Sampling past noise to drive ENSO forecasts

D. Kondrashov, M. D. Chekroun and M. Ghil

Atmospheric and Oceanic Sciences Department and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, USA

1. Motivation

- Improve ENSO long-lead forecasts (beyond 7–8 months), unreliable for current state–of–the–art statistical and dynamical models, by exploiting low-frequency-variability (LFV) episodic connection to the driving "noise" in empirical dynamical model.

2. Empirical Model Reduction (EMR)

- EMR methodology attempts to construct low-order nonlinear system of prognostic equations driven by stochastic forcing, and to estimate both the dynamical operator and properties of the driving noise directly from observations or high-level model simulation:

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}(\mathbf{x},\mathbf{x}) + \mathbf{L}(\mathbf{x},\mathbf{r}_t^k,\xi_t,t), \\ \dot{\mathbf{r}}_t^k &= \mathbf{b}_k(\mathbf{x},\mathbf{r}_t^0,...,\mathbf{r}_t^k) + \mathbf{r}_t^{k+1}, k \in \{0,...,K\} \end{aligned} \quad (1)$$

- $\mathbf{x}$ is a state vector, typically principal components of climate fields. $\mathbf{A}$, $\mathbf{B}$, $\mathbf{L}$ are estimated by multiple linear regression (MLR) by using estimated tendencies $\Delta\mathbf{x}$ as predictants. Multi-level modeled multivariate stochastic forcing $\mathbf{r}_t^k$ represents unresolved processes, and the linear maps $\mathbf{b}_k$ are estimated recursively. The number of levels is determined so that the lag-0 covariance of the regression residual $\mathbf{r}_t^K$ converges to a constant matrix $\mathrm{Q} = \mathrm{cov}(\mathbf{r}_t^K)$, while its lag-1 covariance vanishes. The $\mathbf{r}_t^{K+1} = \xi_t$, called the $(K+1)$th level's residual stochastic forcing reconstructs the original time series **exactly**.

3. EMR ENSO model.

- A two-level ($K = 1$) EMR model with seasonal cycle, obtained by using 20 leading PCs of monthly Tropical Pacific SST anomalies (Kondrashov et al. 2005), is highly competitive in intraseasonal ENSO prediction:
  `iri.columbia.edu/climate/ENSO/currentinfo/modelviews.html`
- ENSO quasi-quadriennial (QQ) 4-year and quasi-biennial (QB) 2-year LFV modes are captured by damped oscillatory eigenmodes of linearized dynamical EMR operator, that are excited by noise (aka nonnormal growth perturbations) and interaction with seasonal cycle.

4. Singular spectrum analysis (SSA)

- SSA is a data-adaptive method for spectral estimation and is extension of classic principal components analysis (PCA) in time domain. SSA is based on diagonalizing time-laggged covariance matrix; the set of its eigenvectors is an optimal set of data-adaptive narrowband filters for decomposing the variance within sliding time window. It is particularly well suited for the analysis of time series exhibiting quasi-periodic LFV behavior. The parts of time series that correspond to trends, oscillatory modes or noise can be identified by using reconstructed components (RCs).

4. Past Noise Forecasting (PNF)

- Since EMR estimates the history of the noise ξ_t that ENSO "lives" on, it offers an opportunity to refine ensemble mean of standard EMR prediction, by exploiting pathwise relation of LFV episodes to the driving noise.
- Since ξ_t is stationary, we can parameterize the space of probability Ω by a one-to-one time-dependent family of transformations. If ξ_t is known over $[0, t^*]$, we can use sliding windows of length δ "scanning" the noise over $[0, t^*]$ to derive a set S_{t^*} of noise "snippets", which can serve as new realizations $\xi_t(\omega)$ to compute EMR prediction: $\mathbf{x}(t^*+t, t^*, \mathbf{x}_{t^*}; \omega)$, $t \in [0, \delta]$.
- Note that ensemble means of EMR forced by spatially correlated white noise distributed according $N(0, \mathrm{Q})$, and by ξ_t snippets of S_{t^*}, are the same. Thus we need to refine S_{t^*}!
- To do so, (S1) SSA is applied to select similar episodes in the history of the LFV phase in PC1 w.r.t. the immediate phase at time t^* –start of prediction, and (S2) based on such occurrences, appropriate noise snippets from the "past noise" are selected to obtain prediction of EMR model.
- S1. To find similar LFV episodes, we compute RC_k as sum of k leading SSA RCs of PC1, and find times $t_j < t^* - \Delta$ such that $RC_k(t_j : t_j + \Delta)$ is close enough to $RC_k(t^* - \Delta : t^*)$ in rms error and correlation over continuous time interval of size Δ:

$$\begin{aligned} \mathfrak{T}'_{t^*} := \Big\{ & t_j \in (0, t^* - \Delta) \ : \\ & \mathrm{rms}(RC_k(t_j : t_j + \Delta) - RC_k(t^* - \Delta : t^*)) \le \alpha, \ \text{and}, \\ & \mathrm{corr}(RC_k(t_j : t_j + \Delta), RC_k(t^* - \Delta : t^*)) \ge 1 - \beta \Big\}, \end{aligned} \quad (2)$$

- S2. The refined set of noise "snippets" to compute EMR prediction from time t^* is given by following subset of S_{t^*},

$$S'_{t^*} := \{\zeta^{t_j} \in S_{t^*} : t_j \in \mathfrak{T}'_{t^*}\}; \quad (3)$$

  where each ζ^{t_j} is a copy of length δ of the residual noise ξ_t in the past.
- The PNF prediction for $\delta = 16$ months ahead is given by the mean over ensemble driven by S'_{t^*}.

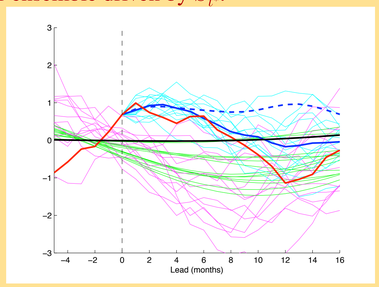

Figure 1: Principle of the PNF method: (vertical dashed) time t^* at which we start prediction, (red) PC1($t^* - \Delta$,$t^* + 16$), with $\Delta = 5$, (**black**) RC_2 of PC1 (that captures most energetic, QQ mode); (green) RC_2 analogues selected by (2); (magenta) corresponding PC1 analogues, (dashed blue) EMR ensemble mean over whole set of snippets S_{t^*}; (cyan) PNF ensemble plume driven by selected subset of snippets S'_{t^*} in (3), $x_1(t^*+t, t^*, \mathbf{x}_{t^*}; \zeta^{t_j}))$; (thick blue) PNF ensemble mean over S'_{t^*}.

5. Numerical Results and Skill

The proof–of–concept PNF prediction consists of fitting EMR model and obtaining full set of noise snippets S_{t^*} for 1950–1999 training period, and performing validation in 2000-2009. Drastic improvement of Niño-3 prediction beyond 6 months by PNF-EMR is due to its ability to predict energetic LFV episodes of QQ+QB in 2000-2009.

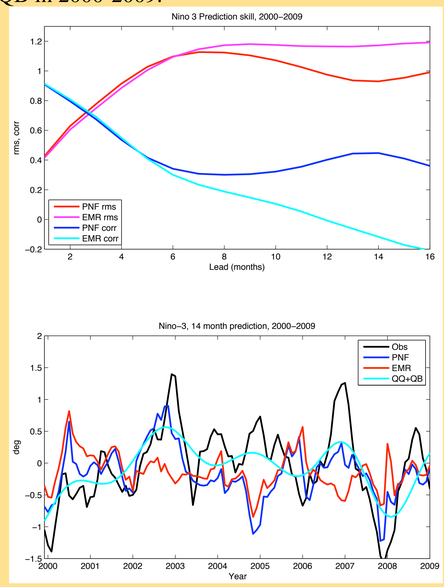

Figure 2: Upper panel: Drastic improvement of Niño-3 prediction skill by PNF. Lower panel: PNF improvement in predicting Niño-3 at 14-month-lead is consistent with energetic phase of low-frequency modes (QQ + QB) components of Niño-3 (cyan), captured by SSA.

Interpretation: In the EMR ENSO model the chaos is "weak" (cf. Fig. 4) and the sensitivity w.r.t. the forcing is reasonably moderate (cf. Fig. 5). The refined set of noise snippets S'_{t^*} corresponds to different initial conditions than at the time t^* of immediate forecasting, but is built on similar LFV phase in the past (training period) preceding this time. Due to the weak long-lead dependence on perturbations in the noise forcing and initial conditions, it is thus natural to expect that *forecasted* trajectories should "synchronize" with *observed* LFV (QQ+QB) in the future at a longer lead as well, thus leading to improvement in prediction skill.

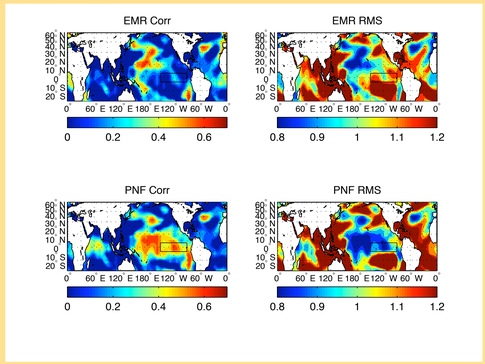

Figure 3: The Niño SST anomaly is defined as the area average over the rectangular box. PNF skill is uniformly better (lower RMS, higher Corr) in the equatorial Tropical Pacific and Indian Ocean, area where ENSO is active the most.

6. Theoretical justification of PNF method: pathwise linear response and weak chaos in EMR ENSO model.

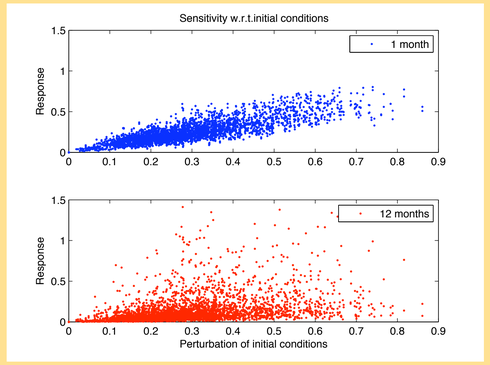

Figure 4: **Weak sensitivity w.r.t. initial data and synchronization by the noise at long lead times**: x-axis is the normalized magnitude $\|\mathbf{x}_0 - \hat{\mathbf{x}}_0\|$ of a perturbation of an initial condition $\mathbf{x}_0$ taken at time s. The y-axis represents normalized difference $\|\mathbf{x}(t, s, \mathbf{x}_0; \omega) - \mathbf{x}(t, s, \hat{\mathbf{x}}_0; \omega)\|$ at 1 month ($t - s = 1$, upper-panel) and at 12 months ($t - s = 12$, lower panel), driven by same noise realization ω.

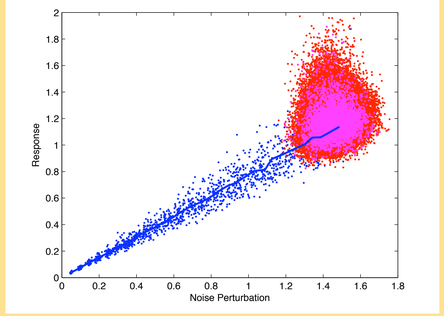

Figure 5: **Pathwise linear response of EMR ENSO model to noise perturbation**: The x-axis represents normalized magnitude of noise perturbation $[\int_s^t \|\chi_u(\omega') - \xi_u(\omega)\| du]$. The blue dots represent the normalized response of the EMR model'solutions $\|\mathbf{x}(t, s, \mathbf{x}_s; \omega) - \mathbf{x}(t, s, \mathbf{x}_s; \omega')\|$ integrated over $t - s = 16$ months, with initial condition $\mathbf{x}_s$ taken at time s, for arbitrary perturbations $\chi_t(\omega') = \xi_t(\omega) + \epsilon\xi_t(\omega')$ ($\epsilon > 0$) of the residual noise from its training period. The red dots correspond to the model's response $\|\mathbf{x}(t^*+t, t^*, \mathbf{x}_{t^*}; \omega') - \mathbf{x}(t^*+t, t^*, \mathbf{x}_{t^*}; \omega)\|$ integrated over $t = 16$ months and over the validation period, for specific perturbations $\chi_t(\omega') \in S_{t^*}$, *i.e.* from the full set of noise snippets of the training period; where $\xi_t(\omega)$ is now the residual noise of the validation period. The magenta dots are the same as the red ones except that $\chi_t(\omega') \in S'_{t^*}$, *i.e.* the response is assessed from the refined set of noise snippets obtained by the PNF procedure (cf. S1 & S2) in 4.

7. References

1. Kondrashov, D., S. Kravtsov, A. W. Robertson, and M. Ghil (2005), A hierarchy of data-based ENSO models, *J. Climate*, *18*, 4425–4444.

2. Chekroun, M.D., E. Simonnet, and M. Ghil (2009), Stochastic climate dynamics: Random attractors and time-dependent invariant measures, submitted.

3. Ghil M. *et al*, (2002), Advanced spectral methods for climatic time series, *Rev. Geophys., 40(1)*, 3.1–3.41.

This work was supported by U.S. Department of Energy grants DE-FG02-07ER64429 and DE-FG02-07ER64439 at UCLA; all grants are from the DOE's Climate Modeling Programs, Office of Science (BER).