

Non-Gaussian stochastic dynamical model for the El Niño southern oscillation

L. T. Giorgini^{1,*}, W. Moon^{2,†}, N. Chen^{3,‡} and J. S. Wettlaufer^{4,§}¹*Nordita, Royal Institute of Technology and Stockholm University, Stockholm 106 91, Sweden*²*Department of Mathematics, Stockholm University 106 91 Stockholm, Sweden*³*Department of Mathematics, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA*⁴*Departments of Earth & Planetary Sciences, Mathematics and Physics, Yale University, New Haven, Connecticut 06520, USA*

(Received 1 December 2020; revised 20 May 2021; accepted 10 May 2022; published 27 June 2022)

A nonautonomous stochastic dynamical model approach is developed to describe the seasonal to interannual variability of the El Niño southern oscillation (ENSO). We determine the model coefficients by systematic statistical estimations using partial observations involving only sea surface temperature data. Our approach reproduces the observed seasonal phase locking and its uncertainty, as well as the highly non-Gaussian statistics of ENSO. Finally, we recover the intermittent time series of the hidden processes, including the thermocline depth and the wind bursts.

DOI: [10.1103/PhysRevResearch.4.L022065](https://doi.org/10.1103/PhysRevResearch.4.L022065)

Air-sea interactions in the tropical Pacific drive the largest interannual variability process in climate called the El Niño southern oscillation (ENSO), with an influence that reaches the higher latitudes via atmospheric and oceanic teleconnections [1,2]. Due to its spatiotemporal impacts, estimating the seasonal to interannual state of ENSO is essential for predicting a wide range of regional and global climate phenomena [3,4]. However, the complexity of ENSO, involving as it does stochastic forcing from atmospheric transients (e.g., Ref. [5]) and nonlinear air-sea interactions (e.g., Ref. [6]), poses great challenges in both modeling and prediction.

There are two traditional methods of modeling ENSO (see, e.g., Ref. [7] for a review). One is to use the state-of-art coupled general circulation models (GCMs), which solve the governing conservation (energy, mass, momentum) equations for the rotating planet with many sub-grid-scale parametrizations of unresolved small-scale processes. The other is to build low-order statistical models, such as linear inverse models (LIMs; [8,9]), which extract the dynamical properties of the system from observed statistics. The former provide better representations of the spatiotemporal patterns and air-sea interactions from intraseasonal to interannual timescales [10] of the coupled ocean-atmosphere system in the tropical Pacific, where ENSO is an emergent phenomenon. The latter, commonly designed to describe and predict particular ENSO indices, characterize the large-scale features of

the system and are much cheaper computationally than GCMs [11]. Nonetheless, despite great progress in developing these methods, there remain limitations to analysis and prediction. On the one hand, GCMs incur errors from random initial conditions generated from an incorrect state estimation of unobserved processes and from poor representation of physical mechanisms and parametrizations [12,13]. On the other hand, although the statistical models based on multivariate regressions may provide information about the phase of ENSO, they lack a physical foundation (e.g., Ref. [7]). One direction for progress is to utilize a hybrid strategy that simultaneously maximizes the advantages of both physically oriented and statistical models [14–16].

The basic theoretical understanding of the tropical air-sea interactions underlying ENSO [17,18] has led to a hierarchy of simple models of the main processes that control the sea surface temperature (SST) in the tropical ocean [19]. A prominent approach is that of Cane and Zebiak (CZ) [20], which unifies different coupled atmosphere-ocean models of the tropical Pacific. By applying a two-strip approximation to the forced shallow water equations of the CZ model, Jin derived a conceptual recharge-discharge oscillator model [21,22], which relates variations of the eastern Pacific SST to western Pacific thermocline depth anomalies. Despite it being a two-dimensional linear model, it (and its generalizations) can capture interannual variability, seasonal variability with time-periodic coefficients [23,24], and a range of stochastic forcing representing weather and intraseasonal processes, such as westerly wind bursts (WWBs) [25].

Here we describe a two-stage stochastic model approach that captures the large-scale dynamical and statistical characteristics of ENSO. Stochastic recharge-discharge oscillator models have used observations to recover statistics on interannual timescales [26,27] and have incorporated the seasonal cycle of the Bjerknes feedback to treat ENSO “phase-locking” and the “spring predictability barrier” [23,24,28]. However, while our approach builds upon the recharge-discharge model,

*ludovico.giorgini@su.se

†woosok.moon@su.se

‡chennan@math.wisc.edu

§john.wettlaufer@yale.edu

we incorporate a slowly varying low-order deterministic component that systematically treats time-varying coefficients and multiplicative noise to accurately describe the central intraseasonal to interannual features of ENSO. We clearly depart from multivariate regression statistical models that require observations of all state variables. Our framework reproduces ENSO's observed seasonal phase locking, uncertainty, and highly non-Gaussian statistics. In particular, we incorporate a subset of observations from which we can infer unobserved quantities and quantify their uncertainties using an efficient and exact data assimilation scheme, and hence we can accurately recover the difficult to observe thermocline depth and wind bursts. By systematically determining time-dependent model coefficients, we precisely simulate the seasonal to interannual ENSO statistics. Our nonautonomous two dimensional stochastic model allows us to estimate the key dynamical quantities for which there are no direct observations, but in a manner more accurate than previous approaches. In particular, using only the SST index data, we (a) recover the time-periodic behavior of the coupling term between the SST and the thermocline depth, (b) estimate the drift and diffusion terms in the thermocline depth equation, and hence (c) accurately reconstruct the phase locking of ENSO without the need to invoke multiplicative colored noise in the SST equation as in Ref. [24]. Finally, using solely the SST data we can include an additional process—the westerly wind bursts which explains the highly non-Gaussian ENSO statistics.

We begin with the following coupled model,

$$\begin{aligned}\dot{x}(t) &= a(t)x(t) + \omega(t)h(t) + N(t)\xi_x(t) \text{ and} \\ \dot{h}(t) &= -\omega(t)x(t) + \lambda h(t) + \sigma \xi_h(t),\end{aligned}\quad (1)$$

in which $x(t)$ is the averaged SST anomaly in the equatorial central-eastern Pacific (the so-called Niño 3.4 region), and $h(t)$ is the thermocline depth, which is a surrogate of the heat content, averaged over the western tropical Pacific. The time-dependent function, $a(t)$, represents the seasonal evolution of the Bjerknes feedback and the noise amplitude, $N(t)$, captures the role of the relatively short timescale forcing of the SST anomaly $x(t)$. Both $\xi_x(t)$ and $\xi_h(t)$ are independent Gaussian white noise processes. Finally, the other time-dependent function, $\omega(t)$, controls the coupling of tropical atmosphere and the upper ocean, which is one of the key parameters that determines the quasioscillatory behavior of the dominant ENSO modes, and in traditional low-order models is treated as a constant [23]. Here, however, treating it as time-dependent facilitates a more accurate description of the intraseasonal and interannual statistics.

We emphasize the role of the seasonality of background fields in the tropical Pacific by making all the time-dependent coefficients in (1) periodic on the annual cycle. We determine the functions $a(t)$ and $N(t)$ using the monthly statistics of the Niño 3.4 index [29]. The coefficient $a(t)$ is related to the relationship between the monthly variance $\langle x^2(t) \rangle$ and the covariance between two adjacent months $\langle x(t)x(t + \Delta t) \rangle$. We use the derived $a(t)$ to estimate $N(t)$ from the statistics of the derived data: $y = x(t + \Delta t) - x(t) - a(t)x(t)\Delta t$. It has been shown previously that $a(t)$ and $N(t)$ can be accurately estimated independent of the other coefficients (see the Supplemental Material in Ref. [29]). The coupling

function, $\omega(t)$, and the constants λ and σ are determined by an expectation-maximization algorithm [30–32], which is a useful and efficient statistical estimation method involving incomplete data. We note that only the Niño 3.4 index data is used as the partial observation for model calibration, whereas the thermocline depth, $h(t)$, is assumed to be unknown since it is not directly available from satellite observations. Therefore, as detailed in the Supplemental Material, we infer $h(t)$ together with unknown coefficients $\omega(t)$, λ , and σ [33].

We evaluate the quality of the model by comparing the model statistics with those of the observed Niño 3.4 index in Fig. 1. A qualitative comparison between a random realization of the model and the observed data show comparable variability in Fig. 1(a) and quantitative comparisons of the power spectra, probability density functions (PDFs), and the autocorrelation functions (ACFs) of the SST anomaly shown in Figs. 1(b)–1(d) demonstrate that the model is able to accurately simulate the interannual quasi-oscillatory behavior of ENSO. Moreover, as seen in Fig. 1(e), the model reproduces monthly standard deviations of the Niño 3.4 index data with high precision. Therefore, two important seasonal features of ENSO, phase locking to the seasonal cycle [43] and the spring predictability barrier [24], are accurately captured in this two-dimensional model. Phase locking is related to the activation of the positive Bjerknes feedback from summer to early winter, which leads to the concentration of abnormal ENSO events during November and December, as is clear from the maximum of the standard deviation near the end of the year. This feature is clearly seen from the model trajectory [Fig. 1(a)], wherein the peaks of nearly all of the major ENSO events occur in the boreal winter, consistent with the observations. The spring predictability barrier is related to the increase in the noise magnitude, $N(t)$, during the spring, which increases model uncertainties during the boreal summer thereby causing the loss of predictability [29]. Indeed, this is reflected by the structural similarity between the model and the observations in the persistence forecast diagrams shown in Figs. 1(f) and 1(g).

Despite its simplicity, the model system of Eqs. (1) recovers the monthly standard deviation and ACFs of the Niño 3.4 index with higher accuracy than coupled GCMs [44], linear and nonlinear statistical models [45], and simpler recharge-discharge based approaches [23]. Our systematic statistical estimation of coefficients enable this stochastic model to regenerate the statistics of the Niño 3.4 index. In particular, as shown in Fig. 1(d), the precise representation of the ACFs beyond 2 years suggests that the overall dynamical features of the thermocline depth, $h(t)$, are well captured in this model. Variations in $h(t)$, and thus of the equatorial warm water volume (WWV) in the western tropical Pacific, influence the SST through vertical advection of temperature anomalies by mean upwelling, a process known as the “thermocline feedback.” Thus, the thermocline depth acts as a precursor for the interannual prediction of SST anomalies [46].

Although the model described above successfully captures the seasonal variability and the interannual large-scale dynamical features of ENSO, it produces only Gaussian PDFs, whereas the observed statistics are highly non-Gaussian. Importantly, a model must include atmospheric wind burst forcing, which plays a critical role in generating the

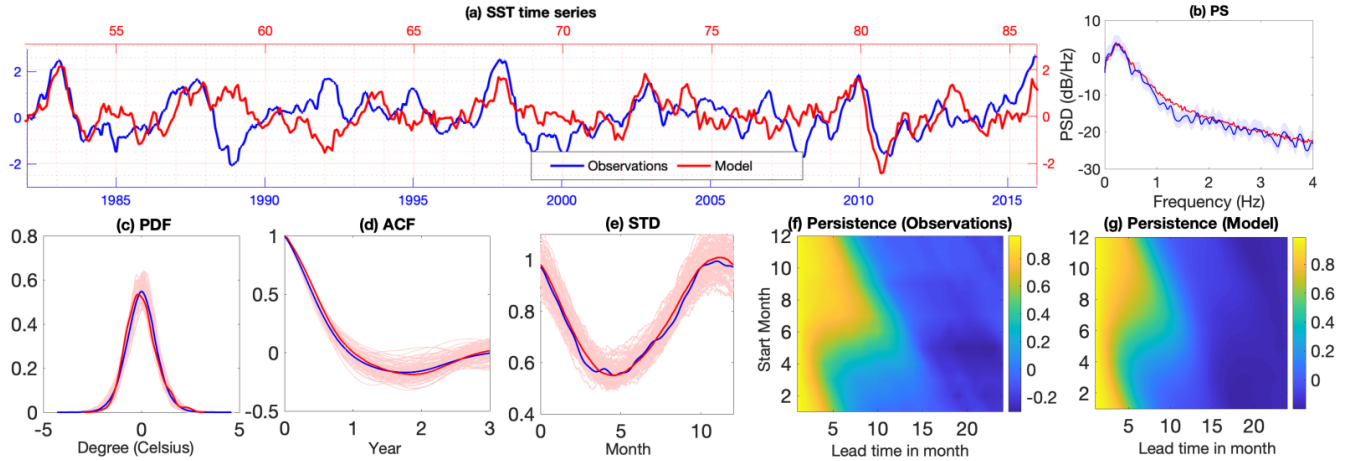


FIG. 1. Comparison of the observed SST (blue; bottom x axis) with the model Eq. (1) (red; top x axis). The SST time series are shown in panel (a), where the model result is one of the random realizations from Eq. (1) that shares the same length as the observational data from 1983 to 2016. The comparisons of the power spectra, PDFs and ACFs are shown in panels (b), (c), and (d). The SST standard deviation as a function of calendar month is shown in panel (e). The blue shaded area in panel (b) are the 95% upper and lower confidence intervals obtained from the observed data. The thick red curve in panels (b)–(e) was obtained from a 10 000-year run of the model, while the thin pink curves were derived by running the model 100 times for the same time duration (146 years) as the observed data set and they give the uncertainty associated to the thick red curves. In panels (f) and (g), the diagram of the persistence forecast from the observational data and that from the model can be compared.

extreme ENSO events responsible for non-Gaussianity [25,47]. Therefore, we include wind bursts in the stochastic model of Eqs. (1) as follows,

$$\begin{aligned}\dot{x}(t) &= a(t)x(t) + \omega(t)h(t) + \alpha_1\tau(t) + N(t)\xi_x(t), \\ \dot{h}(t) &= -\omega(t)x(t) + \lambda h(t) + \alpha_2\tau(t) + \sigma\xi_h(t), \text{ and} \\ \dot{\tau}(t) &= d_\tau\tau(t) + \rho(x)\xi_\tau(t),\end{aligned}\quad (2)$$

where we adopt a simple linear process with multiplicative noise to generate the stochastic wind bursts $\tau(t)$ that drive the SST and thermocline dynamics. The parameters in the wind burst equation are chosen to accurately reproduce the observed wind stress [Figs. 2(d)–2(f)], while the coefficients of Eqs. (1) were slightly modified to take into account the additional process in the model. Despite the non-Gaussianity of the model, the thermocline depth and the wind bursts remain as Gaussian processes conditioned on the observed SST. Moreover, by using the partial observations that involve only the Niño SST data, we have closed analytic formulas for solving the conditional distribution of the unobserved thermocline depth and the wind bursts [48],

$$p(h(t), \tau(t)|x(s), s \leq t) \sim \mathcal{N}(\mu_f(t), R_f(t)), \quad (3)$$

where $\mu_f(t)$ and $R_f(t)$ are the so-called filter mean and filter covariance in data assimilation.

The filter estimate (3) of the non-Gaussian model (2) can be used to carry out an online time series reconstruction of the unobserved variables and analyze the associated uncertainty, which can also serve as the forecast initialization. Note that observational data for the equatorial heat content [the integrated WWV above the 20°C isotherm between 5N and 5S, 120E to 80W [46], which is equivalent to the thermocline

depth $h(t)$ in (2)] is only available since 1983,¹ whereas SST data is available from the preindustrial era. Therefore, the available SST data can be incorporated into the model (2) to reconstruct the thermocline data over a much longer period.

We first validate the accuracy of the recovered $h(t)$ and $\tau(t)$ in the period after 1983. Figure 2(a) shows the filter mean time series of the thermocline depth, $h(t)$ (red), from the non-Gaussian model (2), which compares favorably to the observational value (blue) with a pattern correlation that is nearly 0.9. Moreover, the small error between the observed and the recovered time series lies approximately within one standard deviation of the filter estimated uncertainty (pink-shaded region). In particular, as expected from the physical role of the WWBs, the extreme events are accurately recovered. Similar conclusions are reached for the wind burst time series [Fig. 2(e)]. In addition to a skillful state estimation, the reconstructed wind burst time series has the further influence of filtering out the independent white noise in observations. As shown in Figs. 2(b) and 2(f), the model PDFs capture those of the observations for both the thermocline depth and the wind bursts, particularly the significant non-Gaussian skewness and kurtosis. Moreover, similar accuracy is found in ACFs and the power spectra as shown Figs. 2(c), 2(d), 2(g), and 2(h). The success in recovering unobserved processes is due to the confluence of the model containing the essential physics and properly constrained stochastic forcing and coefficients.

The other approach commonly used to recover the thermocline depth is based on the empirical model reduction (EMR) algorithm [45,50], which extracts the signal of the thermocline depth from the residual of an LIM hindcast. This residual

¹The WWV values are derived from the temperature analyses of the Bureau National Operations Centre (BNOC) at the Australian Bureau of Meteorology [49].

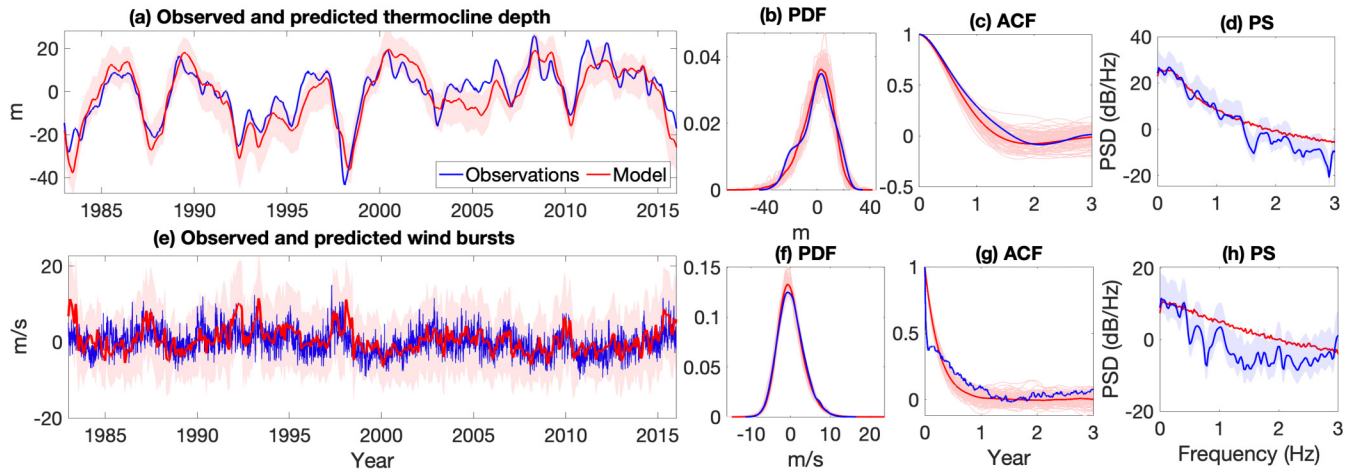


FIG. 2. Comparison of the observational value (blue) and the recovered time series from the filter estimate (red) using Eq. (3), based on the non-Gaussian model (2) for the thermocline depth (a) and the wind bursts (e). The solid red curves are the filter mean while the shaded red areas show one standard deviation in the filter estimate (the square root of the filter variance). The comparison of the PDFs, the ACFs, and the power spectra between the observations and simulations from the non-Gaussian model (2) are shown in panels (b)–(d) for the thermocline depth and panels (f)–(h) for the wind bursts. As in Fig. 1, the shaded blue area in panels (d) and (h) are the 95% upper and lower confidence intervals obtained from the observations, and the thin pink curves in panels (b), (c), (f), and (g) are uncertainty estimates of the thick red curves.

contains both white noise and the discernible footprint of the subsurface dynamics that can be isolated and then used to improve the model forecast skill. However, estimating $h(t)$ from the EMR algorithm is considerably more computationally expensive than using the filter and/or smoother estimate constructed from our dynamical stochastic model. Indeed, in order to perform the LIM hindcast, a principal component analysis must be carried out on the SST data of the entire Pacific and hence a satisfactory estimation of $h(t)$ requires a very large number of principal components. In contrast, we estimate this hidden process using only the SST index data. Moreover, we obtain a superior estimation of $h(t)$; in the western tropical Pacific, we explain 78% of the total variance whereas the EMR algorithm explains only 50% [50].

As shown Fig. 2, the accurate nonlinear filter estimates of Eq. (3) provide an online estimation method of the thermocline depth $h(t)$ and the wind bursts $\tau(t)$ relevant for model

initialization. On the other hand, reconstructing historical data can be achieved by solving a slightly different conditional distribution

$$p(h(t), \tau(t)|x(s), 0 \leq s \leq T) \sim \mathcal{N}(\mu_s(t), R_s(t)), \quad (4)$$

where $t \in [0, T]$. This is the so-called smoother estimate. It differs from the filter in the sense that it exploits all the available information up to the current time. One desirable feature of the nonlinear smoother estimate in Eq. (4), associated with the non-Gaussian system in Eqs. (2), is that it can be solved via closed analytic formulas (see the Supplemental Material [33]). We use the smoother estimate to recover the historical realizations of the two hidden variables conditioned on the entire observational period of the Niño 3.4 index from 1870 to present. The results are shown in Fig. 3. Since the thermocline depth and the wind bursts are the most important precursors for predicting the state of ENSO, these recovered variables

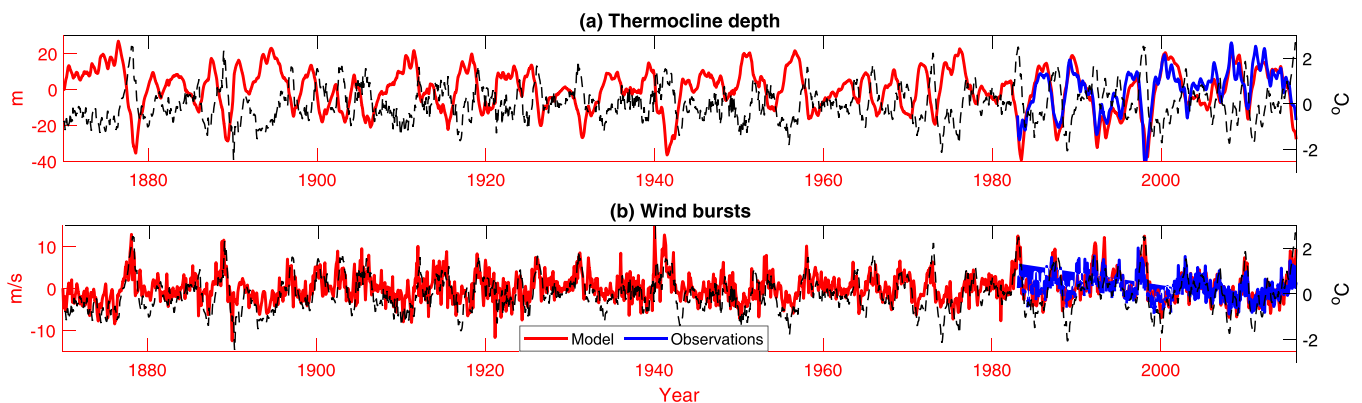


FIG. 3. The recovered time series of the hidden variables (red curves; left y axis) for (a) the thermocline depth and (b) the wind bursts, using the non-Gaussian model, Eqs. (2), with the partial observations involving only SST data (dashed black curves; right y axis) from 1870 to 2016. The method used here is the nonlinear smoother estimate, Eq. (4). The blue curves show the observational data from 1983 to 2016.

can be used to evaluate and improve prediction models and provide a baseline for improving parametrizations in coupled GCMs.

Understanding the complexity of climate dynamics presents a dilemma of completeness, between using solely statistical regression models lacking the essential physics, to global climate models, which have more variables and parametrizations than observables. Our approach presents a compromise through hybridization. We combine a slowly varying low-order dynamical system with an observationally faithful statistical representation of short-time processes. In consequence, we find a realistic description of ENSO as measured by the success in (i) reproducing the observed seasonal phase locking and its uncertainty, (ii) capturing the spring predictability barrier, (iii) simulating the observed highly non-Gaussian statistics, and (iv) accurately recovering the

intermittent time series of the hidden processes via a non-linear, but analytically solvable, data assimilation scheme. Furthermore, we suggest that extending the approach described here to examine climate variability on a wide range of time scales [2,51], wherein proxy data are available, will provide important insights into mechanisms and feedbacks in the climate system. Finally, given the generality of stochastic dynamical systems, it is hoped that our approach can be adopted not only for other problems in climate physics, but far more broadly in the physical sciences.

L.T.G., W.M., and J.S.W. gratefully acknowledge support from the Swedish Research Council Grant No. 638-2013-9243. The research of N.C. is partially funded by the Office of VCRGE at UW-Madison and by ONR N00014-21-1-2904. Nordita is partially supported by Nordforsk.

-
- [1] M. J. McPhaden, S. E. Zebiak, and M. H. Glantz, ENSO as an integrating concept in earth science, *Science* **314**, 1740 (2006).
 - [2] M. Ghil and V. Lucarini, The physics of climate variability and climate change, *Rev. Mod. Phys.* **92**, 035002 (2020).
 - [3] D. S. Battisti and E. S. Sarachik, Understanding and predicting ENSO, *Rev. Geophys.* **33**, 1367 (1995).
 - [4] D. Zhou, A. Gozolchiani, Y. Ashkenazy, and S. Havlin, Teleconnection Paths Via Climate Network Direct Link Detection, *Phys. Rev. Lett.* **115**, 268501 (2015).
 - [5] R. Kleeman and A. M. Moore, A theory for the limitation of ENSO predictability due to stochastic atmospheric transients, *J. Atmos. Sci.* **54**, 753 (1997).
 - [6] S. I. An and F. F. Jin, Nonlinearity and asymmetry of ENSO, *J. Clim.* **17**, 2399 (2004).
 - [7] Y. Tang, R.-H. Zhang, T. Liu, W. Duan, D. Yang, F. Zheng, H. Ren, T. Lian, C. Gao, D. Chen *et al.*, A theory for the limitation of ENSO predictability due to stochastic atmospheric transients, *Natl. Sci. Rev.* **5**, 826 (2018).
 - [8] C. Penland, Random forcing and forecasting using principal oscillation pattern analysis, *Mon. Weather Rev.* **117**, 2165 (1989).
 - [9] C. Penland and P. D. Sardeshmukh, The optimal growth of tropical sea surface temperature anomalies, *J. Climate* **8**, 1999 (1995).
 - [10] G. Jan van Oldenborgh, M. A. Balmaseda, L. Ferranti, T. N. Stockdale, and D. L. Anderson, Did the ECMWF seasonal forecast model outperform statistical ENSO forecast models over the last 15 years? *J. Clim.* **18**, 3240 (2005).
 - [11] D. Petrova, J. Ballester, S. J. Koopman, and X. Rodo, Multiyear statistical prediction of ENSO enhanced by the tropical Pacific observing system, *J. Clim.* **33**, 163 (2020).
 - [12] D. Chen and M. A. Cane, El Niño prediction and predictability, *J. Comput. Phys.* **227**, 3625 (2008).
 - [13] A. G. Barnston, Evolution of ENSO prediction over the past 40 years, in *Science and Technology Infusion Climate Bulletin*, 40th NOAA Annual Climate Diagnostics and Prediction Workshop (NOAA National Weather Service, MD, 2015), pp. 94–101.
 - [14] T. Geng, W. Cai, and L. Wu, Two types of ENSO varying in tandem facilitated by nonlinear atmospheric convection, *Geophys. Res. Lett.* **47**, e2020GL088784 (2020).
 - [15] S. An, S. Kim, and A. Timmermann, Fokker-Planck dynamics of the El Niño southern oscillation, *Sci. Rep.* **10**, 16282 (2020).
 - [16] M. Bianucci and R. Mannella, On the determination of the optimal parameters in the CAM model, *Chaos* **31**, 033113 (2021).
 - [17] M. A. Cane and S. E. Zebiak, A theory for El Niño and the southern oscillation, *Science* **228**, 1085 (1985).
 - [18] F. F. Jin and J. D. Neelin, Modes of interannual tropical ocean-atmosphere interaction: A unified view. Part I: Numerical results, *J. Atmos. Sci.* **50**, 3477 (1993).
 - [19] C. Wang, C. Deser, J. Y. Yu, P. Di Nezio, and A. Clement, in *Coral Reefs of the Eastern Tropical Pacific*, edited by P. Glynn, D. Manzello, and I. Enochs (Springer, Dordrecht, 2017), p. 85.
 - [20] S. E. Zebiak and M. A. Cane, A model El Niño southern oscillation, *Mon. Weather Rev.* **115**, 2262 (1987).
 - [21] F. F. Jin, An equatorial ocean recharge paradigm for ENSO. Part I: Conceptual model, *J. Atmos. Sci.* **54**, 811 (1997).
 - [22] F. F. Jin, An equatorial ocean recharge paradigm for ENSO. Part II: A stripped-down coupled model, *J. Atmos. Sci.* **54**, 830 (1997).
 - [23] K. Stein, N. Schneider, A. Timmermann, and F. F. Jin, Seasonal synchronization of ENSO events in a linear stochastic model, *J. Clim.* **23**, 5629 (2010).
 - [24] A. F. Levine and M. J. McPhaden, The annual cycle in ENSO growth rate as a cause of the spring predictability barrier, *Geophys. Res. Lett.* **42**, 5034 (2015).
 - [25] N. Chen and A. J. Majda, Simple stochastic dynamical models capturing the statistical diversity of El Niño southern oscillation, *Proc. Natl. Acad. Sci. USA* **114**, 1468 (2017).
 - [26] G. Burgers, The El Niño stochastic oscillator, *Clim. Dyn.* **15**, 521 (1999).
 - [27] G. Burgers, F. F. Jin, and G. J. Van Oldenborgh, The simplest ENSO recharge oscillator, *Geophys. Res. Lett.* **32**, L13706 (2005).
 - [28] K. Stein, A. Timmermann, and N. Schneider, Phase Synchronization of the El Niño Southern Oscillation with the Annual Cycle, *Phys. Rev. Lett.* **107**, 128501 (2011).
 - [29] W. Moon and J. S. Wettlaufer, A unified nonlinear stochastic time series analysis for climate science, *Sci. Rep.* **7**, 44228 (2017).

- [30] A. P. Dempster, N. M. Laird, and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *J. R. Stat. Soc., Ser. B* **39**, 1 (1977).
- [31] R. Sundberg, Maximum likelihood theory for incomplete data from an exponential family, *Scand. J. Stat.* **1**, 49 (1974).
- [32] R. Sundberg, An iterative method for solution of the likelihood equations for incomplete data from exponential families, *Commun. Stat. Sim. Comp.* **5**, 55 (1976).
- [33] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevResearch.4.L022065> for observations, assimilation and parameter estimation, which includes Refs. [34–42].
- [34] N. Chen, Learning nonlinear turbulent dynamics from partial observations via analytically solvable conditional statistics, *J. Comput. Phys.* **418**, 109635 (2020).
- [35] The WWV values are derived from the temperature analyses of the Bureau National Operations Centre (BNOC) at the Australian Bureau of Meteorology [49].
- [36] BNOC (Bureau National Operations Centre), Operational implementation of ACCESS-S1 seasonal prediction system, Operations Bulletin No. 120, 2018.
- [37] D. W. Behringer, M. Ji, and A. Leetmaa, An improved coupled model for ENSO prediction and implications for ocean initialization. Part I: The ocean data assimilation system, *Mon. Weather Rev.* **126**, 1013 (1998).
- [38] E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen *et al.*, The NCEP/NCAR 40-year reanalysis project, *Bull. Amer. Meteor. Soc.* **77**, 437 (1996).
- [39] R. E. Kalman and R. S. Bucy, New results in linear filtering and prediction theory, *J. Basic Eng.* **83**, 95 (1961).
- [40] N. Chen and A. Majda, Conditional Gaussian systems for multi-scale nonlinear stochastic systems: Prediction, state estimation, and uncertainty quantification, *Entropy* **20**, 509 (2018).
- [41] N. Chen and A. J. Majda, Efficient nonlinear optimal smoothing and sampling algorithms for complex turbulent nonlinear dynamical systems with partial observations, *J. Comput. Phys.* **410**, 109381 (2020).
- [42] N. Chen, X. Hou, Q. Li, and Y. Li, Understanding and predicting nonlinear turbulent dynamical systems with information theory, *Atmosphere* **10**, 248 (2019).
- [43] E. Tziperman, M. A. Cane, S. E. Zebiak, Y. Xue, and B. Blumenthal, Locking of El Niño’s peak time to the end of the calendar year in the delayed oscillator picture of ENSO, *J. Clim.* **11**, 2191 (1998).
- [44] H. Bellenger, E. Guilyardi, J. Leloup, M. Lengaigne, and J. Vialard, Locking of El Niño’s peak time to the end of the calendar year in the delayed oscillator picture of ENSO, *Clim. Dyn.* **42**, 1999 (2014).
- [45] D. Kondrashov, S. Kravtsov, A. W. Robertson, and M. Ghil, A hierarchy of data-based ENSO models, *J. Clim.* **18**, 4425 (2005).
- [46] C. S. Meinen and M. J. McPhaden, Observations of warm water volume changes in the equatorial Pacific and their relationship to El Niño and La Niña, *J. Clim.* **13**, 3551 (2000).
- [47] S. Thual, A. J. Majda, N. Chen, and S. N. Stechmann, Simple stochastic model for El Niño with westerly wind bursts, *Proc. Natl. Acad. Sci. USA* **113**, 10245 (2016).
- [48] A. J. Majda and N. Chen, Model error, information barriers, state estimation, and prediction in complex multiscale systems, *Entropy* **20**, 644 (2018).
- [49] N. R. Smith, The BMRC ocean thermal analysis system, *Aust. Met. Mag.* **44**, 93 (1995).
- [50] D. Chapman *et al.*, A vector autoregressive ENSO prediction model, *J. Clim.* **28**, 8511 (2015).
- [51] W. Moon, S. Agarwal, and J. S. Wettlaufer, Intrinsic Pink-Noise Multidecadal Global Climate Dynamics Mode, *Phys. Rev. Lett.* **121**, 108701 (2018).