# Linear-Regression-based Models of Nonlinear Processes

Sergey Kravtsov

Department of Mathematical Sciences, UWM

Collaborators:

Dmitri Kondrashov, Andrew Robertson, Michael Ghil

# Nomenclature

Response variable:

$$\{y^{(n)}\}\ (1 \le n \le N) \equiv \{y^{(1)}, \ldots, y^{(N)}\}$$

Predictor variable:

$$\{x^{(n)}\}\ (1 \le n \le N) \equiv \{x^{(1)}, \ldots, x^{(N)}\}$$

- Each $y^{(n)}$ is normally distributed about $\widehat{y}^{(n)}$
- Each $x^{(n)}$ is known exactly

$$\widehat{y} = f(x;\ a_1, \ldots, a_J)$$ – known dependence

REGRESSION: FIND $\{a_j\}\ (1 \le j \le J)$

# Linear Regression – I

$$\chi^2 \equiv \sum_{n=1}^{N} \frac{[y^{(n)} - \hat{y}(x^{(n)}; a_1, a_2, \ldots, a_J)]^2}{\sigma_n^2}$$

- The set of $\mathbf{a} \equiv \{a_1, \ldots, a_J\}$ which minimizes $\chi^2$ is the maximum likelihood estimate of the regression parameters
- Assume $\hat{y}(x^{(n)}; a_1, a_2, \ldots, a_J)$ is linear in ***a***: General Linear Least-Squares
- If $\hat{y}$ is also linear in *x*: Linear Regression

# Linear Regression–II

$$x^{(n)} = \overline{x} + x'^{(n)}; \qquad y^{(n)} = \overline{y} + y'^{(n)}$$

Assume $\widehat{y} = bx + a$ and minimize $\chi^2$:

$$\chi^2 = \frac{1}{N} \sum_{n=1}^{N} \left(y^{(n)} - \widehat{y}(x^{(n)})\right)^2 \equiv \overline{y^{*\,2}}$$

Normal Equations:

$$\begin{aligned} a + b\bar{x} &= \bar{y} \\ a\bar{x} + b\overline{x^2} &= \overline{xy} \end{aligned}$$

# Linear Regression–III

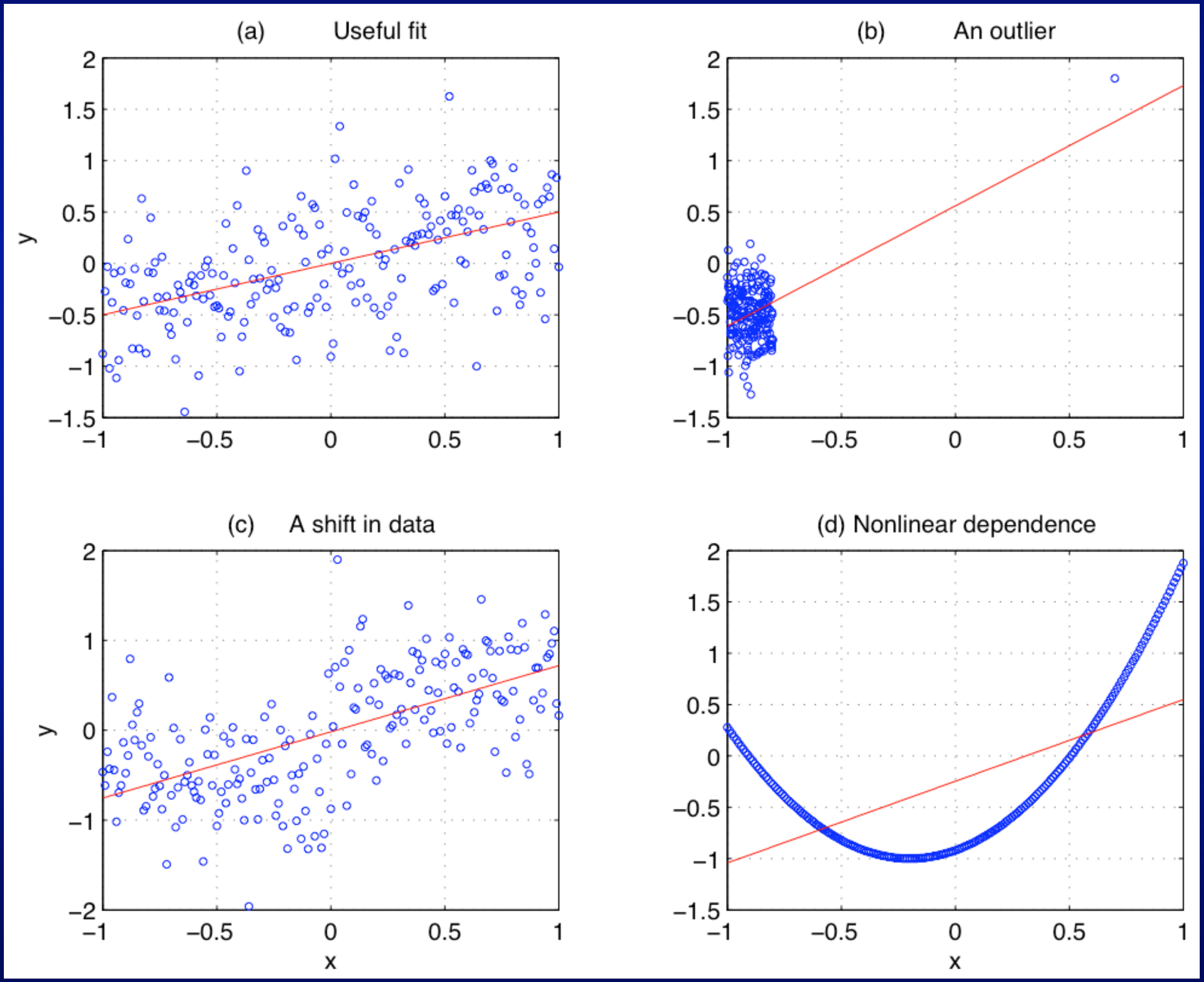
$$1 = \frac{\overline{y^{*2}}}{\overline{y'^2}} + b^2 \frac{\overline{x'^2}}{\overline{y'^2}}; \qquad b = \overline{x'y'}/\overline{x'^2}$$

$$1 = \frac{\overline{y^{*2}}}{\overline{y'^2}} + \frac{(\overline{x'y'})^2}{\overline{x'^2}\,\overline{y'^2}} = \frac{\overline{y^{*2}}}{\overline{y'^2}} + r^2,$$

$$r \equiv \frac{\overline{x'y'}}{\sqrt{\overline{x'^2}}\sqrt{\overline{y'^2}}}$$

– Correlation Coefficient

# Linear Regression–IV



# Multiple Linear Regression–I

- Centered and normalized variables:

$$y^{(n)} \longrightarrow (y^{(n)} - \bar{y})/\sqrt{\overline{y'^2}} \qquad x_j^{(n)} \longrightarrow (x_j^{(n)} - \bar{x}_j)/\sqrt{\overline{x_j'^2}}$$

- Model: $\widehat{y} = a_1 x_1 + \ldots + a_J x_J$

- Normal equations:

$$\sum_{j=1}^{J} r_{kj} a_j = r_k; \qquad r_{kj} \equiv \overline{x_k' x_j'}; \; r_k \equiv \overline{x_k' y'}$$

# Multiple Linear Regression–II

- Two predictors (note degeneracy at $r_{12}$=1):

$$a_1 = \frac{r_1 - r_{12} r_2}{1 - r_{12}^2}; \qquad a_2 = \frac{r_2 - r_{12} r_1}{1 - r_{12}^2}$$

- Multiple correlation coefficient:

$$R^2 = \frac{r_1^2 + r_2^2 - 2 r_1 r_2 r_{12}}{1 - r_{12}^2}$$

$$1 = \frac{\overline{y^{*2}}}{\overline{y'^2}} + R^2$$

- Minimum useful correlation:

$$R^2 > r_1^2 \longrightarrow |r_2| > |r_2^*| \equiv |r_1 r_{12}|$$

# General Linear Least-Squares

$$\mathbf{X} \equiv \begin{pmatrix} \frac{X_1(x^{(1)})}{\sigma_1} & \frac{X_2(x^{(1)})}{\sigma_1} & \cdots & \frac{X_J(x^{(1)})}{\sigma_1} \\ \frac{X_1(x^{(2)})}{\sigma_2} & \frac{X_2(x^{(2)})}{\sigma_2} & \cdots & \frac{X_J(x^{(2)})}{\sigma_2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{X_1(x^{(N)})}{\sigma_N} & \frac{X_2(x^{(N)})}{\sigma_N} & \cdots & \frac{X_J(x^{(N)})}{\sigma_N} \end{pmatrix} \qquad \tilde{\mathbf{y}} \equiv \begin{pmatrix} \frac{y^{(1)}}{\sigma_1} \\ \frac{y^{(2)}}{\sigma_2} \\ \cdots \\ \frac{y^{(N)}}{\sigma_N} \end{pmatrix}$$

$$\mathbf{a} \equiv \begin{pmatrix} a_1 & a_2 & \ldots & a_J \end{pmatrix}$$

- Minimize: $\chi^2 = |\tilde{\mathbf{y}} - \mathbf{X} \cdot \mathbf{a}|^2$

# Regularization via SVD

$$\mathbf{X} = \mathbf{U} \cdot [\text{diag}\ (w_j)] \cdot \mathbf{V}^\top$$

$$\mathbf{X}^{-1} = \mathbf{V} \cdot [\text{diag}\ (1/w_j)] \cdot \mathbf{U}^\top$$

- Least-squares "solution" of $\mathbf{X} \cdot \mathbf{a} = \tilde{\mathbf{y}}$ is

$$\mathbf{a} = \mathbf{V} \cdot [\text{diag}\ (1/w_j)] \cdot (\mathbf{U}^\top \cdot \tilde{\mathbf{y}})$$

- "Principal Component" regularization:

For small $w_j$, set $1/w_j$ to zero!

# Partial Least-Squares

- Involves rotated principal components (PCs): [orthogonal transformation looks for "optimal" linear combinations of PCs]
- "Optimal" = (i) rotated PCs are nearly uncorrelated
  (ii) maximally correlated with response
- Rotation is done in a subspace of *N* leading PCs; *N* is determined by cross-validation
- Canned packages available
- Performs better than PCR on large problems

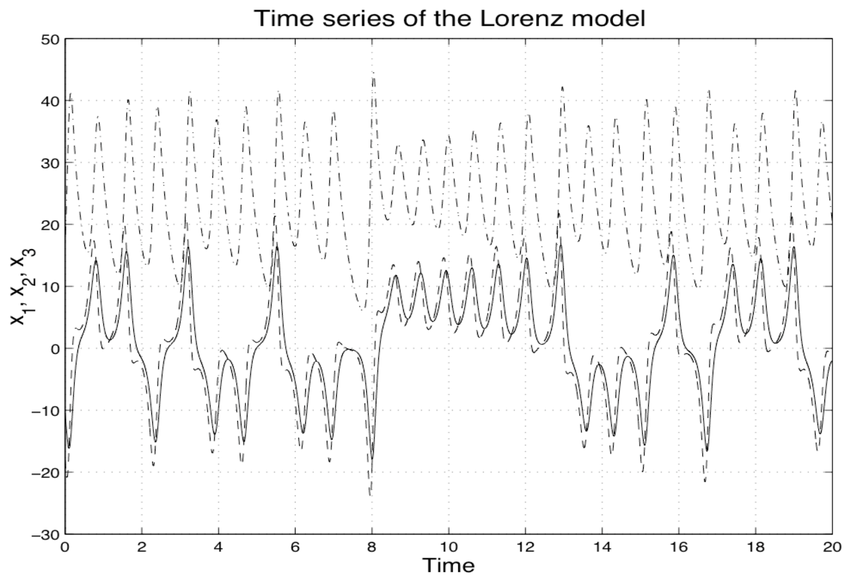
# Didactic Example–I (Lorenz '63)

$$\dot{x}_1 = -sx_1 + sx_2,$$

$$\dot{x}_2 = -x_1x_3 + rx_1 - x_2,$$

$$\dot{x}_3 = x_1x_2 - bx_3.$$

$$s = 10;\ r = 28;\ b = 8/3$$




$$y_i^{(n)} \equiv (x_i^{(n+1)} - x_i^{(n)})/\Delta t$$

$$\widehat{y}_i = a_{0,i} + \sum_{j=1}^{3} a_{j,i}x_j + \sum_{j=1}^{3}\sum_{k\geq j} \tilde{a}_{jk,i}x_jx_k$$

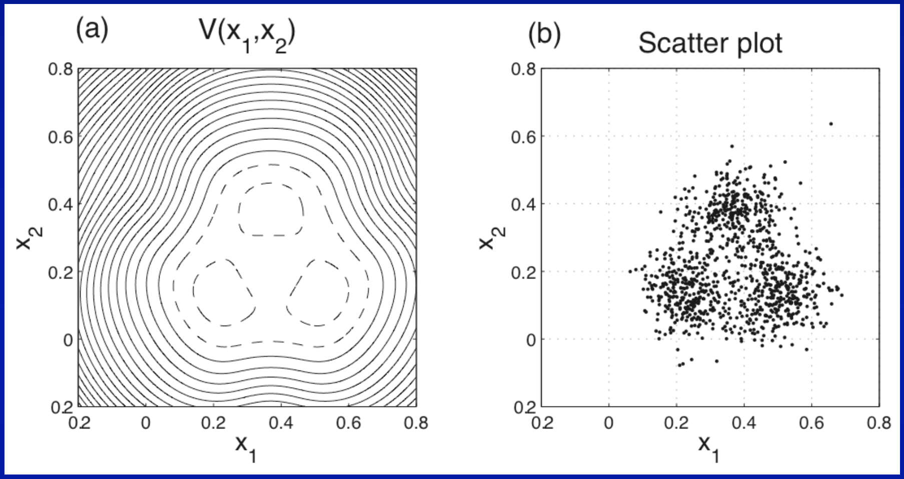
# Lorenz-63 Example (cont'd)

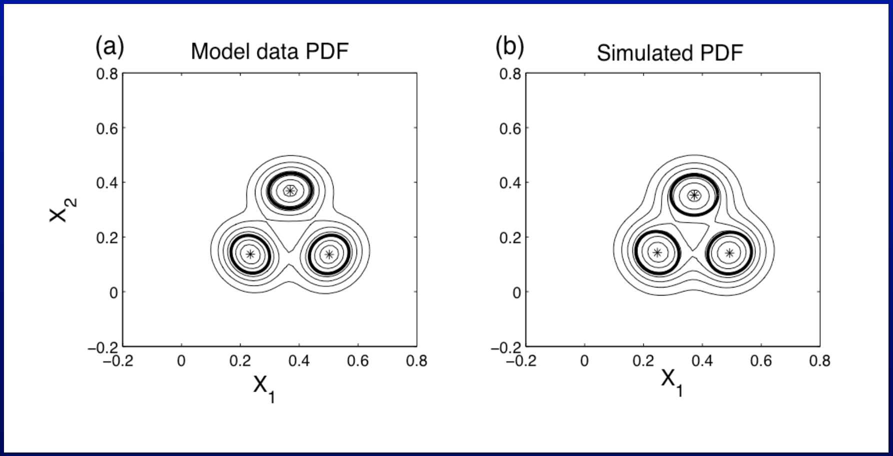
- Given short enough $\Delta t$, coefficients of the Lorenz model are reconstructed with a good accuracy for sample time series of length as short as $T \approx 1$
- These coefficients define a model, whose long integration allows one to infer correct long-term statistics of the system, e.g., PDF (application #1)
- Employing PCR and/or PLS for short samples is advisable
- Hereafter, we will always treat regression models as maps (<u>discrete time</u>), rather than flows (conti-nuous time). <u>Exception</u>: Linear stability analysis

# Didactic Example–II (Triple well)

$$d\mathbf{x}(t) = -\nabla V(\mathbf{x})dt + \sigma\, d\mathbf{b}$$

- $V(x_1,x_2)$ is <u>not polynomial</u>
- <u>Polynomial</u> regression model produces time series, whose statistics are nearly identical to those of the full model
- Regularization required for polynomial models of order $\geq 5$



# Multi-level models – I

- Motivation: serial correlations in the residual

Main (0) level: $(x^{n+1} - x^n)/\Delta t = a_{x,\,0} x^n + r_0^n$

Level 1: $(r_0^{n+1} - r_0^n)/\Delta t = a_{x,\,1} x^n + a_{r_0,\,1} r_0^n + r_1^n$

… and so on …

Level *L*: $r_{L-1}^{n+1} - r_{L-1}^n = \Delta t [a_{x,\,L} x^n + \ \ldots] + \Delta r_L$

- $\Delta r_L$ – Gaussian random deviate with appropriate var.
- If suppress dependence on *x* in levels *1–L*, then the above model is formally identical to an ARMA model

# Multi-level models – II

- Multiple predictors: *N* leading PCs of the field(s) of interest (PCs of data matrix, not design matrix!)

- Response variables: one-step [sampling interval] time differences of predictors

- Each response variable is fit by an *independent* multi-level model. The main level is polynomial in predictors; all others – linear

# Multi-level models – III

- Number of levels *L* is such that each of the last-level residuals (for each channel corresponding to a given response variable) is "white" in time
- Spatial (cross-channel) correlations of the last-level residuals are retained in subsequent regression-model simulations
- Number of PCs (*N*) is chosen to optimize the model's performance
- PLS is used at the main (nonlinear) level of each channel

# NH LFV in MM’93 Model – I

Model (Marshall and Molteni 1993):

- Global QG, T21, 3-level with topography; perpetual-winter forcing; ~1500 degrees of freedom
- Reasonably realistic in terms of LFV (multiple planetary-flow regimes and low-frequency [submonthly-to-intraseasonal] oscillations)
- Extensively studied: A popular laboratory tool for testing out various statistical techniques

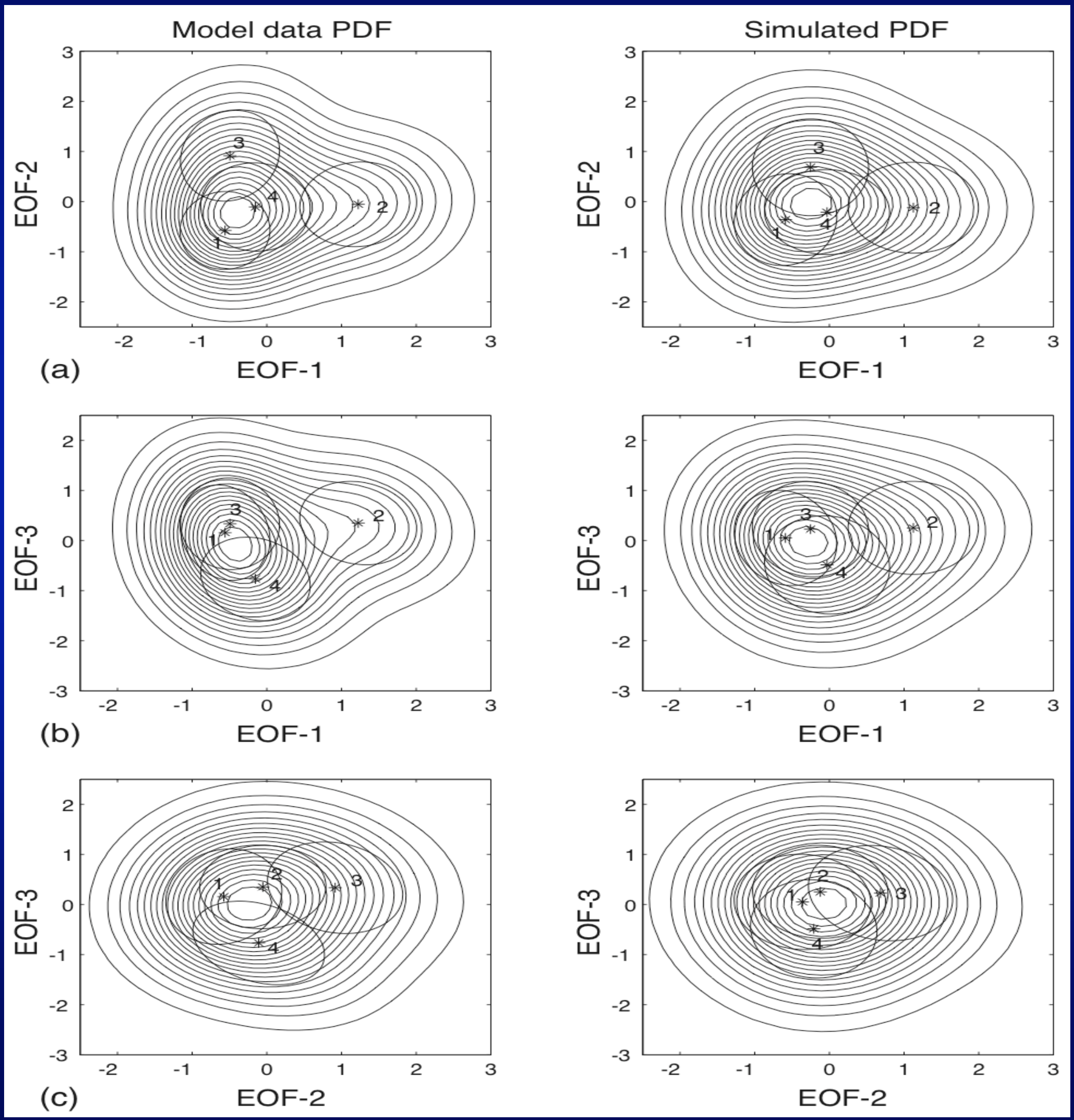
# NH LFV in MM'93 Model – II

Output: daily streamfunction ($\Psi$) fields (≈ $10^5$ days)

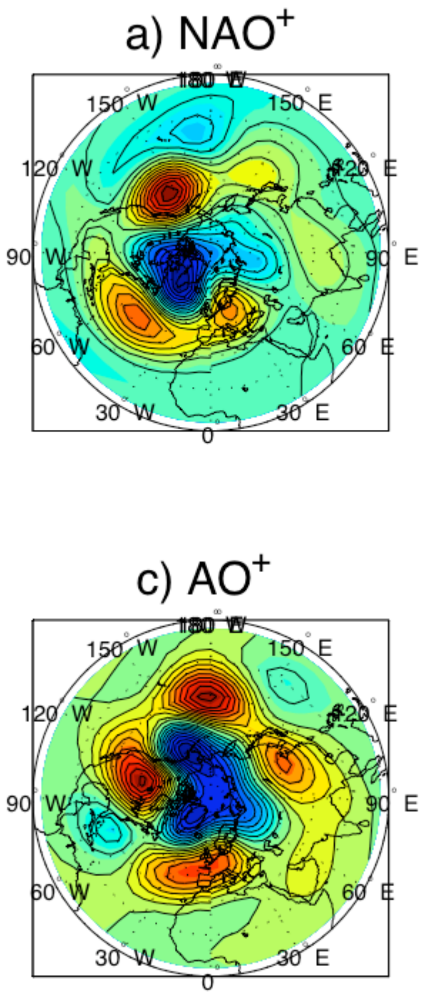
Regression model:

- 15 variables, 3 levels, quadratic at the main level
- Variables: Leading PCs of the middle-level $\Psi$
- Degrees of freedom: 45 (a factor of 40 less than in the MM-93 model)
- Number of regression coefficients: (15+1+15•16/2+30+45)•15=3165 (<< $10^5$)
- PLS applied at the main level

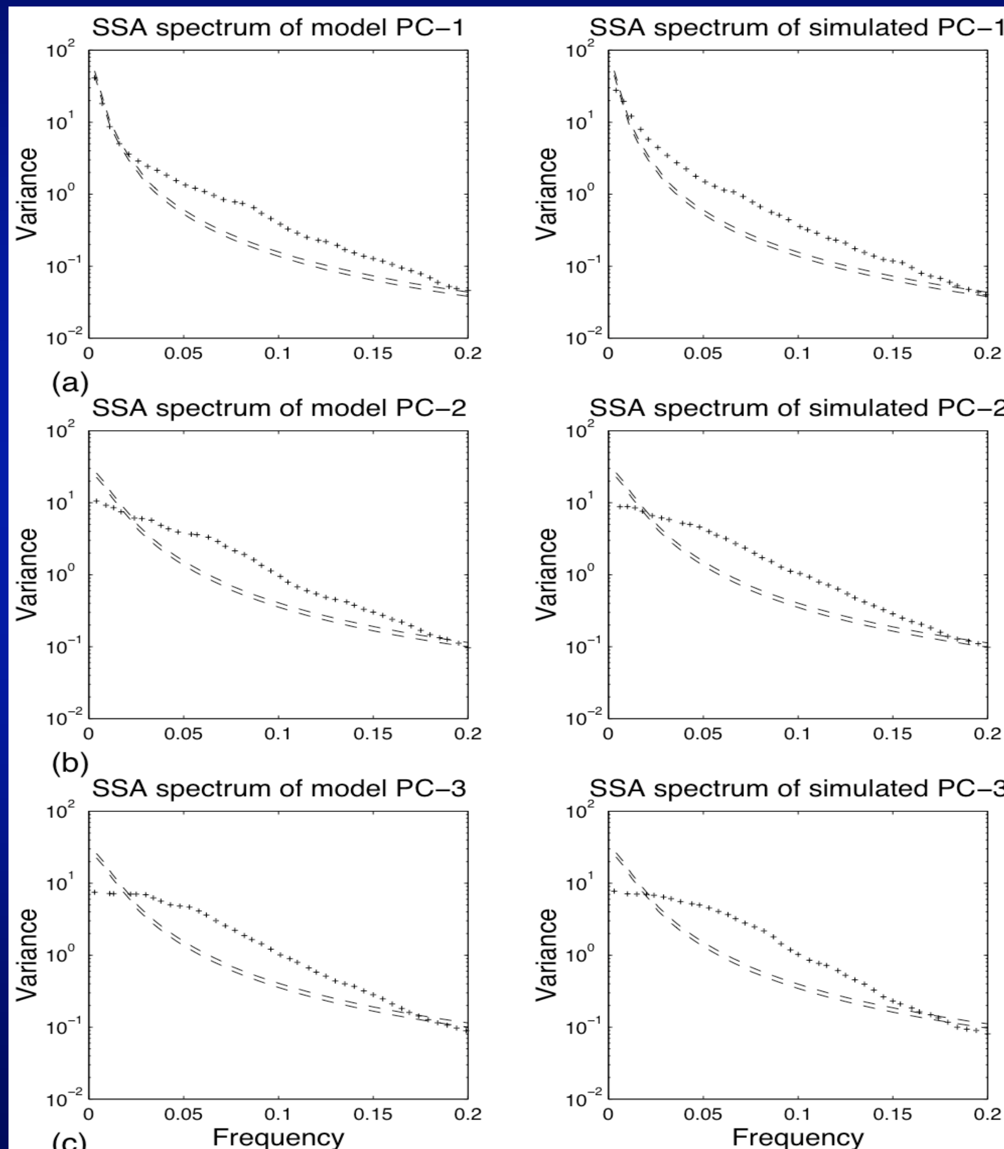
# NH LFV in MM'93 Model – III



# NH LFV in MM’93 Model – IV

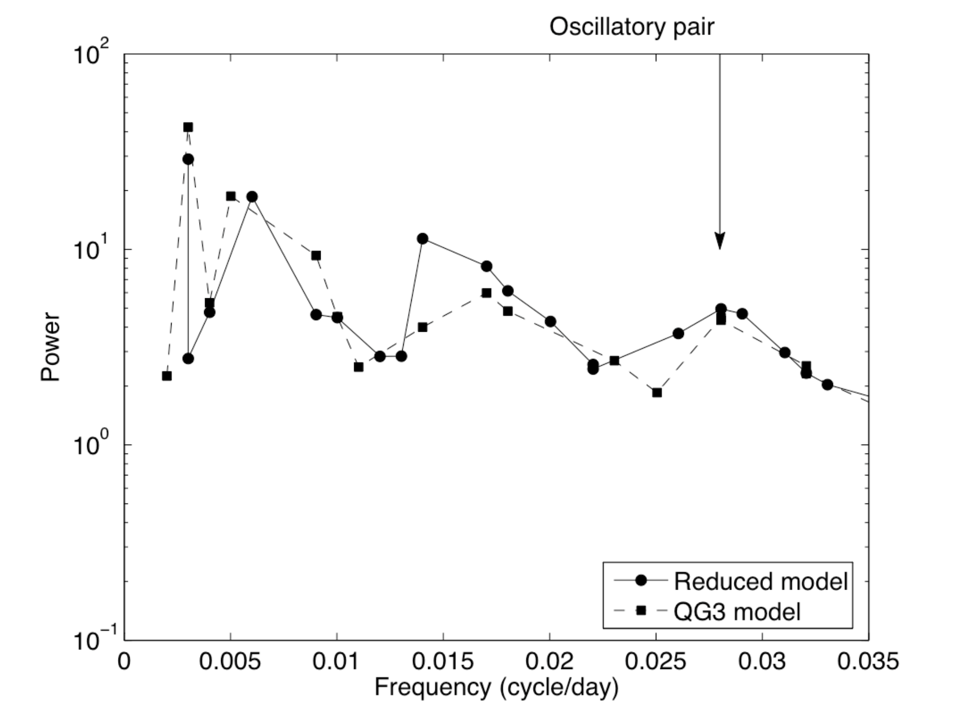


# NH LFV in MM'93 Model – V



# NH LFV in MM’93 Model – VI

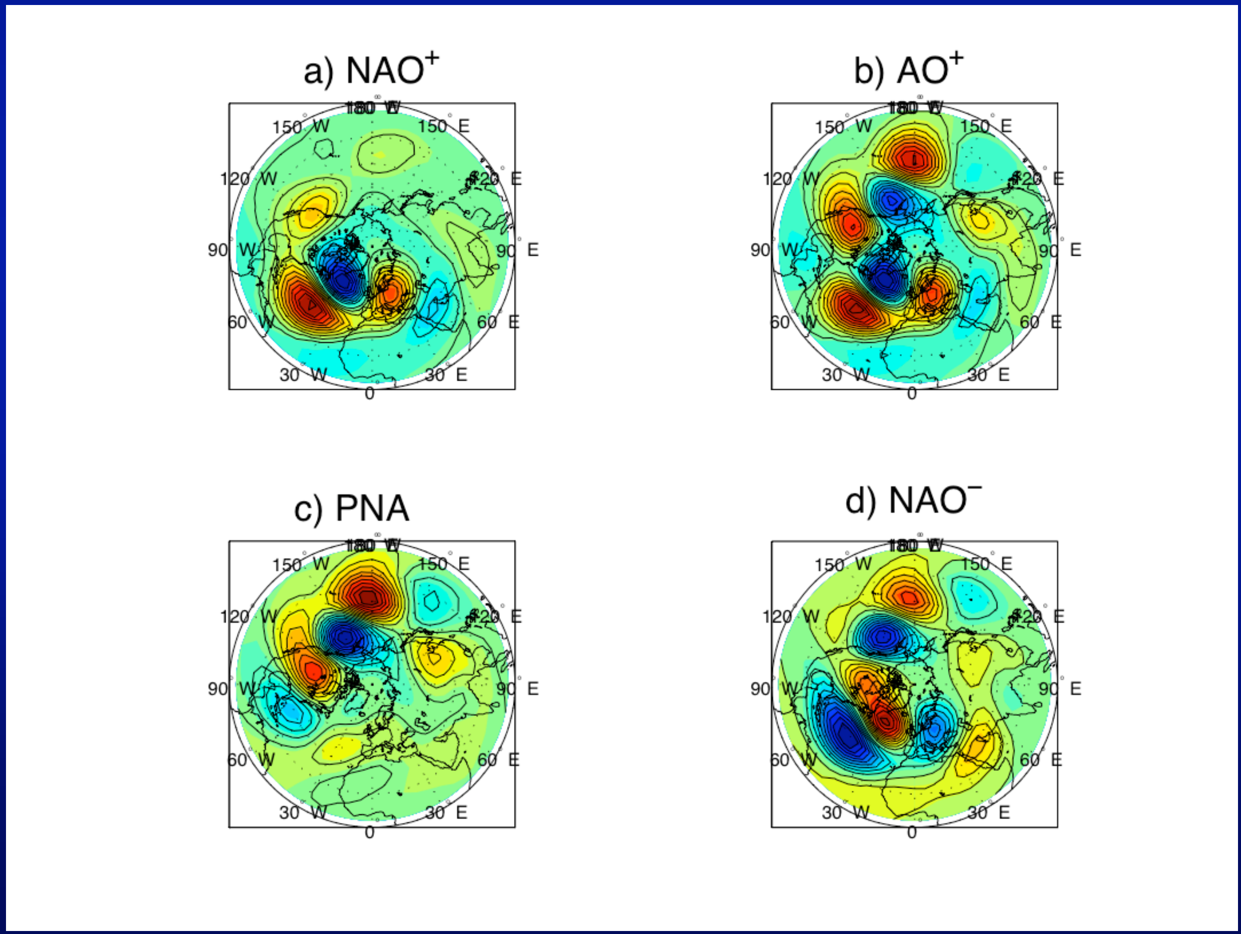
- There are 37- and 20-day oscillatory signals identified by multi-channel SSA




- Composite maps of oscillations are computed by identifying 8 phase categories according to M-SSA reconstruction

# NH LFV in MM'93 Model – VII

Composite 37-day cycle:



# NH LFV in MM'93 Model – VIII

<u>Regimes and Oscillations</u>:

- Fraction of regime days as a function of oscillation phase

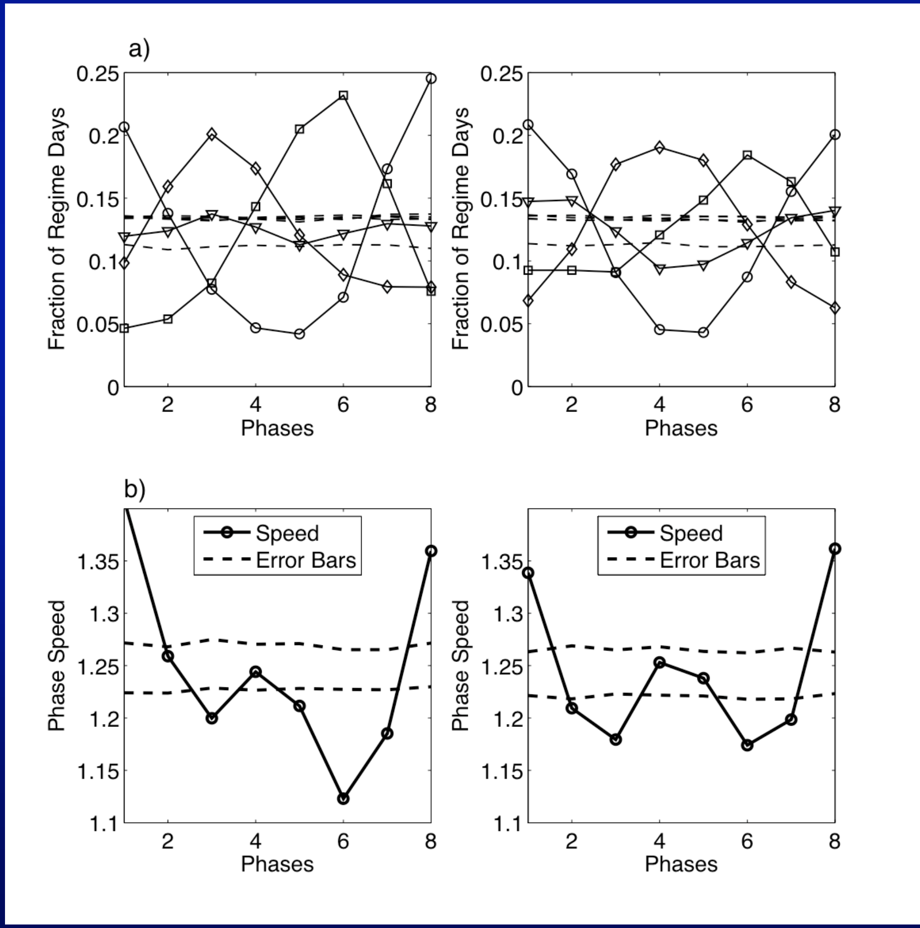
- RC/ΔRC phase speed (both RC and ΔRC are normalized so that linear sinusoidal oscillation would have a constant phase speed)

# NH LFV in MM'93 Model – VIII

## Regimes and Oscillations:

- Fraction of regime days

- Phase speed



# NH LFV in MM'93 Model – IX

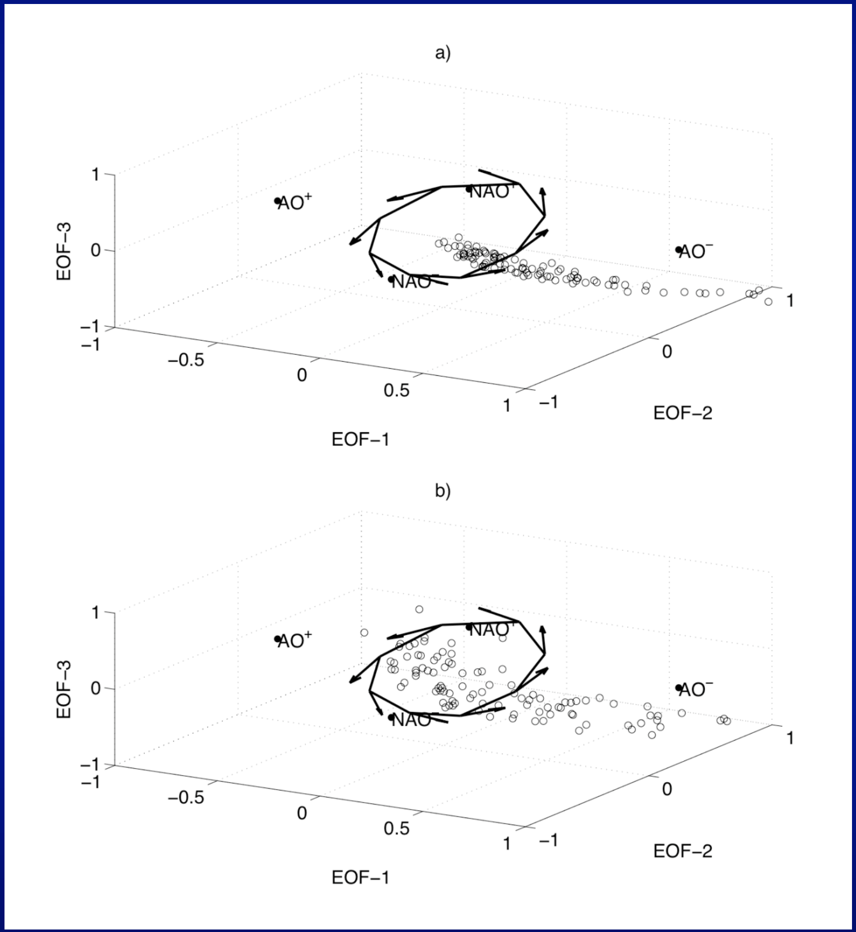
## Regimes and Oscillations:

- $AO^+$ and $NAO^-$ are associated with anomalous (nonlinear) slow-down of a 37-day oscillation trajectory

- $AO^-$ is a stand-alone (not associated with the oscillation) persistent "regime"

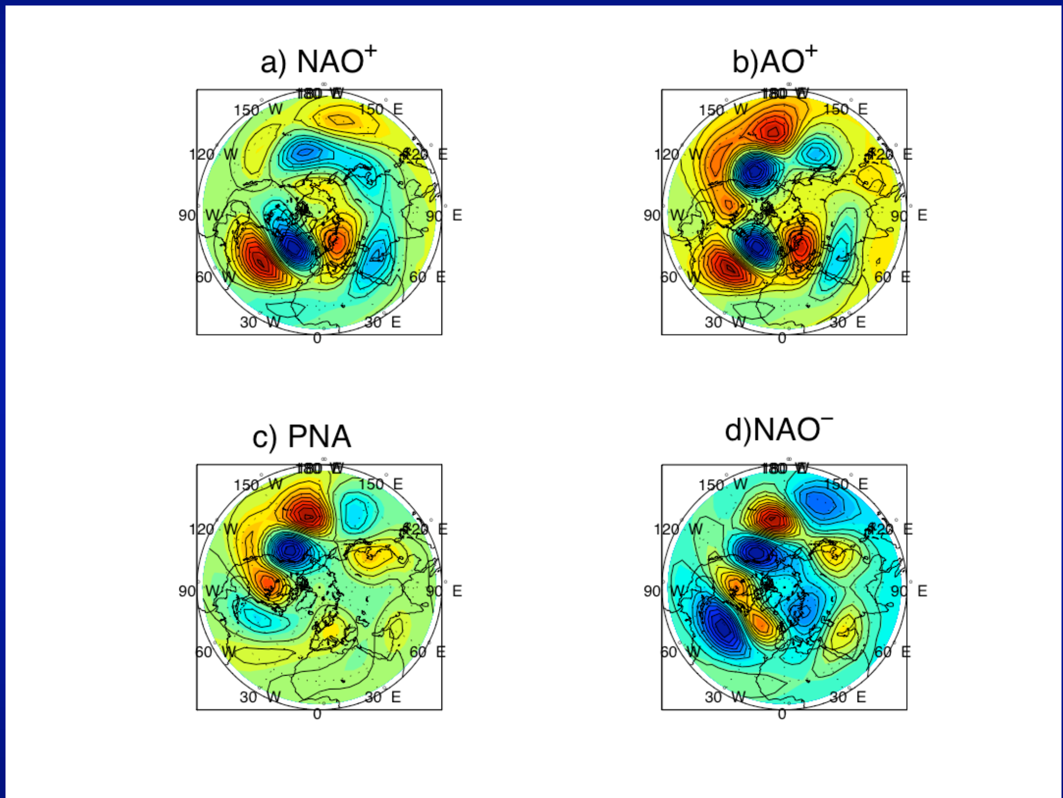
# NH LFV in MM'93 Model – X

Quasi-stationary states of the deterministic component of the regression model



# NH LFV in MM'93 Model – XI

37-day eigenmode of the regression model linearized about climatology*




* Very similar to composite 37-day oscillation

# Conclusions on MM'93 Model

- 15 (45)-variables regression model closely approximates 1500-variables model's major statistical features (PDFs, spectra, regimes, transition matrices, and so on)

- Dynamical analysis of the reduced model identifies $AO^{-}$ as the steady state of this model

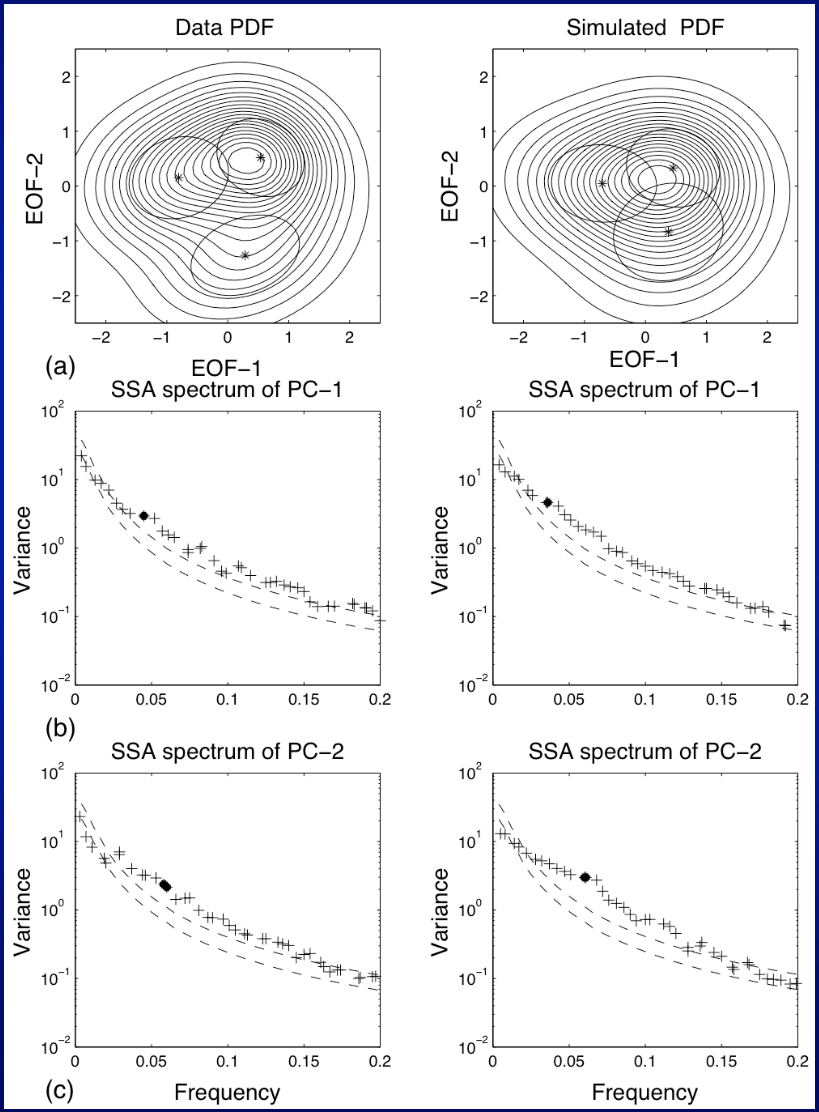
- 37-day mode is associated, in the reduced model, with the least-damped linear eigenmode. Its spatial pattern and eigenvalues are invariant in $AO^{-}$— climatology direction (quasi-stationary "plateau")

- 4 (12)-variable model does not work!!!

# Observed heights

• 44 years of daily 700-mb-height winter data

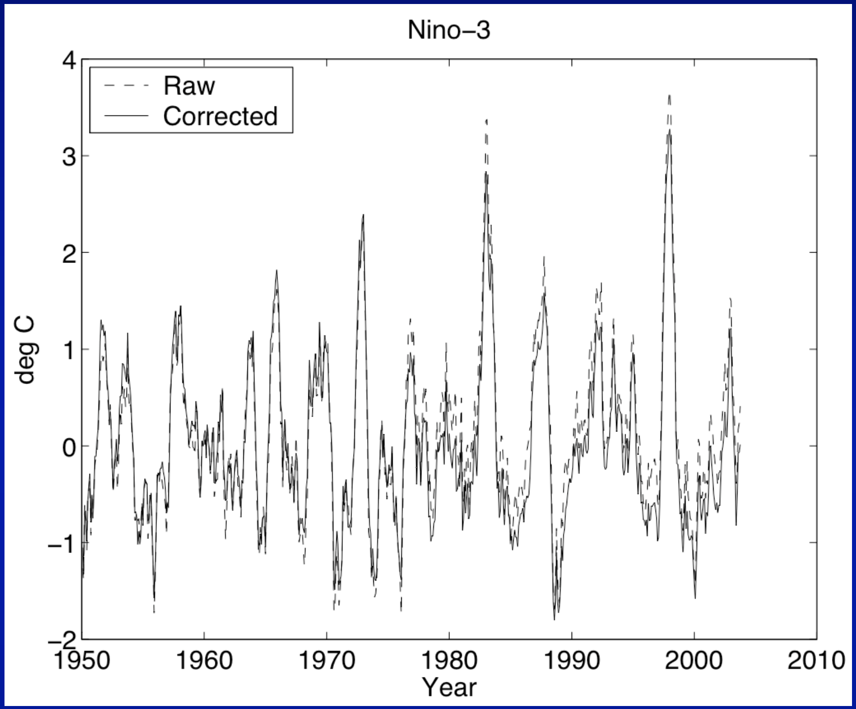
• 12-variable, 2-level model works OK, but dynamical operator has unstable directions: "sanity checks" required



# ENSO – I

Data:

- Monthly SSTs: 1950–2004, 30 S–60 N, 5x5 grid (Kaplan et al.)
- 1976/1977 shift removed

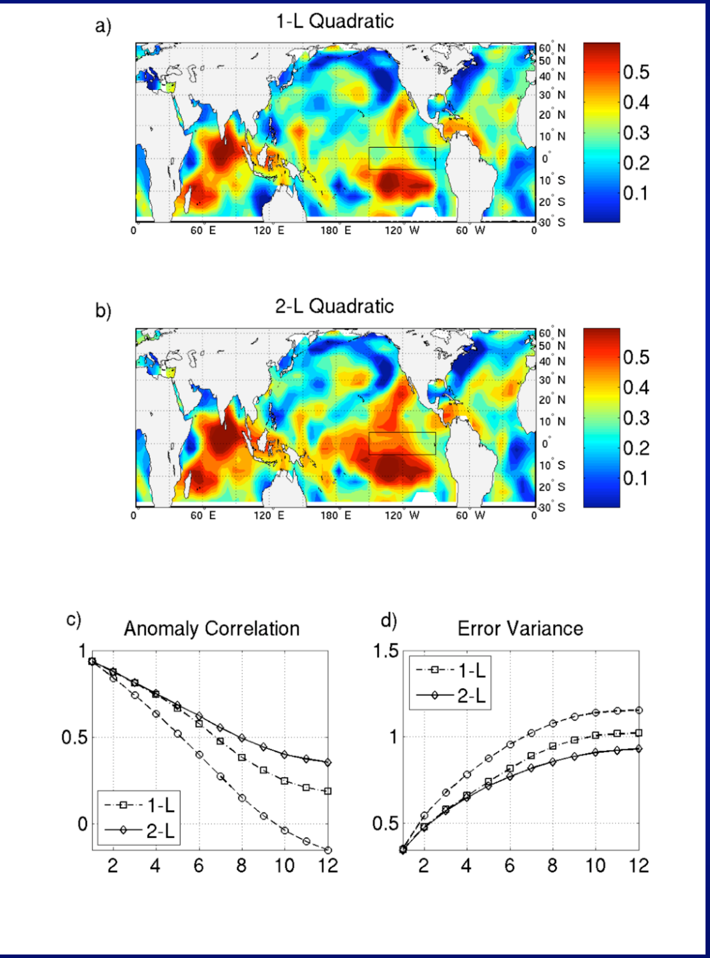



- SST data skewed: Nonlinearity important?

# ENSO – II

Regression model:

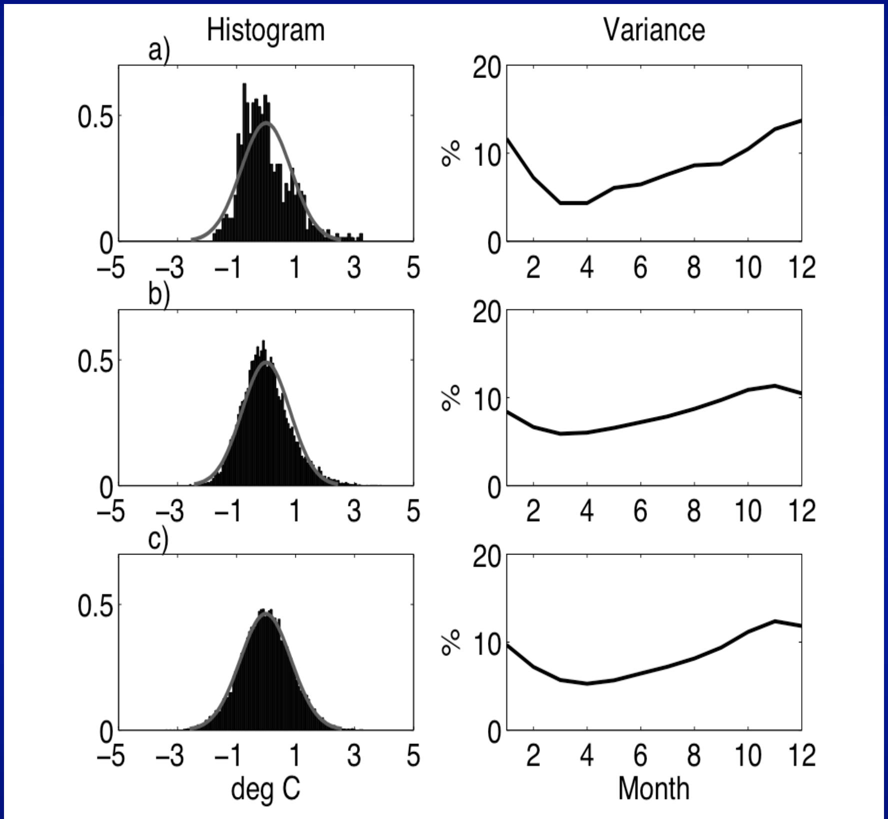
- 2-level, 20-variable (EOFs of SST)
- Seasonal variations in linear part of the main (quadratic) level
- Competitive skill: Currently a member of a multi-model prediction scheme of the IRI




(http://iri.columbia.edu/climate/ENSO/currentinfo/SST_table.html)

# ENSO – III

- Observed
- Quadratic model (100-member ensemble)
- Linear model (100-member ensemble)




Quadratic model has a slightly smaller rms error of extreme-event forecast (not shown)

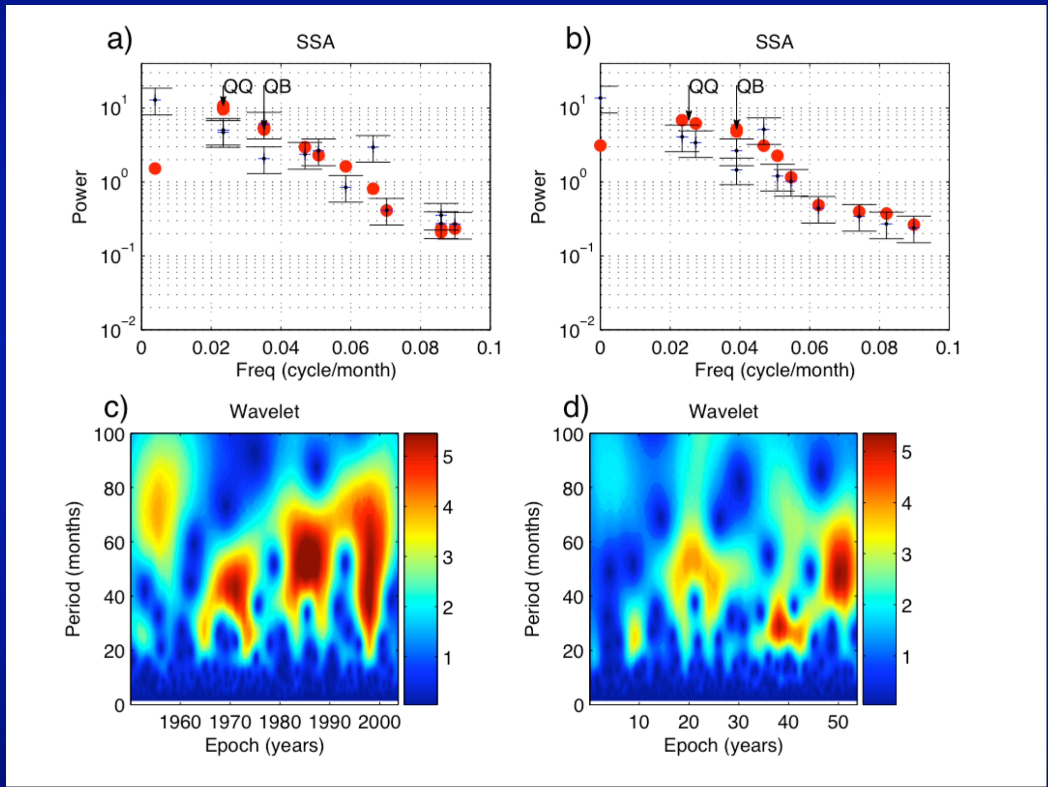
# ENSO – IV

Spectra:

Data Model

- SSA

- Wavelet



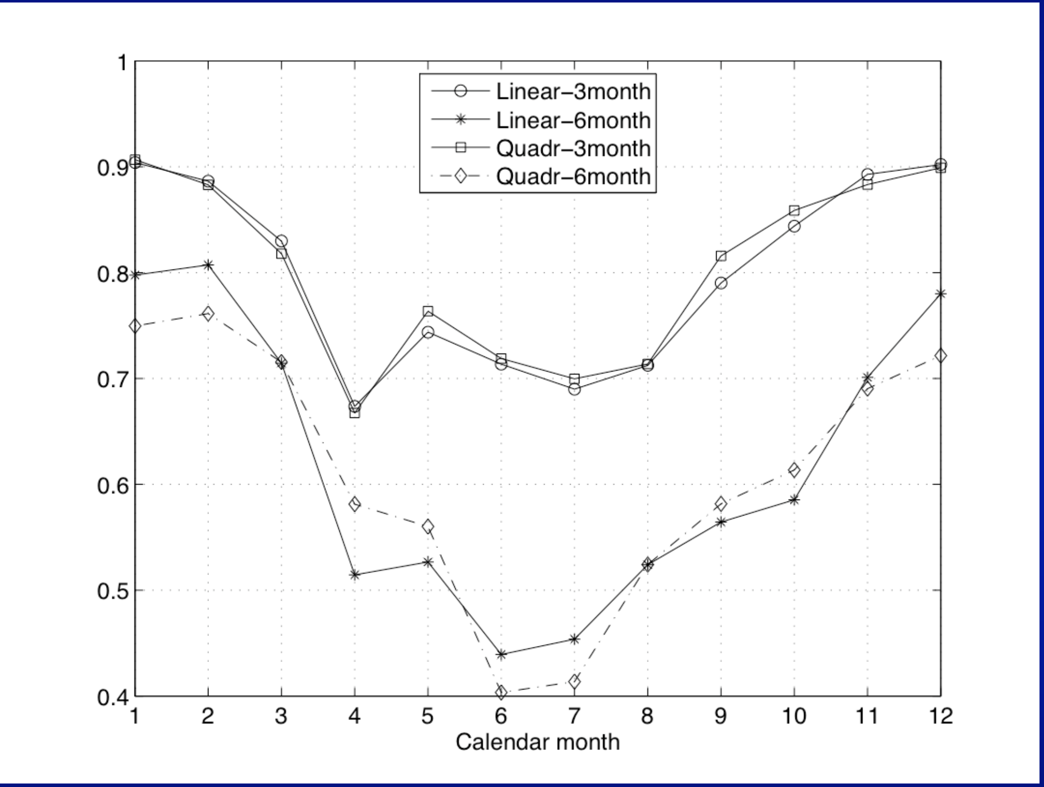

QQ and QB oscillatory modes are reproduced by the model, thus leading to a skillful forecast

# ENSO – V

“Spring Barrier:”

- June’s SSTs are more difficult to predict

- A feature of virtiually all ENSO forecast schemes

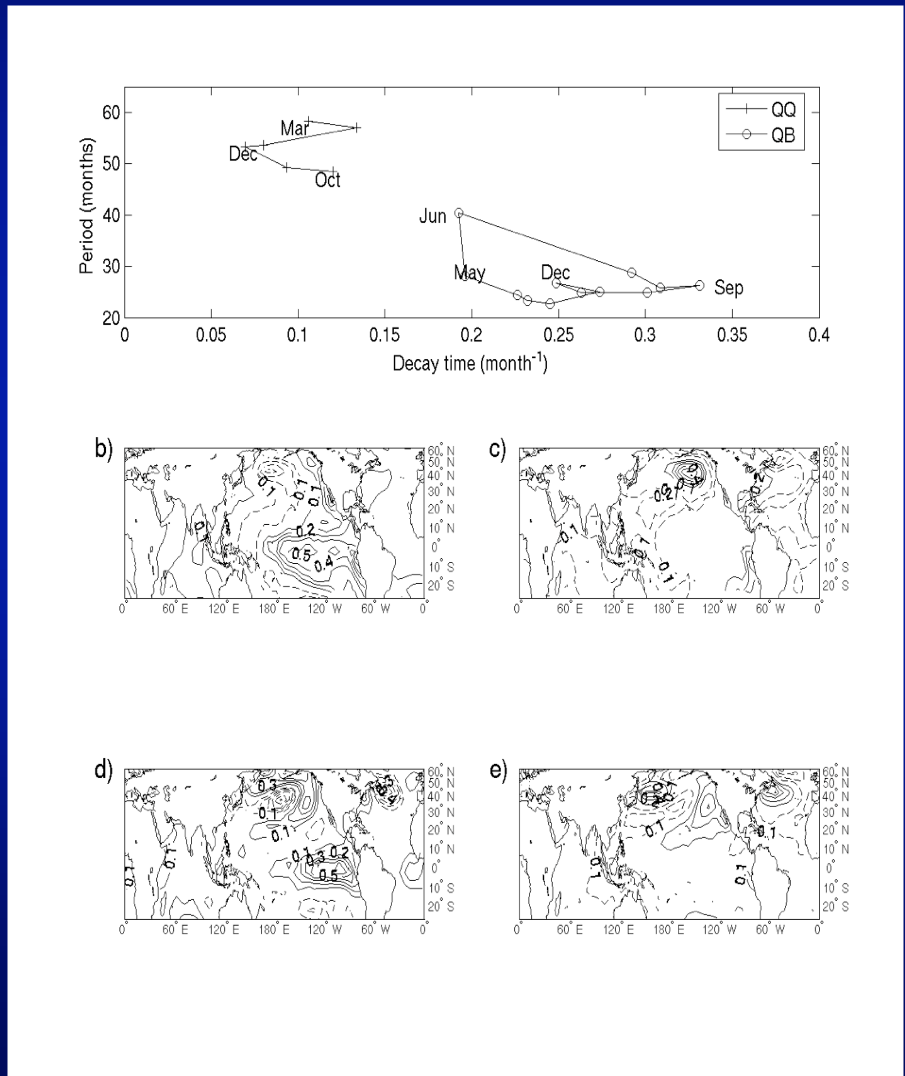



- SST anomalies are weaker in late winter through summer (WHY??), and signal-to-noise ratio is low

# ENSO – VI

- Month-by-month stability analysis of the linearized regression model identifies weakly damped QQ mode (with a period of 48–60 mo), as well as strongly damped QB mode

- QQ mode is least damped in December and is not identifiable in summer!



# ENSO – VII

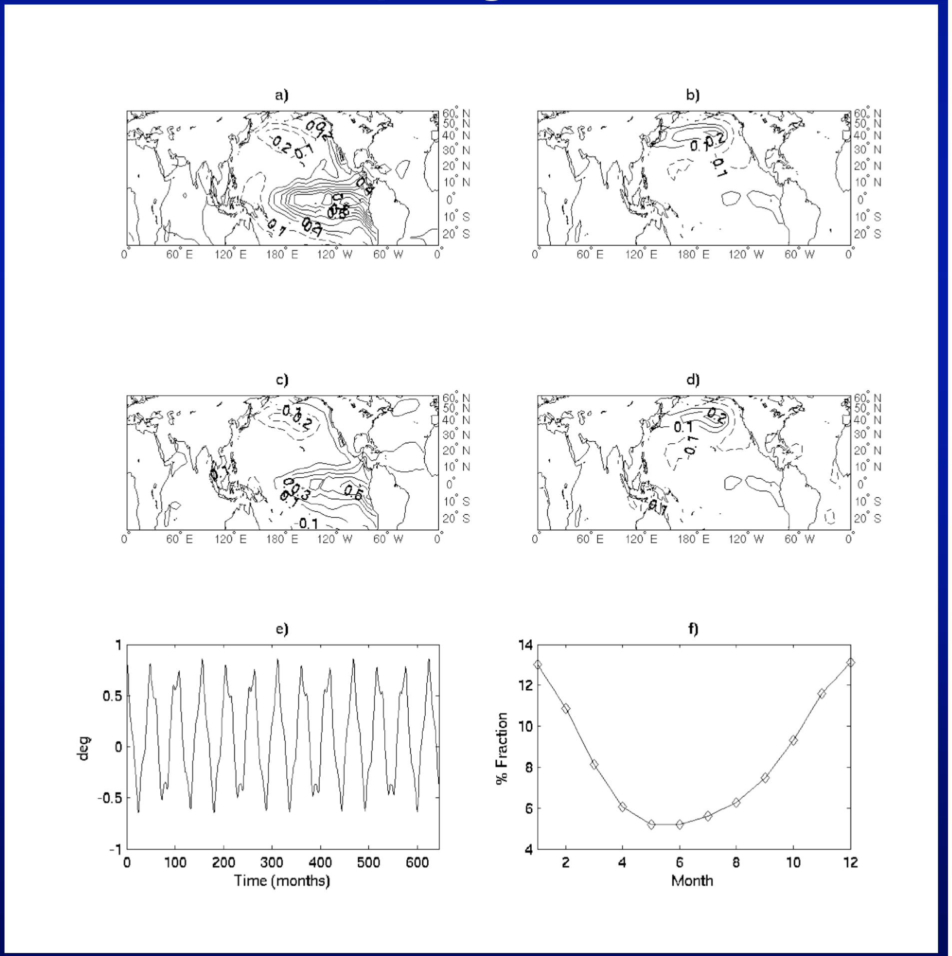
Floquet Analysis ($T$=12 mo):

- Period: 52 months
- Damping: 11 months

$$\dot{\mathbf{x}} = \mathbf{L}(t)\mathbf{x}$$

$$\dot{\mathbf{\Phi}} = \mathbf{L}(t)\mathbf{\Phi}, \quad \mathbf{\Phi}(0) = \mathbf{I}$$

$$\mathbf{M} \equiv \mathbf{\Phi}(T)$$




Floquet modes are related to eigenvectors of monodromy matrix ***M***

# ENSO – VIII

<u>ENSO development and non-normal growth of small perturbations</u>

(Penland and Sardeshmukh 1995; Thompson and Battisti 2000)

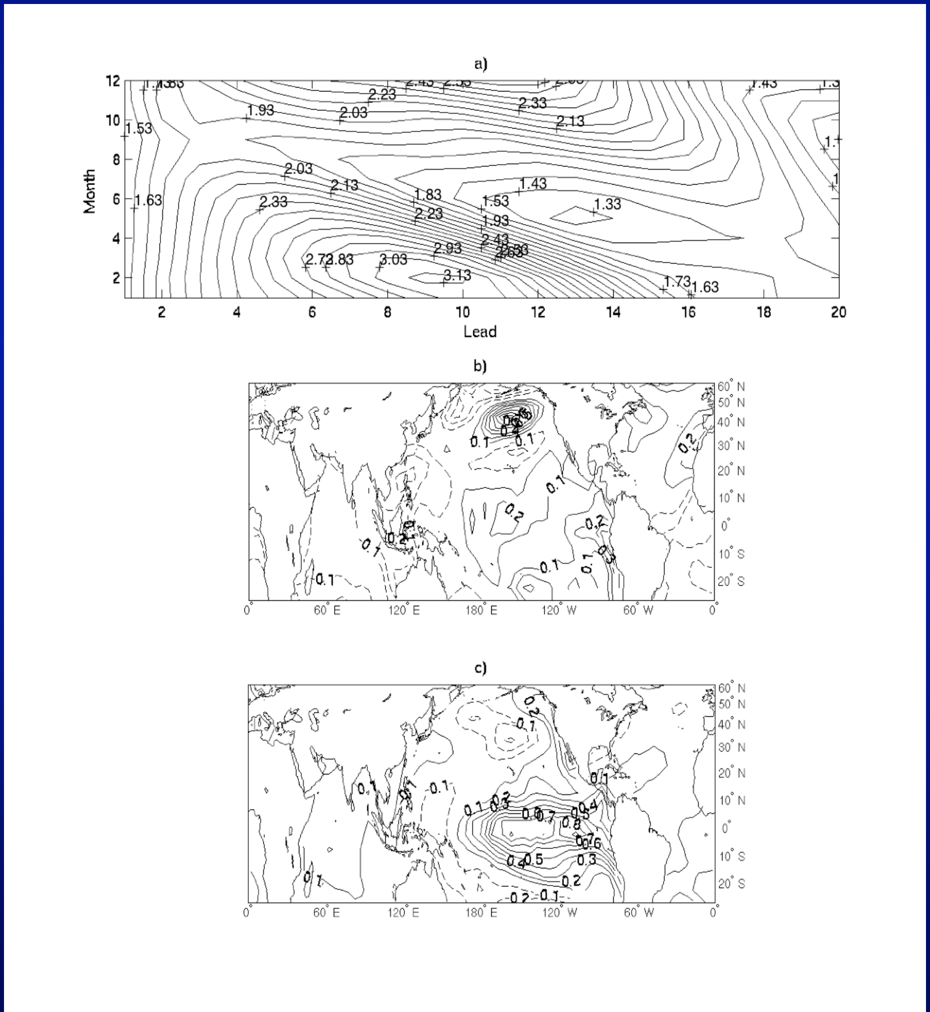
$$\Phi(\tau) = \mathbf{U}\cdot\mathbf{S}\cdot\mathbf{V}^{\top}$$

*V* – optimal initial vectors

*U* – final pattern at $\tau$

• <u>Maximum growth</u>:

Start in Feb., $\tau$=10 mo



# Conclusions on ENSO model

- Competitive skill; 2 levels really matter
- “Linear,” as well as “nonlinear” phenomenology of ENSO is well captured
- Statistical features related to model’s dynamical operator
- SST-only model: other variables? (A. Clarke)

# CONCLUSIONS

- General Linear Least-Squares is method well fit, in combination with regularization techniques such as PCR and PLS, for statistical modeling of geophysical data sets
- Multi-level structure is convenient to implement and provides a framework for dynamical interpretation in terms of the “eddy – mean flow” feedback
- Easy add-ons, such as seasonal cycle
- Analysis of regression models provides conceptual view for possible dynamical causes behind the observed statistics

# CONCLUSIONS (cont'd)

Pitfalls:

- Models are maps: need to have an idea about (time) scales in the system and sample accordingly
- Models are parameteric: functional form is pre-specified
- Choice of predictors is subjective
- No quadratic invariants guaranteed – instability possible

# References

Kravtsov, S., D. Kondrashov, and M. Ghil, 2005: Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability. *J. Climate*, **18**, 4404–4424.

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

Kondrashov, D., S. Kravtsov, and M. Ghil, 2006: Empirical mode reduction in a model of extratropical low-frequency variability. *J. Atmos. Sci*., accepted.