposed method constructs a periodic, colored-noise-forced, quadratic stochastic system from cyclostationary input data. In this formulation, the active nonlinear components can be empirically determined through a sparse identification scheme specifically developed for colored-noise-forced systems. Application to climate prediction demonstrates that CS-Colored-nLIM achieves predictive performance comparable to modern machine learning and statistical models, while offering greater flexibility with moderate computational demand, making it potentially applicable to a broader range of complex systems beyond climate dynamics.

DOI: [10.1103/j318-6j37](https://doi.org/10.1103/j318-6j37)

I. INTRODUCTION

Stochastic inverse modeling is a powerful tool for understanding and predicting complex systems across the physical, biological, social, and environmental sciences [1–7]. It links empirical measurements with theoretical models by inferring underlying dynamics and stochastic forcing from input data. This data-driven approach is particularly useful when direct measurements of internal processes are challenging or first-principles models are incomplete or computationally demanding [8,9]. Advances in data and computation have further enhanced its capability, making it an increasingly versatile method for modeling and prediction in various scientific fields [10–13].

In climate science, the classical linear inverse model (LIM) is a moment-based empirical method that produces a stationary (ST), additive-white-noise-forced linear stochastic process to describe the evolution of input data [14–16]. We will refer to this type of LIM as ST-White-LIM. Despite its versatility in diverse applications, such as mode identification, El Niño Southern Oscillation (ENSO) prediction, and interbasin causal interactions, its inherent simplicity limits its ability to represent the complex dynamics of the climate system [17–21]. To address these limitations, various extensions of ST-White-LIM have been developed, introducing time dependence (Cyclostationary, CS) and different types of stochastic forcing (additive persistent noise or multiplicative state-dependent noise) [22–26]. Incorporating a cyclostationary framework enables the representation of the seasonally modulated evolution of the climate system, while accounting for nonwhite noise can enhance predictive skill or reproduce non-Gaussian features such as ENSO skewness [26–28]. Nevertheless, linear formulations face challenges in representing climate processes with pronounced nonlinearity, such as ENSO [29,30].

Recently, machine learning (ML) has gained increasing attention for studying complex dynamical systems [31–34]. In climate science, ML techniques such as artificial neural

^{*}Contact author: lien.justin.t8@dc.tohoku.ac.jp[†]Present address: Research Institute for Global Change, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

networks or Bayesian methods have demonstrated impressive skill for predicting tropical climate variability, in some cases outperforming advanced climate forecast systems [35–42]. However, these approaches typically require extensive parameter tuning, high computational cost, and large training datasets, conditions that may be difficult to meet in climate studies, where high-quality global observations are available only for the past few decades. Therefore, they are sometimes integrated with dynamical modeling, for instance, by training the ML model on outputs from advanced complex climate model simulations [43,44]. Nevertheless, the limited interpretability of ML models and their susceptibility to overfitting make it challenging to understand the underlying mechanisms of climate variability based on the ML results [45–47].

Nonlinear inverse modeling, on the other hand, balances interpretability and model complexity. These models are often capable of reproducing essential statistical properties of complex systems and have been successfully applied across a range of disciplines [48–59]. However, many formulations rely on the white-noise assumption, under which parameter estimation is more accurate; when this assumption is violated, the presence of nonwhite noise can cause biased parameter estimates [60–63]. An alternative approach is the eXtended nonlinear Recharge Oscillator model (XRO), a lagged-correlation-based nonlinear framework that integrates cyclostationarity, additive-colored-noise forcing, and quadratic nonlinearity [64–67]. In its pioneering application, XRO used ten key climate indices from multiple ocean basins and incorporated nonlinear interactions over the Niño 3.4 region and the Indian Ocean Dipole (IOD). This model successfully characterized essential ENSO mechanisms, such as ENSO asymmetry and differences in the duration and frequency of El Niño and La Niña events, and achieved ENSO prediction skill comparable to leading machine-learning models while requiring substantially lower computational cost [67]. However, the active nonlinear terms in XRO are prescribed based on prior physical knowledge, which limits its general applicability, particularly beyond climate-related systems.

The assignment of active nonlinear terms can be formulated as a sparse identification problem. The Sparse Identification of Nonlinear Dynamics (SINDy) framework is a widely used approach for discovering governing equations directly from data by selecting a minimal set of active terms from a library of candidate functions [68,69]. Extensions of SINDy have been proposed to handle various challenges, such as data scarcity and noisy measurement, and integration with neural networks or partial differential equations, with applications in epidemiological and biological modeling [70–74]. Nevertheless, the most popular algorithms for sparse regression have only been tested on Gaussian white noise, limiting their applicability to real-world processes driven by temporally correlated noise [75–77]. This limitation suggests the need for a sparse identification scheme that can robustly handle colored noise.

Motivated by these developments, we previously extended the moment-based inverse modeling framework to incorporate nonlinear dynamics in a stationary, additive-colored-noise-forced formulation, referred to in this study as ST-Colored-nLIM, and demonstrated its applicability using a simple two-dimensional ENSO recharge oscillator model

[60]. Building on this foundation, the present study advances the inverse modeling framework to a cyclostationary formulation, termed CS-Colored-nLIM. In the parameter estimation algorithm, we develop a simple sparse identification scheme for a colored-noise-forced system, enabling the active nonlinear terms to be determined fully empirically. Consequently, CS-Colored-nLIM shares the same theoretical foundation as XRO—accounting for cyclostationarity, colored-noise forcing, and quadratic nonlinearity—while differing primarily in the specification of active nonlinear terms [67]. Furthermore, previously developed moment-based inverse models with additive noise forcing can be regarded as reduced forms of the proposed CS-Colored-nLIM [14,23,24,26,60]. Hence, CS-Colored-nLIM inherits the fundamental properties and characteristics of its predecessors while offering greater generality and flexibility.

This study first presents in Sec. II the key mathematical formulation for parameter estimation in CS-Colored-nLIM and discusses its methodological similarities and differences with SINDy, XRO, and its reduced forms. An illustrative numerical example for the proposed method is presented in Sec. III. In Sec. IV, the applicability of the proposed model is then demonstrated through climate prediction experiments and compared with the reduced variants and XRO. The derivation and proof of the main formula are provided in Appendix A, while Appendix B discusses a reduced variant called CS-White-nLIM, corresponding to the white-noise-forced case. Appendixes C and D provide additional technical details on the CS-Colored-nLIM and XRO methodologies used in the prediction experiments. In addition, the Supplemental Material [78] includes (1) a brief demonstration illustrating the bias introduced by conventional white-noise-based parameter estimation techniques when applied to colored-noise-forced systems, (2) demonstration code for the ideal example described in Sec. III, and (3) additional numerical details and prediction results presented in Sec. IV [78].

To facilitate a smoother presentation of the methodology, we first clarify the notation and assumptions used throughout this study. Boldface denotes mathematical objects in the forward formulation, while regular font represents estimated quantities from the inverse method. Vectors are assumed to be column vectors. The normalized Gaussian white noise ξ has zero mean and unit variance (i.e., $\langle \xi(\cdot + \tau)\xi(\cdot)^T \rangle = \delta(\tau)\mathbf{I}$), where $\langle \cdot \rangle$, δ , and τ denote the expectation, Dirac delta function, and the lag (or lead) variable, respectively. The normalized Ornstein-Uhlenbeck colored noise η with the noise correlation time $\gamma > 0$ is realized by the steady state of $\frac{d}{dt}\eta = -\frac{1}{\gamma}\eta + \frac{1}{\gamma}\xi$ [79]. As its temporal structure satisfies $\langle \eta(\cdot + \tau)\eta(\cdot)^T \rangle = \frac{1}{2\gamma}e^{\frac{-|\tau|}{\gamma}}\mathbf{I}$, γ can be physically interpreted as the characteristic time for the past dependence of noise. Stationarity or cyclostationarity is assumed implicitly. The stochastic differential equation (SDE) can be interpreted in either the Ito or Stratonovich sense, as the stochastic forcing is modeled by additive noise.

II. MATHEMATICAL FOUNDATIONS OF NONLINEAR INVERSE MODELS

The cyclostationary nonlinear inverse models describe the n -dimensional input dataset $\{x(t)\} \subset \mathbb{R}^n$ that may or may not

have a zero mean with the approximate SDE

$$\frac{dx_i}{dt} = \underbrace{\sum_{j \leq k} B_{ijk} x_j x_k + \sum_j A_{ij} x_j + C_i}_{\text{Dynamics: } F_i(x, t)} + \underbrace{\sum_j (\sqrt{2Q})_{ij} \text{noise}_j}_{\text{Stochastic forcing}}, \quad (1)$$

for $i = 1, \dots, n$. The dynamics $F = [F_1; \dots; F_n]$ are parametrized by (B, A, C) , collectively referred to as the dynamical parameters, represented by periodic 3-, 2-, and 1-tensors corresponding to the quadratic, linear, and constant terms, respectively, with the same periodicity. Without loss of generality, the fundamental period may be normalized to unity. With an abuse of notation, the time variable t and phase variable θ are identified via the canonical projection ω from the real line to the circle with unit perimeter (i.e., $\theta = \omega(t)$) [80]. The tensor B_{ijk} is constrained such that $B_{ijk} = 0$ if $j > k$, and Q is the positive definite stochastic matrix that describes the amplitude and cross-variable structure of the noise.

The stochastic forcing is modeled by either white or colored noise. This section focuses on the colored-noise case, while the white-noise formulation is discussed in Appendix B, as it can be regarded as a limiting case of the CS-Colored-nLIM. We start with the forward formulation before addressing the inverse methods.

A. Forward formulation

The governing equation of a state vector $\mathbf{x}$ in the CS-Colored-nLIM framework reads

$$\frac{d\mathbf{x}_i}{dt} = \underbrace{\sum_{j \leq k} \mathbf{B}_{ijk}(t) \mathbf{x}_j \mathbf{x}_k + \sum_j \mathbf{A}_{ij}(t) \mathbf{x}_j + \mathbf{C}_i(t)}_{\mathbf{F}_i(\mathbf{x}, t)} + \sum_j (\sqrt{2\mathbf{Q}})_{ij} \eta_j, \quad (2)$$

where $\mathbf{Q}$ is time independent. Following Ref. [60], the uncentered correlation functions (also referred to as lag- τ moments) of $\mathbf{x}$ are defined from 1-order to 4-order as $\mathbf{E} = \langle \mathbf{x}(t) \rangle$, $\mathbf{K}(\tau, \theta) = \langle \mathbf{x}(t + \tau) \mathbf{x}(t)^T \rangle$ of $\mathbf{x}$, $\mathbf{M}_{ijk}(\tau, \theta) = \langle \mathbf{x}_i(t + \tau) \mathbf{x}_j(t) \mathbf{x}_k(t) \rangle$, and $\mathbf{S}_{ijkw}(\tau, \theta) = \langle \mathbf{x}_i(t + \tau) \mathbf{x}_j(t) \mathbf{x}_k(t) \mathbf{x}_w(\cdot) \rangle$. These correlation functions ($\mathbf{E}, \mathbf{K}, \mathbf{M}, \mathbf{S}$) of $\mathbf{x}$ may not admit explicit forms, but their derivatives encode the information of the system parameters.

For notational simplicity, we drop the explicit dependence on τ and θ , as all variables are evaluated at $\tau = 0$ for arbitrary θ , for the derivative relations. That is, we write $\mathbf{T}$ for the tensor $\mathbf{T}(\theta)$ or $\mathbf{T}(0, \theta)$. Then, the derivative relations are

$$\vec{\mathbf{E}}_\theta = \mathbf{L}_\mathbf{K}^2 \vec{\mathbf{B}} + \mathbf{L}_\mathbf{E}^1 \vec{\mathbf{A}} + \vec{\mathbf{C}}, \quad (3)$$

$$\begin{aligned} \gamma \vec{\mathbf{M}}_{\tau\tau} + \vec{\mathbf{M}}_\tau &= \gamma (\mathbf{L}_\mathbf{M}^2 \vec{\mathbf{B}}' + \mathbf{L}_\mathbf{K}^1 \vec{\mathbf{A}}' + \mathbf{L}_\mathbf{E}^0 \vec{\mathbf{C}}') \\ &+ \gamma (\mathbf{L}_{\text{Sym}(\mathbf{M}_\tau)}^2 \vec{\mathbf{B}} + \mathbf{L}_\mathbf{K}^1 \vec{\mathbf{A}}) \\ &+ \mathbf{L}_\mathbf{M}^2 \vec{\mathbf{B}} + \mathbf{L}_\mathbf{K}^1 \vec{\mathbf{A}} + \mathbf{L}_\mathbf{E}^0 \vec{\mathbf{C}}, \end{aligned} \quad (4)$$

$$\begin{aligned} \gamma \vec{\mathbf{M}}_{\tau\tau} + \vec{\mathbf{M}}_\tau &= \gamma (\mathbf{L}_\mathbf{S}^2 \vec{\mathbf{B}}' + \mathbf{L}_\mathbf{M}^1 \vec{\mathbf{A}}' + \mathbf{L}_\mathbf{K}^0 \vec{\mathbf{C}}') \\ &+ \gamma (\mathbf{L}_{\text{Sym}(\mathbf{S}_\tau)}^2 \vec{\mathbf{B}} + \mathbf{L}_\mathbf{M}^1 \vec{\mathbf{A}}) \\ &+ \mathbf{L}_\mathbf{S}^2 \vec{\mathbf{B}} + \mathbf{L}_\mathbf{M}^1 \vec{\mathbf{A}} + \mathbf{L}_\mathbf{K}^0 \vec{\mathbf{C}}, \end{aligned} \quad (5)$$

where $\mathbf{L}_\mathbf{T}^m$ is a matrix associated with the tensor $\mathbf{T}$, $\vec{\mathbf{T}}$ [or $\text{vec}(\mathbf{T})$ and $\text{vec} \mathbf{T}$] is the vectorization of $\mathbf{T}$, and $\text{Sym}(\cdot)$ is the symmetrization operator. It is worth noting that Eqs. (3)–(5) do not involve the stochastic matrix $\mathbf{Q}$ and, at most, provide $n_C = n$, $n_A = n^2$, and $n_B = n^2(n+1)/2$ linearly independent scalar constraints for $(\mathbf{B}, \mathbf{A}, \mathbf{C})$, respectively. Moreover, $\mathbf{Q}$ can be explicitly written as a function of dynamical parameters and the correlation functions [see Eq. (A8)]. Consequently, the system parameters $(\mathbf{B}, \mathbf{A}, \mathbf{C}, \mathbf{Q})$ are fully determined by the statistics $(\mathbf{E}, \mathbf{K}, \mathbf{M}, \mathbf{S})$ and γ . Details of these matrices and operators are provided in Appendix A.

B. Inverse method

Before detailing the inverse method, we briefly outline the idea. The objective is to estimate the system parameters (γ, B, A, C, Q) from the correlation functions (E, K, M, S) derived from the input data $\{x(t)\}$. In the ideal case where $\{x(t)\}$ is sufficiently long and perfectly sampled from Eq. (2), solving linear systems determined by Eqs. (3)–(5) yields the true parameters. In practice, however, limited sampling and the subdivision of data used to compute phase-dependent statistics can cause model instability [81]. To mitigate this, a low-pass filter may be applied to the input correlation functions in θ when necessary, a procedure referred to as phase averaging [26,80]. Furthermore, as the number of nonlinear terms increases cubically with system dimension, overfitting can become a significant concern. To address this issue, sparsity is imposed by allowing only a small subset of statistically significant and dynamically relevant B_{ijk} terms to be nonzero. In addition, when no nonlinear terms are active, the inverse method is designed such that the estimated parameters reduce to those determined by the linear model, ensuring a *continuous* transition from linear to nonlinear dynamics as more active components are introduced.

1. Algorithm

Step 1: Compute statistics. Given an n -dimensional input dataset $\{x(t)\}$ sampled at frequency N (i.e., N samples per cycle), the correlation functions are computed. The numerical derivatives with respect to τ and θ are computed by the forward difference method and the center difference scheme, respectively.

Step 2: Sparse identification. For each fixed γ , the matrix representation of Eqs. (3)–(5) can be written as follows after rearrangement:

$$\begin{bmatrix} X_{MB} & X_{MA} & X_{MC} \\ X_{KB} & X_{KA} & X_{KC} \\ X_{EB} & X_{EA} & X_{EC} \end{bmatrix} \begin{bmatrix} \vec{\mathbf{B}} \\ \vec{\mathbf{A}} \\ \vec{\mathbf{C}} \end{bmatrix} = \begin{bmatrix} \vec{\mathbf{Y}}_M \\ \vec{\mathbf{Y}}_K \\ \vec{\mathbf{Y}}_E \end{bmatrix}, \quad (6)$$

where $\vec{B} \in \mathbb{R}^{N \cdot n_B}$ is the phase-wise vertical concatenation of the estimated parameter $\vec{B}(\theta) \in \mathbb{R}^{n_B}$ at the phase θ , $X_{MB} \in \mathbb{R}^{(N \cdot n_B) \times (N \cdot n_B)}$ is a matrix consisting of the matrices acting on $\vec{B}$ in Eq. (5), and $\vec{Y}_M \in \mathbb{R}^{N \cdot n_B}$ is the vector on the left-hand side of Eq. (5). The notation used for X_{MB} and $\vec{Y}_M$ is analogously extended to other blocks and vectors.

Since sparsity identification is only applied to B , and the approximate system should remain consistent with the linear models, Eq. (6) is reduced using the Schur complement. To identify the active terms within B , the sparse identification problem is formulated using the standardized group Lasso:

$$\min_{\vec{B}} \left(\|\widetilde{X}_{MB} \vec{B} - \widetilde{Y}_M\|_2^2 + \lambda \sum_{i,j \leq k} \|\vec{B}_{ijk}\|_2 \right), \quad (7)$$

where $\vec{B}_{ijk} \in \mathbb{R}^N$ denotes the vector containing $B_{ijk}(\theta)$ for all phase samples θ . The standardized quantities are defined as $\widetilde{X}_{MB} = \text{standard}(X_{MB} - X_{12}X_{22}^{-1}X_{21})$ and $\widetilde{Y}_M = \text{standard}(Y_2 - X_{12}X_{22}^{-1}Y_2)$, where X_{ij} and Y_i denote the colored blocks marked in Eq. (6) and $\text{standard}(\cdot)$ denotes columnwise standardization [82–84]. The Lasso parameter λ determines the sparsity of B . The optimization problem is efficiently solved using the fast iterative shrinkage-thresholding algorithm (FISTA), and indices corresponding to zero-valued entries are denoted by idx_0 [85,86]. More details are provided in Sec. S.I of the Supplemental Material [78].

Step 3: Solve dynamical parameters. With idx_0 , the corresponding columns are removed from Eq. (6). To ensure consistency with the linear models, the corresponding rows from the equations are also eliminated; that is,

$$\underbrace{\begin{bmatrix} X_{MB} & X_{MA} & X_{MC} \\ X_{KB} & X_{KA} & X_{KC} \\ X_{EB} & X_{EA} & X_{EC} \end{bmatrix}}_X \underbrace{\begin{bmatrix} \vec{B} \\ \vec{A} \\ \vec{C} \end{bmatrix}}_V = \underbrace{\begin{bmatrix} \vec{Y}_M \\ \vec{Y}_K \\ \vec{Y}_E \end{bmatrix}}_Y. \quad (8)$$

The gray-shaded regions in Eq. (8) indicate the columns and rows associated with idx_0 , and X , V , and Y are the resulting matrices and vectors after this reduction. The dynamical parameters (B , A , C) that are not constrained to zero are obtained by solving $XV = Y$.

Step 4: Solve Q . The stochastic matrix Q can be estimated using several approaches. One common method is to approximate Q from the covariance of the residual term $dx/dt - F(x, t)$ [53,67]. Alternatively, it can be derived from analytical expressions such as Eq. (A8). More directly, it can be determined by minimizing a loss function that accounts for the discrepancy between the input moments and the corresponding model moments evaluated at $\tau = 0$ [60].

Step 5: Solve γ . The noise correlation time γ is determined by minimizing a suitable loss function. One of the simplest loss functions introduced in Ref. [23] is adopted here. For each value of γ , Steps 2–4 are repeated to compute the

second-order correlation function K^γ of the model, estimated from a sufficiently long realization. The optimal γ is obtained by minimizing the discrepancy between the input and model correlation functions:

$$\gamma = \arg \min_{\gamma} \underbrace{\mathcal{E}(K^\gamma, K)}_{\text{Loss function}}, \quad (9)$$

where $\mathcal{E}$ denotes the relative error given by $\mathcal{E}(a, b) = \|a - b\|/\|b\|$.

2. Additional constraints

In practice, a limited sample size may require reducing the number of parameters to be estimated, in order to mitigate issues such as overfitting or unstable estimates of correlation functions and their derivatives. A practical approach is to represent the periodic variations of the dynamical parameters as linear combinations of a few dominant Fourier modes (the phase-average approach), rather than allowing them to be fully determined by the data (the fixed-phase approach) [26]. This alternative reformulates Eqs. (6)–(8) as least-square problems, thereby improving numerical stability and preventing overfitting. The detailed implementation is provided in Appendix C.

In addition, input data may be subject to constraints, which can be incorporated into the algorithm to ensure that the reproduced statistical quantities remain consistent with the input data. For example, when the input data have zero mean (i.e., $E = 0$), this imposes constraints $\mathbf{L}_K^2 \vec{B} + \vec{C} = 0$ and $\mathbf{L}_{K_\theta}^2 \vec{B} + \mathbf{L}_K^2 \vec{B}' + \vec{C}' = 0$ for all θ . Following the same procedure of Steps 2 and 3, C can be recovered by $\mathbf{L}_K^2 \vec{B} + \vec{C} = 0$ for all θ after B and A are estimated. When the model covariance closely matches the input covariance, the reproduced mean E^γ remains near zero, provided that A is full ranked.

Moreover, if the active nonlinear components are prescribed from prior knowledge, Step 2 can be skipped. Likewise, if γ is predetermined, then Step 5 can be omitted.

3. Model properties, limiting cases, and comparisons

The equations used in Step 2 [i.e., Eqs. (6) and (7)] are closely related to the SINDy framework. In the cyclostationary setting, SINDy can be formulated as solving the sparse regression problem from the residual-based perspective

$$\arg \min_{(B,A,C)} \left(\sum_{\theta} \left\| \underbrace{\dot{X}(\theta) - [X^2(\theta) \ X(\theta) \ I]}_{\text{Residual}} \begin{bmatrix} \vec{B}(\theta) \\ \vec{A}(\theta) \\ \vec{C}(\theta) \end{bmatrix} \right\|_2^2 + \lambda \sum_{i,j \leq k} \|\vec{B}_{ijk}\|_2 \right), \quad (10)$$

where $\dot{X}(\theta)$, $X^2(\theta)$, and $X(\theta)$ are matrices whose columns correspond to dx/dt , the quadratic, and linear components of x at a fixed θ , respectively. The matrix I contains entries of either 0 or 1 to account for constant terms.

Conceptually, this formulation is similar to solving the sparse identification problem from the moment-based perspective under the white-noise assumption; that is, the second block row of Eq. (6)—the block row corresponding to the subscript K —is exclusively used to reformulate the minimization problem as Eq. (7) with $\gamma = 0$. Therefore, the additional terms involving γ in Eq. (7) can be interpreted as correction terms that account for the persistence of colored noise, and the first and third block rows provide supplementary information to determine the system parameter in the ideal case. It should be noted, however, that these correction terms depend on the assumed functional form of the dynamics F .

Moreover, in the CS-Colored-nLIM framework, reducing Eq. (6) to Eq. (7) not only preserves consistency with the linear models but also decreases the system size in the FISTA implementation, thereby improving computational efficiency in Step 2. In addition, the minimizer obtained in Eq. (7) is not directly regarded as the estimate of B . When B is constrained to be highly sparse—with only a small subset of active components—or when the system dimension is moderate or large (e.g., $n \geq 4$), the solution tends to exhibit bias and coefficients of negligible magnitude. Consequently, the nonlinear components in such cases are unlikely to play a significant and meaningful role in the resulting approximate system and instead act primarily as numerical noise [87]. For this reason, a two-step procedure is employed: nonlinear coefficients are first identified in Step 2, after which the equation is solved in Step 3 using only the selected variables [88].

As mentioned in the introduction, CS-Colored-nLIM shares a broadly similar governing equation with XRO, but differs in three key aspects beyond the specification of active nonlinear terms [67]. First, CS-Colored-nLIM explicitly incorporates a constant term in the deterministic dynamics. This ensures that the reproduced statistics are consistent with data constraints when applicable, such as the zero-mean condition, and eliminates the need for postprocessing adjustment, which is sometimes required in XRO implementations.

Second, the parameter estimation procedures differ. CS-Colored-nLIM uses the derivatives of the correlation functions at $\tau = 0$, employing analytically derived moment equations to account for the noise memory via the γ -dependent terms. This formulation corrects for the instantaneous correlation between the colored noise and the state variables, thereby avoiding the biased estimates that typically arise when methods designed for white-noise forcing are applied to colored-noise-forced systems (see Sec. S.II of the Supplemental Material) [60,78]. In contrast, XRO is mathematically equivalent to evaluating the derivative of the second-order correlation function at a sufficiently large lag. At such lags, the colored noise is effectively decorrelated from the state vector, yielding robust parameter estimates.

Finally, the two frameworks differ in how noise memory is parametrized. CS-Colored-nLIM assumes a common γ for all state variables. As discussed in Ref. [24], although Eqs. (3)–(5) hold whenever $\mathbf{Q}\mathbf{\Gamma} = \mathbf{\Gamma}\mathbf{Q}$ (with $\mathbf{\Gamma}$ denoting the diagonal matrix of variable-specific noise correlation times), allowing variable-dependent noise memory would substantially enlarge the search space in Eq. (9) due to the γ -dependent terms, making the method numerically infeasible. On the other hand, XRO, in principle, accommodates

variable-dependent noise correlation times more easily: when $\tau \rightarrow \infty$, all colored-noise components—regardless of their individual memories—decorrelate from the state vector. However, in practice, accurate estimation of derivatives at large lags requires sufficiently long and high-quality datasets so that the large-lag correlation faithfully represents the system without contamination from other factors.

CS-Colored-nLIM can be viewed as an extension of previously developed moment-based inverse modeling [15,23,24,26,60]. When the input data are assumed to be stationary (i.e., when no periodicity is imposed on the system parameters), Eqs. (3)–(5) simplify such that all partial derivatives with respect to θ vanish, and the standardized group Lasso in Eq. (7) reduces to the regular standardized Lasso. Moreover, as discussed in Ref. [60], in the white-noise limit of the inverse method, the estimated parameters converge to those obtained in the white-noise case. In particular, setting $\gamma = 0$ recovers the white-noise-forced approximate system. Furthermore, by design, when no active nonlinear components are present (i.e., $B = 0$), the higher-order correlation functions are excluded from the parameter estimation procedure, and the approximate system becomes identical to that determined by the conventional LIM, which yields a Gaussian distribution.

Consequently, depending on whether the approximate system is stationary or cyclostationary, driven by white or colored noise, and linear or nonlinear, a total of eight model variants can be derived. These variants are named based on the assumptions. For example, ST-White-LIM produces a stationary, white-noise-forced, linear system. As the most general formulation, CS-Colored-nLIM retains both the strengths and limitations of its reduced variants. Its principal advantages include conceptual simplicity, interpretability, and the ability to provide unbiased parameter estimates under colored-noise forcing. In addition, as a moment-based approach, the computational cost is largely independent of the sample size, aside from the marginal increase associated with computing the correlation functions. This property highlights its efficiency when dealing with a long input time series.

However, several challenges remain. The primary computational bottleneck in estimating the dynamical parameters lies in achieving the desired level of sparsity. Although FISTA is generally efficient, the initial stage of model implementation often requires multiple runs to determine a suitable Lasso parameter (see Sec. S.III.A of the Supplemental Material for a runtime estimate in a practical setting) [78]. This initial cost can be reduced by parallel computation, and once a suitable parameter range is identified, subsequent computations become more efficient. Moreover, the model performance depends on the quality of input data and the numerical accuracy of derivatives of the correlation function. Solving the inverse equations, such as Eq. (8) may encounter numerical instability due to feature collinearity, and unconstrained periodic dynamics can cause vulnerability under data sparsity.

III. AN IDEAL NUMERICAL EXAMPLE

We present an instructive synthetic example to clarify the algorithmic workflow. Consider a two-dimensional instance

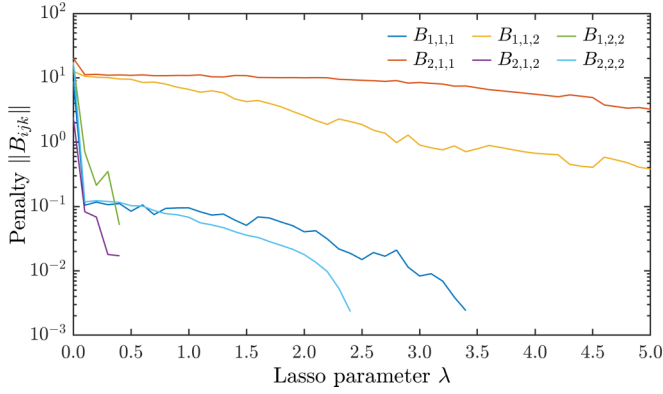


FIG. 1. Vector norms of the group Lasso solutions B_{ijk} of Eq. (7) for the ideal numerical example (11) as a function of the Lasso parameter λ . The y axis is shown on a logarithmic scale. Vanishing B_{ijk} terms are not displayed.

of Eq. (2), given by

$$\begin{aligned} \mathbf{B}_{112}(t) &= -1 - 0.2 \cos(2\pi t), \\ \mathbf{B}_{211}(t) &= -\mathbf{B}_{112}(t), \\ \mathbf{B}_{ijk}(t) &= 0, \text{ for all other } (i, j, k), \\ \mathbf{A}(t) &= (1 + 0.3 \sin(2\pi t)) \begin{bmatrix} -1 & 2 \\ -1 & -2 \end{bmatrix}, \\ \mathbf{C}(t) &= \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}, \quad \text{and } \mathbf{Q} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}. \end{aligned} \quad (11)$$

The colored noise η is characterized by the noise correlation time $\gamma = 0.5$. Equation (11) is numerically integrated using the Euler scheme with an integration time step of $dt = 0.001$. The resulting trajectory is then sampled at intervals of $\Delta t = 0.01$ (i.e., $N = 100$) for a total duration of 10 000 to generate $\{x(t)\}$. The objective is to reconstruct the system parameters $(\gamma, \mathbf{B}, \mathbf{A}, \mathbf{C}, \mathbf{Q})$ from the synthetic data $\{x(t)\}$. We begin with the fixed-phase approach.

With the statistical quantities computed as described in Step 1, for a fixed noise correlation time γ , we first identify the active nonlinear components using Eq. (7). Figure 1 shows an example of the group Lasso solution paths for $\gamma = 0.5$ (i.e., the vector norm of $\|\text{vec}(B_{ijk})\|$ as a function of the Lasso parameter λ). As λ increases, only the true active components $B_{1,1,2}$ and $B_{2,1,1}$ remain nonzero, thereby isolating the correct nonlinear terms.

With the identified nonlinear components for this γ , the estimated dynamical parameters (B, A, C) are obtained by solving Eq. (8), and the stochastic matrix Q is obtained from Eq. (A8). Next, the second-order model correlation function K^γ is estimated from a numerical realization of a duration of 10 000. Repeating this procedure over a range of γ gives the optimal $\gamma = 0.53$. The corresponding loss curve is shown in Fig. 2.

The final parameter estimates are taken to be those associated with the optimal γ . A direct comparison with the true system parameters shows that the relative errors in B , A , C , and Q are 3.4%, 3.2%, 8.4%, and 6.9%, respectively.

The code used to generate this example is provided in the Supplemental Material, together with the implementation of

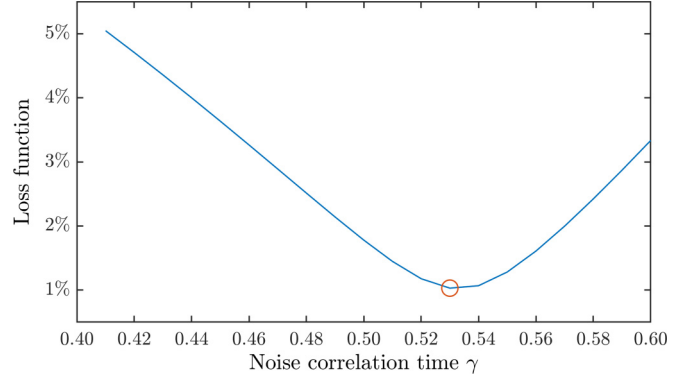


FIG. 2. The loss function in Eq. (9) for the ideal numerical example (11). The marker indicates the optimal γ .

the phase-averaged approach [78]. See also Fig. S.1 of the Supplemental Material for the corresponding runtime for the FISTA solver.

IV. APPLICATION TO CLIMATE PREDICTION

We conduct seasonal climate prediction experiments with CS-Colored-nLIM, using both observational and simulation datasets. Observational records provide a real-world benchmark but are relatively short for training and evaluating statistical models of interannual variability. This limitation is mitigated by long simulation datasets, which span several centuries and are produced by comprehensive dynamical global climate models (GCMs) governed by the laws of fluid dynamics and parametrizations of unresolved physical processes. Despite some systematic errors, many of these models produce fairly realistic representations of observed climate variability, making their output suitable as a surrogate of observational data.

The predictive performance of CS-Colored-nLIM is evaluated relative to seven reduced variants and XRO [67]. For convenience, the approximate system is referred to by the same name as the algorithm used to construct it whenever no ambiguity arises. The experimental setup for the moment-based models is described below, while the implementation details specific to XRO are provided in Appendix D.

A. Data preparation

All datasets consist of ten climate indices with monthly resolution, listed and symbolized in Table I. The observational dataset, obtained from Ref. [67], spans the period from 1979 to 2022, corresponding to the satellite observation era. The simulation datasets are the GCM outputs covering 1850–2024, including the historical period (1850–2014) and future projections from 2015 onward.

These indices represent spatially averaged geophysical variables, primarily sea surface temperature anomalies (SSTAs), that characterize major modes of oceanic variability. Specifically, the ENSO, D20, NPM, and SPMM indices describe variability in the Pacific Ocean; the IOB, IOD, and SIOD indices capture variability in the Indian Ocean; and the TNA, ATL3, and SASD indices represent variability in the

TABLE I. The ten climate indices used in this study. The column “Id.” denotes the index identifier, corresponding to the component of the state vector in the model implementation. The column “Field” indicates the spatially averaged geophysical field associated with each index. For more details, see Table 1 of the Supplemental Material in Ref. [67].

Id.	Acronym	Full name	Field
1	ENSO	El Niño Southern Oscillation	SSTA
2	D20	–	Thermocline depth anomlies
3	NPM	North Pacific Meridional Mode	SSTA
4	SPMM	South Pacific Meridional Mode	SSTA
5	IOB	Indian Ocean Basin mode	SSTA
6	IOD	Indian Ocean Dipole mode	SSTA gradient
7	SIOD	Southern Indian Ocean Dipole mode	SSTA gradient
8	TNA	Tropical North Atlantic variability	SSTA
9	ATL3	Atlantic Niño	SSTA
10	SASD	South Atlantic Subtropical Dipole	SSTA gradient

Atlantic Ocean. The variables τ and θ are regarded as lead and calendar months, respectively.

The data preprocessing for all eight moment-based models follows the same procedure as described in Ref. [24]. For the observation dataset, the monthly climatological mean over the whole period (1979–2022) and the linear trend are removed, followed by the application of a 3-month running average and normalization. For the simulation datasets, the monthly climatological mean over 1980–2010 and the quadratic trend are removed, after which the same 3-month running average and normalization procedures are applied. The preprocessed datasets serve as the training and testing data for the numerical implementation. For cyclostationary models, the phase averaging is applied when computing the correlation functions.

B. Model configuration

The model configurations for CS-Colored-nLIM and its reduced variants are summarized in Table II. For nonlinear models, since the input data consist of anomalies, a zero-mean constraint is imposed, as mentioned in Sec. II B 2. The stochastic matrices are further assumed to be diagonal and are estimated by a loss function based on the discrepancies between the input and model moments (up to third order) evaluated at $\tau = 0$. To mitigate model instability and overfitting arising from the limited sample size, the dynamical parameters are expressed as a Fourier series up to the second order (i.e., phase averaged) for the cyclostationary variants,

consistent with the default XRO configuration, as the climate system typically exhibits dominant annual and seasonal cycles [67]. The number of active nonlinear terms, k , is constrained to between three and eight by tuning the Lasso parameter or prescribed to match the configuration of XRO (hereafter referred to as $k = \text{XRO}$), where in the present notation, $B_{1,1,1}$, $B_{1,1,2}$, and $B_{6,6,6}$ are active. These terms represent the nonlinear processes related to ENSO, D20, and IOD, reflecting prior physical knowledge of interbasin coupling [67].

The noise correlation time γ for the colored-noise models is not estimated empirically but rather prescribed over 0.03–0.08 years. This choice follows previous studies, which independently reported submonthly noise-memory scales [24,60,67]. Prescribing γ in this range enables a fair assessment of the model sensitivity to parameter selection.

C. Prediction setup

Given an initial condition x_t at time t and a lead month τ , the forecasted vector $\hat{x}_{t+\tau}$ is obtained from the linear evolution operator for the linear models, which admits an analytic expression [20,24,26]. In contrast, for nonlinear models, the evolution operator is neither linear nor available in closed form. Therefore, Eq. (1) is numerically integrated using the Euler method with an integration time step $dt = 1/12$ months to generate a 1000-member ensemble. Then, the forecasted state $\hat{x}_{t+\tau}$ is estimated as the ensemble mean. The ensemble

TABLE II. Summary of the eight model variants, indicating periodicity (ST/CS), noise type (W/C), and nonlinearity (LIM/nLIM), together with additional model features described in the “Note” column and key references where the mathematical formulations are presented. The “Note” column contains only additional features not described in Sec. IV B.

	Model		Note	Ref.
ST	W	LIM	A is estimated by Eq. (A2) in Ref. [60].	[15,60]
		nLIM	A time-independent wall is introduced.	[60]
	C	LIM	A is estimated by Eq. (A4) in Ref. [60].	[23,60]
		nLIM	A time-independent wall is introduced.	[60]
CS	W	LIM	A phase-averaged approach is adopted, and Q is time independent and diagonal.	[26]
		nLIM	A periodic wall is introduced via Eq. (12), and Q is time independent and diagonal.	–
	C	LIM	A phase-averaged approach is adopted, and Q is time independent and diagonal.	[24]
		nLIM	A periodic wall is introduced via Eq. (12).	–

size is chosen such that further increases produce negligible changes in prediction skill.

Additional considerations are required for integrating nonlinear models. The approximate system does not guarantee stationarity or bounded trajectories [53,89]. Thus, an artificial boundary condition is applied to state components whose dynamics $F_i(x, t)$ include active nonlinear terms. More precisely, the set of such components is denoted as $G = \{i | \text{vec}(B_{ijk}) \neq 0 \text{ for some } j, k\}$. The boundary condition is implemented by adding a restoring force $\mathbf{V}$, whose nonzero components are defined by

$$[\mathbf{V}(\|x\|, t)]_G = -w(\|z_G\| - 3.5) \cdot z_G, \quad (12)$$

where z is the phasewise z -score vector given by $z_j(t) = x_j(t)/\sqrt{K_{jj}(0, \theta)}$ and w is the smooth weight function defined as $w(z) = e^{mz} \cdot e^{-z^{-1}}$ for $z \geq 0$ and for some $m \gg 1$, and zero otherwise [60]. See Sec. S.III B of the Supplemental Material for more details about Eq. (12) [78].

For the colored-noise-forced model, as noted in Ref. [24], the conditional expected colored noise $\eta(t) = \mathbb{E}(\eta | x(t) = x_t)$ contributes nontrivially to the evolution. Because $\eta(t)$ does not admit a closed form, it is approximated by the conditional expected colored noise in the corresponding linear setup. This approximation is valid since the nonlinear terms are sparse compared to the linear response, thereby exerting minimal influence on the expected colored noise.

In cases where the initial condition lies outside the imposed boundary, predictions are performed using the corresponding linear model. For example, for CS-Colored-nLIM, the prediction is carried out by CS-Colored-LIM with the same γ .

The prediction skill is evaluated using the correlation coefficient between the forecasted and target values. In particular, the effective prediction range, referred to as the 0.5 correlation skill, is defined as the maximum lead month up to which the correlation coefficient remains above 0.5. Since the data are 3-month running means this metric corresponds to seasonal forecast skill.

D. Observation dataset

We start with the discussion of the in-sample prediction skill, where the training and evaluation periods are identical (1979–2022).

A representative case with noise correlation time $\gamma = 0.07$ years and the number of active components $k = 7$ is first examined (see Sec. S.III C of the Supplemental Material for the rationale behind this choice) [78]. Figure 3 presents the probability distributions of the input data and those reproduced by all eight models listed in Table II. Since the linear models are constructed solely from the input second-order correlation functions (i.e., the lag- τ covariance), they inherently tend to reproduce the input covariance more faithfully. In contrast, the nonlinear models incorporate higher-order correlation functions, placing emphasis on both skewness and covariance. As a result, the reproduced covariance may not perfectly align with the input covariance. Nonetheless, all models respect the zero-mean condition by construction, thereby avoiding systematic bias toward either warm or cool events in the simulations without postprocessing.

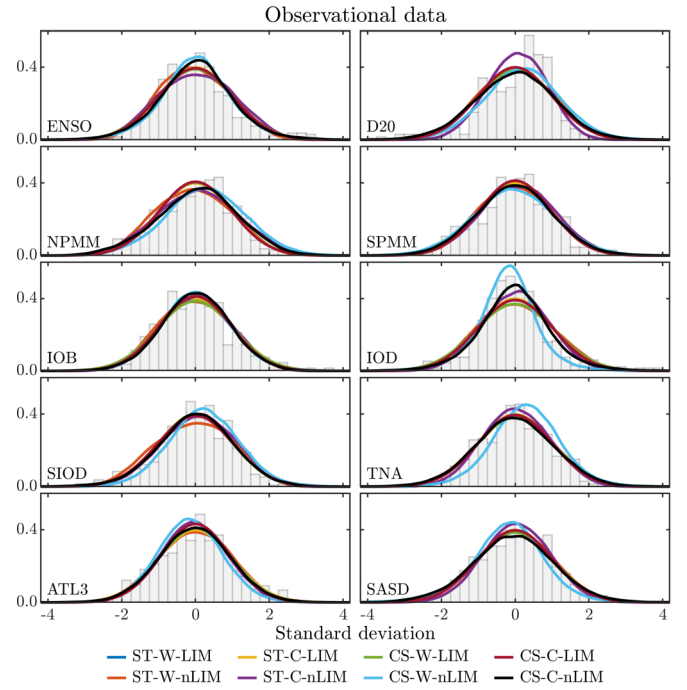


FIG. 3. The probability distribution of the observation dataset (bars) and those reproduced by all eight models (colored line) listed in Table II, with $\gamma = 0.07$ years and $k = 7$. For linear models, the distributions are computed analytically by the fluctuation-dissipation relation, whereas for the nonlinear models, they are obtained by 5000-year simulations. The distributions of ST-White-LIM and ST-Colored-LIM are indistinguishable because, by construction, the variance of the Gaussian distribution equals the input variance.

Figure 4 shows the correlation skills as a function of lead months. As discussed in Ref. [24], irrespective of nonlinearity, the cyclostationary formulation significantly enhances short-term skill ($\tau \leq 12$ months). This improvement arises because climate dynamics are strongly modulated by seasonality, leading to phase locking of climate modes during specific seasons (i.e., seasonal variance peaks) that stationary models cannot capture [90–93]. In contrast, the colored-noise representation improves the skill beyond 12 lead months, as the seasonal effect is averaged out and persistent stochastic forcing becomes more dominant. This behavior again aligns with Ref. [24], demonstrating that atmospheric processes on monthly to submonthly scales, potentially related to Madden-Julian Oscillation or westerly wind burst, can contribute over 10% in the system evolution in the linear setup [24]. Consequently, the colored-noise models capture this persistent forcing more effectively, leading to improved prediction skill at longer lead times.

Furthermore, incorporating nonlinear dynamics further enhances correlation skill, particularly when the baseline skill exceeds 0.5. Figure 5 shows that the nonlinear models outperform their linear counterparts, with CS-Colored-nLIM achieving skill gains of up to 0.05. Table III indicates that the nonlinear formulation leads to an overall improvement in 0.5 correlation skill by approximately 1–2 months across most indices, though the degree of improvement varies among indices and is not uniformly significant. In addition, a Fourier

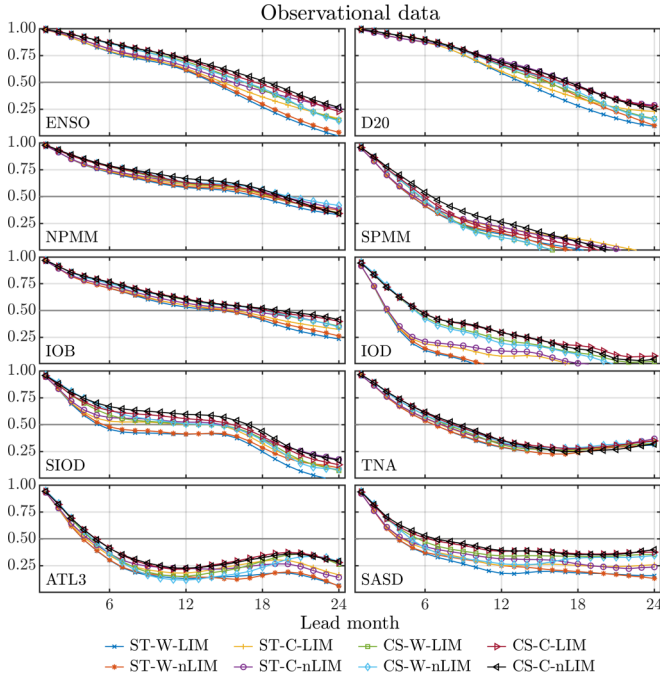


FIG. 4. The in-sample correlation skills for the observation dataset, with $\gamma = 0.07$ years and $k = 7$. The target values are the τ month lead (i.e., the three-month running averaged climate indices). The skill lower than 0 is not shown.

transform surrogate test, presented in Sec. S.III D of the Supplemental Material [78], indicates that the improvement likely arises from actual nonlinearity that is absent in the linear formulations, reflecting the inherently complex nature of the climate system [78,94]. See also Sec. S.III E of the Supplemental Material for an example illustrating changes in system evolution due to the inclusion of nonlinearity [78].

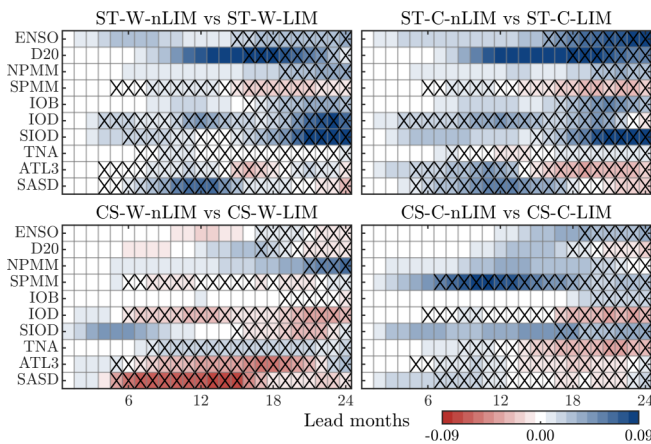


FIG. 5. Differences in in-sample correlation skills for the observation-based dataset, with $\gamma = 0.07$ years and $k = 7$. Each panel compares a nonlinear model with its corresponding linear counterpart, as indicated by the panel titles. The heatmap shows the correlation-skill differences (nonlinear minus linear), with an interval of 0.01. Cells marked with $\times$ indicate cases where the correlation skill of the nonlinear model is below 0.5.

TABLE III. The in-sample 0.5 correlation skills for the observation dataset with $\gamma = 0.07$ years and $k = 7$. The active nonlinear components are empirically determined by Eq. (7). Bold font indicates the best performance for each variable.

Variable	Stationary				Cyclostationary			
	White		Colored		White		Colored	
	LIM	nLIM	LIM	nLIM	LIM	nLIM	LIM	nLIM
ENSO	14	14	14	15	16	16	17	18
D20	13	15	14	17	15	16	16	17
NPMM	17	18	18	19	18	20	18	19
SPMM	4	4	5	5	5	5	5	6
IOB	14	15	15	16	18	18	18	19
IOD	2	3	3	3	5	5	5	5
SIOD	5	5	14	14	13	14	15	16
TNA	6	6	7	7	7	7	8	8
ATL3	3	3	4	4	4	4	4	4
SASD	3	3	4	4	5	4	6	6

Among all indices, the discussion focuses on ENSO, IOD, and ATL3, which are representative of the equatorial Pacific, Indian, and Atlantic Oceans, respectively. The ENSO dynamics are characterized by a stable oscillatory mode, primarily governed by SST-thermocline feedbacks with a typical period of 2–7 years, and are known to interact with other major climate modes across basins [95–98]. As the climate indices selected to construct the state vectors in this study are specifically designed to capture these features, CS-Colored-nLIM attains a 0.5 correlation skill of 18 months [67]. This is comparable to the performance of state-of-the-art statistical models and ML-based approaches, which typically achieve 0.5 correlation skills of 17–19 months when trained on monthly data after 1979 [43,67,99]. In fact, these benchmarks can be extended to 20–24 months when trained on additional datasets, such as advanced climate forecast products, historical simulations, or longer observational records [100,101]. Even with a relatively short 44-year dataset, CS-Colored-nLIM demonstrates strong predictive capability for ENSO.

In contrast, the IOD and ATL3 indices tend to exhibit lower predictability, with typical 0.5 correlation skills of 4–7 and 3–5 months, respectively [39,102–106]. Notably, ATL3 shows a weak rebound in skill at longer lead times, despite remaining below 0.5. This may reflect delayed, longer-timescale interactions, such as remote influences from ENSO (see also Fig. 2(k) of Ref. [67]), which introduce additional memory into ATL3 [107–110]. Overall, compared with ENSO, the improvement in predictive skill remains limited even when statistical or ML-based models are trained on advanced forecast products, such as the North American Multi-Model Ensemble [111–113]. This limitation may arise from multiple coupled mechanisms or localized processes that contribute to the irregular variability of these modes [114–120]. In addition, the input dataset used here was primarily optimized for ENSO prediction, which may further constrain the performance of CS-Colored-nLIM for other climate indices. Improving the representation of these processes through better feature selection, data assimilation, or basin-specific model tuning could enhance prediction skill. Despite these challenges,

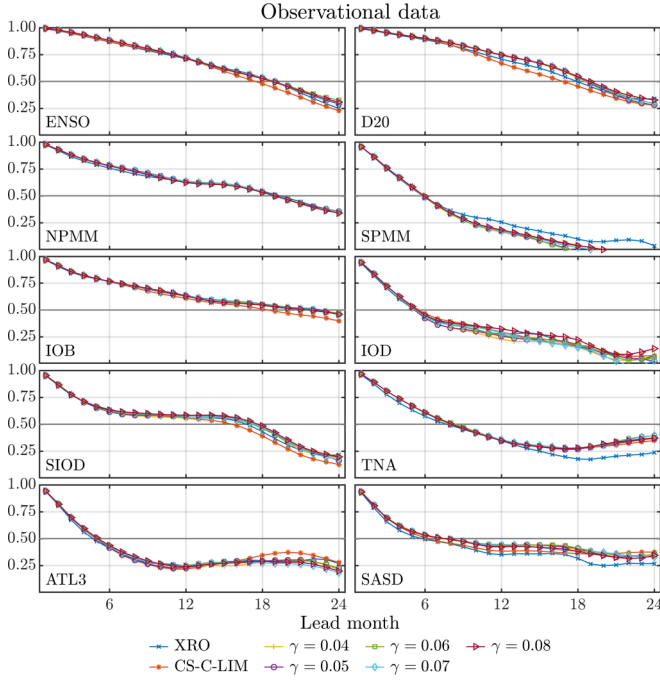


FIG. 6. The in-sample correlation skills of XRO and CS-Colored-nLIM ($k = \text{XRO}$, $\gamma = 0.03\text{--}0.08$ years) for the observation dataset.

CS-Colored-nLIM still maintains a reasonable level of in-sample predictive skill for both indices.

Regardless of the predictive performance achieved by CS-Colored-nLIM, the empirically identified active nonlinear terms differ from the XRO configuration. For instance, CS-Colored-nLIM for $k = 7$ exhibits active terms $B_{3,1,1}$, $B_{4,1,2}$, $B_{4,7,7}$, $B_{7,1,6}$, $B_{7,6,8}$, $B_{7,7,8}$, and $B_{10,5,6}$. Among these, $B_{10,5,6}$ has a largest magnitude (measured as $\|\text{vec } B_{ijk}\|$). This term represents the combined influence of the IOB and IOD on the SASD, suggesting a cross-basin interaction between variability in the Indian Ocean and the South Atlantic. One possible interpretation is that it reflects the influence of the Indian Ocean on South Atlantic variability, for example, through processes related to Agulhas Leakage [121–123].

To enable a fair comparison, CS-Colored-nLIM is also implemented with $k = \text{XRO}$. Figure 6 shows the correlation skills and, Table IV presents the 0.5 correlation skills of XRO and CS-Colored-nLIM across γ . Overall, CS-Colored-nLIM and XRO exhibit comparable performance, with the relative advantage varying among SST indices and ocean basins. For instance, CS-Colored-nLIM generally performs better for D20, SIOD, and Atlantic (ATL3, TNA, SASD) Ocean indices, whereas XRO achieves slightly higher skill for NPMM and IOB. In addition, the results indicate that CS-Colored-nLIM for $k = \text{XRO}$ achieves performance comparable to the representative case where active nonlinear terms are empirically determined, shown in Table III.

Indeed, the predictive performance is consistent across parameter settings. Table V shows the results for different combinations of γ and k , demonstrating a consistent improvement in prediction skill compared to CS-Colored-LIM results, shown in Table III. These results further suggest that

TABLE IV. The in-sample 0.5 correlation skills of CS-Colored-nLIM, evaluated for $\gamma = 0.03\text{--}0.08$ years and $k = \text{XRO}$, and those of XRO for the observation dataset. The two models share the same active nonlinear terms, $B_{1,1,1}$, $B_{1,1,2}$, and $B_{6,6,6}$. Bold font indicates the best performance for each variable.

Observation dataset						
Variable	CS-Colored-nLIM					XRO
	0.04	0.05	0.06	0.07	0.08	
ENSO	18	18	18	18	18	18
D20	18	18	18	18	18	17
NPMM	18	18	18	18	18	19
SPMM	5	5	5	5	5	5
IOB	21	21	21	22	21	22
IOD	5	5	5	5	5	5
SIOD	17	17	17	17	17	16
TNA	8	8	8	8	8	7
ATL3	4	4	4	5	5	4
SASD	7	7	7	7	7	5

incorporating nonlinearity can enhance prediction skill, while the precise configuration of active nonlinear components is of secondary importance—either because it has little influence on performance or because the models have already reached the practical limit of predictability.

E. Simulation datasets

Because the 44-year observational record is relatively short for assessing out-of-sample forecast skill, high-quality simulation data are used instead. As a representative example, the Community Earth System Model version 2 (CESM2) simulation dataset under the SSP370 scenario is employed [124]. SSP370 represents a future high-emission pathway branching from the historical simulation; the transition between the historical and scenario periods is implemented using the smoothed CMIP6 BMB protocol, which was designed to avoid spurious drifts and maintain a more realistic long-term balance [125]. In this experiment, the models are trained using the CESM2 simulation over 1850–1999 and evaluated over 2000–2024.

The results, shown in Fig. 7 and Table VI, indicate that incorporating nonlinear dynamics enhances prediction skill, although the improvement is less consistent than in the in-sample case based on the observational dataset, as shown in Fig. 4 and Table III. This reduced consistency likely reflects the fact that the identified nonlinear terms do not fully capture the evolving nonlinear behavior of the climate system during the later evaluation period. Such difficulty is expected when predicting previously unseen data and is further amplified by the temporal evolution of climate nonlinearity, particularly under the accelerated global warming observed in recent decades [126–128].

In addition, Fig. 8 and Table VII compare the prediction skills obtained when the nonlinear coefficients in CS-Colored-nLIM are prescribed to match those used in XRO. The two models show comparable predictive performance, with their relative skill varying across indices and lead months.

TABLE V. The in-sample 0.5 correlation skills of CS-Colored-nLIM for the observation-based dataset, evaluated for $\gamma = 0.04$ – 0.08 years and $k = 3$ – 8 . Bold font indicates the best performance for each variable.

Variable	$\gamma = 0.04$						$\gamma = 0.05$						$\gamma = 0.06$						$\gamma = 0.07$						$\gamma = 0.08$					
	3	4	5	6	7	8	3	4	5	6	7	8	3	4	5	6	7	8	3	4	5	6	7	8	3	4	5	6	7	8
ENSO	17	18	18	17	18	18	17	17	17	18	18	18	17	17	18	18	18	18	17	17	18	18	18	18	17	18	18	18	18	18
D20	16	16	16	16	17	17	17	17	16	17	17	17	16	16	17	17	17	17	16	16	17	17	17	17	16	17	17	17	17	17
NPMM	19	19	19	19	20	20	19	19	19	20	20	20	19	19	19	20	20	19	19	19	19	19	19	20	19	19	19	19	19	19
SPMM	5	5	5	6	5	5	5	5	6	6	6	6	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	
IOB	18	19	19	19	19	20	18	19	19	20	20	20	18	18	19	20	20	20	18	18	19	19	19	21	18	18	19	19	19	19
IOD	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
SIOD	14	14	14	15	16	16	14	15	14	16	16	16	14	15	16	16	17	17	15	15	16	16	16	17	15	15	16	16	16	16
TNA	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	
ATL3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	4	4	4	4	4	4	4	4	4	4
SASD	6	6	6	6	6	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	

Compared with the cases in which active nonlinear terms are determined empirically, CS-Colored-nLIM ($k = \text{XRO}$) performs similarly, suggesting that a physically motivated assignment of nonlinear terms can be particularly advantageous in out-of-sample predictions. Furthermore, relative to the in-sample skills obtained using the observational dataset, the skill for NPMM is notably reduced, whereas those for IOD and TNA are substantially extended. These differences may reflect variations in climate characteristics across training and evaluation periods, as well as between observational and simulation datasets.

Indeed, climate models may vary in their representation of the climate system, especially due to differences in the parametrization of unresolved processes. It is therefore important to test the performance of CS-Colored-nLIM on multiple GCM outputs. Four additional GCMs from CMIP6

(Coupled Model Intercomparison Project Phase 6) are therefore examined ACCESS-ESM1-5, CanESM5, MIROC6, and MPI-ESM1-2-LR [129–132]. Similar to the CESM2 output, historical simulations from 1850 to 2014 are merged with future projections for 2015–2024, here using the SSP585 scenario. Even though SSP585 features higher emissions than SSP370, the difference is negligible over the 10-year period considered.

Figures 9(a) and 9(c) show the pairs of 0.5 correlation skills for CS-Colored-nLIMs ($k = 7$ and $k = \text{XRO}$) and their linear counterpart CS-Colored-LIM, using a noise correlation time of $\gamma = 0.07$ years. In general, model performance varies across indices and datasets. Among the four GCMs examined, the results suggest that MIROC6 likely exhibits stronger nonlinearity and higher predictability, whereas ACCESS-ESM1-5 appears more linear and less predictable, reflecting differences in GCM physics and resulting climate variability.

Overall, the inverse models perform comparably, but the two CS-Colored-nLIM configurations show slightly more consistent improvements for ENSO and D20. This behavior likely reflects the fact that the input set of climate indices was originally selected for ENSO studies [67]. As a result,

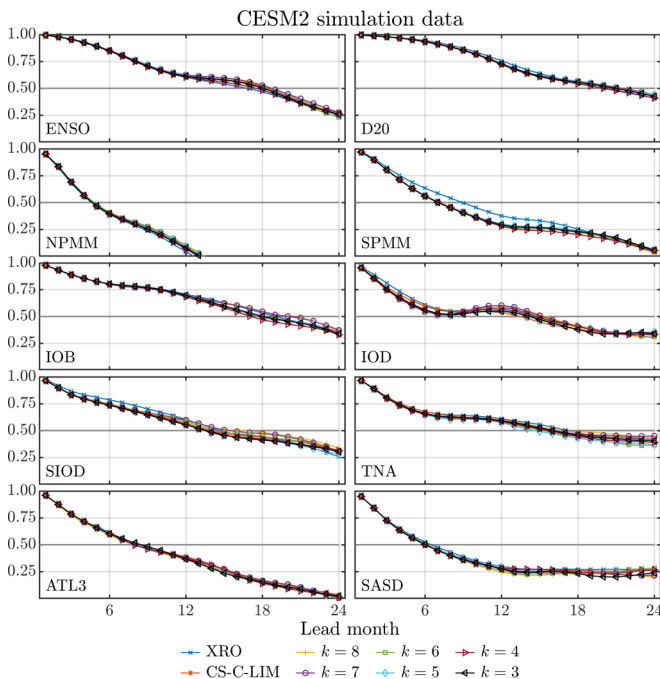


FIG. 7. The out-of-sample correlation skills for the CESM2 simulation dataset, with $\gamma = 0.07$ years and $k = 3, \dots, 8$.

TABLE VI. The out-of-sample 0.5 correlation skills for the CESM2 simulation dataset, with $\gamma = 0.07$ years and $k = 3, \dots, 8$. Bold font indicates the best performance for each variable.

CESM2 simulation dataset							
Variable	CS-Colored-nLIM						CS-Colored-LIM
	3	4	5	6	7	8	
ENSO	18	17	18	18	18	18	18
D20	20	20	20	21	21	21	21
NPMM	4	4	4	4	4	4	4
SPMM	7	7	7	6	7	7	7
IOB	18	17	17	17	20	19	18
IOD	14	14	14	14	15	15	13
SIOD	13	13	13	14	15	16	13
TNA	15	15	14	15	16	16	15
ATL3	8	7	8	8	8	7	8
SASD	6	6	6	5	6	6	6

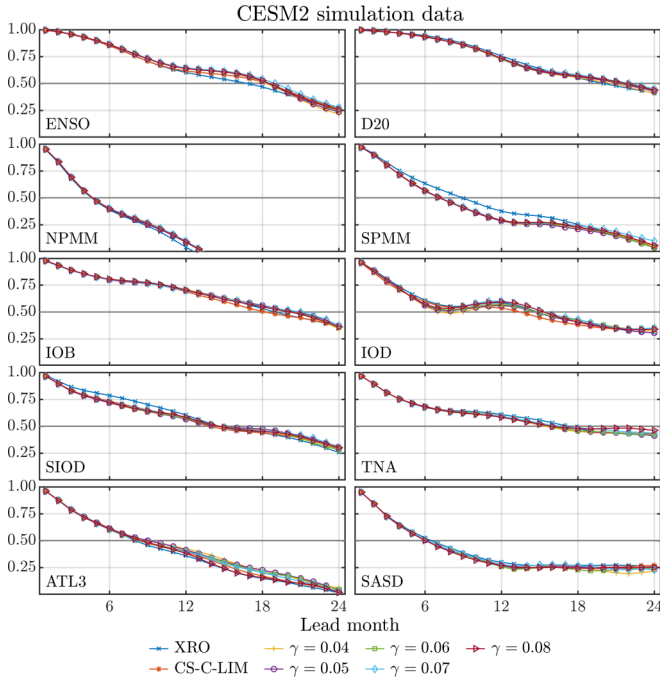


FIG. 8. The out-of-sample correlation skills for the CESM2 simulation dataset, with $\gamma = 0.04$ – 0.08 years and $k = \text{XRO}$.

increases in model complexity tend to manifest in ENSO-related indices. At the same time, the results also suggest that ENSO variability in these simulation datasets is already captured to a large extent by linear dynamics. Therefore, the linear model approaches the practical predictability limit, leaving limited room for additional improvement by the nonlinear formulation in out-of-sample prediction.

Figures 9(b) and 9(d) show pairwise skill comparisons among the nonlinear models. No single nonlinear formulation or model consistently outperforms the others across all indices or all simulation datasets. Taken together, the nonlinear models tend to perform similarly to, and in some cases

TABLE VII. The out-of-sample 0.5 correlation skills of CS-Colored-nLIM, evaluated for $\gamma = 0.04$ – 0.08 years and $k = \text{XRO}$, and those of XRO for the CESM2 simulation dataset. Bold font indicates the best performance for each variable.

CESM2 simulation dataset						
Variable	CS-Colored-nLIM					XRO
	0.04	0.05	0.06	0.07	0.08	
ENSO	18	18	18	19	18	16
D20	20	21	21	22	21	19
NPMM	4	4	4	4	4	4
SPMM	7	7	7	7	7	8
IOB	20	20	20	21	20	19
IOD	7	14	14	15	15	14
SIOD	14	14	14	14	14	14
TNA	15	15	15	16	15	17
ATL3	9	8	8	8	8	7
SASD	6	6	6	6	6	6

better than, the linear model, CS-Colored-LIM. These results indicate that CS-Colored-nLIM provides an inverse-modeling framework that has the potential to exceed the linear baseline while achieving predictive skill comparable to an established nonlinear method. More detailed results are provided in Secs. S.IV and S.V of the Supplemental Material [78].

V. CONCLUDING REMARKS

This study extends the moment-based inverse modeling framework by incorporating cyclostationarity, temporally correlated noise, and nonlinearity, along with a simple sparse identification scheme that empirically determines the active nonlinear components, resulting in the proposed CS-Colored-nLIM. This method generalizes previous formulations, which appear as its reduced variants, and naturally inherits many of their advantages and limitations. In addition, CS-Colored-nLIM is designed to preserve data constraints when applicable and to empirically determine variations in periodic dynamics, enabling a more accurate and realistic representation of data evolution.

When applied to climate prediction to a 44-year observational dataset and extended historical datasets, the proposed model demonstrates comparable or superior performance relative to its reduced variants and to existing ML-based and statistical modeling approaches. The results also suggest that the precise configuration of nonlinear components does not yield a systematic influence on the annual prediction skill. In particular, CS-Colored-nLIM configured with the nonlinear terms used in XRO performs similarly to the fully data-driven formulation, with each variant outperforming the other in different indices or lead times, but without any consistent pattern. Nevertheless, different configurations may correspond to distinct underlying physical mechanisms, which require further analysis to be properly validated.

In addition to the limitations inherited from its reduced variants, CS-Colored-nLIM introduces a few challenges. The approximate system is more sensitive to sample size because of its cyclostationary formulation and reliance on higher-order correlation functions, indicating that robust estimation requires sufficiently long training data and, in some cases, parameter reduction. In climate applications, this can be mitigated by training the model on high-quality climate model outputs or extended historical simulations, and by representing periodic variations using the leading Fourier modes. Moreover, the empirical determination of active nonlinear terms creates a trade-off between interpretability and flexibility. Tuning the Lasso parameter λ in the sparse identification process also increases computational cost, particularly for systems with high dimensionality or with a fast sampling rate relative to the length of a complete cycle.

A further limitation is that the method may not always preserve diagonal third moments (i.e., skewness of anomalous input data), as their contributions can be overshadowed by nondiagonal terms in the sparse identification step. While formulating Eq. (7) with a subset of preselected third moments could emphasize them, doing so may compromise the uniqueness of the group Lasso solution. This highlights the value of physically motivated nonlinear structures, though such considerations lie beyond the present scope [67]. Despite these

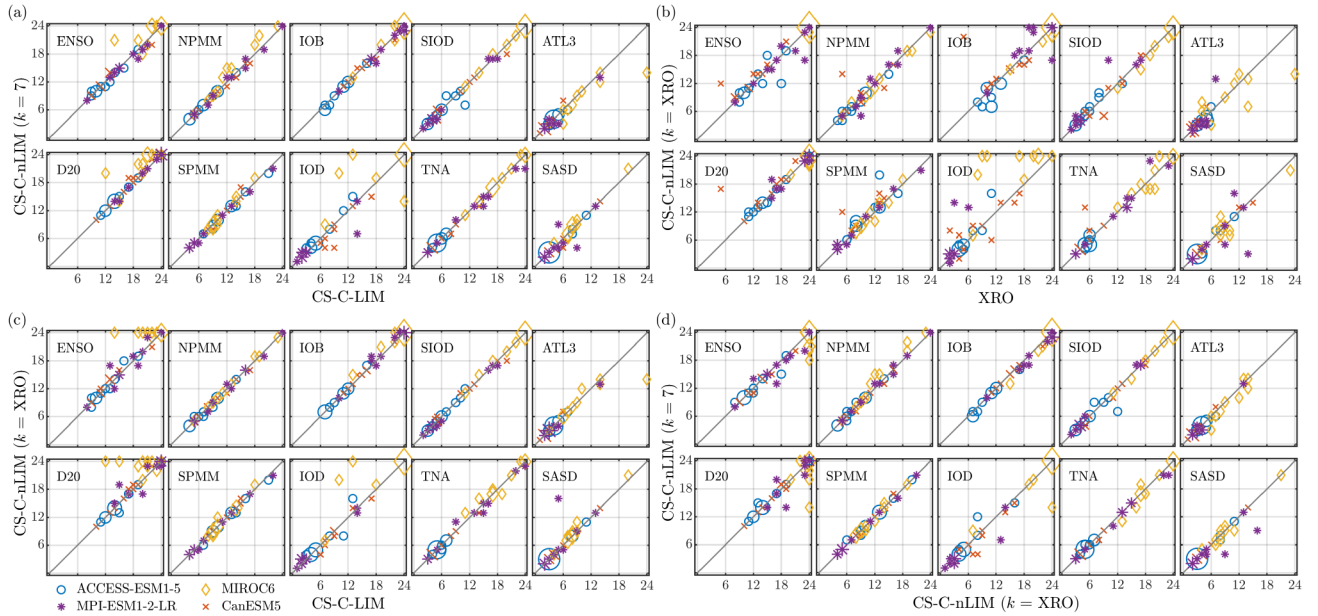


FIG. 9. Pairwise comparisons of the 0.5-correlation skills for (a) CS-Colored-LIM and CS-Colored-nLIM ($k = 7$), (b) XRO and CS-Colored-nLIM ($k = \text{XRO}$), (c) CS-Colored-LIM and CS-Colored-nLIM ($k = \text{XRO}$), and (d) CS-Colored-nLIM ($k = \text{XRO}$) and CS-Colored-nLIM ($k = 7$), evaluated using ten ensemble members from four global climate models: ACCESS-ESM1-5 (blue, circles), CanESM5 (orange, crosses), MIROC6 (yellow, diamonds), and MPI-ESM1-2-LR (purple, stars). The colored-noise models use a noise correlation time of $\gamma = 0.07$ years. Each marker represents a pair of skill values produced by the two models shown on the x and y axes, respectively. Marker size indicates the number of ensemble members yielding the same pair of skill values. The diagonal line denotes cases in which the two models achieve identical 0.5 correlation skills; points above the diagonal correspond to cases where the model on the y axis performs better.

challenges, the added complexity enhances the adaptability of CS-Colored-nLIM, offering potential for application to a wide range of problems beyond climate science.

ACKNOWLEDGMENTS

J.L. and H.A. were supported by the Council for Science, Technology and Innovation (CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), the 3rd period of SIP “Smart Energy Management System” (Grant No. JPJ012207, funding agency: JST). Y.-Y.K. was supported by the Institute for Basic Science (IBS) of the Republic of Korea under Grant No. IBS-R028-D1 and by the European Unions Horizon Europe research and innovation programme under grant agreement No. 101180125 (SEACLIM). I.R. was supported by the Japan Society for the Promotion of Science through Grant-in-Aid for Scientific Research (JSPS KAKENHI), Grant No. JP23K25946, and the Kyushu University Program for Collaborative Research, Grant No. 2025CR-AO-15. The authors acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. The authors thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies that support CMIP6 and ESGF.

DATA AVAILABILITY

The data that support the findings of this article are openly available [78,124,129–133].

APPENDIX A: PROOF OF CS-COLORED-NLIM

In this Appendix, we give a proof of CS-Colored-nLIM [i.e., Eqs. (3)–(5)], along with some direct consequences. The derivation follows the approach used in Refs. [24, 60], which utilizes the adjoint Fokker-Planck operators. For notational compactness, the mathematical derivation is written in tensor notation as in Ref. [60], summarized below.

1. Tensor notation

An a -tensor $\mathbf{T}$ is an element in $\mathbb{R}^{\overbrace{d \times \cdots \times d}^{a \text{ times}}}$. The tensor multiplication of an a -tensor $\mathbf{T}$ and a b -tensor $\mathbf{U}$ produces an $(a + b - 2m)$ -tensor $\mathbf{T} \times_m \mathbf{U}$, by summing over the last m indices of $\mathbf{T}$ and the first m indices of $\mathbf{U}$. For example, given a 3-tensor $\mathbf{T}$ and a 4-tensor $\mathbf{U}$, their product $\mathbf{T} \times_2 \mathbf{U}$ is a 3-tensor given by $[\mathbf{T} \times_2 \mathbf{U}]_{ijk} = \sum_{p,q} \mathbf{T}_{ipq} \mathbf{U}_{pqjk}$. The transpose $\mathbf{T}^*$ of $\mathbf{T}$ is given by switching the first and second indices while leaving the remaining indices unchanged (i.e., $[\mathbf{T}^*]_{i_1 i_2 i_3 \dots i_a} = \mathbf{T}_{i_2 i_1 i_3 \dots i_a}$) and the symmetrization operator Sym is given by $\text{Sym}(\mathbf{T}) = \mathbf{T}^* + \mathbf{T}$, without dividing the factor of 2.

2. Derivation

We first introduce the mathematical ingredients required for the proof. The governing Eq. (2) can be written as the augmented system for $(\mathbf{x}; \boldsymbol{\eta})$:

$$\begin{cases} \frac{d\mathbf{x}}{dt} = \underbrace{\mathbf{B}(t) \times_2 (\mathbf{x}\mathbf{x}^T) + \mathbf{A}(t) \times_1 \mathbf{x} + \mathbf{C}(t)}_{\mathbf{F}(\mathbf{x}, t)} + \sqrt{2\mathbf{Q}} \times_1 \boldsymbol{\eta}, \\ \frac{d\boldsymbol{\eta}}{dt} = -\frac{1}{\gamma} \boldsymbol{\eta} + \frac{1}{\gamma} \boldsymbol{\xi} \end{cases} \quad (\text{A1})$$

The Fokker-Planck equation for the probability distribution P of Eq. (A1) reads

$$\begin{aligned} \frac{\partial}{\partial t} P &= - \sum_i \frac{\partial}{\partial x_i} \mathbf{F}_i(x, t) P - \sum_{i,j} (\sqrt{2\mathbf{Q}})_{ij} \frac{\partial}{\partial x_i} \eta_j P \\ &\quad + \frac{1}{\gamma} \sum_i \frac{\partial}{\partial \eta_i} \eta_i P + \frac{1}{2\gamma^2} \sum_i \frac{\partial^2}{\partial \eta_i^2} P, \\ &= \mathbf{L}_{\text{FP}} P, \end{aligned}$$

where $\mathbf{L}_{\text{FP}}(t)$ denotes the Fokker-Planck operator.

Let P_{ss} denote the cyclostationary (steady-state) distribution and $\mathbf{L}_{\text{FP}}^*$ denote the adjoint Fokker-Planck operator given

$$\begin{aligned} \mathbf{L}_{\text{FP}}^*(t) &= \sum_i \mathbf{F}_i(x, t) \frac{\partial}{\partial x_i} + \sum_{i,j} (\sqrt{2\mathbf{Q}})_{ij} \eta_j \frac{\partial}{\partial x_i} \\ &\quad - \frac{1}{\gamma} \sum_i \eta_i \frac{\partial}{\partial \eta_i} + \frac{1}{2\gamma^2} \sum_i \frac{\partial^2}{\partial \eta_i^2}. \end{aligned}$$

In addition to the statistics of the state vector as introduced in the main text, we denote the noise-state correlation tensors $\mathbf{K}_{\eta\mathbf{x}}(\theta) = \langle \eta(t) \mathbf{x}(t)^T \rangle$ and $\mathbf{M}_{\eta\mathbf{x}}(\theta) = \langle \eta(t) \times_1 (\mathbf{x}(t) \mathbf{x}(t)^T) \rangle$, which are nontrivial.

The partial derivatives with respect to τ are right derivatives, whereas those with respect to θ are two-sided derivatives.

We are now ready to present the main result.

Theorem A1. The derivatives of the statistics of the state vector satisfy

$$\mathbf{E}_\theta(\theta) = \mathbf{B}(\theta) \times_2 \mathbf{K}(0, \theta) + \mathbf{A}(\theta) \times_1 \mathbf{E}(\theta) + \mathbf{C}(\theta), \quad (\text{A2})$$

$$\mathbf{K}_\tau(0, \theta) = \mathbf{B}(\theta) \times_2 \mathbf{M}(0, \theta) + \mathbf{A}(\theta) \times_1 \mathbf{K}(0, \theta) + \sqrt{2\mathbf{Q}} \times_1 \mathbf{K}_{\eta\mathbf{x}}(\theta) + \mathbf{C}(\theta) \times_0 \mathbf{E}(\theta), \quad (\text{A3})$$

$$\begin{aligned} \gamma \mathbf{K}_{\tau\tau}(0, \theta) &= \gamma (\mathbf{B}'(\theta) \times_2 \mathbf{M}(0, \theta) + \mathbf{A}'(\theta) \times_1 \mathbf{K}(0, \theta) + \mathbf{C}'(\theta) \times_0 \mathbf{E}(\theta)) \\ &\quad + \gamma (\mathbf{B}(\theta) \times_2 \text{Sym}(\mathbf{M}_\tau(0, \theta)) + \mathbf{A}(\theta) \times_1 \mathbf{K}_\tau(0, \theta)) - \sqrt{2\mathbf{Q}} \times_1 \mathbf{K}_{\eta\mathbf{x}}(\theta), \end{aligned} \quad (\text{A4})$$

$$\mathbf{M}_\tau(0, \theta) = \mathbf{B}(\theta) \times_2 \mathbf{S}(0, \theta) + \mathbf{A}(\theta) \times_1 \mathbf{M}(0, \theta) + \sqrt{2\mathbf{Q}} \times_1 \mathbf{M}_{\eta\mathbf{x}}(\theta) + \mathbf{C}(\theta) \times_0 \mathbf{K}(0, \theta), \quad (\text{A5})$$

$$\begin{aligned} \gamma \mathbf{M}_{\tau\tau}(0, \theta) &= \gamma (\mathbf{B}'(\theta) \times_2 \mathbf{S}(0, \theta) + \mathbf{A}'(\theta) \times_1 \mathbf{M}(0, \theta) + \mathbf{C}'(\theta) \times_0 \mathbf{K}(0, \theta)) \\ &\quad + \gamma (\mathbf{B}(\theta) \times_2 \text{Sym}(\mathbf{S}_\tau(0, \theta)) + \mathbf{A}(\theta) \times_1 \mathbf{M}_\tau(0, \theta)) - \sqrt{2\mathbf{Q}} \times_1 \mathbf{M}_{\eta\mathbf{x}}(\theta). \end{aligned} \quad (\text{A6})$$

Proof. Following Refs. [24,60,134,135], for L^2 functions f and g , we have

$$\begin{aligned} \langle f(\mathbf{x}(t+\tau), \eta(t+\tau)), g(\mathbf{x}(t), \eta(t)) \rangle &= \int_{\mathbb{R}^{2n}} f(x, \eta) e^{\int_0^\tau \mathbf{L}_{\text{FP}}(t+\tau') d\tau'} P_{\text{ss}}(x, \eta, t) g(x, \eta) dx d\eta \\ &= \int_{\mathbb{R}^{2n}} g(x, \eta) P_{\text{ss}}(x, \eta, t) e^{\int_0^\tau \mathbf{L}_{\text{FP}}^*(t+\tau') d\tau'} f(x, \eta) dx d\eta \\ &= \int_{\mathbb{R}^{2n}} g(x, \eta) P_{\text{ss}}(x, \eta, t) \left(\mathbf{I} + \tau \mathbf{L}_{\text{FP}}^*(t) + \frac{\tau^2}{2} \left(\mathbf{L}_{\text{FP}}^*(t) \mathbf{L}_{\text{FP}}^*(t) + \frac{d}{dt} \mathbf{L}_{\text{FP}}^*(t) \right) \right) \\ &\quad \times f(x, \eta) dx d\eta + O(\tau^3). \end{aligned} \quad (\text{A7})$$

Therefore, by Eq. (A7), we compute the first-order right derivative of the (p, q) entry of $\mathbf{K}$ with respect to τ as follows:

$$\begin{aligned} \frac{\partial}{\partial \tau} \mathbf{K}_{pq}(0, \theta) &= \int_{\mathbb{R}^{2n}} x_q P_{\text{ss}}(x, \eta, t) \mathbf{L}_{\text{FP}}^*(\theta) x_p dx d\eta \\ &= \int_{\mathbb{R}^{2n}} x_q P_{\text{ss}}(x, \eta) \left(\sum_{j,k} \mathbf{B}_{pjk}(\theta) x_j x_k + \sum_j \mathbf{A}_{pj}(\theta) x_j + \mathbf{C}_p(\theta) + \sum_j \eta_j (\sqrt{2\mathbf{Q}})_{pj} \right) dx d\eta \\ &= \sum_{j,k} \mathbf{B}_{pjk}(\theta) \mathbf{M}_{jkq}(0, \theta) + \sum_j \mathbf{A}_{pj}(\theta) \mathbf{K}_{jq}(0, \theta) + \mathbf{C}_p(\theta) \mathbf{E}_q(\theta) + \sum_j (\sqrt{2\mathbf{Q}})_{pj} (\mathbf{K}_{\eta\mathbf{x}}(\theta))_{jq}. \end{aligned}$$

This is equivalent to Eq. (A3) in tensor form.

The computations for $\mathbf{K}_{\tau\tau}(0, \theta)$, $\mathbf{M}_\tau(0, \theta)$, and $\mathbf{M}_{\tau\tau}(0, \theta)$ proceed analogously, and the detailed steps are omitted for brevity. Finally, with the fact that $\langle \eta(\cdot) \rangle = 0$, Eq. (A2) can be obtained from

$$0 = \int_{\mathbb{R}^{2n}} x_j \mathbf{L}_{\text{FP}} P_{\text{ss}} dx d\eta = \int_{\mathbb{R}^{2n}} P_{\text{ss}} \mathbf{L}_{\text{FP}}^* x_j dx d\eta.$$

For implementation purposes, we transform the tensor equations into matrix form. For a tensor $\mathbf{T}$, let $\text{vec } \mathbf{T} = \text{vec}(\mathbf{T}) = \vec{\mathbf{T}}$ denote its vectorized form. The tensor multiplication $\mathbf{T} \times_m \mathbf{U}$ can be written in terms of matrix multiplication as $\mathbf{T} \times_m \mathbf{U} = \mathbf{L}_U^m \vec{\mathbf{T}}$, where $\mathbf{L}_U^m$ represents the matrix corresponding to $\mathbf{U}$. Then, eliminating $\sqrt{2\mathbf{Q}} \times_1 \mathbf{K}_{\eta\mathbf{x}}$ in Eqs. (A3) and (A4) yields Eq. (4) in matrix form. Similarly, Eqs. (3) and (5) can be obtained through the same procedure.

Finally, there are two immediate results.

Corollary A1. The stochastic matrix $\mathbf{Q}$ can be written as in terms of $(\boldsymbol{\gamma}, \mathbf{B}, \mathbf{A}, \mathbf{C})$ and the statistics of $\mathbf{x}$:

$$\begin{aligned} \mathbf{Q} = & \boldsymbol{\gamma} \frac{d}{d\theta} (\mathbf{K}_\tau(0, \theta) - \mathbf{B}(\theta) \times_2 \mathbf{M}(0, \theta) - \mathbf{A}(\theta) \times_1 \mathbf{K}(0, \theta) - \mathbf{C}(\theta) \times_0 \mathbf{E}(\theta)) \\ & - \boldsymbol{\gamma} (\mathbf{M}_\tau(0, \theta) - \mathbf{B}(\theta) \times_2 \mathbf{S}(0, \theta) - \mathbf{A}(\theta) \times_1 \mathbf{M}(0, \theta) - \mathbf{C}(\theta) \times_0 \mathbf{K}(0, \theta)) \times_2 \mathbf{B}(\theta)^* \\ & - \boldsymbol{\gamma} (\mathbf{K}_\tau(0, \theta) - \mathbf{B}(\theta) \times_2 \mathbf{M}(0, \theta) - \mathbf{A}(\theta) \times_1 \mathbf{K}(0, \theta) - \mathbf{C}(\theta) \times_0 \mathbf{E}(\theta)) \times_1 \left(\mathbf{A}(\theta)^* - \frac{1}{\boldsymbol{\gamma}} \mathbf{I} \right). \end{aligned} \quad (\text{A8})$$

The cyclostationary balance equation in the nonlinear framework is given by

$$\mathbf{K}_\theta(0, \theta) = \text{Sym}(\mathbf{B}(\theta) \times_2 \mathbf{M}(0, \theta) + \mathbf{A}(\theta) \times_1 \mathbf{K}(0, \theta) + \mathbf{C}(\theta) \times_0 \mathbf{E}(\theta) + \sqrt{2\mathbf{Q}} \times_1 \mathbf{K}_{\eta\mathbf{x}}(\theta)). \quad (\text{A9})$$

Proof. Equation (A8) can be obtained by a direct computation of

$$\frac{d}{d\theta} (\mathbf{K}_{\eta\mathbf{x}})_{pq} = \int_{\mathbb{R}^{2n}} \eta_p x_q \mathbf{L}_{\text{FP}} P_{ss} dx d\eta = \int_{\mathbb{R}^{2n}} P_{ss} \mathbf{L}_{\text{FP}}^* \eta_p x_q dx d\eta,$$

leading to

$$\frac{d}{d\theta} \mathbf{K}_{\eta\mathbf{x}} = \mathbf{M}_{\eta\mathbf{x}}(\theta) \times_2 \mathbf{B}(\theta)^* + \mathbf{K}_{\eta\mathbf{x}}(\theta) \times_1 \left(\mathbf{A}(\theta)^* - \frac{1}{\boldsymbol{\gamma}} \mathbf{I} \right) + \frac{1}{2\boldsymbol{\gamma}} \sqrt{2\mathbf{Q}}. \quad (\text{A10})$$

Then, multiplying $\sqrt{2\mathbf{Q}}$ from the left and eliminating $\sqrt{2\mathbf{Q}} \times_1 \mathbf{K}_{\eta\mathbf{x}}$ and $\sqrt{2\mathbf{Q}} \times_1 \mathbf{M}_{\eta\mathbf{x}}$ using Eqs. (A3) and (A5) yields Eq. (A8). Finally, Eq. (A9) can be obtained by computing

$$\frac{d}{d\theta} \mathbf{K}_{pq}(0, \theta) = \int_{\mathbb{R}^{2n}} x_p x_q \mathbf{L}_{\text{FP}} P_{ss} dx d\eta = \int_{\mathbb{R}^{2n}} P_{ss} \mathbf{L}_{\text{FP}}^* x_p x_q dx d\eta.$$

APPENDIX B: CS-WHITE-NLIM

The governing equation of CS-White-nLIM is given by

$$\frac{d\mathbf{x}}{dt} = \mathbf{B}(t) \times_2 (\mathbf{x}\mathbf{x}^T) + \mathbf{A}(t) \times_1 \mathbf{x} + \mathbf{C}(t) + \sqrt{2\mathbf{Q}(t)} \times_1 \boldsymbol{\xi}, \quad (\text{B1})$$

where in contrast to CS-Colored-nLIM, the stochastic matrix $\mathbf{Q}$ is allowed to be periodic.

Because the white-noise limit is formal, the derivatives of the correlation functions of Eq. (2) do not converge to those of Eq. (B1) as $\gamma \rightarrow 0$, as noted in Ref. [60]. Therefore, the derivation must be done separately.

Indeed, with the Fokker-Planck equation of Eq. (B1) and following a similar computation, we have

$$\mathbf{E}_\theta(\theta) = \mathbf{B}(\theta) \times_2 \mathbf{K}(0, \theta) + \mathbf{A}(\theta) \times_1 \mathbf{E}(\theta) + \mathbf{C}(\theta), \quad (\text{B2})$$

$$\begin{aligned} \mathbf{K}_\tau(0, \theta) = & \mathbf{B}(\theta) \times_2 \mathbf{M}(0, \theta) + \mathbf{A}(\theta) \times_1 \mathbf{K}(0, \theta) + \mathbf{C}(\theta) \\ & \times_0 \mathbf{E}(\theta), \end{aligned} \quad (\text{B3})$$

$$\begin{aligned} \mathbf{M}_\tau(0, \theta) = & \mathbf{B}(\theta) \times_2 \mathbf{S}(0, \theta) + \mathbf{A}(\theta) \times_1 \mathbf{M}(0, \theta) + \mathbf{C}(\theta) \\ & \times_0 \mathbf{K}(0, \theta). \end{aligned} \quad (\text{B4})$$

Moreover, the balance equation is given by

$$\begin{aligned} \frac{d}{d\theta} \mathbf{K}(0, \theta) = & \text{Sym}(\mathbf{B}(\theta) \times_2 \mathbf{M}(0, \theta) + \mathbf{A}(\theta) \times_1 \mathbf{K}(0, \theta) \\ & + \mathbf{C}(\theta) \times_0 \mathbf{E}(\theta) + \mathbf{Q}(\theta)). \end{aligned} \quad (\text{B5})$$

Eqs. (B2)–(B4) are Eqs. (A2)–(A6) when $\gamma = 0$. This result again indicates that in the inverse formulation, the parameter estimation procedure is continuous for $\gamma \geq 0$ near 0.

APPENDIX C: PHASE-AVERAGED APPROACH

Compared to the fixed-phase approach, the phase-averaged approach constrains the estimated parameters to the leading Fourier modes up to order L . To avoid confusion, let $B^{(1)}$ denote the estimated parameters from the fixed-phase approach [i.e., Eqs. (7) and (8)] and $B^{(2)}$ those from the phase-averaged approach. The same notation applies to A and C .

Let N be the sampling frequency and Q_0 be a tall matrix whose columns consist of the Fourier basis of unit length up to order L . That is, Q_0 is a matrix of size N -by- $(2L+1)$ if $N > 2L+1$, or an N -by- N matrix if all Fourier modes are included. Define the leading Fourier coefficients of $\text{vec } \hat{B}_{ijk}^{(2)}$ as $\text{vec } B_{ijk}^{(2)} = Q_0 \text{vec } \hat{B}_{ijk}^{(2)}$. Then Eq. (7) is reformulated in the Fourier domain by projecting it onto the truncated Fourier basis as

$$\min_{\text{vec } \hat{B}^{(2)}} \left(\left\| \widetilde{X}_{MB} \widetilde{Q} \text{vec } \hat{B}^{(2)} - \widetilde{Y}_M \right\|_2^2 + \lambda \sum_{i,j \leq k} \left\| \text{vec } \hat{B}_{ijk}^{(2)} \right\|_2 \right), \quad (\text{C1})$$

where $\widetilde{Q}$ is a block diagonal matrix constructed from Q_0 and is compatible to $\widetilde{X}_{MB}$ and $\text{vec } \hat{B}^{(2)}$, the vertical concatenation of all $\text{vec } \hat{B}_{ijk}^{(2)}$.

With the identified active nonlinear components, the dynamical parameters are solved by reformulating Eq. (8) in the Fourier domain:

$$XQ\widehat{V}^{(2)} = Y,$$

where $\widehat{V}^{(2)}$ represents the leading Fourier coefficients of the estimated dynamical parameters $(B^{(2)}, A^{(2)}, C^{(2)})$ and Q the block diagonal constructed from Q_0 that is compatible to both X and $\widehat{V}^{(2)}$. Then the dynamical parameters are recovered from $Q\widehat{V}^{(2)}$.

It is worth noting that the fixed-phase approach can be viewed as a special case of the phase-averaged approach. Suppose that $\widehat{X}_{MB}$ has full rank so that the group-lasso solution is unique. When all Fourier modes are retained in the phase-averaged formulation, Eq. (C1) reduces to Eq. (7) since the Fourier transformation is orthogonal. Because the configurations of active nonlinear coefficients are identical, the two approaches become equivalent:

$$(B^{(1)}, A^{(1)}, C^{(1)}) = (B^{(2)}, A^{(2)}, C^{(2)}).$$

APPENDIX D: XRO SETUP AND FORECAST

This appendix describes the experimental configuration of XRO and discusses its methodological differences from CS-Colored-nLIM in prediction.

The data preprocessing for XRO is essentially identical to that used for the moment-based models, except that the three-month running mean is not applied. The dynamical parameters are also expressed as a Fourier series up to the second order, following the default configuration in Ref. [67]. Although XRO is trained on unaveraged data, its prediction skills are evaluated using three-month running-mean data to ensure consistency with the results from the moment-based models in Secs. IV D and IV E.

Prediction in XRO is performed by numerically integrating the approximate system (1) without the stochastic term:

$$\frac{dx_i}{dt} = \underbrace{\sum_{j \leq k} B_{ijk}(t)x_j x_k + \sum_j A_{ij}(t)x_j}_{\text{Dynamics: } F_i(x,t)}. \quad (\text{D1})$$

This deterministic formulation may introduce bias from the formal solution obtained by integrating the full stochastic system in Eq. (1) with initialized colored noise and subsequently taking the ensemble mean [24]. Nevertheless, this simplification significantly reduces computational cost.

On the other hand, CS-Colored-nLIM initializes the colored noise by approximating its distribution with that of the corresponding linear system, integrates the approximate SDE (A1), and computes the ensemble mean. Although one could instead integrate only the deterministic component of Eq. (A1) to avoid ensemble generation, we adopt the ensemble-based approach to access the probability distribution and ensemble spread of the system (see Figs. 3 and S.5 of the Supplemental Material [78]), despite the cost of increased computational demand. Moreover, while this initialization strategy for colored noise still yields minor biased forecasted values from the formal solution—even when evaluated over an ensemble of infinite members—it effectively corrects the initial tendency dx/dt associated with colored forcing.

It should also be noted that, in the linear formulation, the effect of colored noise can be incorporated into the evolution operator, thereby achieving both accurate representation of colored-noise effects and computational efficiency, as discussed in Ref. [24].

-
- [1] M. Barati Moghaddam, M. Mazaheri, and J. Mohammad Vali Samani, Inverse modeling of contaminant transport for pollution source identification in surface and groundwaters: A review, *Groundwater Sustainable Dev.* **15**, 100651 (2021).
 - [2] H. Bonakdari, A. D. Binns, and B. Gharabaghi, A comparative study of linear stochastic with nonlinear daily river discharge forecast models, *Water Resour. Manage.* **34**, 3689 (2020).
 - [3] H. Ghorbanidehno, A. Kokkinaki, J. Lee, and E. Darve, Recent developments in fast and scalable inverse modeling and data assimilation methods in hydrology, *J. Hydrol.* **591**, 125266 (2020).
 - [4] J. Grundmann, S. Hörning, and A. Bárdossy, Stochastic reconstruction of spatio-temporal rainfall patterns by inverse hydrologic modelling, *Hydrol. Earth Syst. Sci.* **23**, 225 (2019).
 - [5] Y.-Y. Kim, J. Lien, K. Rodgers, J.-Y. Lee, and F.-F. Jin, Exploring marine biogeochemical prediction with linear dynamical frameworks (unpublished).
 - [6] D. Nolte and C. Bertoglio, Inverse problems in blood flow modeling: A review, *Int. J. Numer. Methods Biomed. Eng.* **38**, e3613 (2022).
 - [7] R. Sternfels and C. J. Earls, Reduced-order model tracking and interpolation to solve PDE-based Bayesian inverse problems, *Inverse Probl.* **29**, 075014 (2013).
 - [8] X. Li, S.-L. Chen, C. S. Teo, and K. K. Tan, Data-based tuning of reduced-order inverse model in both disturbance observer and feedforward with application to tray indexing, *IEEE Trans. Ind. Electron.* **64**, 5492 (2017).
 - [9] X. Liu, Q. Zhou, J. Birkholzer, and W. A. Illman, Geostatistical reduced-order models in underdetermined inverse problems, *Water Resour. Res.* **49**, 6587 (2013).
 - [10] L. Jiang and N. Ou, Multiscale model reduction method for Bayesian inverse problems of subsurface flow, *J. Comput. Appl. Math.* **319**, 188 (2017).
 - [11] L. Lu, Y. Li, and X. Liu, Data-driven approach for extracting the most probable exit trajectory of stochastic dynamical systems with non-Gaussian Lévy noise, *Acta Mech. Sin.* **40**, 523094 (2024).
 - [12] X. Wang, J. Feng, Q. Liu, and Y. Xu, Noise-induced alternations and data-driven parameter estimation of a stochastic perceptual model, *Eur. Phys. J.: Spec. Top.* **234**, 707 (2025).
 - [13] Y. Zhang, Y. Li, and Y. Jin, Data-driven modelling and dynamic analysis of the multistable energy harvester with non-Gaussian Lévy noise, *Appl. Math. Modell.* **125**, 390 (2024).

- [14] C. Penland and T. Magorian, Prediction of Niño 3 sea surface temperatures using linear inverse modeling, *J. Clim.* **6**, 1067 (1993).
- [15] C. Penland and L. Matrosova, A balance condition for stochastic numerical models with application to the El Niño-southern oscillation, *J. Clim.* **7**, 1352 (1994).
- [16] C. Penland, A stochastic model of IndoPacific sea surface temperature anomalies, *Phys. D* **98**, 534 (1996).
- [17] F. Kwasniok, Linear inverse modeling of large-scale atmospheric flow using optimal mode decomposition, *J. Atmos. Sci.* **79**, 2181 (2022).
- [18] J. Lien, H. Ando, Y.-Y. Kim, and T. Tozuka, Linear-inverse-modeling approach to estimating Liang-Kleeman information flow in a cyclostationary process under memoryless and persistent noise, *Phys. Rev. E* **112**, 054101 (2025).
- [19] M. Newman, S.-I. Shin, and M. A. Alexander, Natural variation in ENSO flavors, *Geophys. Res. Lett.* **38**, L14705 (2011).
- [20] C. Penland and P. D. Sardeshmukh, The optimal growth of tropical sea surface temperature anomalies, *J. Clim.* **8**, 1999 (1995).
- [21] Y. Zhao, F. Song, D. Sun, L. Dong, and A. Capotondi, Different roles of Indian Ocean basin and dipole modes in tropical Pacific climate variability, *npj Clim. Atmos. Sci.* **8**, 163 (2025).
- [22] H.-C. Chen and F.-F. Jin, Simulations of ENSO phase-locking in CMIP5 and CMIP6, *J. Clim.* **34**, 5135 (2021).
- [23] J. Lien, Y.-N. Kuo, H. Ando, and S. Kido, Colored linear inverse model: A data-driven method for studying dynamical systems with temporally correlated stochasticity, *Phys. Rev. Res.* **7**, 023042 (2025).
- [24] J. Lien, H. Ando, I. Richter, and S. Kido, Noise persistence in a cyclostationary linear inverse model framework and its impact on ENSO evolution and prediction skills, *J. Atmos. Sci.* **83**, 391 (2026).
- [25] C. Martinez-Villalobos, D. J. Vimont, C. Penland, M. Newman, and J. D. Neelin, Calculating state-dependent noise in a linear inverse model framework, *J. Atmos. Sci.* **75**, 479 (2018).
- [26] S.-I. Shin, P. D. Sardeshmukh, M. Newman, C. Penland, and M. A. Alexander, Impact of annual cycle on ENSO variability and predictability, *J. Clim.* **34**, 171 (2021).
- [27] C. Martinez-Villalobos, M. Newman, D. J. Vimont, C. Penland, and J. David Neelin, Observed El Niño-La Niña asymmetry in a linear model, *Geophys. Res. Lett.* **46**, 9909 (2019).
- [28] S. Kido, I. Richter, T. Tozuka, and P. Chang, Understanding the interplay between ENSO and related tropical SST variability using linear inverse models, *Clim. Dyn.* **61**, 1029 (2023).
- [29] X. Fang, H. Dijkstra, C. Wieners, and F. Guardamagna, A nonlinear full-field conceptual model for ENSO diversity, *J. Clim.* **37**, 3759 (2024).
- [30] C. Wang, A review of ENSO theories, *Natl. Sci. Rev.* **5**, 813 (2018).
- [31] H. Almukhalafi, A. Noor, and T. H. Noor, Traffic management approaches using machine learning and deep learning techniques: A survey, *Eng. Appl. Artif. Intell.* **133**, 108147 (2024).
- [32] D. M. Camacho, K. M. Collins, R. K. Powers, J. C. Costello, and J. J. Collins, Next-generation machine learning for biological networks, *Cell* **173**, 1581 (2018).
- [33] A. Rajkomar, J. Dean, and I. Kohane, Machine learning in medicine, *N. Engl. J. Med.* **380**, 1347 (2019).
- [34] R. Séférian, L. Bopp, M. Gehlen, D. Swingedouw, J. Mignot, E. Guilyardi, and J. Servonnat, Multiyear predictability of tropical marine productivity, *Proc. Natl. Acad. Sci. USA* **111**, 11646 (2014).
- [35] M.-L. Bachélery, J. Brajard, M. Patacchiola, S. Illig, and N. Keenlyside, Predicting atlantic and benguela Niño events with deep learning, *Sci. Adv.* **11**, eads5185 (2025).
- [36] I. Colfescu, H. Christensen, and D. J. Gagne, A machine learning-based approach to quantify ENSO sources of predictability, *Geophys. Res. Lett.* **51**, e2023GL105194 (2024).
- [37] M. Hashemi, Forecasting El Niño and La Niña using spatially and temporally structured predictors and a convolutional neural network, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **14**, 3438 (2021).
- [38] X. Li, W. Tan, Z.-Z. Hu, and N. C. Johnson, Evolution and prediction of two extremely strong Atlantic Niños in 2019–2021: Impact of Benguela warming, *Geophys. Res. Lett.* **50**, e2023GL104215 (2023).
- [39] F. Ling, J.-J. Luo, Y. Li, T. Tang, L. Bai, W. Ouyang, and T. Yamagata, Multi-task machine learning improves multi-seasonal prediction of the Indian Ocean Dipole, *Nat. Commun.* **13**, 7681 (2022).
- [40] H. Liu, Y. Tang, D. Chen, and T. Lian, Predictability of the Indian Ocean Dipole in the coupled models, *Clim. Dyn.* **48**, 2005 (2017).
- [41] J. Liu, Y. Tang, Y. Wu, T. Li, Q. Wang, and D. Chen, Forecasting the Indian Ocean Dipole with deep learning techniques, *Geophys. Res. Lett.* **48**, e2021GL094407 (2021).
- [42] K. A. Silva, G. de Souza Rolim, and L. E. de Oliveira Aparecido, Forecasting El Niño and La Niña events using decision tree classifier, *Theor. Appl. Climatol.* **148**, 1279 (2022).
- [43] Y.-G. Ham, J.-H. Kim, and J.-J. Luo, Deep learning for multi-year ENSO forecasts, *Nature (London)* **573**, 568 (2019).
- [44] I. Richter, P. Chang, and X. Liu, Impact of systematic GCM errors on prediction skill as estimated by linear inverse modeling, *J. Clim.* **33**, 10073 (2020).
- [45] W. W. Hsieh, *Machine Learning in Environmental and Climate Science: Overview and Introduction*, Oxford Research Encyclopedia of Climate Science (Oxford University Press, 2025).
- [46] P. Lyu, T. Tang, F. Ling, J.-J. Luo, N. Boers, W. Ouyang, and L. Bai, Resonet: Robust and explainable ENSO forecasts with hybrid convolution and transformer networks, *Adv. Atmos. Sci.* **41**, 1289 (2024).
- [47] R. Yang, J. Hu, Z. Li, J. Mu, T. Yu, J. Xia, X. Li, A. Dasgupta, and H. Xiong, Interpretable machine learning for weather and climate prediction: A review, *Atmos. Environ.* **338**, 120797 (2024).
- [48] C. Chen, M. A. Cane, N. Henderson, D. E. Lee, D. Chapman, D. Kondrashov, and M. D. Chekroun, Diversity, nonlinearity, seasonality, and memory effect in ENSO simulation and prediction using empirical model reduction, *J. Clim.* **29**, 1809 (2016).
- [49] S. De, N. K. Agarwal, J. S. Chowdary, and D. Patekar, Scale interactions during ENSO—A new dynamical mechanism on unveiling ENSO-monsoon nonlinearity, *Theor. Appl. Climatol.* **155**, 8119 (2024).
- [50] C. Fronk and L. Petzold, Interpretable polynomial neural ordinary differential equations, *Chaos* **33**, 043101 (2023).

- [51] D. Kondrashov, S. Kravtsov, A. W. Robertson, and M. Ghil, A hierarchy of data-based ENSO models, *J. Clim.* **18**, 4425 (2005).
- [52] D. Kondrashov, M. D. Chekroun, and M. Ghil, Data-driven non-Markovian closure models, *Physica D* **297**, 33 (2015).
- [53] S. Kravtsov, D. Kondrashov, and M. Ghil, Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability, *J. Clim.* **18**, 4404 (2005).
- [54] A. Lück and V. Wolf, Generalized method of moments for estimating parameters of stochastic reaction networks, *BMC Syst. Biol.* **10**, 98 (2016).
- [55] Y. Li and Z. Wei, Dynamics and optimal control of a stochastic coronavirus (COVID-19) epidemic model with diffusion, *Nonlinear Dyn.* **109**, 91 (2022).
- [56] D. Olabode, J. Culp, A. Fisher, A. Tower, D. Hull-Nye, and X. Wang, Deterministic and stochastic models for the epidemic dynamics of COVID-19 in Wuhan, China, *Math. Biosci. Eng.* **18**, 950 (2021).
- [57] G. Schneider, P. F. Craigmile, and R. Herbei, Maximum likelihood estimation for stochastic differential equations using sequential Gaussian-process-based optimization, *Technometrics* **59**, 178 (2017).
- [58] J. Timmer, Parameter estimation in nonlinear stochastic differential equations, *Chaos Solitons Fract.* **11**, 2571 (2000).
- [59] C. Qiu and M. E. Orazem, A weighted nonlinear regression-based inverse model for interpretation of pipeline survey data, *Electrochim. Acta* **49**, 3965 (2004).
- [60] J. Lien and H. Ando, Nonlinear extensions of linear inverse models under memoryless or persistent random forcing, *Phys. Rev. Res.* **7**, L032038 (2025).
- [61] L. Ljung, *System identification: Theory for the user*, Prentice Hall information and system sciences series (Prentice Hall PTR, Upper Saddle River, New Jersey, 1999).
- [62] A. J. Majda and J. Harlim, Physics constrained nonlinear regression models for time series, *Nonlinearity* **26**, 201 (2013).
- [63] X. Wang, J. Feng, Q. Liu, Y. Li, and Y. Xu, Neural network-based parameter estimation of stochastic differential equations driven by Lévy noise, *Physica A* **606**, 128146 (2022).
- [64] T. Bódai, Enhanced El Niño predictability from climate mode interactions, *npj Clim. Atmos. Sci.* **8**, 278 (2025).
- [65] H.-J. Park, S.-I. An, J.-H. Park, and C. Liu, Unraveling the interactions between tropical Indo-Pacific climate modes using a simple model framework, *npj Clim. Atmos. Sci.* **8**, 336 (2025).
- [66] M. F. Stuecker, S. Zhao, A. Timmermann, R. Ghosh, T. Semmler, S.-S. Lee, J.-Y. Moon, F.-F. Jin, and T. Jung, Global climate mode resonance due to rapidly intensifying El Niño-Southern Oscillation, *Nat. Commun.* **16**, 9013 (2025).
- [67] S. Zhao, F.-F. Jin, M. F. Stuecker, P. R. Thompson, J.-S. Kug, M. J. McPhaden, M. A. Cane, A. T. Wittenberg, and W. Cai, Explainable El Niño predictability from climate mode interactions, *Nature (London)* **630**, 891 (2024).
- [68] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, *Proc. Natl. Acad. Sci. USA* **113**, 3932 (2016).
- [69] L. Zhang and H. Schaeffer, On the convergence of the SINDy algorithm, *Multiscale Model. Simul.* **17**, 948 (2019).
- [70] M. Anvari, H. Marasi, and H. Kheiri, Implicit Runge-Kutta based sparse identification of governing equations in biologically motivated systems, *Sci. Rep.* **15**, 32286 (2025).
- [71] K. Champion, B. Lusch, J. N. Kutz, and S. L. Brunton, Data-driven discovery of coordinates and governing equations, *Proc. Natl. Acad. Sci. USA* **116**, 22445 (2019).
- [72] U. Fasel, J. N. Kutz, B. W. Brunton, and S. L. Brunton, Ensemble-SINDy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control, *Proc. R. Soc. A* **478**, 20210904 (2022).
- [73] Y. Jiang and J. Sun, Evo-SINDy: Universal discovery of partial differential equations using cooperative evolutionary computation, in *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '25 (Association for Computing Machinery, New York, NY, 2025), pp. 791–799.
- [74] N. M. Mangan, T. Askham, S. L. Brunton, J. N. Kutz, and J. L. Proctor, Model selection for hybrid dynamical systems via sparse regression, *Proc. R. Soc. A* **475**, 20180534 (2019).
- [75] K. Champion, P. Zheng, A. Y. Aravkin, S. L. Brunton, and J. N. Kutz, A unified sparse optimization framework to learn parsimonious physics-informed models from data, *IEEE Access* **8**, 169259 (2020).
- [76] X. Wu, M. McDermott, and A. L. MacLean, Data-driven model discovery and model selection for noisy biological systems, *PLoS Comput. Biol.* **21**, e1012762 (2025).
- [77] P. Zheng, T. Askham, S. L. Brunton, J. N. Kutz, and A. Y. Aravkin, A unified framework for sparse relaxed regularized regression: SR3, *IEEE Access* **7**, 1404 (2019).
- [78] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/j3l8-6j37> for a discussion of biases introduced by conventional white-noise-based parameter estimation techniques, demonstration code for the ideal example, and additional numerical details and results for the climate prediction experiments.
- [79] G. Gradziuk, G. Torregrosa, and C. P. Broedersz, Irreversibility in linear systems with colored noise, *Phys. Rev. E* **105**, 024118 (2022).
- [80] J. Lien, Y.-N. Kuo, H. Ando, and S. Kido, On cyclostationary linear inverse models: A mathematical insight and implication, *Front. Complex Syst.* **3**, 1563687 (2025).
- [81] T.-H. Kim and H. White, On more robust estimation of skewness and kurtosis, *Finance Res. Lett.* **1**, 56 (2004).
- [82] N. Simon and R. Tibshirani, Standardization and the group LASSO penalty, *Stat. Sin.* **22**, 983 (2012).
- [83] R. Tibshirani, Regression shrinkage and selection via the LASSO, *J. R. Stat. Soc. B* **58**, 267 (1996).
- [84] M. Yuan and Y. Lin, Model selection and estimation in regression with grouped variables, *J. R. Stat. Soc. B* **68**, 49 (2006).
- [85] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imag. Sci.* **2**, 183 (2009).
- [86] D. Kim and J. A. Fessler, Another look at the fast iterative shrinkage/thresholding algorithm (FISTA), *SIAM J. Optim.* **28**, 223 (2018).
- [87] L. Freijeiro-González, M. Febrero-Bande, and W. González-Manteiga, A critical review of LASSO and its derivatives for variable selection under dependence among covariates, *Int. Stat. Rev.* **90**, 118 (2022).

- [88] A. Belloni and V. Chernozhukov, Least squares after model selection in high-dimensional sparse models, *Bernoulli* **19**, 521 (2013).
- [89] C. Martinez-Villalobos, B. Dewitte, R. D. Garreaud, and L. Loyola, Extreme coastal El Niño events are tightly linked to the development of the Pacific Meridional Modes, *npj Clim. Atmos. Sci.* **7**, 123 (2024).
- [90] H.-C. Chen and F.-F. Jin, Fundamental behavior of ENSO phase locking, *J. Clim.* **33**, 1953 (2020).
- [91] I. Richter, S.-P. Xie, Y. Morioka, T. Doi, B. Taguchi, and S. Behera, Phase locking of equatorial Atlantic variability through the seasonal migration of the ITCZ, *Clim. Dyn.* **48**, 3615 (2017).
- [92] M. F. Stuecker, The climate variability trio: stochastic fluctuations, El Niño, and the seasonal cycle, *Geosci. Lett.* **10**, 51 (2023).
- [93] E. Tziperman, S. E. Zebiak, and M. A. Cane, Mechanisms of Seasonal—ENSO Interaction, *J. Atmos. Sci.* **54**, 61 (1997).
- [94] G. Lancaster, D. Iatsenko, A. Pidde, V. Ticcinelli, and A. Stefanovska, Surrogate data for hypothesis testing of physical systems, *Phys. Rep.* **748**, 1 (2018).
- [95] G. Burgers, F.-F. Jin, and G. J. van Oldenborgh, The simplest ENSO recharge oscillator, *Geophys. Res. Lett.* **32**, 2005GL022951 (2005).
- [96] E. Exarchou, P. Ortega, B. Rodríguez-Fonseca, T. Losada, I. Polo, and C. Prodhomme, Impact of equatorial Atlantic variability on ENSO predictive skill, *Nat. Commun.* **12**, 1612 (2021).
- [97] T. Li, Y. Tang, T. Lian, and A. Hu, Quantifying the relative contributions of the global oceans to ENSO predictability with deep learning, *Geophys. Res. Lett.* **51**, e2023GL106584 (2024).
- [98] J.-J. Luo, R. Zhang, S. K. Behera, Y. Masumoto, F.-F. Jin, R. Lukas, and T. Yamagata, Interaction between El Niño and extreme Indian Ocean dipole, *J. Clim.* **23**, 726 (2010).
- [99] X. Liang, Z. Hu, and X. You, ENSONet: A data-driven ENSO forecasting model with concise spatial location learning parameters and temporal embedding, *Clim. Dyn.* **62**, 4081 (2024).
- [100] Q. Chen, Y. Cui, G. Hong, K. Ashok, Y. Pu, X. Zheng, X. Zhang, W. Zhong, P. Zhan, and Z. Wang, Toward long-range ENSO prediction with an explainable deep learning model, *npj Clim. Atmos. Sci.* **8**, 259 (2025).
- [101] H. Wang, S. Hu, and X. Li, An interpretable deep learning ENSO forecasting model, *Ocean-Land-Atmos. Res.* **2**, 0012 (2023).
- [102] H. C. Nnamchi and M. Latif, Predictable equatorial Atlantic variability from atmospheric convection-ocean coupling, *npj Clim. Atmos. Sci.* **8**, 149 (2025).
- [103] L. Shi, H. H. Hendon, O. Alves, J.-J. Luo, M. Balmaseda, and D. Anderson, How predictable is the Indian Ocean Dipole? *Mon. Weather Rev.* **140**, 3867 (2012).
- [104] T. N. Stockdale, M. A. Balmaseda, and A. Vidard, Tropical Atlantic SST prediction with coupled ocean-atmosphere GCMs, *J. Clim.* **19**, 6047 (2006).
- [105] R. Wang, L. Chen, T. Li, and J.-J. Luo, Atlantic Niño/Niña Prediction skills in NMME models, *Atmosphere* **12**, 803 (2021).
- [106] Y. Wu, Y. Ding, Y. Tang, and D. Yang, Impact of cold tongue bias on Indian Ocean Dipole prediction skills, *Geophys. Res. Lett.* **52**, e2025GL114690 (2025).
- [107] G. R. Foltz and M. J. McPhaden, Interaction between the Atlantic meridional and Niño modes, *Geophys. Res. Lett.* **37**, 2010GL044001 (2010).
- [108] M. Martín-Rey, I. Polo, B. Rodríguez-Fonseca, T. Losada, and A. Lazar, Is there evidence of changes in tropical Atlantic variability modes under AMO phases in the observational record? *J. Clim.* **31**, 515 (2018).
- [109] H. C. Nnamchi, J. Li, and R. N. C. Anyadike, Does a dipole mode really exist in the South Atlantic Ocean? *J. Geophys. Res.: Atmos.* **116**, D15104 (2011).
- [110] I. Richter, S. Kido, T. Tozuka, Y. Kosaka, H. Tokinaga, and P. Chang, Revisiting the inconsistent influence of El Niño—Southern Oscillation on the equatorial Atlantic, *J. Clim.* **38**, 481 (2025).
- [111] D. S. Harnos, J.-K. E. Schemm, H. Wang, and C. A. Finan, NMME-based hybrid prediction of Atlantic hurricane season activity, *Clim. Dyn.* **53**, 7267 (2019).
- [112] X. Song, Y. Tang, T. Liu, and X. Li, Predictability of Indian Ocean Dipole over 138 years using a CESM ensemble-prediction system, *J. Geophys. Res.: Oceans* **127**, e2021JC018210 (2022).
- [113] Y. Tao, C. Qiu, D. Wang, M. Li, and G. Zhang, Indian Ocean Dipole (IOD) forecasts based on convolutional neural network with sea level pressure precursor, *Environ. Res. Lett.* **19**, 104045 (2024).
- [114] N. S. Keenlyside and M. Latif, Understanding equatorial Atlantic interannual variability, *J. Clim.* **20**, 131 (2007).
- [115] J. F. Lübbecke and M. J. McPhaden, A comparative stability analysis of Atlantic and Pacific Niño modes, *J. Clim.* **26**, 5965 (2013).
- [116] M. Martín-Rey, B. Rodríguez-Fonseca, I. Polo, and F. Kucharski, On the Atlantic–Pacific Niños connection: A multidecadal modulated mode, *Clim. Dyn.* **43**, 3163 (2014).
- [117] H.-J. Park, S.-I. An, J.-H. Park, M. F. Stuecker, C. Liu, and S.-W. Yeh, Local feedback and ENSO govern decadal changes in variability and seasonal synchronization of the Indian Ocean Dipole, *Commun. Earth Environ.* **5**, 357 (2024).
- [118] I. Richter, S. K. Behera, Y. Masumoto, B. Taguchi, H. Sasaki, and T. Yamagata, Multiple causes of interannual sea surface temperature variability in the equatorial Atlantic Ocean, *Nat. Geosci.* **6**, 43 (2013).
- [119] I. Richter, T. Doi, S. K. Behera, and N. Keenlyside, On the link between mean state biases and prediction skill in the tropics: An atmospheric perspective, *Clim. Dyn.* **50**, 3355 (2018).
- [120] H. Tokinaga, I. Richter, and Y. Kosaka, ENSO influence on the Atlantic Niño, revisited: Multi-year versus single-year ENSO events, *J. Clim.* **32**, 4585 (2019).
- [121] A. Biastoch, J. V. Durgadoo, A. K. Morrison, E. van Sebille, W. Weijer, and S. M. Griffies, Atlantic multi-decadal oscillation covaries with Agulhas leakage, *Nat. Commun.* **6**, 10082 (2015).
- [122] J. F. Lübbecke, J. V. Durgadoo, and A. Biastoch, Contribution of increased Agulhas leakage to tropical Atlantic warming, *J. Clim.* **28**, 9697 (2015).
- [123] L. Stramma and J. R. E. Lutjeharms, The flow field of the subtropical gyre of the South Indian Ocean, *J. Geophys. Res.: Oceans* **102**, 5513 (1997).
- [124] K. Rodgers, S. Lee, N. Rosenbloom, A. Timmermann, G. Danabasoglu, C. Deser, J. Edwards, J. Kim, I. Simpson,

- K. Stein, M. Stuecker, R. Yamaguchi, T. Bódai, E. Chung, L. Huang, W. Kim, J. Lamarque, D. Lombardozzi, W. Wieder, and S. Yeager, Ubiquity of human-induced changes in climate variability, *Earth Syst. Dyn.* **12**, 1393 (2021).
- [125] Y.-Y. Kim, J.-Y. Lee, Y. Chikamoto, A. Timmermann, S.-S. Lee, E. Y. Kwon, W. Park, N. A. Hasan, I. Bethke, F. Fransner, A. Karwat, A. R. Subrahmanian, and C. L. E. Franzke, Robust estimates of Earth system predictability of the first kind using the CESM2 multiyear prediction system (CESM2-MP), *Bull. Am. Meteorol. Soc.* **106**, E2087 (2025).
- [126] W. W. Hay, The accelerating rate of global change, *Rend. Lincei* **25**, 29 (2014).
- [127] R. Hugonnet, R. McNabb, E. Berthier, B. Menounos, C. Nuth, L. Girod, D. Farinotti, M. Huss, I. Dussaillant, F. Brun, and A. Kääb, Accelerated global glacier mass loss in the early twenty-first century, *Nature (London)* **592**, 726 (2021).
- [128] S. J. Smith, J. Edmonds, C. A. Hartin, A. Mundra, and K. Calvin, Near-term acceleration in the rate of temperature change, *Nat. Clim. Change* **5**, 333 (2015).
- [129] H. Shiogama, M. Abe, and H. Tatebe, MIROC MIROC6 model output prepared for CMIP6 ScenarioMIP ssp585 (2019), accessed: 13 November 2025, <https://www.wdc-climate.de/ui/cmip6?input=CMIP6.ScenarioMIP.MIROC.MIROC6.ssp585>.
- [130] N. C. Swart, J. N. Cole, V. V. Kharin, M. Lazare, J. F. Scinocca, N. P. Gillett, J. Anstey, V. Arora, J. R. Christian, Y. Jiao, W. G. Lee, F. Majaess, O. A. Saenko, C. Seiler, C. Seinen, A. Shao, L. Solheim, K. von Salzen, D. Yang, B. Winter, *et al.*, CCCma CanESM5 model output prepared for CMIP6 ScenarioMIP ssp585 (2019), accessed: 13 November 2025, <https://www.wdc-climate.de/ui/cmip6?input=CMIP6.ScenarioMIP.CCCma.CanESM5.ssp585>.
- [131] K.-H. Wieners, M. Giorgetta, J. Jungclaus, C. Reick, M. Esch, M. Bittner, V. Gayler, H. Haak, P. de Vrese, T. Raddatz *et al.*, MPI-M MPI-ESM1.2-LR model output prepared for CMIP6 ScenarioMIP ssp585 (2019), accessed: 13 November 2025, <https://www.wdc-climate.de/ui/cmip6?input=CMIP6.ScenarioMIP.MPI-M.MPI-ESM1-2-LR.ssp585>.
- [132] T. Ziehn, M. Chamberlain, A. Lenton, R. Law, R. Bodman, M. Dix, Y. Wang, P. Dobrohotoff, J. Srbinovsky, L. Stevens, P. Vohralik, C. Mackallah, A. Sullivan, S. O’Farrell, and K. Druken, CSIRO ACCESS-ESM1.5 model output prepared for CMIP6 ScenarioMIP ssp585 (2019), accessed: 13 November 2025, <https://www.wdc-climate.de/ui/cmip6?input=CMIP6.ScenarioMIP.CSIRO.ACCESS-ESM1-5.ssp585>.
- [133] S. Zhao, Extended nonlinear recharge oscillator (xro) model for “explainable El Niño predictability from climate mode interactions”, Zenodo (2024), doi:10.5281/zenodo.10681114.
- [134] P. Jung and H. Risken, Motion in a double-well potential with additive colored Gaussian noise, *Z. Phys. B: Condens. Matter* **61**, 367 (1985).
- [135] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, 2nd ed., Springer Series in Synergetics (Springer, Berlin, 1989), Vol. 18, pp. xiv+472.