Predicting stochastic systems by noise sampling, and application to the El Niño-Southern Oscillation

Mickaël David Chekroun[a,b,1], Dmitri Kondrashov[b], and Michael Ghil[a,b]

[a]Environmental Research and Teaching Institute, École Normale Supérieure, F-75230 Paris Cedex 05, France; and [b]Department of Atmospheric and Oceanic Sciences, and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90095

Edited by* Alexandre J. Chorin, University of California, Berkeley, CA, and approved May 25, 2011 (received for review October 21, 2010)

Interannual and interdecadal prediction are major challenges of climate dynamics. In this article we develop a prediction method for climate processes that exhibit low-frequency variability (LFV). The method constructs a nonlinear stochastic model from past observations and estimates a path of the "weather" noise that drives this model over previous finite-time windows. The method has two steps: (*i*) select noise samples—or "snippets"—from the past noise, which have forced the system during short-time intervals that resemble the LFV phase just preceding the currently observed state; and (*ii*) use these snippets to drive the system from the current state into the future. The method is placed in the framework of pathwise linear-response theory and is then applied to an El Niño–Southern Oscillation (ENSO) model derived by the empirical model reduction (EMR) methodology; this nonlinear model has 40 coupled, slow, and fast variables. The domain of validity of this forecasting procedure depends on the nature of the system's pathwise response; it is shown numerically that the ENSO model's response is linear on interannual time scales. As a result, the method's skill at a 6- to 16-month lead is highly competitive when compared with currently used dynamic and statistic prediction methods for the Niño-3 index and the global sea surface temperature field.

ENSO forecasting has a decade-long history and relies mainly on two classes of models: dynamical and statistical (1, 2). Still, a further distinction has to be made within the latter class: Some of the statistical models do not make any use of dynamical information, like Lorenz's method of analogues (3) and its followers (4–6), while others do use a dynamical model—previously fitted to the observations from the past—to drive the statistics in the future (2, 7, 8). Empirical stochastic models belong to this hybrid category, and linear versions of such models have been used in ENSO forecasting for two decades; see ref. 9 for a survey. More recently, Kravtsov et al. (10) have extended this approach to nonlinear models by developing an EMR methodology that can include quadratic nonlinearities as well as state-dependent noise that parameterizes small-scale effects, without assuming a priori scale separation (11).

The purpose of this paper is to show that, under suitable circumstances, a better understanding of the role of the fast processes, weather or noise, can help predict the slow ones—namely, the climate. To achieve this purpose, we proceed in two steps: (*i*) develop a special prediction methodology, called past noise forecasting (PNF), using EMR models; and (*ii*) provide a theoretical framework for applying the PNF method—or any other forecasting method based on perturbations of the noise—to other empirical stochastic models. Of late, probabilistic forecasts in weather and climate prediction have become fairly widespread: They are grounded in an estimation of the probability density function (PDF: 12–14).

We take here a distinct, pathwise approach instead, and will show that this approach is particularly well adapted to empirical stochastic models and to phenomena in which a considerable part of the variability exhibits some form of repetitive regularity (15). We focus mainly on the EMR models of ENSO introduced and studied by Kondrashov et al. (16).

The paper is organized as follows. We present in the next section the concept of *pathwise sensitivity*, which describes the response of the stochastic system at hand with respect to perturbations of the noise realization. If this pathwise response is linear, then the system's response to small perturbations of the noise may be small. In the subsequent section, we verify numerically that this linear-response property is fulfilled by the EMR model of ENSO used here. These two sections provide the theoretical framework for our PNF method, which is outlined in the fourth section.

The method's key idea consists of selecting appropriate noise samplings—or snippets—from the past of the noise realization derived during the EMR model's fitting procedure, and then using a subset of these snippets to drive the model's dynamics into the future. An effective subset of snippets is selected by relying on ENSO's well-known LFV modes (17) to key the selection to the analogous phases of these modes.

Essentially, the PNF method is a forecasting method based on perturbation techniques of the actual path of the future noise. The effects of changing the magnitude of these perturbations are analyzed by relying on the pathwise linear-response property. This property provides, in the fifth section, a clear theoretical explanation of the relatively good performance of the PNF method in the case at hand, as illustrated further by the statistical-significance tests in *SI Text*. The PNF approach is quite flexible, and further extensions are suggested in *Concluding Remarks*.

Predictability and Pathwise Sensitivity

Random forcing in a climate model may increase the robustness of its long-term behavior (18), but it adds another source of uncertainty to short- or medium-term prediction. In stochastic linear models of the form $d\mathbf{x} = \mathbf{A}\mathbf{x}dt + d\xi_t$ (9)—where $\mathbf{x}$ is the state vector, t is time, and ξ_t is a vector-valued Wiener process—sensitivity to the initial state may be observed in the short run. In the long run, however, because stationarity of the stochastic process requires the linear operator $\mathbf{A}$ to be exponentially stable, there is no such sensitivity: To the contrary, trajectories started from different initial states $\mathbf{x}_0$ and driven by the same noise realization $d\xi_t(\omega)$ are completely synchronized. Such *pathwise synchronization* characterizes all stochastic dynamical systems having only negative Lyapunov exponents (18, 19). We shall see that a weaker form of synchronization, namely so-called on-off synchronization, characterizes certain systems that do have a positive exponent (19).

Assuming the driving noise $d\xi_t$ is white, Ornstein–Uhlenbeck theory can be used to estimate $\mathbf{A}$ and the covariance of $d\xi_t$. The Green's function of the process can then be used for predicting

Author contributions: M.D.C., D.K., and M.G. designed research; M.D.C. and D.K. performed research; M.D.C. and D.K. contributed new analytic tools; M.D.C., D.K., and M.G. analyzed data; and M.D.C., D.K., and M.G. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence should be addressed. E-mail: mchekroun@atmos.ucla.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1015753108/-/DCSupplemental.

the expected evolution of the state $\mathbf{x}(t)$ (9). Still, there is no information on the future of the sample path $t \mapsto d\xi_t(\omega)$ that has been fitted from the observations over a time window $(0,t^*)$; where $\omega \in \Omega$ marks the sampled realization, and Ω is the appropriate probability space for the sample paths.

For a nonlinear stochastic model that exhibits a positive Lyapunov exponent when the noise is turned off, we are thus faced—once the noise is turned on and besides the sensitivity to the initial state—with the additional uncertainty in the realization of the noise. Hence the forecasting problem becomes, at first sight, even more difficult than in the purely deterministic case. In the remainder of this section, we show that for certain chaotic systems—which are of geophysical interest, as seen in the next section—the noise can help, rather than hinder, the prediction of the state.

We consider here a nonlinear stochastic model of the form,

$$d\mathbf{x} = \mathbf{f}(\mathbf{x},t)dt + \mathbf{g}(\mathbf{x},t)d\xi_t; \qquad [1]$$

the model is used to simulate a multivariate time series over a time interval $(0,t^*)$ and to predict its future evolution over the interval $(t^*, t^* + T)$. In Eq. **1** and the sequel, the noise $d\xi_t$ is not necessarily white and hereafter will be denoted by ξ_t. One assumes that the deterministic functions $\mathbf{f}$ and $\mathbf{g}$ have been estimated, by EMR or in some other way. Unless one knows, though, how to approximate the sample path of the noise $t \mapsto \xi_t(\omega)$ for $t \in (t^*, t^* + T)$—with ω the unknown but fixed realization—one has to rely on predictions of the system's PDF (12–14).

To exploit knowledge of past noise, as we propose to do here, requires first an estimate of the model's response to a "perturbation" of ω. Consider an *observable* ψ of the system governed by Eq. **1**—i.e., a real-valued continuous function $X \to \mathbb{R}$, where X is the system's phase-space. For definiteness and clarity, let $\xi_t(\omega')$ be a perturbation of ξ_t over $(0,t^*)$, with ω' a different realization and $0 < \epsilon \ll 1$; the perturbed path of the noise is $\chi_t^\epsilon(\omega') = \xi_t(\omega) + \epsilon\xi_t(\omega')$, although much more general perturbations can be considered. Denote by $\Phi(t,s,\mathbf{x}_0;\omega)$ the solution of Eq. **1** emanating from $\mathbf{x}_0$ at time $s < t$, when the system is driven by the path $t \mapsto \xi_t(\omega)$, and by $\Phi_\epsilon(t,s,\mathbf{x}_0;\omega')$ the one driven by $\chi_t^\epsilon(\omega')$ and still emanating from $\mathbf{x}(s) = \mathbf{x}_0$.

Consider now a perturbation $\chi_t^\epsilon(\omega')$ applied over the interval (s,t). We define the (local) deviation $\delta_{(s,t)}^{\mathbf{x}_0;\omega}(\psi,\omega')$ of ψ at t, when starting from $\mathbf{x}_0$ at time s and driven by the two noise paths $\chi_t^\epsilon(\omega')$ and $\xi_t(\omega)$, respectively, from s to t:

$$\delta_{(s,t)}^{\mathbf{x}_0;\omega}(\psi,\omega') := \psi(\Phi_\epsilon(t,s,\mathbf{x}_0;\omega')) - \psi(\Phi(t,s,\mathbf{x}_0;\omega)); \qquad [2]$$

ω here refers to the realization that drives the unperturbed system. The corresponding mean response at time t is provided by the expectation $\mathbb{E}_{\omega' \in \Omega}[\delta_{(s,t)}^{\mathbf{x}_0;\omega}(\psi,\omega')]$; when no confusion is possible, we drop the indexing over $\omega' \in \Omega$. We will be mainly interested in the expected response, averaged over (s,t) and given by $\mathbb{E}[\bar{\delta}_{(s,t)}^{\mathbf{x}_0;\omega}(\psi,\cdot)]$, where $\bar{\delta}_{(s,t)}^{\mathbf{x}_0;\omega}(\psi,\cdot) := (s-t)^{-1}\int_s^t \delta_{(s,u)}^{\mathbf{x}_0;\omega}(\psi,\cdot)du$.

At this point, we take a brief excursion beyond the scope of the present paper and note that, by taking an ensemble average over $\mathbf{x}_0 \in X$ and letting $s \to -\infty$, one recovers, in our stochastic context, quantities that are analogous to those considered in Ruelle's (20) response theory for smooth, time-dependent perturbations $F(\mathbf{x},t)$ of autonomous systems with chaotic behavior. In that theory, the nature of the response—whether linear or nonlinear—is independent of the respective choices of the observable ψ and the perturbation in the (deterministic) forcing F.

There is no room to discuss here a rigorous framework in which to assess the response of a stochastic system governed by Eq. **1** to perturbations of the noise path $t \mapsto \xi_t(\omega)$. Only numerical results for the short-range response can be given in the present paper; these will be justified rigorously elsewhere. Applications of linear-response theory to climate sensitivity have been recently investigated in a deterministic (21) and in a stochastic context (22), but only for the model's PDF, while pathwise statistics are discussed in ref. 19.

To which extent does the local, short-range response depend on $\mathbf{x}_0$, ω, s and t? It appears that the linear or nonlinear nature of $\mathbb{E}[\bar{\delta}_{(s,t)}^{\mathbf{x}_0;\omega}(\psi,\cdot)]$, over a fixed interval (s,t), is unchanged for almost all $\mathbf{x}_0$ and almost all ω; as a consequence, we will drop the dependence in $\mathbf{x}_0$ and ω. This statement holds, for instance, in the case of the stochastic Lorenz system forced by $\sigma\mathbf{x}d\xi_t$, with $\sigma > 0$, as considered in ref. 19: For $t-s$ fixed—and almost surely in $\mathbf{x}_0$ and ω—$\mathbb{E}[\bar{\delta}_{(s,t)}^{\mathbf{x}_0;\omega}(\psi,\cdot)]$ depends nonlinearly on the perturbation; this nonlinear response sets in above a certain threshold ϵ_0 of the perturbation size, which is much smaller than the variance of the noise.

To the contrary, other nonlinear (and chaotic) stochastic systems can exhibit—over a wide range of perturbations and still for almost all $\mathbf{x}_0$ and ω—a linear response. For $t-s$ fixed, such a response is visualized through a slope of $\mathbb{E}[\bar{\delta}_{(s,t)}(\psi,\cdot)]$ vs. ϵ that is constant or changes only slightly with $\mathbf{x}_0$ and s. In certain cases, a loss of this linear dependence may occur above a relatively large perturbation size. In practice, the knowledge of $\mathbb{E}[\bar{\delta}_{(s,t)}(\psi,\cdot)]$, and the threshold ϵ_0 past which the response becomes nonlinear, gives a quantitative assessment of the pathwise sensitivity: The smaller ϵ_0 the more sensitive the system may be to a perturbation $\epsilon\xi_t(\omega')$ of the path. It follows also that a "reasonable" value of $\epsilon \leq \epsilon_0$ could serve as a good indicator of an admissible noise perturbation. In the next section, we illustrate what is reasonable for an EMR-model of ENSO.

APPLIED MATHEMATICS

Pathwise Linear Response of an EMR-ENSO Model

EMR models (10, 11, 16) are a subset of the stochastic systems described by Eq. **1**; they can be compactly written as

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}(\mathbf{x},\mathbf{x}) + \mathbf{L}(\mathbf{x},\mathbf{r}_t^l,\xi_t,t), \qquad 0 \leq l \leq L-1. \qquad [3]$$

Here $\mathbf{x}$ represents the slowest and most energetic modes, $\mathbf{B}(\mathbf{x},\mathbf{x})$ is a quadratic nonlinearity, and $\mathbf{L}$ is a time-dependent linear operator obtained by integrating recursively from the Lth level to the top level, $l = 0$, the linear stochastic equations that relate the auxiliary stochastic forcings $\mathbf{r}_t^l$ and $\mathbf{r}_t^{l+1}$—i.e., $d\mathbf{r}_t^l = \mathbf{b}_l(\mathbf{x},\mathbf{r}_t^0,\ldots,\mathbf{r}_t^l)dt + \mathbf{r}_t^{l+1}dt$; $\mathbf{A}$ and $\mathbf{B}$ are estimated by a least-square procedure, and $\mathbf{b}_l$ are linear maps estimated recursively along the same lines. The procedure is stopped when $\mathbf{r}_t^{L-1} = \xi_t$, called the Lth-level residual forcing, has a lag-1 vanishing autocorrelation. The stochastic forcings $\mathbf{r}_t^l$ are ordered from the one with strongest memory, $\mathbf{r}_t^0$, to the most weakly autocorrelated one, ξ_t. Here and throughout this section, ξ_t in Eq. **3** is to be understood thus as an approximation over the sampling interval of $d\xi_t$ in Eq. **1**. This procedure allows one to parameterize the "fast" modes in terms of the "slow" ones; see *SI Text* for further details.

Kondrashov et al. (16) have shown that a two-level EMR model—i.e., $L = 2$ in Eq. **3**—can simulate key features of the global sea surface temperature (SST) field's LFV and is quite competitive in predicting ENSO events on the seasonal-to-interannual scale. Let us assume that such a model has been fitted to the SST observations over some interval $(0,T_0)$ and measure its pathwise sensitivity to changes in the model's "weather"; i.e., in the EMR procedure's residual noise $\mathbf{r}_t^1 = \xi_t$.

Consider now the observable $\psi(\mathbf{x}) := \|\mathbf{x}\|$, where $\|\cdot\|$ denotes the Euclidean norm of $\mathbf{x}$. Our goal is to obtain an estimate of $\mathbb{E}[\bar{\delta}_{(s,t)}(\psi,\cdot)]du$ in terms of ϵ; to do so, we calculate

$$\mathbb{E}\left[\frac{1}{(t-s)\sigma_{0,T_0}(\|\Phi\|)}\int_s^t \|\Phi_\epsilon(u,s,\mathbf{x}_0;\cdot) - \Phi(u,s,\mathbf{x}_0;\omega)\|du\right], \qquad [4]$$

as a function of $\mathbb{E}_{\omega' \in \Omega}[(\sigma_{0,T_0}(\|\xi\|))^{-1}\int_s^t \|\chi_u^\epsilon(\omega') - \xi_u(\omega)\|du]$. Here $(t-s) \ll T_0$, $\Phi(t)$ is the reference solution driven by the reference path of the noise $\xi_t(\omega)$, $\sigma_{(0,T_0)}(\|\xi\|)$ is the standard deviation on $(0,T_0)$ of the norm of the residual noise $\xi_t(\omega)$ at the second level, while $\sigma_{0,T_0}(\|\Phi\|)$ is the standard deviation of the norm $\|\Phi\|$ on $(0,T_0)$.

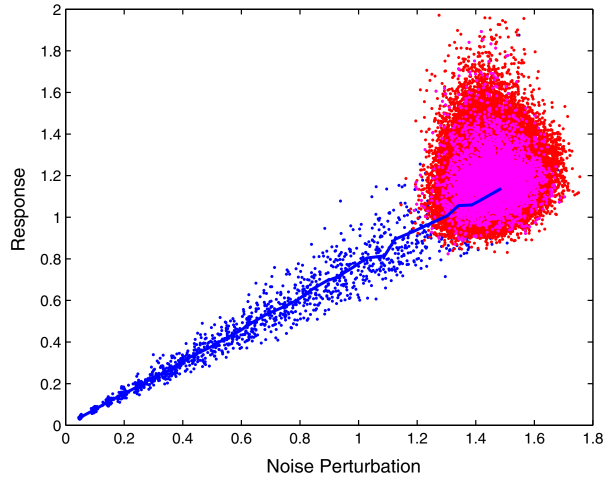

Fig. 1. Pathwise linear response of the two-level EMR model of ENSO (16) for $T = 16$ months ahead. Each blue dot represents an averaged deviation calculated from **2** over a 16-month interval; the perturbation ω' is chosen among one hundred arbitrary ones, for each value of ϵ. These 100 deviations are plotted against $\mathbb{E}_{\omega' \in \Omega}[((t-s)\sigma_{0,T})^{-1}(\|\xi\|)\int_s^t \|\chi_u^\epsilon(\omega') - \xi_u(\omega)\| du]$; see text for details. The straight blue line corresponds to the estimation of the response given by **4** against the same "metric." The magenta and red dots are introduced and discussed in the paper's last-but-one section.

The initial state at time s is chosen as a representative $\mathbf{x}_0$, associated with a state found in the past of the SST field, and t is chosen so that $t - s = T = 16$ months, while the state of the system is known up to time t. The results are depicted as the blue points in Fig. 1, which clearly demonstrates a pathwise linear response, with a slope of about 0.8 for our two-level EMR model (blue curve in the figure) over a large range of perturbations of the noise.

This stochastic EMR model has a leading Lyapunov exponent that is positive, and it exhibits the intermittency phenomenon described by Chekroun et al. (19). This phenomenon is characterized by on-off synchronization of the trajectories for almost all fixed realizations; it is also associated with a sensitive dependence to initial states that is considerably weaker than observed in a stochastically perturbed Lorenz system (19).

In order to perform a more detailed analysis of the sensitivity to initial states of the ENSO model used here, we propose a simple methodology that is consistent with the forecasting method developed in this paper (see below). To be precise, we have considered 60 reference initial states spread over 60 different initial times s; to each of these 60 states, we have applied 120 perturbations corresponding each to a state occupied by the system over 10 yr prior to s. The system was driven by the same realization ω of the noise, and the divergences between the trajectories $\|\mathbf{x}(t,s,\mathbf{x_0};\omega) - \mathbf{x}(t,s,\tilde{\mathbf{x}}_0;\omega)\|$ emanating from the reference initial state $\mathbf{x_0}$ and the 120 perturbed ones $\tilde{\mathbf{x}}_0$, have been plotted in Fig. 2 *A* and *B* at 1 mo ($t - s = 1$, blue points) and 16 mo ($t - s = 16$, red points), respectively.

The divergence at 1 mo exhibits a linear dependence with respect to the magnitude $\|\mathbf{x_0} - \tilde{\mathbf{x}}_0\|$ of the perturbation (*A*), but this dependence becomes nonlinear at 16 mo (*B*). Fig. 2*B*, however, clearly demonstrates that two distinct clusters of red points may be identified: Either the model responds actively to a perturbation of the initial state, and the scatter is large, or the response is weak; in the latter case, the divergence is less than 0.25—i.e., less than 25% of the SST field's standard deviation. Fig. 2*B* shows, furthermore, that the "mass" of this compact cluster is larger than the mass of the former cluster.

This feature demonstrates that, at 16 mo, the trajectories have a stronger tendency to synchronize than to diverge, which indicates on-off synchronization with a stronger on mode than the off mode (19). It thus appears that a fairly strong tendency of trajectories to synchronize for given, fixed stochastic forcing at intermediate range is associated with a weaker dependence on perturbations of this forcing, on the one hand, and with smaller divergence of trajectories at short times, on the other.

The results in Figs. 1 and 2 are thus fairly encouraging in terms of developing a predictive methodology based on estimates of future noise, because errors in such estimates do not seem to produce unduly large errors in the forecast. We develop such a methodology in the next section.

Past Noise Forecasting (PNF)

Sampling Stationary Stochastic Processes. A well-known fact from the theory of stochastic processes is that—given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ associated with a stochastic process $\{t \mapsto \xi_t(\omega)\}_{\omega \in \Omega}$, with $\mathscr{F}$ the σ-algebra of events and $\mathbb{P}$ the probability measure—a time-parameterization of Ω is obtained by introducing the shift operator $\theta_t \xi_s(\omega) := \xi_{t+s}(\omega)$ and letting $\xi_s(\omega) = \omega(s)$; see refs. 18 and 19 and references therein. If the stochastic process is stationary or with stationary increments, then the noise statistics are invariant under the transformation θ_t or its helix analogue (19)—i.e., $\theta_t \mathbb{P} = \mathbb{P}$.

It follows that shifting in time the path of such a stochastic process $t \mapsto \xi_t(\omega)$, given for a fixed realization ω, leads to another path, $t \mapsto \xi_t(\omega')$, of the same stochastic process but for another realization ω'. An immediate consequence is that, given a path $t \mapsto \xi_t(\omega)$ of length $t^* > 0$, each continuous segment or snippet of length T of this path, with $T < t^*$, provides another path $t \mapsto \xi_t(\omega')$ of length T, for a realization ω' of the same stochastic process. This property allows one, therefore, to select several sample paths of length T—associated with other realizations $\{\omega'\}$—from a single path of length $t^* \gg T$.

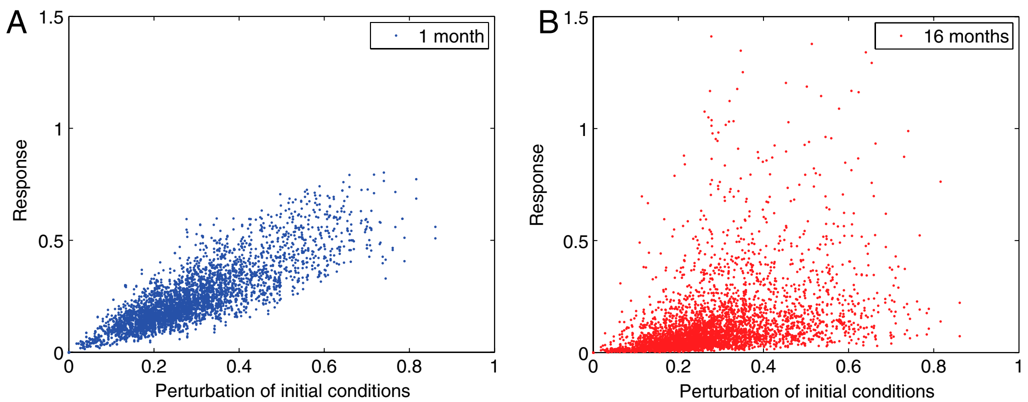

Fig. 2. Sensitivity with respect to initial states. The x-axis represents the magnitude $\|\mathbf{x_0} - \tilde{\mathbf{x}}_0\|$ of a perturbation $\tilde{\mathbf{x}}_0$ of an initial state $\mathbf{x_0}$ taken at time s, while the y axis represents the divergence $\|\mathbf{x}(t,s,\mathbf{x}_0;\omega) - \mathbf{x}(t,s,\tilde{\mathbf{x}}_0;\omega)\|$ between the corresponding trajectories (*A*) at 1 mo ($t - s = 1$); and (*B*) at 16 mo ($t - s = 16$), driven by the same noise realization ω. The results are plotted here for 60 different times s, and 120 perturbations $\tilde{\mathbf{x}}_0$ applied at each time s, corresponding to 120 states occupied by the system over 10 years prior to s. Note that the variance of the field is about 1, and therefore the perturbations considered here are quite substantial.

Noise Sampling Refinement Based on LFV Analogs. We consider now $0 < T \ll t^*$ and select N continuous, possibly overlapping segments of length T each, from a time series of the noise ξ_t of length t^*, given for a fixed realization ω. Note that $N \leq t^*$ depends on t^*, and will be denoted by N_{t^*} when necessary. Starting the copying at time $t = 0$, this procedure leads to a set $\mathcal{S}_{t^*}$ of N snippets of size T, denoted by $\{\zeta^{t_i}: i = 0,\ldots,N-1\}$, with $t_i \in [0,t^* - T]$, $t_0 = 0$, and $t_i < t_{i+1}$. Each of these ζ^{t_i} constitutes a path of length T of the noise for a realization ω_i that could potentially drive, from time t^* to time $t^* + T$, any stochastic system driven by the noise ξ_t; note that $\omega_0 = \omega$.

First, we notice that—by running the two-level EMR–ENSO model of ref. 16 from an arbitrary t^*, and driven by $\zeta^{t_i} \in \mathcal{S}_{t^*}$—the average over a large enough number N_{t^*} of simulations is essentially the same as the one obtained from a large ensemble of arbitrary realizations of the residual noise ξ_t, as given by Eq. **3**. This numerical observation points to two facts: (*i*) The ensemble $\mathcal{S}_{t^*}$ contains sufficient statistics to recover the mean obtained from a larger ensemble; and (*ii*) no improvement is obtained with respect to forecasting by the mean, thus motivating the need to refine the set $\mathcal{S}_{t^*}$ of snippets in some way. In other words, we ask if there exists a subset $\mathcal{S}'_{t^*}$ of $\mathcal{S}_{t^*}$ that allows one to improve the forecast by simply taking the mean over this subset.

Because $\mathcal{S}_{t^*}$ is derived merely by sampling the stochastic forcing from the past, we examine next the phase of the system that corresponds to a snippet ζ^{t_i}, in order to select those snippets that correspond to the phase of the system at time t^*. This selection is motivated by the fact that, for the EMR–ENSO model at hand, sensitivity with respect to the initial state is linear or weak on time scales of $T = 1 - 16$ months (cf. Fig. 2), and therefore the system's phase is mainly determined by the stochastic forcing over these time scales. We thus select noise samples of length T from the past that have forced the system when it occupied a phase that resembles the one just preceding the currently observed state at t^*, and use these to drive the system from t^* on till $t^* + T$. The aggregate forecast skill by relying on the individual forecasts so obtained will depend mainly upon the distribution of the selected snippets as perturbations of the actual path of the forcing into the future, provided the system exhibits a linear pathwise response to such perturbations. The latter point is discussed below, just before *Concluding Remarks*.

We still have to define the precise meaning we attribute to the system's being in a given phase. In fact, several methods exist to examine the instantaneous phase of a stochastic, as well as a deterministic, system. In order to refine the set $\mathcal{S}_{t^*}$ by considering the phase of our two-level EMR model of ENSO, we adopt here a heuristic approach and look at the phase of a smoothed version of the SST field's time series. The smoothing is obtained by singular spectrum analysis (SSA) (23) and turns out to provide a sharp enough selection criterion for the snippets of interest in our PNF.

To simplify the discussion, we only consider here x_1—i.e., the first component of our EMR model—which gives the time evolution of the leading principal component (PC1) of the SST field (16) and captures about 80% of the variance (24). The idea of smoothing PC1 by SSA relies on the fact that the subset formed by the leading reconstructed components (RCs) of PC1 captures the LFV of this time series, while filtering out the spurious high frequencies. Ghil and Jiang (2) showed that much of ENSO prediction skill does lie in correctly forecasting its quasi-quadrennial (QQ) and quasi-biennial (QB) modes (25, 26).

Our approach consists of finding analogues for a continuous segment of a selected RC—of length $0 < \Delta \leq T$ months and ending at t^*, the epoch from which we want to predict—by scanning the record of that RC over the available time series. The closeness of our analogues is determined by using both the rms (RMS) and the correlation coefficient (Corr) between two such segments. Consider the sum RC_K of the $k = 1,\ldots,K$ leading RCs of PC1 (cf. ref. 23 and *SI Text*). The selection procedure for analogues that correspond to prediction from an epoch t^* is based on the following two steps:

Step S1.
Let $\mathrm{RC}_K(t_j,t_j+\Delta)$ be the segment of RC_K that starts at t_j and has length Δ. We first select the times $t_j < t^* - \Delta$ such that $\mathrm{RC}_K(t_j,t_j+\Delta)$ is close enough to $\mathrm{RC}_K(t^*-\Delta,t^*)$ in both RMS and Corr. More precisely, we compute the set:

$$\begin{aligned}&\mathcal{T}'_{t^*}(\alpha,\beta) := \{t_j \in (0,t^*-\Delta):\\ &\mathrm{rms}(\mathrm{RC}_K(t_j,t_j+\Delta) - \mathrm{RC}_K(t^*-\Delta,t^*)) \leq \alpha, \quad \text{and,}\\ &\mathrm{corr}(\mathrm{RC}_K(t_j,t_j+\Delta),\mathrm{RC}_K(t^*-\Delta,t^*)) \geq 1-\beta\},\end{aligned} \qquad [5]$$

with prescribed parameters $0 < \alpha, \beta \ll 1$. In theory, $\mathcal{T}'_{t^*}$ could be the empty set for very small α and β. In practice, α and β have to be tuned so to avoid this, while $\mathrm{card}(\mathcal{T}'_{t^*}) < N_{t^*}$ to ensure that $\mathcal{S}'_{t^*}$ is a strict subset of $\mathcal{S}_{t^*}$; see step 2.
Step S2.
Define next the following subset of $\mathcal{S}_{t^*}$,

$$\mathcal{S}'_{t^*} := \{\zeta^{t_i} \in \mathcal{S}_{t^*} : t_i \in \mathcal{T}'_{t^*}\}. \qquad [6]$$

The set $\mathcal{S}'_{t^*}$ is entirely based on the past observations and will serve us to drive the system into the future, as follows.

For any noise snippet ζ^{t_i}, we denote by $\mathbf{x}(t^*+t,t^*,\mathbf{x}_{t^*};\zeta^{t_i})$ the solution of our two-level EMR model of ENSO at time t^*+t starting from the current observation $\mathbf{x}_{t^*}$ at time t^*, and forced by ζ^{t_i} over the interval (t^*, t^*+t). Note that, by construction, $0 < t \leq T$.

Our predictor—from t^* for t months ahead—will be:

$$\hat{\mathbf{x}}(t^*+t,t^*,\mathbf{x}_{t^*}) := \sum_{t_j \in \mathcal{T}'_{t^*}} \mu_j \mathbf{x}(t^*+t,t^*,\mathbf{x}_{t^*};\zeta^{t_j}), \qquad [7]$$

where $\mu_j \in [0,1]$ and $\sum \mu_j = 1$. In the sequel, we simply illustrate the method with $\mu_j = 1/\mathrm{card}(\mathcal{T}'_{t^*})$, for all j, and postpone the discussion of optimizing the choice of the μ_js for future work. Note that, in Eq. **7**, ζ^{t_j} can be replaced by a shifted ζ^{t_j+s}, with $0 < s \leq \Delta$, because we consider only an "average phase" of the system over Δ-months, rather than an instantaneous phase.

We call any method that consists of using Eq. **7** as a predictor, and steps **S1** and **S2** for the selection procedure of the noise snippets ζ^{t_j} a PNF method; in other words, EMR fitting is but one way to determine the driving noise ζ^t from which the snippets are selected. A schematic illustration of the concept is given in *SI Text*.

In practice, the parameters K and Δ—needed to fully specify a PNF (K,Δ) method—have to be determined either by tuning, as discussed in *SI Text* or by a more systematic optimization method. PNF methods are being presented here in the context of ENSO and for a two-level EMR model, but they are quite general and can be used, in principle, for any empirical stochastic model fitted to some time-evolving dataset. Their forecasting skill depends, of course, on the nature of the model's pathwise response introduced in the previous section; see also the last-but-one section.

Numerical Results and Forecast Skill

We use here a two-level EMR model—i.e., $L = 2$ in Eq. **3**—of ENSO variability (16), embedded in the subspace of the 20 leading PCs of a time series of SST anomalies that is $t_0^* = 585$-month long (January 1950–September 1998), given on a $5° \times 5°$ grid over the 30 S–60 N latitude band (27) and has the annual cycle removed. We predicted starting on successive months $t_q^* = t_0^* + q\delta t^*$, with $\delta t^* = 1$ month, and $q \in \{0,\ldots,Q-1\}$, with $Q = 114$, via (**7**), and the maximum number of months over which we do so is fixed at $T = 16$ months. Note that the total number of possible $N_{t_q}^* \equiv N_{t_0^*} = t_0^* - T = 569$ snippets here is the same for each $t_q^* \geq t_0^*$—i.e., we do not update the EMR model from one value of t_q^* to another; keeping it fixed only makes the test for the PNF method even harder.

For $\Delta = 5$ months and using selection criterion **S1**—applied to RC_K of PC1, with $K = 2$, $\alpha = 0.5$ and $\beta = 0.05$ in Eq. **5**—the PNF (2,5) method selects snippets ζ^{t_i} that allow us to reach the goal set in the preceding section: improve the forecast obtained by taking the mean of the EMR model by refining the subsets $\mathcal{S}_{t^*}$. Fig. 3*A* illustrates this improvement in predicting the Niño-3 index by comparing the mean of the EMR model prediction (red curve) with the predictor given by (**7**) (blue curve). These predictions are issued each month t^* using only data prior to that month, for $Q = 114$ values of t_q^*, running from October 1998 to March 2008.

We note in Fig. 3*A* that the PNF method is able to capture episodes of strong anomalies in the evolution of the Niño-3 index—especially the large El Niño in 2003 and the large La Niña in 2008—much better than the mean EMR method does. While the PNF performance is not uniformly better, it is at no time substantially worse that the mean EMR either. The PNF improvement is most striking during the energetic phases of constructive interference between the LFV modes, QQ and QB (2, 26).

Fig. 3 *B* and *C* shows that the PNF method yields significantly better skill in Niño-3 prediction beyond 6 mo, compared with the standard EMR method of Kondrashov et al. (16). The optimal improvement occurs at relatively long lead times of 12–16 mo. This improvement is due, in part, to the fact that, in our EMR–ENSO model, the chaotic behavior is relatively weak and most trajectories, while starting from different initial states, synchronize for a fixed realization of the noise at longer lead times, as shown in Fig. 2*B*. The PNF improvement is also consistent with the characteristic decay time τ of the ENSO eigenmode associated with its most energetic LFV mode, the QQ mode; this decay time of $\tau \approx 14$ months (16) might ultimately determine the empirical limits of long-term ENSO predictability.

Even more strikingly, Fig. 4 shows that PNF skill in predicting the SST field itself at a 14-mo lead is uniformly better over the entire equatorial Tropical Pacific and Indian Ocean area, where

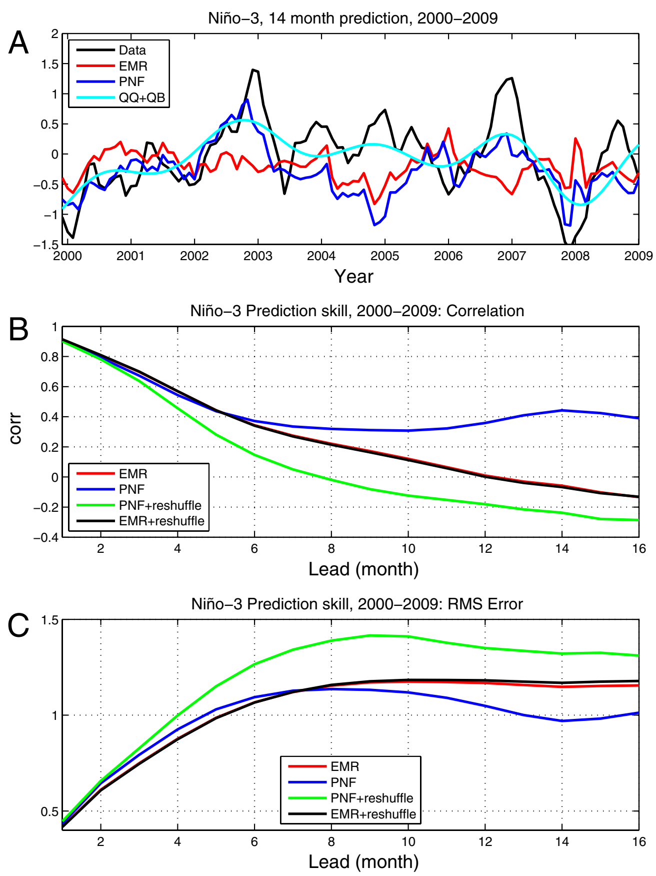

Fig. 3. Niño-3 prediction skill for years 2000–2009. (*A*) PNF forecasts (blue curve) validated at 14-mo lead, compared to the actual data (black), EMR forecast (red), and the quasi-quadriennal (QQ) + quasi-biennial (QB) components of SSA reconstruction (2, 26) (cyan); see text for details. (*B*) RMS and (*C*) Corr forecast skill obtained by using the mean of a two-level EMR–ENSO model driven by a large set S_{t*} of noise realizations (red), and driven by a smaller subset S'_{t*} of realizations selected by the PNF(2,5) method (blue); the average is taken over the $Q = 114$ forecasts issued each month from October 1998 to March 2008. The forecasts use only data prior to that month—i.e., no "look-ahead" whatsoever is involved, and the skill is plotted for leads of 1–16 mo—with the seasonal cycle removed and normalized by the variability. For each forecast, the EMR–ENSO model is initialized at the observed state for the time t^* at which the forecast is started. These results are compared with a reshuffled version $\hat{\xi}_t$ (green) of the residual noise ξ_t.

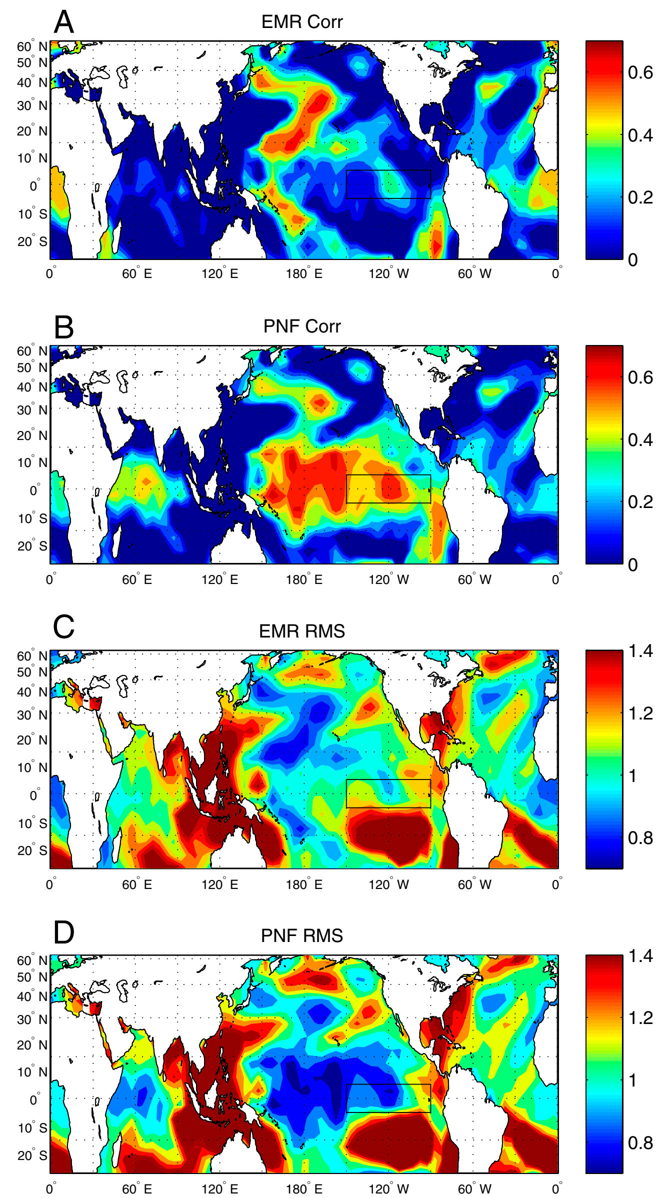

Fig. 4. Prediction skill for 14-mo-lead forecasts of the SST field by (*A* and *C*) an EMR model and (*B* and *D*) the PNF method. The skill is evaluated in terms of maps of (*A* and *B*) anomaly correlation Corr and (*C* and *D*) normalized rms errors RMS. Both EMR and PNF were trained on $T_0 = 50$ years of SST data (1950–2000), and validated on 10 years (2000–2009) of data, with the seasonal cycle subtracted. The Niño-3.4 index is defined as the mean SST anomaly over the rectangular box. PNF skill is uniformly better (lower RMS and higher Corr) in the equatorial Tropical Pacific and Indian Ocean, where ENSO is most active.

ENSO is most active. This improvement is manifest in both forecast skill measures, Corr (*A* and *B*) as well as RMS (C and D).

In order to understand better the advantage of selecting appropriate noise snippets to drive the PNF forecasts, we compare the results in Fig. 3 *B* and *C* with those obtained by using a reshuffled version $\hat{\xi}_t$ (green) of the residual noise ξ_t, obtained at the second level. The reshuffling modifies the autocorrelations of ξ_t but changes neither its mean nor its variance. This comparison demonstrates that the EMR forecasts with reshuffled noise $\hat{\xi}_t$ are identical to those with the original EMR—i.e., the ensemble mean, over $N_{t_0^*}$ snippets, of the future paths is insensitive to random permutations of the residual noise.

To the contrary, the PNF(2,5) predictor in Eq. **7** is sensitive to being driven by $\hat{\xi}_t$ snippets, and the skill of this reshuffled PNF drops dramatically, because it does not respect the relation between the driving noise and the LFV phases in the past. Statistical-significance tests in *SI Text* buttress the numerical results on PNF and EMR prediction skill presented here. Furthermore, the PNF methodology is illustrated in *SI Text* by introducing and analyzing a simple ecological "toy model" forced by white noise.

PNF Method and Pathwise Linear Response

It remains to show that the relatively good skill obtained by the PNF method is in agreement with the pathwise sensitivity results obtained for the two-level EMR model of ENSO (cf. Fig. 1) and the fact that sensitivity with respect to initial data is weak at longer lead times (cf. Fig. 2*B*). Going back to Fig. 1, we have plotted in magenta and for the 114 epochs $t^* = t_0^* + q\delta^*$, $q = 1,\ldots,Q$ used in obtaining Fig. 3, the corresponding outputs, $\mathbf{x}(t^*+t,t^*,\mathbf{x}_{t^*};\zeta^{t_j})$, with $\zeta^{t_j} \in \mathcal{S}'_{t^*}$. The red dots are for the outputs driven by all the elements of $\mathcal{S}_{t^*}$, with no selection of a subset $\mathcal{S}'_{t^*}$.

Clearly the PNF predictors (magenta) form a substantially smaller cluster than the EMR predictors (red cluster): The red outliers correspond to the strongest response and are associated with the worst forecasting skill. We observe furthermore in Fig. 1 that the magenta cluster is located within the linear-response regime, fairly close to a slope of about 0.8, but relatively far from the origin, for almost all initial states.

This result illustrates three facts: (*i*) Our LFV-conditioned PNF method selects snippets that correspond to small enough perturbations of the future noise realization $\{\zeta_t: t^* < t \le t^* + T\}$, and are thus associated with a pathwise linear response of the system; (*ii*) still, this method does not provide paths that are close enough, in general, to ζ_t in order to allow one to recover the system's actual future evolution; and (*iii*) certain extreme episodes are relatively well captured. Note, however, that we used here a relatively rough version of the PNF method, because the purpose was just to demonstrate the usefulness of considering past noise in forecasting future system evolution.

Concluding Remarks

In this article, we have developed a forecasting method for dynamical systems that exhibit (*i*) oscillatory LFV modes and (*ii*) noise effects. The method relies on the use of an empirical stochastic model in which a path of the noise is estimated over a finite-time window. This method uses only information from past observations, and it was illustrated by using an EMR model of ENSO. We showed that the domain of validity of such a method depends strongly on the nature of the system's pathwise response, and that this response is both weak-to-moderate and linear at finite times for the model at hand; see Figs. 1 and 2. The fact that even the "approximately right" noise can help, rather than hinder, prediction is surprising, to say the least.

The forecast skill of the method was evaluated in terms of rms error RMS and anomaly correlations Corr, for lead times of 6–16 months, against a currently operational method, the standard EMR prediction. The latter has been tested over the last 5 yr against 15 dynamical and another 7 statistical methods, and has proven quite competitive at lead times of 6–12 mo, in the context of ENSO prediction at the International Research Institute for climate and society (IRI), http://iri.columbia.edu/climate/ENSO/currentinfo/.

The PNF method seems to improve on the EMR method at 12 mo and beyond (Figs. 3 and 4). Further improvements seem possible and include use of multivariate, rather than univariate, SSA, and better criteria for choosing the most appropriate weather noise by LFV phase. These practical aspects—along with the very intriguing theoretical ones associated with the method's basic concepts—could present interesting topics for future studies.

ACKNOWLEDGMENTS. We are grateful for discussions with and encouragement by S. Kravtsov, J. C. McWilliams, J. D. Neelin, A. W. Robertson, E. Simonnet, and I. Zaliapin. Two anonymous referees provided constructive and insightful suggestions. This study was supported by US Department of Energy grants DE-FG02-07ER64439 and DE-FG02-02ER63413 and by National Science Foundation grant DMS-1049253.

1. Barnston AG, Glantz MH, He Y (1999) Predictive skill of statistical and dynamical climate models in SST forecasts during the 1997–98 El Niño episode and the 1998 La Niña onset. *Bull Am Meteorol Soc* 80:217–243.
2. Ghil M, Jiang N (1998) Recent forecast skill for the El Niño/Southern Oscillation. *Geophys Res Lett* 25:171–174.
3. Lorenz EN (1969) Atmospheric predictability as revealed by naturally occurring analogues. *J Atmos Sci* 26(4):636–646.
4. Barnett TP, Preisendorfer RW (1978) Multifield analog prediction of short-term climate fluctuations using a climate state vector. *J Atmos Sci* 35:1771–1787.
5. Toth Z (1991) Estimation of atmospheric predictability by circulation analogs. *Mon Weather Rev* 119(1):65–72.
6. Van den Dool HM (1994) Searching for analogues, how long must we wait? *Tellus* 46A:314–324.
7. Huang JP, et al. (1993) An analogue-dynamical long-range numerical weather prediction system incorporating historical evolution. *Q J R Meteorol Soc* 119:547–565.
8. Keppenne CL, Ghil M (1992) Adaptive filtering and prediction of the Southern Oscillation index. *J Geophys Res* 97:20449–20454.
9. Penland C (1996) A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D* 98:534–558.
10. Kravtsov S, Kondrashov D, Robertson AW, Ghil M (2005) Multi-level regression modeling of nonlinear processes: Derivation and applications to climatic variability. *J Clim* 18:4404–4424.
11. Kravtsov S, Kondrashov D, Ghil M (2009) *Stochastic Physics and Climate Modelling*, eds TN Palmer and P Williams (Cambridge Univ Press, Cambridge, UK), pp 35–72.
12. Chorin AJ, Hald OH, Kupferman R (2002) Optimal prediction with memory. *Physica D* 166:239–257.
13. Kleeman R (2002) Measuring dynamical prediction utility using relative entropy. *J Atmos Sci* 59:2057–2072.
14. Kleeman R, Majda A, Timofeyev I (2002) Quantifying predictability in a model with statistical features of the atmosphere. *Proc Natl Acad Sci USA* 99:15291–15296.
15. Ghil M, Robertson AW (2002) "Waves" vs. "particles" in the atmosphere's phase space: A pathway to long-range forecasting? *Proc Natl Acad Sci USA* 99:2493–2500.
16. Kondrashov D, Kravtsov S, Robertson AW, Ghil M (2005) A hierarchy of data-based ENSO models. *J Clim* 18:4425–4444.
17. Neelin JD, et al. (1998) ENSO theory. *J Geophys Res* 103:14261–14290.
18. Ghil M, Chekroun MD, Simonnet E (2008) Climate dynamics and fluid mechanics: Natural variability and related uncertainties. *Physica D* 237:2111–2126.
19. Chekroun MD, Simonnet E, Ghil M (2011) Stochastic climate dynamics: Random attractors and time-dependent invariant measures. *Physica D*, 10.1016/j.physd.2011.06.005.
20. Ruelle D (2009) A review of linear response theory for general differentiable dynamical systems. *Nonlinearity* 22:855–870.
21. Abramov R, Majda AJ (2009) A new algorithm for low-frequency climate response. *J Atmos Sci* 66:286–309.
22. Majda AJ, Wang X (2010) Linear response theory for statistical ensembles in complex systems with time-periodic forcing. *Commun Math Sci* 8:145–172.
23. Ghil M, et al. (2002) Advanced spectral methods for climatic time series. *Rev Geophys* 40:3.1–3.41.
24. Preisendorfer RW (1988) *Principal Component Analysis in Meteorology and Oceanography* (Elsevier, New York) p 425.
25. Jiang N, Ghil M, Neelin D (1995) Quasi-quadrennial and quasi-biennial variability in the equatorial Pacific. *Clim Dyn* 12:101–112.
26. Jiang N, Ghil M, Neelin D (1995) Forecasts of Equatorial Pacific SST anomalies using an autoregressive process using singular spectrum analysis. *Experimental Long-Lead Forecast Bulletin* 4:24–27.
27. Kaplan A, et al. (1998) Analyses of global sea-surface temperature 1856–1991. *J Geophys Res* 103:18567–18589.

Supporting Information

Chekroun et al. 10.1073/pnas.1015753108

SI Text

Singular Spectrum Analysis (SSA). SSA is essentially a form of principal component analysis in the time domain: It extracts information from short and noisy time series without prior knowledge of the dynamics generating the time series (1). For a centered and normalized time series $\{x_i: 1,\ldots,N\}$ and a maximum lag (or window size) M, we construct an $M\times M$ lagged covariance matrix $\mathbf{C}_x$. This matrix has Toeplitz form (i.e., each diagonal has constant entries that correspond to equal lags)

$$c_j=\frac{1}{N-j}\sum_{i=1}^{N-j}x_ix_{i+j},\qquad 0\le j\le M-1; \qquad \textbf{[S1]}$$

other formulae have also been proposed to estimate the entries of $\mathbf{C}_x$ (1, 2).

The M eigenvalues λ_k of this $M\times M$ matrix and its eigenvectors, called empirical orthogonal functions (EOFs), are computed and sorted in descending order of the variances λ_k associated with each EOF. The kth principal component, PC_k, is given by

$$a_i^k=\sum_{j=1}^{M}x_{i+j}E_j^k,\qquad 0\le i\le N-M. \qquad \textbf{[S2]}$$

Each PC_k has length $M-N$, but the so-called reconstructed components (RCs) (3–5) are given by

$$x_i^k=\frac{1}{M}\sum_{j=1}^{M}a_{i-j}^kE_j^k,\qquad M\le i\le N-M+1. \qquad \textbf{[S3]}$$

and each RC_k has length N; formulae for x_i^k near the endpoints (i.e., for $1\le i<M$ and $N-M<i\le N$) are given in refs. 1 and 5.

Most of the variance is typically contained in the first few RCs and the remaining RCs contain mostly noise. SSA decomposes a time series into modes that are robust to reasonable changes in M or N and those that are not. The former usually contain one or more pairs of high-variance RCs—with nearly equal λ_k—that capture oscillatory modes with periods less than M; often there are also one or two high-variance RCs that contain variations with periods greater than M. These "aperiodic" RCs may be characterized as "nonlinear trends," while the pairs of EOFs associated with the oscillatory behavior are data-adaptive "sines and cosines." The RCs that are not associated with either the trend or the oscillatory behavior—whether of noisy or not-so-noisy appearance—are usually not robust.

SSA is particularly well suited for the analysis of time series that exhibit nearly periodic, oscillatory behavior in the presence of phase or amplitude modulation; it provides a robust method to identify and reconstruct statistically significant modes of low-frequency variability (LFV). To avoid smooth-looking but spurious oscillations that might arise from finitely sampled noise processes, a statistical-significance test against red noise is performed by comparing the projection of the data onto a given EOF with the corresponding projections of a large ensemble of red-noise surrogates (1, 6). The SSA method's ability to extract robust, albeit modulated, oscillations was used to help determine the phase of noisy time series in the PNF method, as described further below in the present SI; it also played a key role in the study of phase-locking phenomena between the North Atlantic Oscillation and Mid-East temperatures and precipitation (7, 8).

Empirical Model Reduction (EMR). ***Methodology and the two-level EMR-ENSO model.*** EMR models are obtained by constructing a hierarchy of time-dependent, nonlinear, and stochastically forced inverse models to describe the dynamics underlying a time-evolving spatial field (9–13). In EMR, multiple polynomial regression is used to estimate the deterministic propagator of the dynamics, as well as "warm-colored" additive stochastic forcing, directly from the observations or from a *reanalysis* dataset.

More precisely, consider N regularly sampled observations of the state vector $\mathbf{x}(t)=\{x_j(t): j=1,\ldots,J\}$, $t=1,\ldots,N$. Here $\mathbf{x}(t)$ is obtained by carrying out a principal component analysis in the spatial domain or in the phase space of a much higher-dimensional state vector $\mathbf{X}(t)$ and projecting this dataset onto its J leading principal components. If we consider quadratic nonlinear terms only, the main level of the EMR equations for modeling the increments Δx_i is given by

$$\Delta x_i=(N_{ijk}x_jx_k+L_{ij}^{(0)}x_j+F_i)\Delta t+r_{0,i}\Delta t. \qquad \textbf{[S4]}$$

Multiple linear regression is applied recursively to find coefficients N_{ijk}, $L_{ij}^{(0)}$, and F_i that fit the predictand variables Δx_i. The regression residuals $r_{0,i}$ define the time-dependent stochastic forcing; in practice, these residuals are not white—as is often assumed in other inverse modeling methodologies—but have long-tailed autocorrelations. Additional, linear model levels are then included to simulate this correlated noise recursively, by using state-dependent coefficients. The increments $\Delta\mathbf{r}_l$,

$$\Delta\mathbf{r}_l=\mathbf{b}_l(\mathbf{x},\mathbf{r}_0,\ldots,\mathbf{r}_l)\Delta t+\mathbf{r}_{l+1}\Delta t, \qquad \textbf{[S5]}$$

are obtained for $0\le l\le L-2$ with $\mathbf{b}_l$ the estimated linear operator, and the number L of levels is chosen so that the stochastic forcing at the last level, $\mathbf{r}_{L-1}$, is well approximated by a spatially correlated white-noise process.

The number of variables—as well as the order of nonlinearity, which can be higher than quadratic—is determined by optimizing inverse model performance (13). Table S1 below illustrates the fact that—for the leading two PCs (x_1,x_2) of the sea surface temperature (SST) field used in this paper—the residual noise ξ_t of our two-level EMR–ENSO model is essentially indistinguishable from white; here $\xi_t=\mathbf{r}_1$ in the notation of Eqs. **S4** and **S5**—i.e., $L=2$. Recall that ξ_t in the EMR formulation is to be understood as an approximation over the sampling interval Δt of $d\xi_t$ in Eq. **1** of the main text. The table lists the first five lagged autocorrelations of ξ_t for PC_1 and PC_2, along with those of its reshuffled version $\hat{\xi}_t$ of ξ_t, used in Fig. 3 of the main text (green solid in Fig. 3 *B* and *C*).

The lagged correlations of $\hat{\xi}_t$ and ξ_t in the table indicate that neither is truly white: They both have signs that are distributed randomly and magnitudes that fluctuate between roughly 0.01 and 0.1, again rather randomly. We conclude that both $\hat{\xi}_t$ and ξ_t are quite reasonable approximations of finitely sampled white-noise realizations. The failure of the PNF method when using $\hat{\xi}_t$, rather than ξ_t, is thus mainly due to the fact that $\hat{\xi}_t$ gives rise to large perturbations when its segments are selected at inappropriate phases of the LFV and not to the changes in the autocorrelations of the residual noise when reshuffled.

The EMR approach has been successfully applied to tropical SST variability on the seasonal-to-interannual time scale (9). Our EMR–ENSO model captures essential nonlinear aspects of ENSO development, such as the observed asymmetry between the statistical distribution of warm El Niño and cold La Niña events. The estimated noise plays an important role in driving

ENSO events; this role is due, at least in part, to the non-normal growth of small perturbations (9, 14).

Our EMR–ENSO model, in and of itself, has been successfully applied to ENSO prediction; see real-time ENSO forecasts by EMR among the plume of dynamical and statistical models from the International Research Institute (IRI) for Climate and Society, http://iri.columbia.edu/climate /ENSO/currentinfo/modelviews.html.

Low-frequency ENSO modes of variability are captured in the EMR–ENSO model as damped oscillatory eigenmodes of the linearized dynamical EMR operator; these are also known as empirical normal modes (15) and are excited by the noise, as well as by the interaction with the seasonal forcing. The interannual modes identified by EMR are in good agreement with those described by SSA, in particular with the two dominant modes of ENSO variability, the quasi-quadrennial (QQ) and quasi-biennial (QB) modes (1, 9, 16).

EMR as a non-Markovian model. It is possible to rewrite a multilevel EMR model as a single equation that involves time-lagged values of x_i and $r_{l,i}$; the resulting equation is similar to a multivariate version of an autoregressive-moving average (ARMA) model. There are two key differences with respect to standard ARMA models: (*i*) We do allow here for nonlinear dependence; and (*ii*) the multilevel EMR structure for modeling the noise allows one to capture feedback between high- and low-frequency components of the variability, thus parameterizing the "fast" motions (often referred to as the "noise") in terms of the "slow" ones (the "signal").

To illustrate this feature of EMR methodology, consider the following stochastic system that represents the simplest EMR model:

$$dx = f(x)dt + rdt, \qquad dr = (\gamma x - \alpha r)dt + dW_t; \qquad \textbf{[S6]}$$

stochastic differential equations are understood throughout the present SI in the Itô sense, W_t is a standard Wiener process, f is a nonlinear function, $\alpha > 0$, and $\gamma \in \mathbb{R}$. In Eq. **S6**, x is the slow variable and r is the fast one; it is the latter that we want to parameterize.

If we integrate the r-equation in Eq. **S6** from $r_0 \in \mathbb{R}$ at $t = 0$, we obtain, for a fixed realization $\omega \in \Omega$:

$$r(t;\omega) = e^{-\alpha t} r_0 + \gamma \int_0^t e^{-\alpha(t-s)} x(s) ds + W_t(\omega).$$

The latter yields, when substituted into the x-equation in Eq. **S6**, the following randomly forced integro-differential equation,

$$\dot{x} = f(x) + e^{-\alpha t} r_0 + \gamma \int_0^t e^{-\alpha(t-s)} x(s) ds + W_t(\omega). \qquad \textbf{[S7]}$$

In the particular case of $r_0 = 0$ and $f(x) = -\beta x$, with $\beta > 0$, Eq. **S7** is a modified version of the so-called generalized Langevin equation, in which the Wiener process W is the forcing, instead of the "white-noise process $\dot{W}$," as in the classical formulation. A change of variables in the convolution product that appears in the integral of Eq. **S7** clarifies then, in a continuous-time framework, the analogy with an ARMA model.

Formally Eq. **S7** is equivalent to the system **S6**, and we deduce therefore that the dynamics of $x(t)$ in the latter system is in fact non-Markovian and depends on the past history of x. This observation can be extended to the general form **S4** of EMR models and the general situation will be presented and discussed elsewhere. EMR methodology thus assumes, in a sense, that the slow motions—captured by the reduced variables—are non-Markovian. The delayed-oscillator theory of ENSO, in particular, relies on such an assumption (17, 18).

From a theoretical point of view, if we assume furthermore that $x(t)$ is the observed macro state of the system **S6**—i.e., it is an observable of this system—one may note the analogy between Eq. **S7** and the general evolution equation of any observable within the Mori–Zwanzig formalism; see equation 2.7 in refs. 19 and 20. Both are given by random integro-differential equations, with the integral part equal to a convolution between the macro-state evolution and the effects of the unresolved variables; the latter are expressed here by the exponential term in the integrand of Eq. **S7**.

Further links between EMR methodology and model reduction strategies using the Mori–Zwanzig formalism will be pursued elsewhere.

The PNF Method for the EMR–ENSO Model. ***A detailed illustration of PNF skill.*** We refer to the main text for the notations, as well as the description of the PNF method. Fig. S1 here illustrates the method's main features, for a particular epoch t^* at which the numerical predictions of PNF ($K = 2$, $\Delta = 5$) are particularly good.

In what follows, we refer to the equations in the main text by numbers in square brackets. The red curve in Fig S1 is the PC_1 of the actual observations from $t^* - \Delta$ to $t^* + T$, where $T = 16$ months. The heavy blue curve is the PNF(2,5) predictor from t^*—obtained by Eq. [**7**] applied to the first component, x_1, of the EMR state vector $\mathbf{x}$, as the mean over the cyan curves; these curves are given by $x_1(t^* + t, t^*, \mathbf{x}_{t^*}; \zeta^{t_j})$, for $K = 2$ and $\Delta = 5$ in Eq. [**5**]. The dashed blue curve is the prediction given by the EMR model's mean over a large number n of realizations, $n \simeq 600$. The green curves represent segments of length $\Delta = 5$ months of $RC_{K=2}$—as selected from the set $\mathcal{T}'_{t*}(\alpha,\beta)$ according to Eq. [**5**], with $\alpha = 0.5$ and $\beta = 0.04$—and continued up to $t_j + \Delta + T$. In magenta are the corresponding x_1 forecasts associated with the green segments—i.e., given by $x_1(t_j + t, t_j, \mathbf{x}_{t_j}; \zeta^{t_j})$, with $t_j \in \mathcal{T}'_{t*}$—while in black is RC_2 of PC_1 for Δ months prior to t^* and T-months after t^*.

The discrepancy between the magenta curves and the cyan curves clearly demonstrates the subtle relationships between the three main ingredients used in the PNF method: (*i*) the phase of the LFV variability, used in the selection of the noise segments; (*ii*) the actual realization of the noise forcing; and (*iii*) the initial state of the EMR state vector. The two sets of curves are similar with respect to the first two ingredients, but their initial states differ significantly from one set to the other. As a result, the predictive skill of the magenta forecasts is clearly poor when compared with the cyan ones, the latter having been obtained by the PNF method.

The role of the K and Δ parameters. In Fig. S2, we illustrate the dependence of the skill of the PNF method with respect to the number Δ of months over which the analogues in the time series of the RC_K of PC_1 are selected, according to the criteria of Eq. [**5**] as well as the dependence on the choice of K leading RC's considered to do this task. In this figure, a curve labeled for instance PNF(2,5) corresponds to a selection of noise segments based on the proximity—in RMS and Corr—to the most recent history over $t^* - \Delta \le t \le t^*$ of $RC_{K=2}$ over a sliding window of size $\Delta = 5$ months.

The ordering of the skills for PNF (K,Δ) as applied to the Niño–3.4 index in Fig. S3 is comparable with the one obtained for Niño–3. For instance PNF(6,2) and PNF(4,3) perform better than PNF(2,5) at 5 months, but PNF(2,5) gives clearly the best results at 12 months and beyond; see also Fig. 3 in the main text.

Statistical-significance test. We present here a statistical-significance test for the PNF(2,5) method on a longer validation period than in the main text. To do so, we used synthetic time series obtained from EMR-ENSO model simulations. For these simulations, the EMR-ENSO model was trained on the same time interval of $T_0 = 585$ months (January 1950–September 1998) as in the main text (21). Next, however, a 100-year-long control run of this model was obtained, driven by spatially correlated white noise. The control run was then divided into two 50-year portions: 50 years for noise-snippet selection and 50 years for PNF validation.

The results for the longer validation period of $T_f' = 50$ years, using the synthetic EMR–ENSO time series, are plotted in Fig. S3. These results clearly confirm the improvement in PNF (2,5) forecast skill over straightforward EMR forecasting, in both Corr and RMS skill; they should be compared with the results shown for PNF(2,5) in Fig. 3 *B* and *C* of the main text, respectively. The latter skill scores could only be evaluated for $T_f = 10$ years, by using the available observations. The SI results here also demonstrate that the PNF method can work quite well when the driving noise of the model it uses is white in time, rather than weakly autocorrelated, as in the main text.

The extended validation test illustrated in Fig. S3 here also confirms that selecting appropriate past "weather" episodes that match the "climate" phase at time t^* from which a climate prediction has to be made can markedly improve the prediction skill over that provided by large-ensemble averages of a stochastic model fitted to the same data. Provided, that is, that the model on which the PNF method operates exhibit a linear pathwise response on time scales that are comparable to or exceed the prediction lead time T. The latter condition was shown to hold for our EMR–ENSO model in Figs. 1 and 2 of the main text.

Further improvements in PNF prediction are possible by using, for instance, not just PC1 of the dataset, which captures mainly the QQ mode of ENSO, but also PC2. The latter contains substantial QB variance, and it is—as previously stated—mainly the positive interference of the two oscillatory modes, QQ and QB, that yields large events of either warm, El Niño or cold, La Niña type (1, 9, 16).

The PNF Method for a Toy Model with White Noise. The purpose of this section is to further document the fairly surprising properties of the PNF method, in as simple a setting as possible. The model we use is a stochastically and periodically driven version of an ecological model, the Holling model (22). This version has three key properties: (*i*) It is nonlinear and stochastic but does not belong to the class of EMR models; (*ii*) it is forced by white noise, thus emphasizing that the PNF results in the main text are not due to the presence of any sort of memory in the stochastic forcing; and (*iii*) it can be subjected to an easily reproducible battery of numerical tests. We shall see, in effect, that the skill in the PNF forecasts is due, instead, to the model's LFV and to its pathwise linear response as described in the main text.

The deterministic Holling model. The Holling model (22) deals with predator–prey dynamics and assumes, specifically, that predators seek patches that offer enough prey to be worthwhile; it differs, therewith, from the classical Lotka–Volterra model (23–25), which assumes basically that consuming prey does not interfere with searching for it. The original Holling model (22) is governed by two autonomous ordinary differential equations:

$$\dot{x_1} = rx_1(1-x_1) - c\frac{x_1x_2}{\alpha + x_1}, \qquad \dot{x_2} = -mx_2 + c\frac{x_1x_2}{\alpha + x_1}; \qquad \textbf{[S8]}$$

it has served to study, in particular, the so-called "paradox of enrichment" in mathematical ecology (26, 27).

In this model, x_1 and x_2 represent the prey and predator populations, respectively, in terms of the numbers of individuals or of suitable and equivalent measures of their density or biomass. The preys obey logistic growth in the absence of predators, with an *exponential growth factor r* at low densities. The parameter m denotes the predators' *death rate per capita*; c is the *maximum harvest rate* of each predator, while α is the *half-saturation constant*, namely the density of prey at which the predation rate is half the maximum; all of these parameters are positive.

The classical, two-variable Lotka–Volterra model (23–25) is of Hamiltonian type, while the Holling model of Eq. **S8** is dissipative—i.e., volumes contract in phase space. The original Holling model thus exhibits either a globally stable equilibrium or a globally stable limit cycle. Note that more elaborate Lotka–Volterra models, with a larger number of variables, may also exhibit dissipative as well as Hamiltonian dynamics, depending on the interactions between these variables (28, 29).

The parameter-dependent attractor set of Eq. **S8** can be derived by multiplying this equation by $\alpha + x_1$ and scaling the time according to $dt = (\alpha + x_1)d\tau$. The differential equations so obtained are:

$$\begin{aligned} \dot{x_1} &= rx_1(\alpha + x_1)(1 - x_1) - cx_1x_2, \\ \dot{x_2} &= -m\alpha x_2 + (c-m)x_1x_2; \end{aligned} \qquad \textbf{[S9]}$$

here the unknowns are still denoted by x_1 and x_2 and $\dot{x}$ now denotes the derivative of x with respect to τ.

It can be shown that system **9** is *orbitally equivalent* to the original system **8** for $x_1 > -\alpha$; i.e., the phase portraits of the two systems are qualitatively the same in this range. It is then easy to prove that system **S9** undergoes a supercritical Hopf bifurcation from the nontrivial equilibrium E_α, given by

$$E_\alpha := \left(\frac{\alpha m}{c-m}; \frac{r\alpha}{c-m}\left(1 - \frac{\alpha m}{c-m}\right)\right),$$

to a limit cycle, as the parameter α crosses from above the critical value $\alpha^* = (c-m)/(c+m)$, while the other parameters $r > 0$ and $c > m > 0$ remain fixed (30). The frequency ω of the oscillations that emerge at α^* is then given by

$$\omega^2(\alpha^*) = \frac{rc^2m(c-m)}{(c+m)^3}. \qquad \textbf{[S10]}$$

A stochastically driven Holling model. In spite of very substantial observational, theoretical and modeling studies of ENSO-related phenomena and processes, several competing mechanisms for the generation and maintenance of this variability still coexist; see ref. 31 for a review and refs. 17 and 32–34 for some complementary points of view. In one of these approaches, it has become fairly common, in modeling the contributions to ENSO variability of noise-sustained oscillations, to add multiplicative noise to the bifurcation parameter, just ahead of the Hopf bifurcation (34–36).

Another possible mechanism of ENSO variability relies on the idea of non-normal modes being triggered and sustained by noise. Such modes are present in the nonlinear EMR–ENSO model of ref. 9 as well as in the linear inverse model of ref. 37. We chose to study here an even simpler version of the latter mechanism by introducing the following periodically and stochastically forced version of the Holling model (**S9**):

$$\begin{aligned} dx_1 &= \{rx_1(\alpha + x_1)(1 - x_1) - cx_1x_2 + a\sin(2\pi ft)\}dt \\ &\quad + \sigma\{rx_1(\alpha + x_1)(1 - x_1)\}dW_t, \\ dx_2 &= \{-m\alpha x_2 + (c-m)x_1x_2\}dt. \end{aligned} \qquad \textbf{[S11]}$$

This model could, in fact, be used to analyze the multiplicative perturbation of its Hopf bifurcation parameter α as well; we leave such a study for future work.

The stochastic term $\sigma\{rx_1(\alpha+x_1)(1-x_1)\}dW_t$ in Eq. **S11** represents instead a random perturbation of the parameter r by white noise. When considered separately, the effects of this term lead to a random perturbation of the equilibrium E_α for $\alpha > \alpha^*$ and to a modulation of the emerging period, given by Eq. **S10**, just after the Hopf bifurcation—i.e., for $\alpha < \alpha^*$.

The preys x_1 are also assumed to exhibit a seasonal variation—due, for instance, to migration effects—as modeled by the presence of the deterministic, additive forcing $a\sin(2\pi ft)$. This periodic forcing is mathematically similar to that of our EMR–ENSO model, in which a seasonal cycle is also present. Otherwise the modified Holling model of Eq. **S11** is clearly distinct from the class of EMR models, as described by Eq. **S6**.

Numerical formulation and LFV. We integrate the system Eq. **S11** from $t = 0$ to $t = T_f = 2{,}000$ (in dimensionless units) by using a stochastic Euler–Maruyama scheme with step size $\Delta t = 0.1$, in the context of the Itô calculus (cf. ref. 38). The values of the parameters are $\sigma = 0.3, m = r = 1, c = 1.5, \alpha = 0.3, a = 0.05$, and $f = 0.25$, while the initial state is $(x_1(0)x_2(0)) = (0.5{,}0.5)$. For these parameter values, $\alpha^* = 0.2$ and so $\alpha = 0.3 > \alpha^*$; hence, when both periodic forcing and noise are absent, the model has only one globally stable equilibrium.

In Fig. S4 *A* and *B*, respectively, we plot the autocorrelation functions for these parameter values, without and with multiplicative noise forcing. When turning on the periodic forcing, with amplitude $a = 0.05$ and frequency $f = 0.25$, but no noise (Fig. S4*A*), the system exhibits only one periodic orbit of period 4, which is globally stable. In the presence of noise, though (Fig. S4*B*), a low-frequency mode of period equal to approximately 25 units—i.e., a frequency $f' = 0.04$—becomes dominant, while in Fig. S4*A* it is rapidly damped and is visible only during the transient that leads up to the unique attracting periodic orbit.

When the noise is turned on, $\sigma = 0.3$, the SSA spectrum in Fig. S5*B* shows that this mode becomes clearly the more energetic one: Its variance lies within a broad band of frequencies that rise above a 99% confidence limit against the null hypothesis of red noise. It is this band that is responsible for the frequency modulation observed in the corresponding autocorrelation function of Fig. S4*B*; its presence is due to damped non-normal modes that are maintained by the noise.

This behavior is reminiscent of the non-normal modes triggered by noise that have been advocated as a possible mechanism of ENSO variability. Such modes are present in the nonlinear EMR–ENSO model of ref. 9 as well as in the linear inverse model of ref. 37. When applying SSA to the prey population x_2 with a window width of 30 units, RC_2 captures the above-mentioned low-frequency mode with $f' = 0.04$, which contains in fact most of the variance of x_2, as illustrated in Fig. S6.

Numerical PNF results for the toy model. The periodically and stochastically forced system **11** thus exhibits interesting LFV with frequency $f' = 0.04 < f$, where $f = 0.25$ is the forcing frequency. It is easy to check numerically that this system also exhibits a pathwise linear response, in the sense described in the main text. We thus have at our disposal a simple and still quite rich model to test the prediction capabilities of the PNF method against a brute-force, ensemble-mean forecasting approach.

The white-noise assumption in formulating Eq. **S11** gets rid of any source of memory that could come from the stochastic forcing and could obscure the origin of the PNF method's skill. As in the main text, the selection procedure of the "snippets" here is based on the phase of a smoothed version, as determined by SSA prefiltering, of the time series in which the LFV modes are most energetic, namely x_2 in this model. At this stage, the LFV plays a fundamental role in defining similar phases of the time series, while the model's linear pathwise response guarantees that its trajectories are not too sensitive to a change in the realization of its forcing, at least over a reasonable lead time.

The latter property is essential, because the selection procedure of the snippets will always introduce deviations with respect to the actual path of the noise. Linear pathwise response guarantees that this deviation will behave linearly with the "distance" between the selected snippets and the actual path. Even so, depending on the slope of this linear response, large deviations—i.e., prediction errors—may still occur.

For consistency with the ENSO predictions in the main text, we will be interested hereafter in the prediction of the *anomalies* x_1^c and x_2^c of the population variables x_1 and x_2 given by Eq. **S11**, where the superscript $(\cdot)^c$ refers to fact that these anomalies are *centered*—i.e., they have zero mean. Obviously, the PNF method performs similarly for centered or noncentered data: One just has to add the mean back in.

We describe now the technical details of the PNF method used to predict future trajectories of the system Eq. **S11**. This method is based on the phase analysis of an SSA-smoothed version of the x_2^c anomaly—namely, of RC_2, which has period 25 and is the more energetic mode of variability in x_2^c. This choice is analogous to the choice of PC_1 for our EMR–ENSO model, where the QQ mode was the more energetic one in PC_1. The parameter Δ is set equal to 3, which represents $3/4$ of the forcing period, while the lead time ranges from 1 to 10.

Skill is reported in Figs. S7 and S8, while the validation interval and other technical details are given in the figure captions. The results clearly demonstrate the superiority—in terms of both correlation coefficient Corr and RMS error—of the PNF method in forecasting x_2^c against a large, brute-force ensemble of forecasts, for lead times longer than 4, *i.e.*, longer than the forcing period.

The PNF method beats ensemble prediction (ENS) for both predator density anomaly x_2^c and prey density anomaly x_1^c, but the gain in forecasting the predator anomalies is much more substantial. This difference is clearly due to the fact that the snippets were selected based on the smoothed version of x_2^c alone. Unlike in the EMR-ENSO model analyzed in the main text, x_2^c does not contain most of the covariance of the vector $(x_1^c x_2^c)$: the overall variance here is more evenly distributed among x_1 and x_2, and is actually rather larger in x_1^c (cf. Fig. S8). Extension of the present PNF method to one that would deal with LFV distributed over several energetic PCs is under investigation, as pointed out in *Concluding Remarks* of the main text.

1. Ghil M. et al. (2002) Advanced spectral methods for climatic time series, *Rev Geophys*, 40:1–41.
2. Broomhead DS, King GP (1986) Extracting qualitative dynamics from experimental data. *Physica D* 20:217–236.
3. Ghil M, Vautard R (1991) Interdecadal oscillations and the warming trend in global temperature time series. *Nature* 350:324–327.
4. Vautard R, Ghil M (1989) Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D* 35:395–424.
5. Vautard R, Yiou P, Ghil M (1992) Singular-spectrum analysis: A toolkit for short, noisy, chaotic signals. *Physica D* 58:95–126.
6. Allen MR, Robertson AW (1996) Distinguishing modulated oscillations from colored noise in multivariate datasets. *Clim Dyn* 12:775–784.
7. Feliks Y, Ghil M, Robertson AW (2010) Oscillatory climate modes in the Eastern Mediterranean and their synchronization with the North Atlantic Oscillation. *J Clim* 23:4060–4079.
8. Kondrashov D, Ghil M (2006) Spatio-temporal filling of missing points in geophysical datasets. *Nonlin Processes Geophys* 13:151–159.
9. Kondrashov D, Kravtsov S, Robertson AW, Ghil M (2005) A hierarchy of data-based ENSO models. *J Clim*18:4425–4444.
10. Kondrashov D, Kravtsov S, Ghil M (2006) Empirical mode reduction in a model of extratropical low-frequency variability. *J Atmos Sci* 63:1859–1877.
11. Kondrashov D, Kravtsov S, Ghil M (2011) Signatures of nonlinear dynamics in an idealized atmospheric model. *J Atmos Sci* 68:3–12.
12. Kravtsov S, Kondrashov D, Ghil M (2005) Multi-level regression modeling of nonlinear processes: Derivation and applications to climatic variability. *J Clim* 18:4404–4424.

13. Kravtsov S, Kondrashov D, Ghil M (2009) Empirical model reduction and the modeling hierarchy in climate dynamics. *Stochastic Physics and Climate Modelling* Palmer TN, Williams P (Eds.), Cambridge Univ Press, Cambridge, UK, pp 35–72.
14. Moore AM, Kleeman R (1999) The non-normal nature of El Niño and intraseasonal variability. *J Clim* 12:2965–2982.
15. Penland C (1996) A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D* 98:534–558.
16. Jin F-F, Neelin JD, Ghil M (1994) El Niño on the Devil's Staircase: Annual subharmonic steps to chaos. *Science* 264:70–72.
17. Ghil M, Zaliapin I, Thompson S (2008) A delay differential model of ENSO variability: Parametric instability and the distribution of extremes. *Nonlin Processes Geophys* 15:417–433.
18. Wang C, Picaut J (2004) Understanding ENSO physics *Earth's Climate: The Ocean-Atmosphere Interaction* Geophysical Monograph Series, Vol. 147, Wang C, Xie S-P, Carton JA (Eds.) American Geophysical Union, Washington, DC, pp 21–48.
19. Chorin AJ, Hald OH, Kupferman R (2002) Optimal prediction with memory. *Physica D* 166:239–257.
20. Chorin AJ, Hald OH (2006) Stochastic tools in mathematics and science. *Surveys and Tutorials in the Applied Mathematical Sciences* Vol. 1. Springer, New York, p 147.
21. Kaplan A, et al. (1998) Analyses of global sea-surface temperature 1856–1991. *J Geophys Res* 103:18567–18589.
22. Holling C (1965) The functional response of predators to prey density and its role in mimicry and population regulation. *Mem Entomol Soc Can* 45:5–60.
23. Lotka AJ (1925) *Elements of Physical Biology*, Williams & Wilkins, Baltimore.
24. Volterra V (1926) Fluctuations in the abundance of a species considered mathematically. *Nature* 118:558–560.
25. Volterra V (1931) *Leçons sur la Théorie Mathématique de la Lutte pour la Vie* Gauthier-Villars, Paris.
26. Rosenzweig M (1971) The paradox of enrichment. *Science* 171:385–387.
27. Kot M (2001) *Elements of Mathematical Ecology* (Cambridge Univ Press, Cambridge, UK), p 453.
28. Plank M (1995) Hamiltonian structure for the *n*-dimensional Lotka–Volterra equations. *J Math Phys* 36:3520–3534.
29. Roques L, Chekroun MD (2011) Probing chaos and biodiversity in a simple competition model. *Ecological Complexity* 8:98–104.
30. Kuznetsov YA (2004) *Elements of Applied Bifurcation Theory*, Springer-Verlag, New York, 3rd Ed, p 631.
31. Neelin JD, et al. (1998) ENSO theory. *J Geophys Res* 103:14261–14290.
32. Neelin JD, Jin F-F, Syu H-H (2000) Variations in ENSO phase locking. *J Clim* 13:2570–2590.
33. Saunders A, Ghil M (2001) A Boolean delay equation model of ENSO variability. *Physica D*160:54–78.
34. Timmermann A, Jin F-F (2002) A nonlinear mechanism for decadal El Niño amplitude changes. *Geophys Res Lett* 29:1003.
35. Chekroun MD, Simonnet E, Ghil M (2011) Stochastic climate dynamics: Random attractors and time-dependent invariant measures. *Physica D*, 10.1016/j.physd.2011.06.005.
36. Ghil M, Chekroun MD, Simonnet E (2008) Climate dynamics and fluid mechanics: Natural variability and related uncertainties. *Physica D* 237:2111–2126.
37. Penland C, Sardeshmukh PD (1995) The optimal growth of tropical sea-surface temperature anomalies. *J Clim* 8:1999–2024.
38. Kloeden PE, Platen E (1992) *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin.

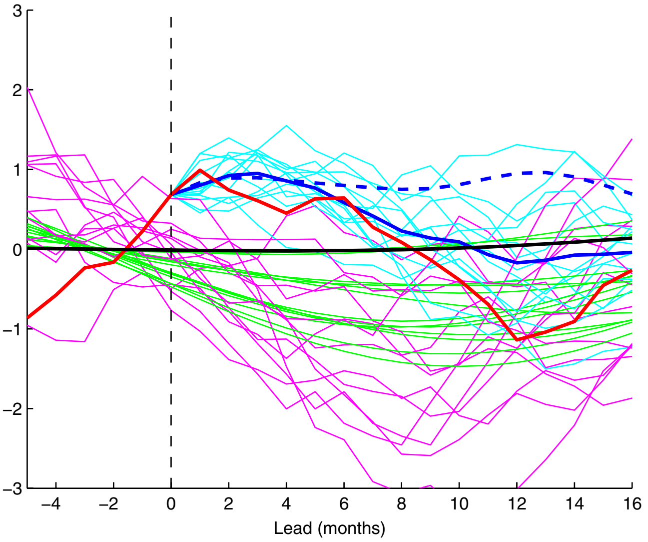

Fig. S1. Principle of the PNF method: The time t^* at which we start forecasting is indicated by the vertical dashed line. All the time series here are centered at time $t^* = 0$. The colors of the trajectories are explained in *SI Text*.

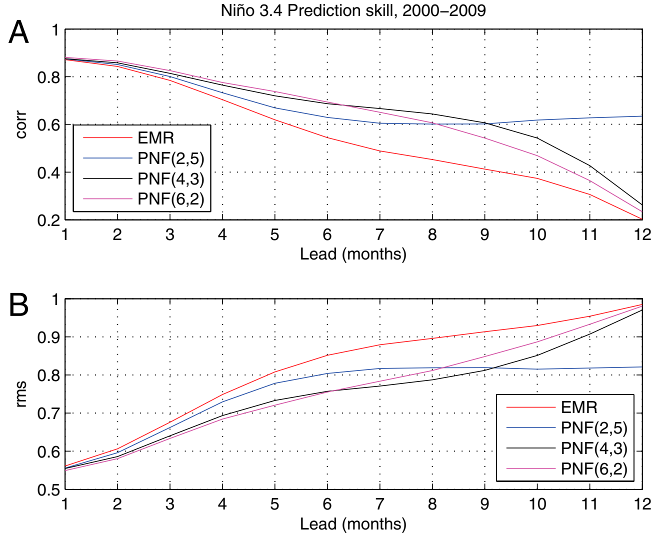

Fig. S2. Skill comparison between PNF(K,Δ) methods as applied to the Niño-3.4 index; the legend is explained in *SI Text*. (*A*) Correlations Corr between forecast and validation; and (*B*) rmsd between the two.

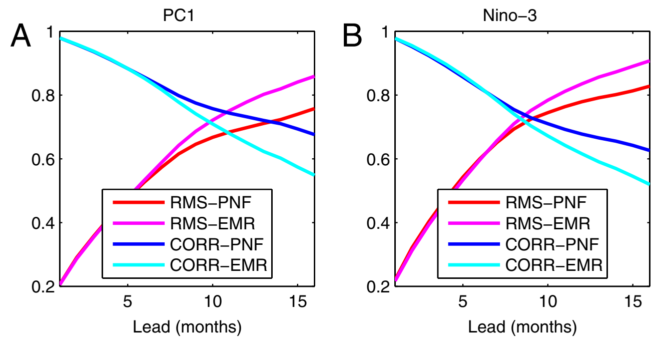

Fig. S3. Skill comparison between EMR predictions and PNF(2,5) predictions in (*A*) PC1, and (*B*) Niño-3 SSTs. The validation interval here is $T_f' = 50$ years long, with $50 \cdot 12 = 600$ monthly values of t^* taken in the interval to make a 16-month prediction out to the respective $t^* + T$; these predictions are made according to either the PNF method or by using a large ensemble average of the EMR–ENSO model. The qualitative improvements obtained by the PNF(2,5) method are independent of the model trajectory used as reference in the validation procedure.

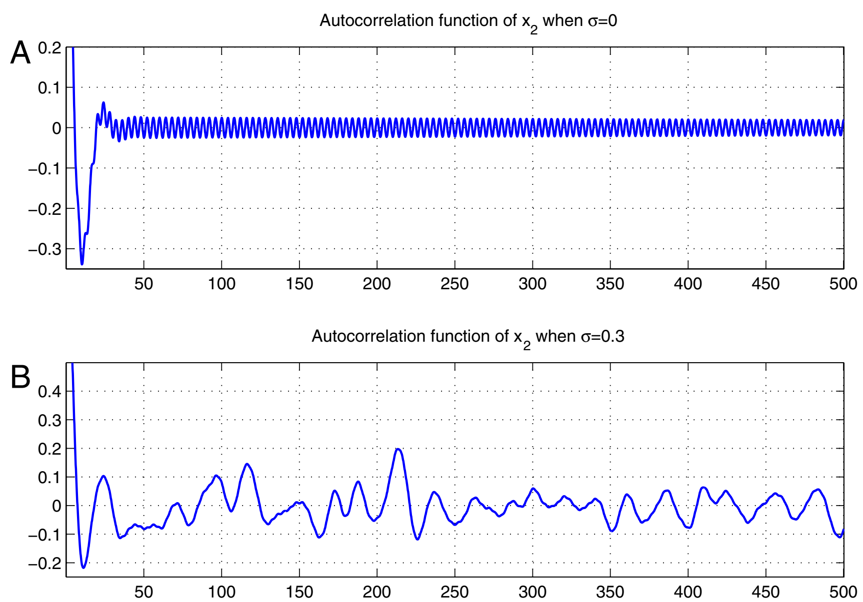

Fig. S4. Autocorrelation functions of the predator population x_2 in system **11**: (*A*) without noise, $\sigma = 0$; and (*B*) with noise, $\sigma = 0.3$. In *B*, the dominant mode corresponds to a broad-band peak, located at a frequency $f' = 0.04$, which stands out above a 99% confidence interval; see Fig. S5.

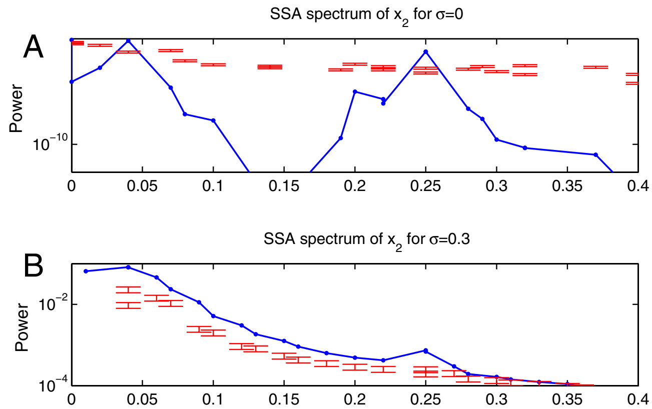

Fig. S5. SSA spectra for the periodically forced system **11**: (*A*) without noise, $\sigma = 0$; and (*B*) with noise, $\sigma = 0.3$. In *A*, the only mode of variability is due to the periodic forcing, $f = 0.25$; in *B*, the dominant mode of variability is located at $f' = 0.04$ and it rises above a "red pair"; see ref. 6.

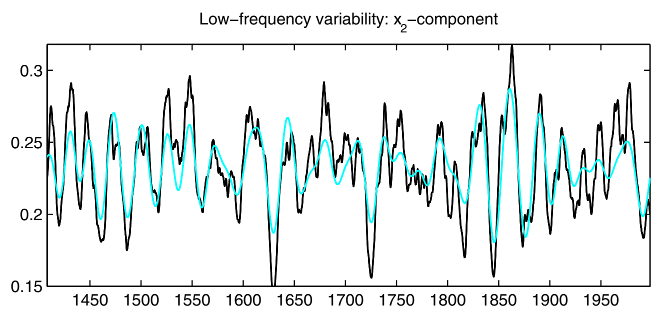

Fig. S6. Time series of the component x_2 (black) and its SSA reconstruction given by RC_2 with the mean of x_2 added (cyan); RC_2 has a dominant period of 25 (nondimensional units) and captures most of the variance in the prey population x_2. This mode is responsible for the LFV present in the signal.

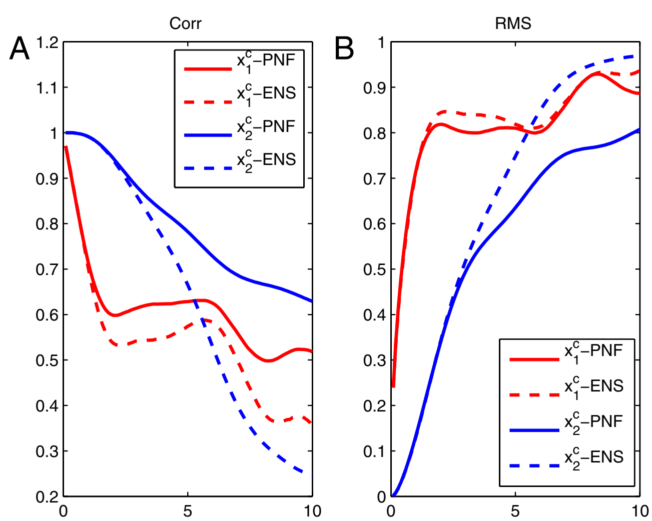

Fig. S7. Skill comparison between ensemble predictions (ENS) and PNF(2,3) predictions for version **11** of the Holling model; the ensemble has $N = 2{,}000$ members: (*A*) Corr, and (*B*) RMS. The validation interval is $T'_f = 590$ units long, from 1,400 to 1,990, with 5,900 values of t^* issued in steps of $\Delta t = 0.1$ throughout the interval, to make a 10-unit prediction out to the respective $t^* + T$. The snippets are selected in the interval that extends from $t = 0$ to $t = 1{,}400$. The variables predicted are the anomalies x_1^c and x_2^c of the prey and predator densities, x_1 and x_2, respectively.

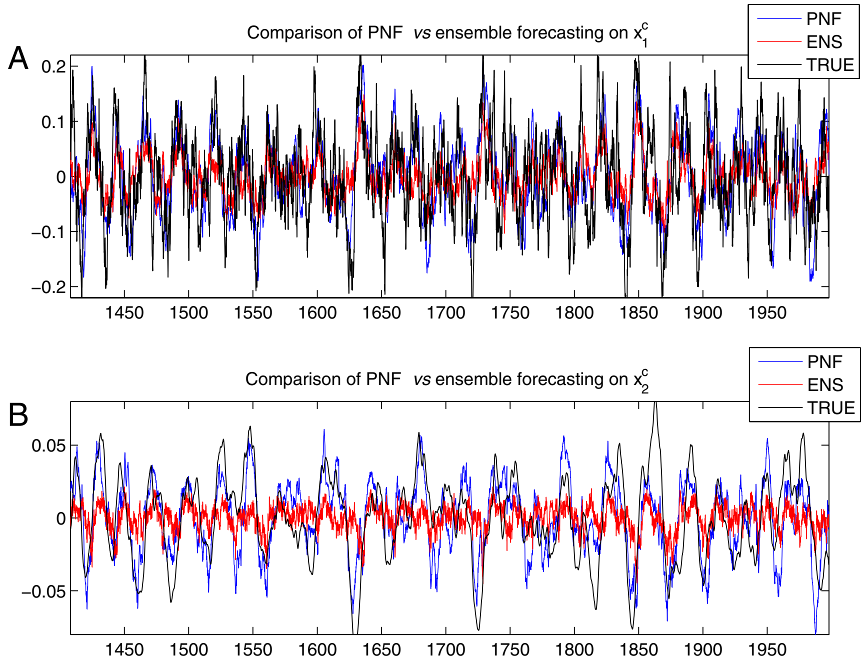

Fig. S8. Comparison between ensemble predictions (ENS) and PNF(2,3) predictions in the time domain: (*A*) for the prey density anomalies x_1^c; and (*B*) the predator density anomalies x_2^c. This plot shows clearly the superior skill of the PNF(2,3) method—for both variables but especially for x_2^c—when compared with brute-force forecasting with a large ensemble. The lead time T in this plot is $T = 8$.

Table S1. Autocorrelations at successive lags of residual noise ξ_t and its reshuffled version $\hat{\xi}_t$

	PC$_1$		PC$_2$	
Lag	ξ_t	$\hat{\xi}_t$	ξ_t	$\hat{\xi}_t$
1	−0.0099	−0.0133	−0.0032	−0.0266
2	−0.1334	0.0049	−0.1289	−0.0199
3	−0.1043	0.0019	−0.0016	0.0679
4	0.0241	−0.0441	0.0025	0.0042
5	0.1177	0.0034	−0.0373	0.0111

The first five integer lags, in units of $t = n\Delta t$; $n = 1,\ldots,5$, with $\Delta t = 1$ month, are listed here for the first two PCs of the SST anomaly field; these two capture about 82% of the total variance of the field.