Nonlinear Multivariate Projections and long range ENSO predictability

Carlos Lima[1], Upmanu Lall[2], Tony Jebara[3] and Tony Barnston[4]

1 Federal Univ of Brasilia, www.columbia.edu/~chr2107
2 Columbia University, www.columbia.edu/~ula2, water.columbia.edu
3 Columbia University, www.cs.columbia.edu/~jebara/
4 Columbia University, iri.columbia.edu/staff/tonyb

Lima, Carlos H. R., Upmanu Lall, Tony Jebara, Anthony G. Barnston, 2009: Statistical Prediction of ENSO from Subsurface Sea Temperature Using a Nonlinear Dimensionality Reduction. *J. Climate*, **22**, 4501–4519. doi: http://dx.doi.org/10.1175/2009JCLI2524.1

Motivation

- ENSO is often understood as a nonlinear dynamical system, yet most multivariate analyses of SSTs are linear
- Upper ocean heat content is considered to be a carrier of the longer term “ENSO signal”
- Can a nonlinear analysis of the spatially distributed thermocline depth data enlighten us as to the space-time evolution of this signal, and help improve long range prediction?

Data

Thermocline

- Thermocline Data
  - Derived from a model-based ocean analysis system (Ji et al., 1995; Ji and Smith, 1995; Behringer et al., 1998)
  - Tropical Pacific (bounded by 26°N and 28°S)
  - Period from Jan 1980 to Nov 2007
  - 4541 grid cells (335 months)
  - Available at

    http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.EMC/.CMB/.Pacific/.monthly/.D20eq
- NINO3 index
  - Kaplan et al. (1998); Reynolds and Smith (1994))
  - *http://iridl.ldeo.columbia.edu/SOURCES/.Indices/.nino/.EXTENDED/.NINO3/*

PCA, Kernel PCA and Maximum Variance Unfolding

$$
\begin{aligned}
\mathbf{X}^{\mathbf{T}} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \quad &= \text{centered matrix of inputs with } N \text{ points in } \Re^M \\
\mathbf{C} = \mathbf{X}^{\mathbf{T}}\mathbf{X} \quad &= \; MxM \text{ covariance matrix} \\
\mathbf{U} \quad &= \; MxL \text{ eigenvector matrix of } \mathbf{C} \\
\mathbf{Y} = \mathbf{U}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}} \quad &= LxN \text{ matrix of main modes}
\end{aligned}
$$

When $M >> N$, we can use SVD:

$\mathbf{X}^{\mathbf{T}} = \mathbf{U}\Sigma\mathbf{V}^{\mathbf{T}} \rightarrow \mathbf{Y} = \Sigma\mathbf{V}^{\mathbf{T}}$ where

$$
\begin{aligned}
\mathbf{V} \quad &= NxL\text{matrix of eigenvectors of the Gram matrix } \mathbf{G} = \mathbf{X}\mathbf{X}^{\mathbf{T}} \\
\Sigma \quad &= \text{Diag. matrix of sq. roots of the top } L \text{ eigenvectors of } \mathbf{G}
\end{aligned}
$$

PCA, Kernel PCA and Maximum Variance Unfolding

If $\mathbf{X}$ shows nonlinear behavior: PCA may not lead to reliable results
Possible alternative : seek nonlinear transformations of $\mathbf{X}$ so that feature space is linear. Consider then a feature space $\mathcal{H}$ and a nonlinear mapping function Φ:

$$\begin{aligned} \Phi &: \Re^M \rightarrow \mathcal{H} \\ \mathbf{x}_i &\longmapsto \Phi(\mathbf{x}_i), \quad i = 1, \cdots, N. \end{aligned}$$

$\Phi(\mathbf{x}_i)$ can be defined by any nonlinear basis function (e.g. $\Phi(\mathbf{x}_i) = \mathbf{x}_i^2$)
Idea: apply PCA in the space defined by $\Phi(\mathbf{X})$ rather than $\mathbf{X}$:
$\Phi(\mathbf{X})^\mathbf{T} = \mathbf{U}\Sigma\mathbf{V}^\mathbf{T}$

PCA, Kernel PCA and Maximum Variance Unfolding

However, $\mathcal{H}$ can have a very large dimension depending on $\Phi \rightarrow$ not computationally feasible

Solution: Kernel Trick $\rightarrow$ do not need to compute the mapping explicitly, but only the dot products, e.g. for $\Phi(\mathbf{w}) = \mathbf{w^2}$:
$K(\mathbf{w}, \mathbf{z}) = \Phi(\mathbf{w}) \cdot \Phi(\mathbf{z}) = (\mathbf{w} \cdot \mathbf{z})^2$
Hence
$K_{i,j} = \Phi(\mathbf{x}_i)\Phi(\mathbf{x}_j)^\mathbf{T}$

Principal components of $\mathbf{X}$ are obtained similarly to PCA, but substituing the Gram matrix $\mathbf{G}$ of the original space by the kernel function K.

PCA, Kernel PCA and Maximum Variance Unfolding

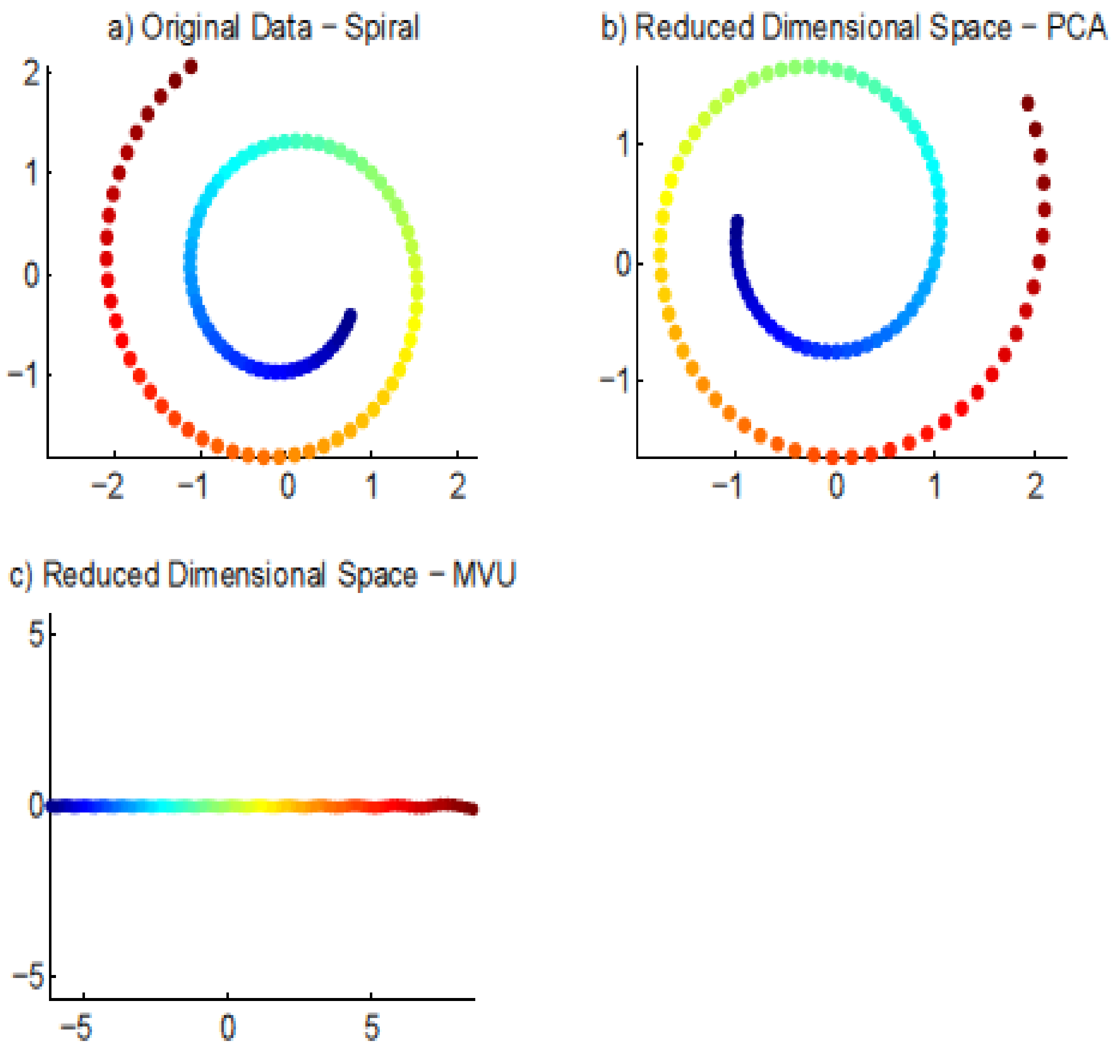
Variants of Kernel PCA: Maximum Variance Unfolding (Weinberger et al, 2004) and Minimum Volume Embedding (Shaw and Jebara, 2007).

Question: Given n high dimensional inputs $\mathbf{x_i} \in \Re^p, i = 1, \ldots, n$, how can we compute outputs $\mathbf{y_i} \in \Re^d$, where $d < p$, such that nearby points remain nearby and distant ones remain distant?

Maximize Trace(K) s.t.:
$K \succeq 0.$
$\sum_{ij} K_{ij} = 0.$

$$K_{ii} + K_{jj} - K_{ij} - K_{ji} = G_{ii} + G_{jj} - G_{ij} - G_{ji},$$
$$\forall i, j \rightarrow \eta_{ij} = 1 \text{ or } [\eta^T \eta]_{ij} > 0.$$

Example

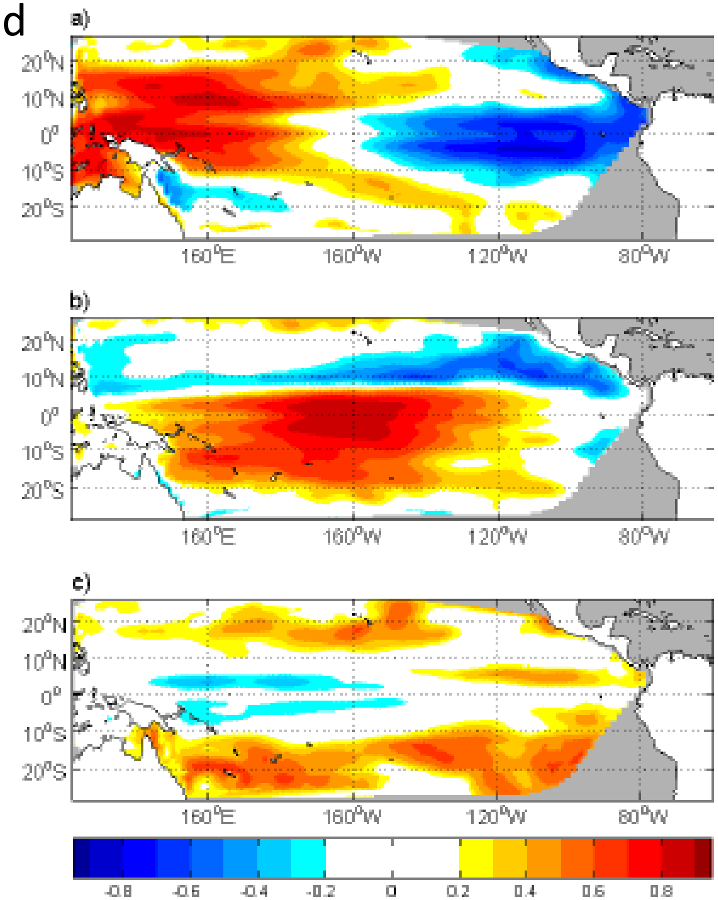

a) Original Data – Spiral
2
1
0
−1
−2 −1 0 1 2
b) Reduced Dimensional Space – PCA
1
0
−1
−1 0 1 2
c) Reduced Dimensional Space – MVU
5
0
−5
−5 0 5

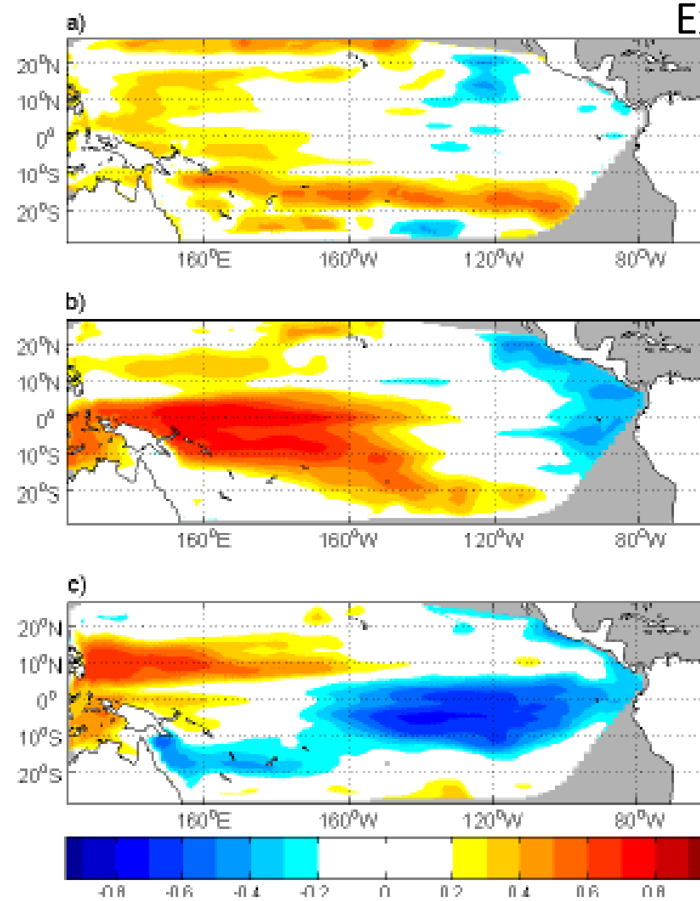
Results

Thermocline Depth Data

Spatial Structure

Variance Explained	PCA	MVU	Variance Explained
23%			52%
18%			25%
7%			12%
Total 48%			**Total 89%**

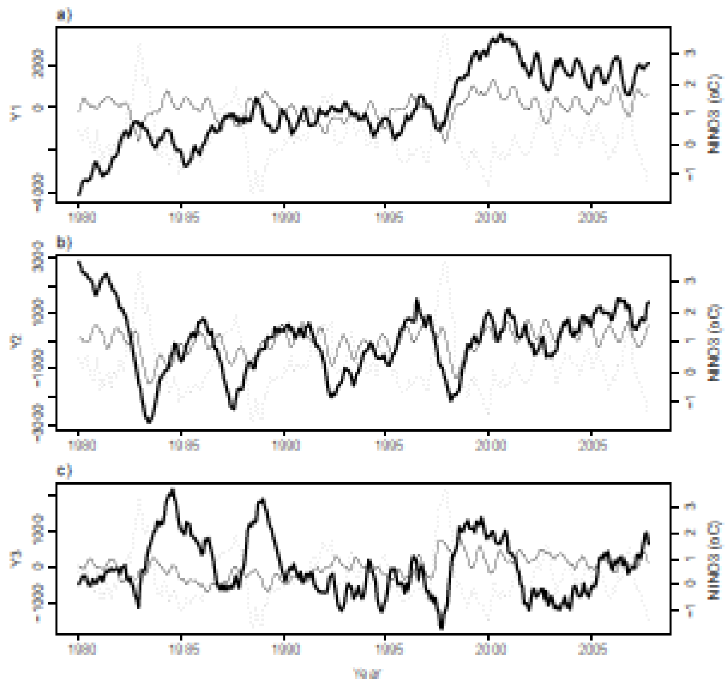




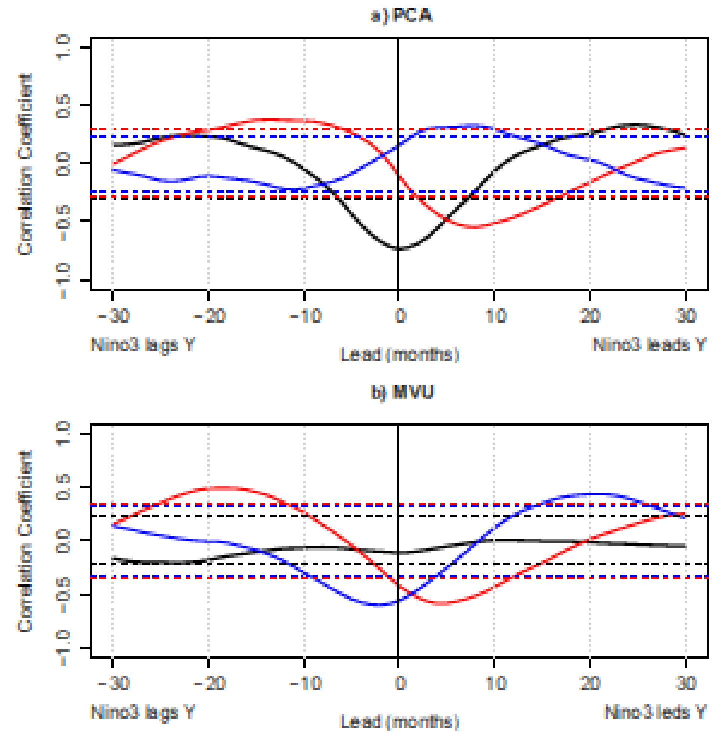
Results

NINO3 Correlations

Series



Cross-Correlation Function



1st 2nd 3rd

PC1 peak corr lag 0 and MV3 leads by 2 months
PC2 leads by 11 months, MV2 **by 20 months**
PC3, PC1 paired?, MV1 weak, persistent correlation

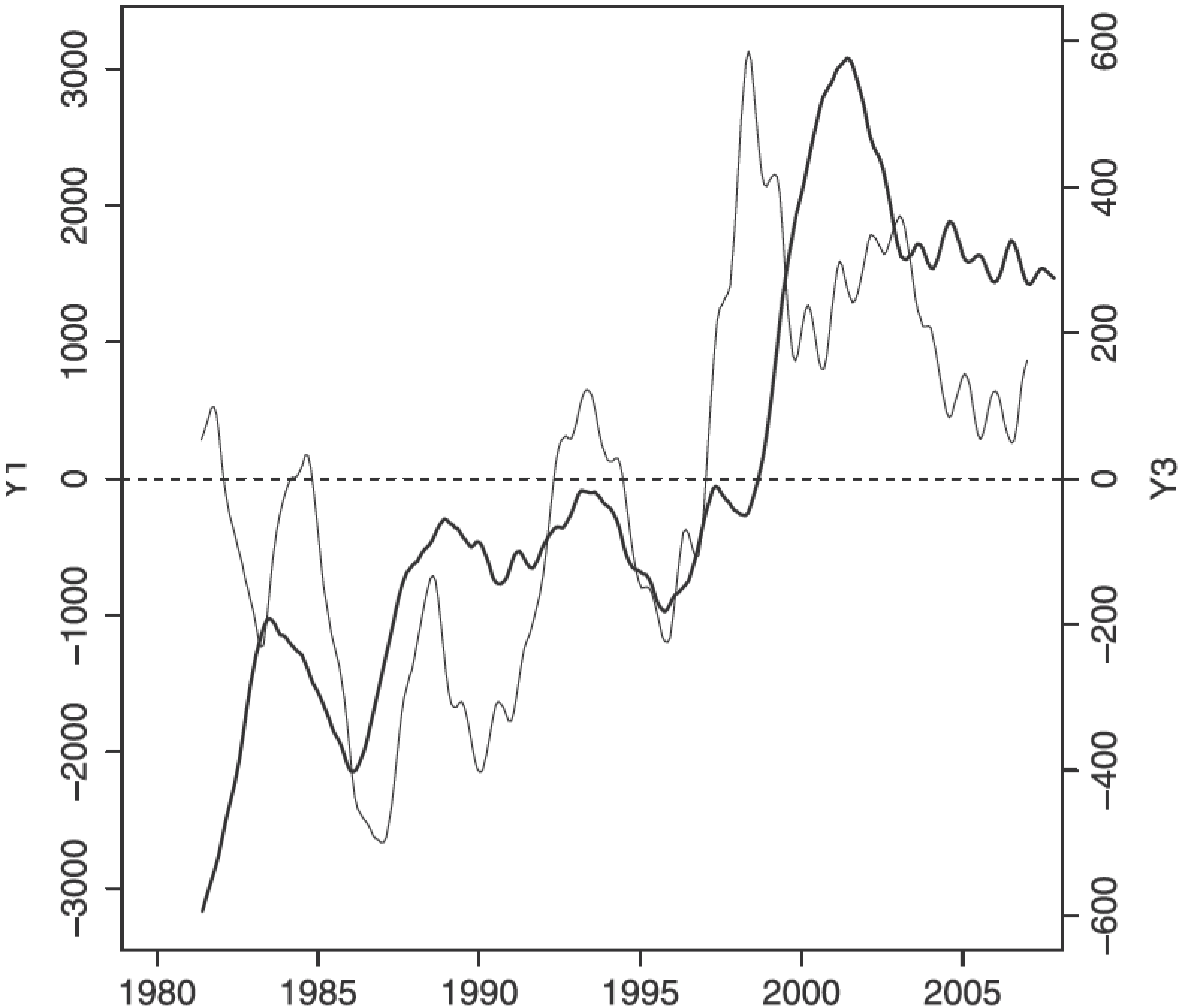


FIG. 6. First MVU leading mode (black thick line; scale on the left y axis) and third PC (black thin line; scale on the right y axis) smoothed by an 18-month filter.

MV1 – thick line
PC3 – think line
Both **filtered** at 18 months to emphasize low frequency variability

Indicative of change in the baseline state

MV1 carries much more variance

Kim, Baek-Min, Soon-Il An, 2011: *J. Climate*, **24**, 1438–1450. suggest base SST increase leads to period doubling bifurcation and amplitude modulation in ENSO dynamics

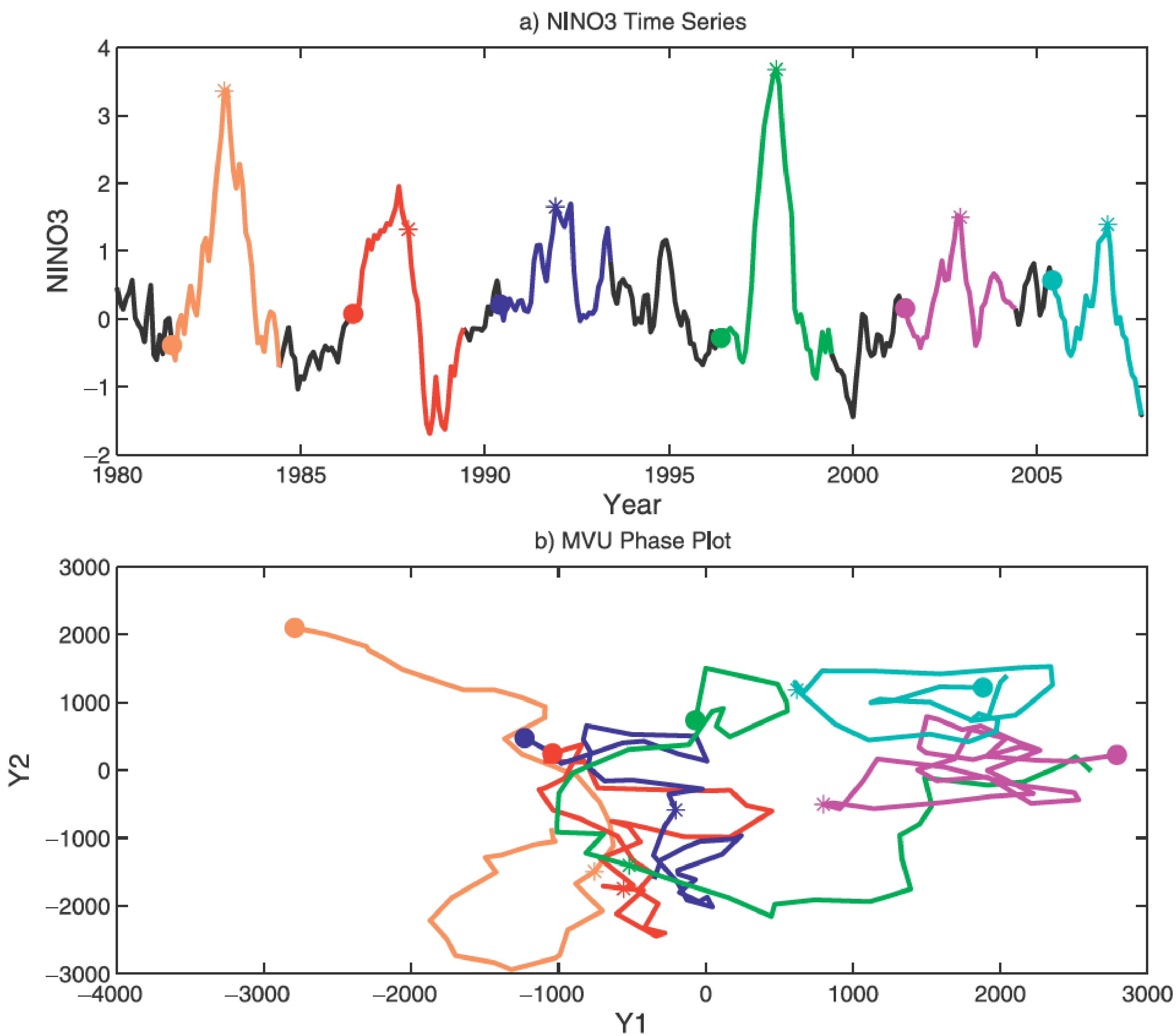


FIG. 7. (a) Niño-3 time series and (b) phase plot of the first two MVU components $\mathbf{y}_1$ and $\mathbf{y}_2$. The asterisks indicate the six El Niñc events (December 1982, December 1987, December 1991, December 1997, December 2002, and December 2006) during the period ol 1980–2007. Solid circles denote the phasing 18 months before those El Niño events took place.

The recent ENSO events are separated in the phase plot of the two leading MVU modes, reflecting potentially different underlying dynamics

Consistent with recent notions of the different flavors of ENSO

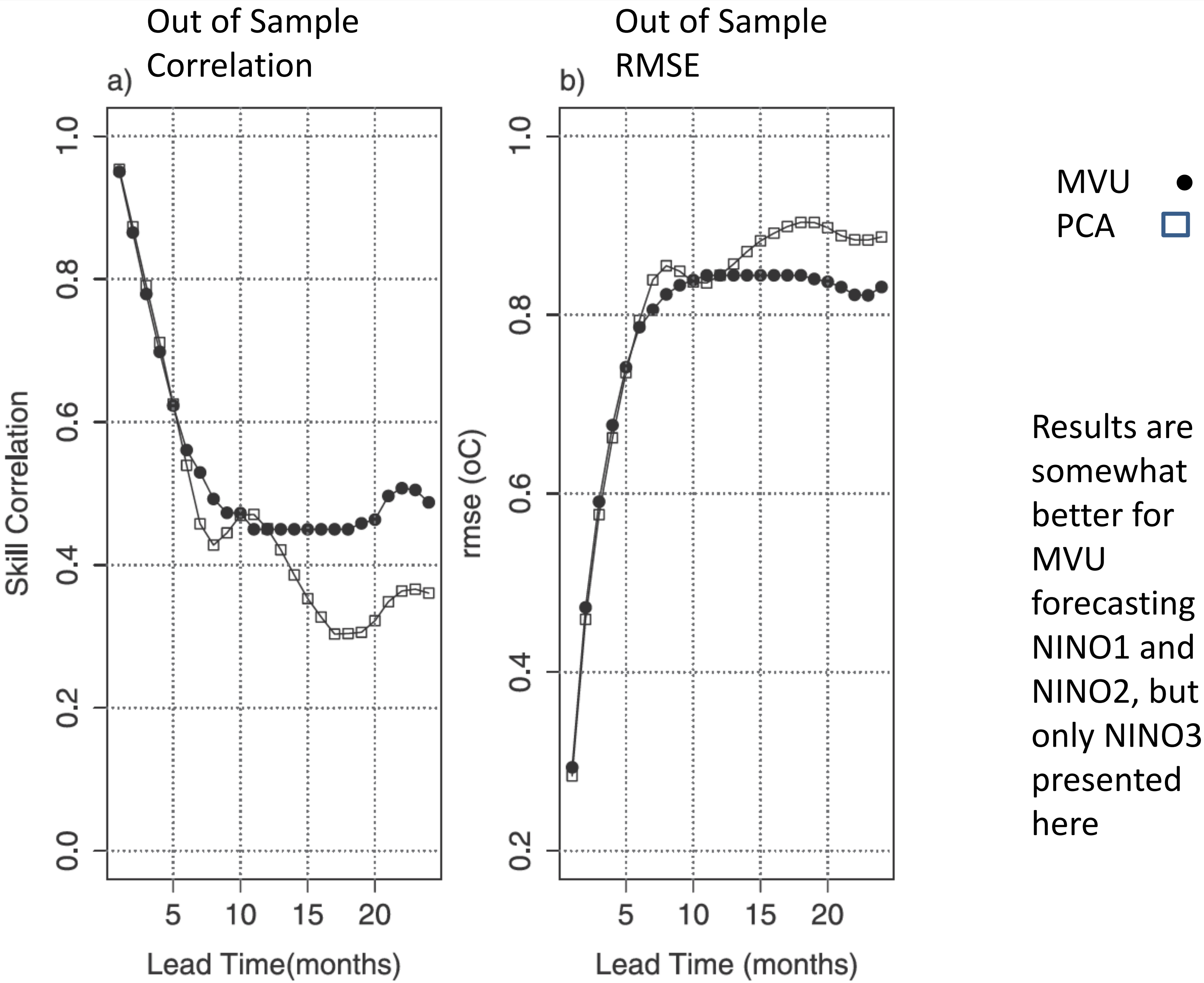
Forecast Results

Model

$$\text{NINO3}(t) = \begin{cases} f(\text{NINO3}(t-\tau), y_1(t-\tau), y_2(t-\tau), y_3(t-\tau), y_2(t-18)) + \epsilon_t, & \tau < 18 \\ f(\text{NINO3}(t-\tau), y_1(t-\tau), y_2(t-\tau), y_3(t-\tau)) + \epsilon_t, & 18 \leq \tau \leq 24 \end{cases}$$

- Simple linear regression
- Model selection through 10-fold cross-validation scheme
- For each lead time τ, one set of predictors that minimizes the RMSE across all models

Forecast Results



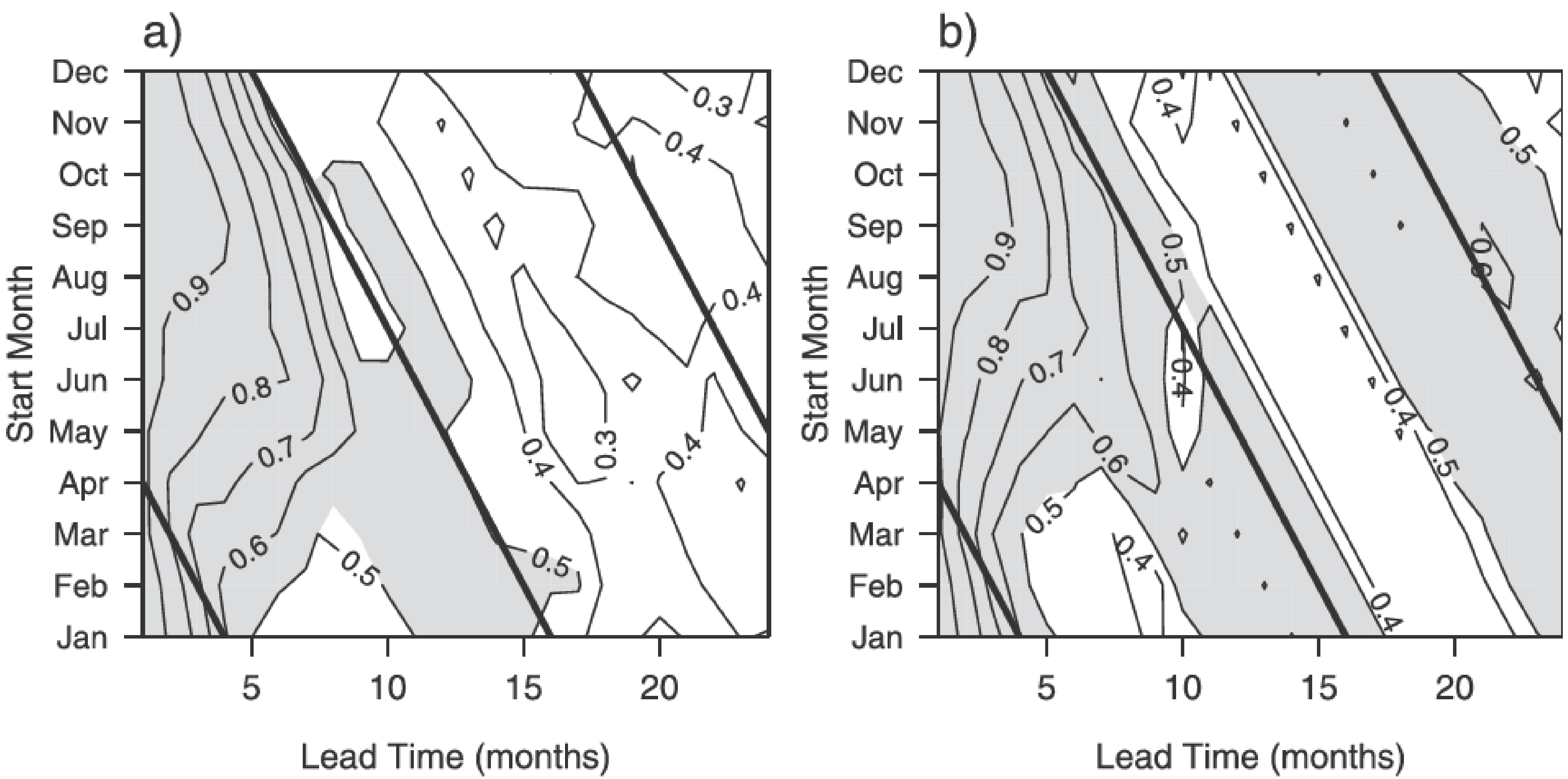
Results are somewhat better for MVU forecasting NINO1 and NINO2, but only NINO3 is presented here

Forecast Results

Correlation of Forecasts and Observations as a Function of Start Month and Lead Time

PCA Based

MVU Based



MVU skill seems to be less dependent on starting month, especially for > 1 year forecasts

Conclusions and Future Research

- The nonlinear modes are quite different from the linear modes in spatial and temporal expression
  - though there are similarities, the patterns that explain the most variance in the field are different
- NINO3 prediction in the 1st year is quite similar for both, though the MVU based modes do not show differences in predictability starting from different months for a given lead time of prediction, and hence do not run into the spring barrier for prediction that the PCA modes suffer from
- NINO3 prediction skill for MVU based predictors in the second year is consistently superior to that from PCA based predictors.
- The recharge-discharge oscillator theory of ENSO appears to be supported by both analyses. However, the MVU appears to better separate the different flavors of ENSO, and may also be better at revealing a pronounced recent trend or shift in ENSO dynamics. These aspects call for further analytical and theoretical study of ENSO dynamics.