Advanced Spectral Methods and Nonlinear Dynamics

Michael Ghil

Ecole Normale Supérieure, Paris, and
University of California, Los Angeles

Motivation

1. Climatic time series have typically broad peaks on top of a continuous, "warm-colored" background ➔ ***Method***

2. Connections to nonlinear dynamics ➔ ***Theory***

3. Need for stringent statistical significance tests ➔ ***Toolkit***

4. Applications to analysis and prediction ➔ ***Examples***

Joint work with: M. R. Allen, M. D. Dettinger, K. Ide, N. Jiang, C. L. Keppene, D. Kondrashov, M. Kimoto, M. E. Mann, J. D. Neelin, M. C. Penland, G. Plaut, A. W. Robertson, A. Saunders, D. Sornette, S. Speich, C. M. Strong, C. Taricco, Y.-d. Tian, Y. S. Unal, R. Vautard, & P. Yiou (on 3 continents).

Motivation & Outline

1. Data sets in the geosciences are often short and contain errors: this is both an obstacle and an incentive.

2. Phenomena in the geosciences often have both regular ("cycles") and irregular ("noise") aspects.

3. Different spatial and temporal scales:
   one person's noise is another person's signal.

4. Need both deterministic and stochastic modeling.

5. Regularities include (quasi-)periodicity ➔ spectral analysis via "classical" methods (see SSA-MTM Toolkit).

6. Irregularities include scaling and (multi-)fractality ➔ "spectral analysis" via Hurst exponents, dimensions, etc. (see Multi-Trend Analysis, MTA)

7. Does some combination of the two, + deterministic and stochastic modeling, provide a pathway to prediction?

For details and publications, please visit these two Web sites:

TCD http://www.atmos.ucla.edu/tcd/ ➔ key person – Dmitri Kondrashov!

E2-C2 http://www.ipsl.jussieu.fr/~ypsce/py_E2C2.html

Climatic Trends & Variability

- Standard view — Binary thinking, dichotomy:

  Trend — Predictable (completely), deterministic, reassuring, good;

  Variability — Unpredictable (totally), stochastic, disconcerting, bad.

- In fact, these two are but extremes of a spectrum of, more or less predictable, types of climatic behavior, between the totally boring & the utterly surprising.

- (Linear) Trend = Stationary >

  Periodic > Quasi-periodic >

  Deterministically aperiodic >

  Random Noise

- Here “>” means “better, more predictable”, &

  Variability = Periodic + Quasi-periodic +

  Aperiodic + Random

Time Series in Nonlinear Dynamics

The 1980s — decade of greed & fast results

(LBOs, junk bonds, fractal dimension).

Packard *et al*. (1980), Roux *et al*. (1980);

Mañe (1981), Ruelle (1981), Takens (1981);

- Method of delays: $\ddot{x}_i = f_i(x_1, \ldots.., x_n) \Leftrightarrow x^{(n)} = F(x^{(n-1)}, \ldots.., x)$

$$\ddot{x} = F(x, \dot{x}) \Rightarrow \begin{cases} \dot{x} = y, \\ \dot{y} = F(x, y) \end{cases}$$

Differentiation ill-posed ⇒ use differences instead!

1st Problem — smoothness:

Whitney embedding lemma doesn't apply to most attractors (e.g.,Lorenz)

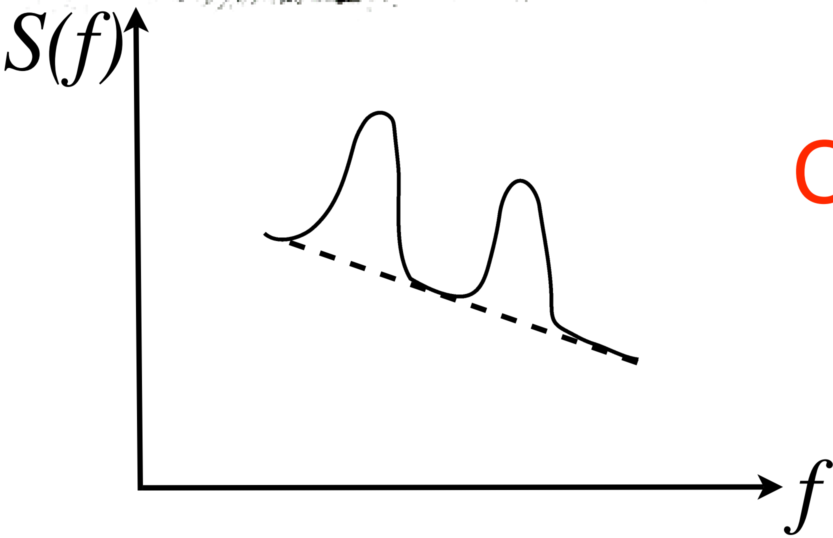
2nd Problem — noise;

3rd Problem — sampling: long recurrence times.

- Some rigorous results on convergence:

Smith (1988, *Phys. Lett. A*), Hunt (1990, *SIAM J. Appl. Math*.)

Spectral Density (Math)/Power Spectrum (Science & Engng.)



Continuous background
+ peaks

- **Wiener-Khinchin (Bochner) Theorem** **Blackman-Tukey Method**

$$R(s) = \lim_{L\to\infty} \frac{1}{2L} \int_{-L}^{L} x(t)x(t+s)dt$$

$$S(f) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(s)e^{-ifs}ds \equiv \hat{R}(s)$$

i.e., the **lag-autocorrelation function** & the **spectral density**

are **Fourier transforms** of each other.

Power Law for Spectrum

$$S(f) \sim f^{-p} + \text{poles}$$

i.e. linear in log-log coordinates

For a 1st-order Markov process or “red noise” $p = 2$

“Pink” noise, $p = 1$ ($1/f$, flicker noise)

“White” noise, $p = 0$

Low-order dynamical (deterministic) systems

have exponential decay of $S(f)$ (linear in log-linear coordinates)

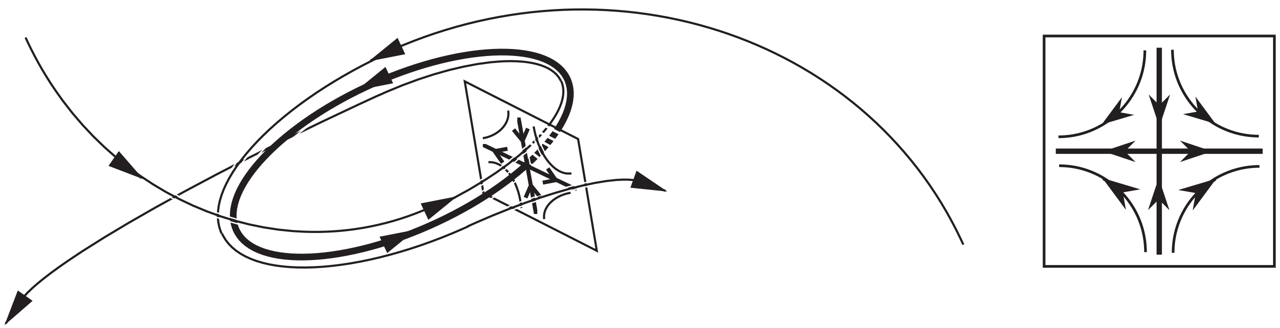
e.g. for Smale horseshoe $\forall k \quad \exists 2^k$ unstable orbits of period k

N.B. Bhattacharaya, Ghil & Vulis (1982, *J. Atmos. Sci*.) showed a spectrum $S \sim f^{-2}$ for a nonlinear PDE with delay (doubly infinite-dimensional)

Power Law for Spectrum (cont'd)

◦ Hypothesis: "***Poles***" correspond to the least unstable periodic orbits

"***unstable limit cycles***" "***Poincaré section***"



◦ Major clue to the physics

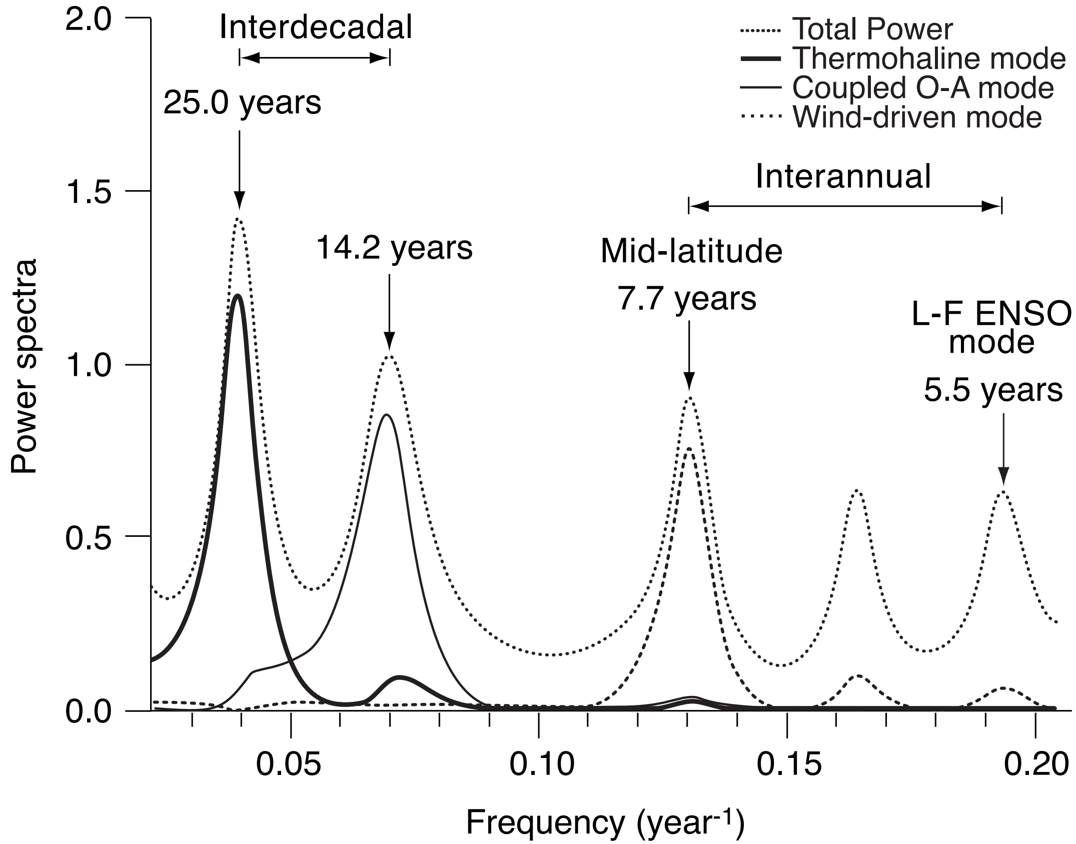
that underlies the dynamics

◦ N.B. Limit cycle not necessarily elliptic, i.e. not

$$(x, y) = (a_f sin(ft), b_f cos(ft))$$

SSA (prefilter) + (low-order) MEM

- "Stack" spectrum



In good agreement with MTM peaks of **Ghil & Vautard (1991, *Nature*)** for the Jones *et al.* (1986) temperatures & stack spectra of Vautard *et al.* (1992, *Physica D*) for the IPCC "consensus" record (both global), to wit 26.3, 14.5, 9.6, 7.5 and 5.2 years.

Peaks at 27 & 14 years also in Koch sea-ice index off Iceland (Stocker & Mysak, 1992), etc.
Plaut, Ghil & Vautard (1995, *Science*)

Power Spectra & Reconstruction

- **A. Transform pair**:

$$X(t+s) = \sum_{k=1}^{M} a_k(t) e_k(s), e_k(s) - EOF$$

The e_k's are **adaptive filters**,

$$a_k(t) = \sum_{s=1}^{M} X(t+s) e_k(s), a_k(t) - PC$$

the a_k's are **filtered time series**.

B. Power spectra

$$S_X(f) = \sum_{k=1}^{M} S_k(f); \quad S_k(f) = R_k(s); \quad R_k(s) \approx \frac{1}{T} \int_0^T a_k(t) a_k(t+s) dt$$

C. Partial reconstruction

$$X^K(t) = \frac{1}{M} \sum_{k \in K} \sum_{s=1}^{M} a_k(t-s) e_k(s);$$

in particular: $K = \{1, 2, \ldots.., S\}$ or $K = \{k\}$ or $K = \{l, l+1; \lambda_l \approx \lambda_{l+1}\}$

Singular

Spectrum

Analysis

Singular Spectrum Analysis (SSA)

Spatial EOFs

x -- *space*

$$\phi(x,t) = \sum a_k(t) e_k(x)$$

$$C_\phi(x,y) = E\phi(x,\omega)\phi(y,\omega) = \frac{1}{T}\int_o^T \phi(x,t)\phi(y,t)\,dt$$

$$C_\phi e_k(x) = \lambda_k e_k(x)$$

SSA

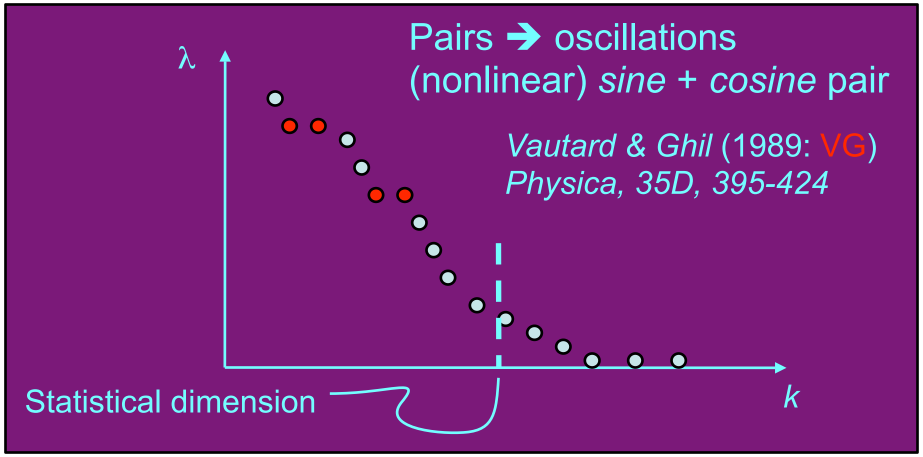
s -- *lag*

$$X(x+s) = \sum a_k(t) e_k(s)$$

$$C_X(s) = EX(t+s,\omega)\phi(s,\omega) = \frac{1}{T}\int_o^T X(t)X(t+s)\,dt$$

$$C_X e_k(s) = \lambda_k e_k(s)$$

Colebrook (1978); *Weare & Nasstrom* (1982);
Broomhead & King (1986: BK); *Fraedrich* (1986)

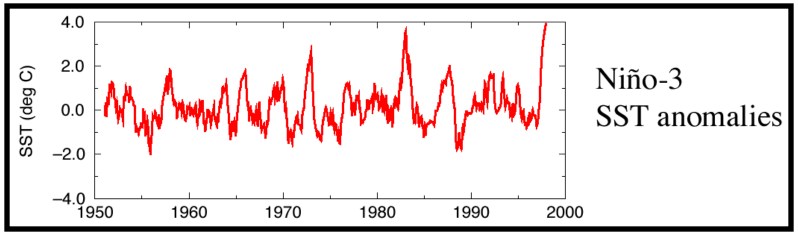


BK+VG: Analogy between Mañe-Takens embedding and the Wiener-Khinchin theorem

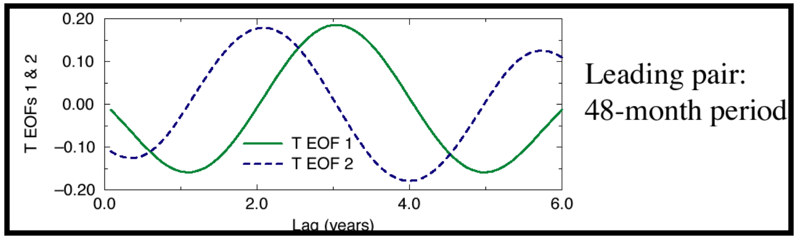
Singular Spectrum Analysis (SSA)

SSA decomposes (geophysical & other) time series into ***Temporal EOFs*** (T-EOFs) and ***Temporal Principal Components*** (T-PCs), based on the series' lag-covariance matrix

Time series

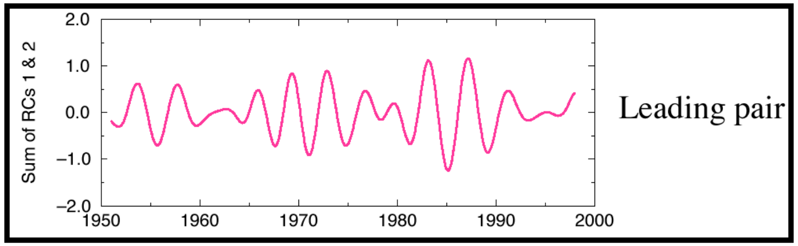


T-EOFs



Selected parts of the series can be reconstructed, via ***Reconstructed Components*** (RCs)

RCs

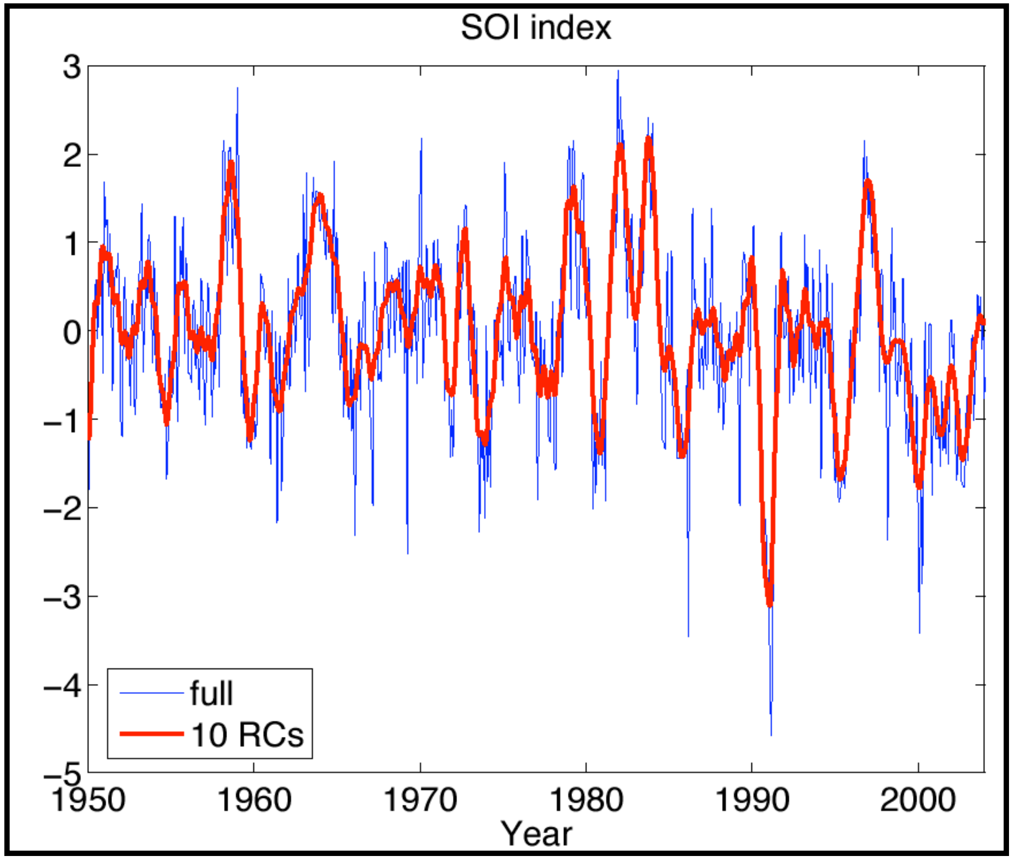


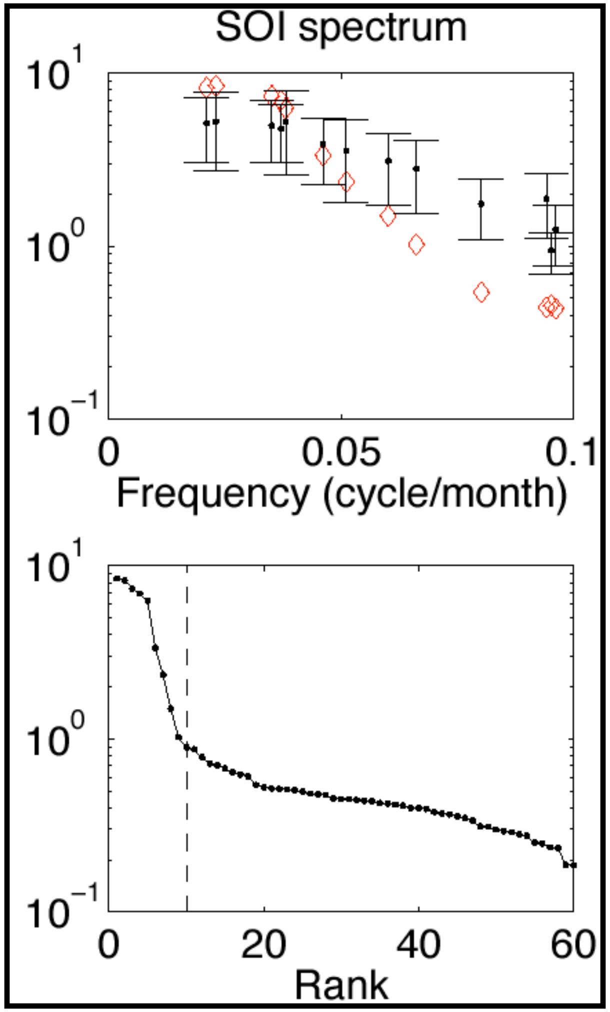
- SSA is good at isolating oscillatory behavior via paired eigenelements.
- SSA tends to lump signals that are longer-term than the window into
  - one or two trend components.

Selected References:

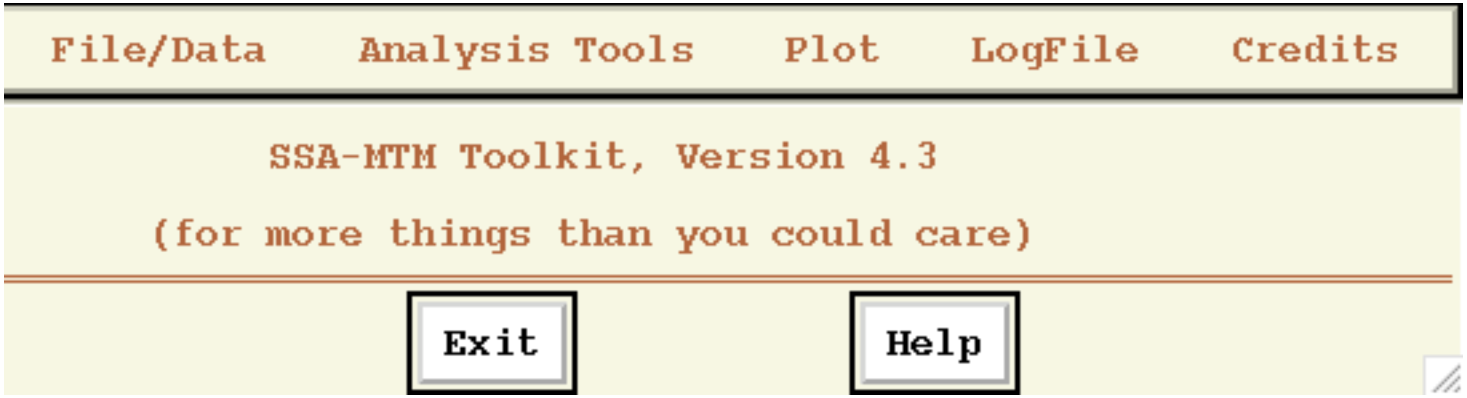
Vautard & Ghil (1989, *Physica* D);
Ghil *et al*. (2002, *Rev. Geophys.)*

Singular Spectrum Analysis (SSA) and M-SSA (cont'd)

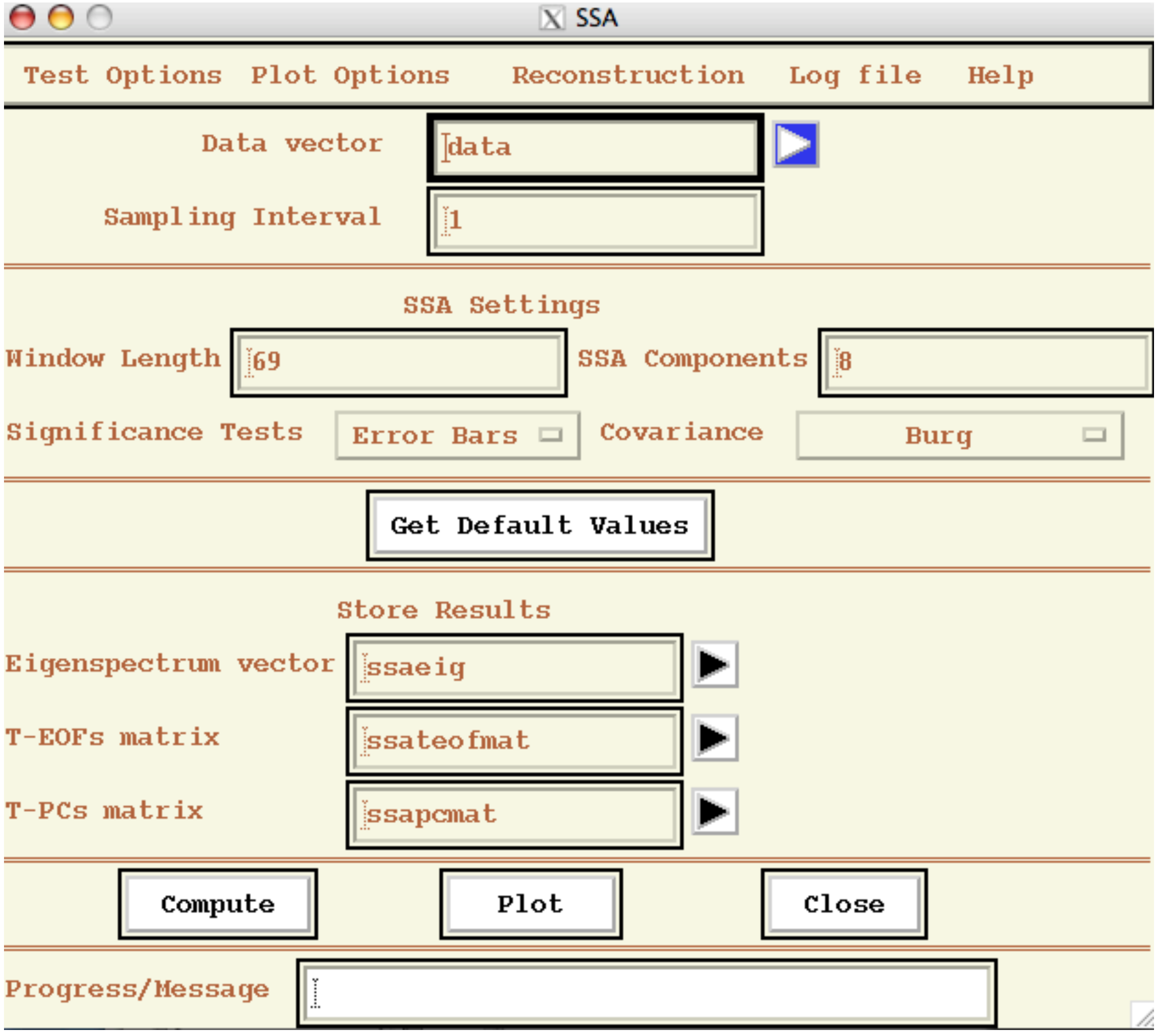




- Break in slope of SSA spectrum distinguishes "***significant***" from "***noise***" EOFs
- Formal Monte-Carlo test (Allen and Smith, 1994) identifies 4-yr and 2-yr ENSO oscillatory modes. A window size of *M* = 60 is enough to "resolve" these modes in a monthly SOI time series

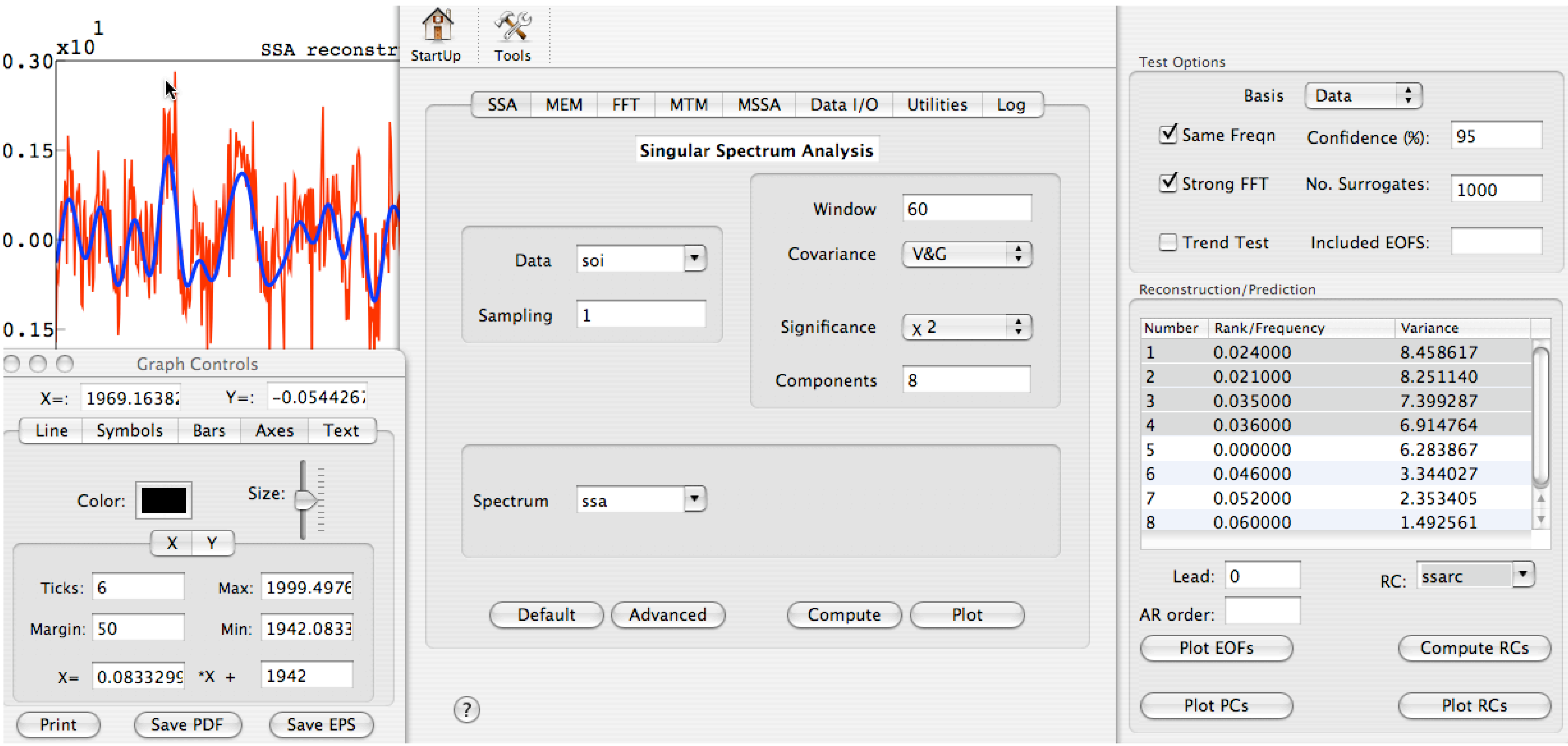


- Ported to Sun, Dec, SGI, PC Linux, and Mac OS X
- Graphics support for IDL and Grace
- Precompiled binaries are available at www.atmos.ucla.edu/ tcd/ssa
- Includes ***Blackman-Tukey FFT***, ***Maximum Entropy Method***, ***Multi-Taper Method (MTM)***, ***SSA and M-SSA***.
- Spectral estimation, decomposition, reconstruction & prediction.
- Significance tests of "**oscillatory modes**" vs. "**noise.**"



- **Free!!!**
- Data management with ***named vectors & matrices.***
- ***Default values*** button.

kSpectra Toolkit for Mac OS X



- **$$ … but: *Project files*, *Automator WorkFlows*, *Spotlight* and more!**
- **www.spectraworks.com**

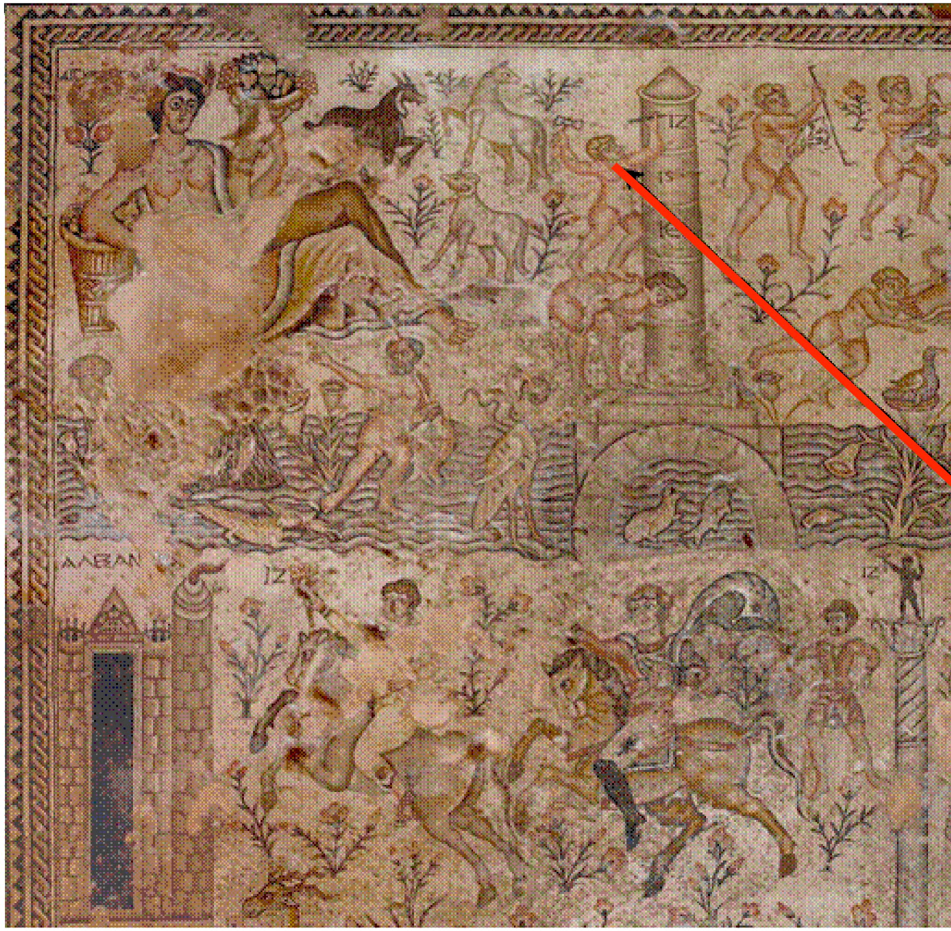
The Nile River Records Revisited:

How good were Joseph's predictions?

Michael Ghil, ENS & UCLA

Yizhak Feliks, IIBR & UCLA, Dmitri Kondrashov, UCLA

Why are there data missing?



Hard Work

- Byzantine-period mosaic from Zippori, the capital of Galilee (1st century B.C. to 4th century A.D.); photo by Yigal Feliks, with permission from the Israel Nature and Parks Protection Authority)

What to do about gaps?

- Most of the advanced ***filling-in*** methods are different flavors of ***Optimal Interpolation*** (**OI**: Reynolds & Smith, 1994; Kaplan 1998).

Drawbacks: they either (i) require error statistics to be specified *a priori*; or (ii) derive it **only** from the interval of dense data coverage.

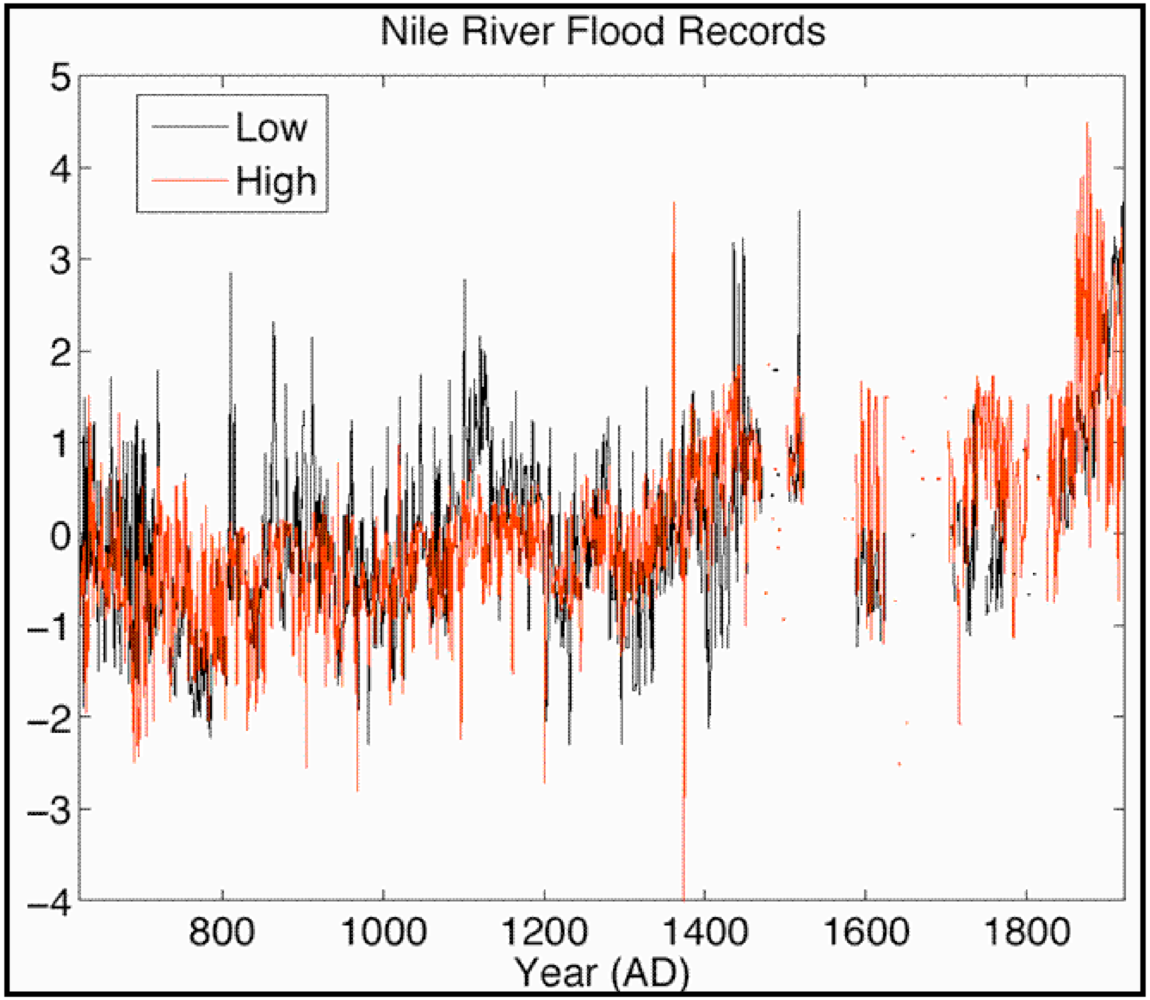
EOF Reconstruction (Beckers & Rixen, 2003): (i) iteratively compute **spatial-covariance** matrix using **all the data**; (ii) determine via cross-validation "**signal**" EOFs and use them to fill in the missing data; accuracy is similar to or better than **OI** (Alvera-Azcarate *et al*. 2004).

Drawbacks: uses **only** spatial correlations **=>** cannot be applied to very **gappy** data.

We propose ***filling in*** **gaps by applying iterative SSA (or M-SSA):**

Utilize both spatial and temporal correlations of data => can be used for highly **gappy** data sets; simple and easy to implement!

Historical records are full of "gaps"....



Annual maxima and minima of the water level at the nilometer on Rodah Island, Cairo.

SSA (M-SSA) Gap Filling

Main idea: utilize both spatial and temporal correlations to iteratively compute self-consistent lag-covariance matrix; M-SSA with M = 1 is the same as the EOF reconstruction method of Beckers & Rixen (2003)

Goal: keep "signal" and truncate "noise" — usually a few leading EOFs correspond to the dominant oscillatory modes, while the rest is noise.

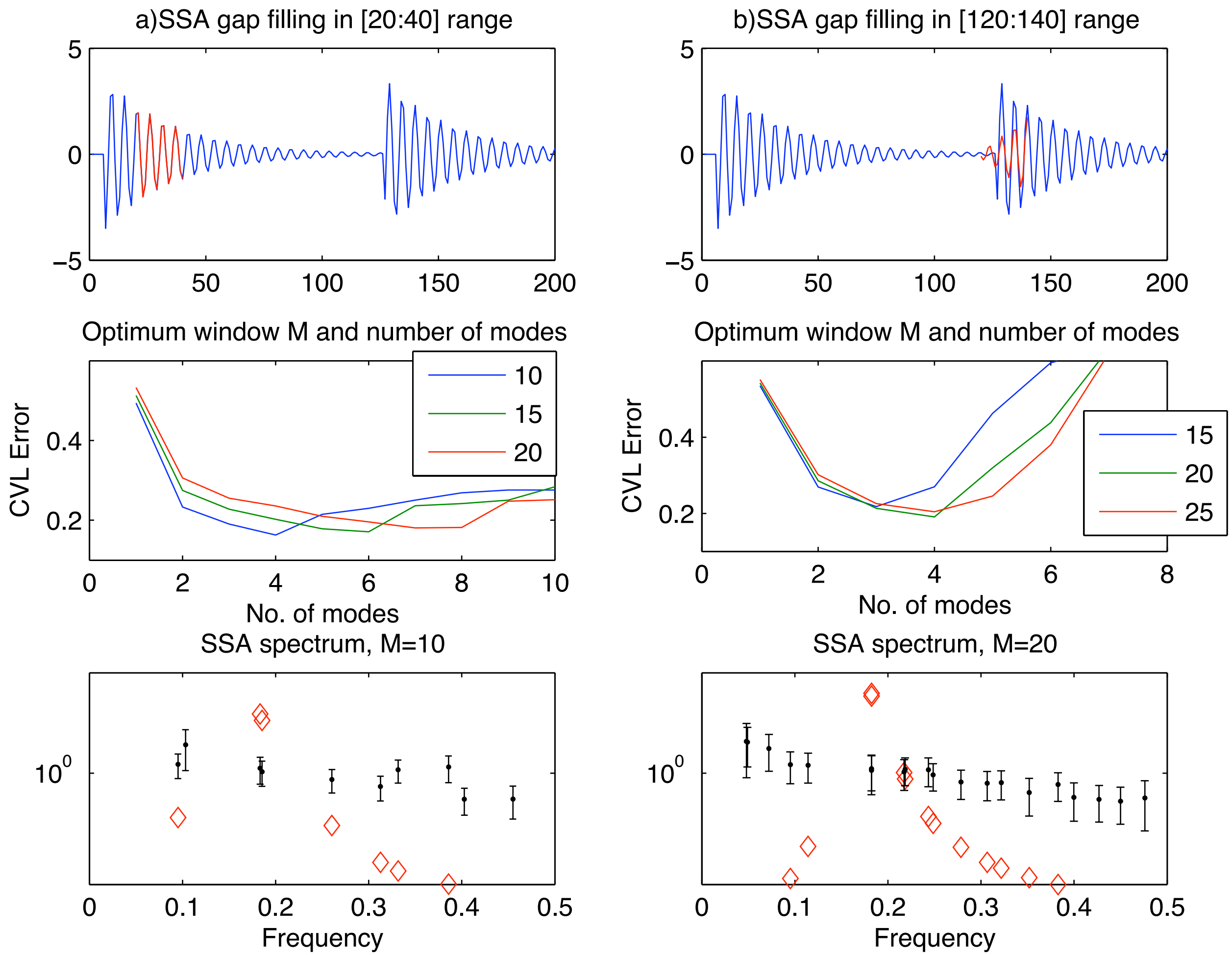
(1) for a given window width M: center the original data by computing the unbiased value of the mean and set the missing-data values to zero.

(2) start iteration with the first EOF, and replace the missing points with the reconstructed component (RC) of that EOF; repeat the SSA algorithm on the new time series, until convergence is achieved.

(3) repeat steps (1) and (2) with two leading EOFs, and so on.

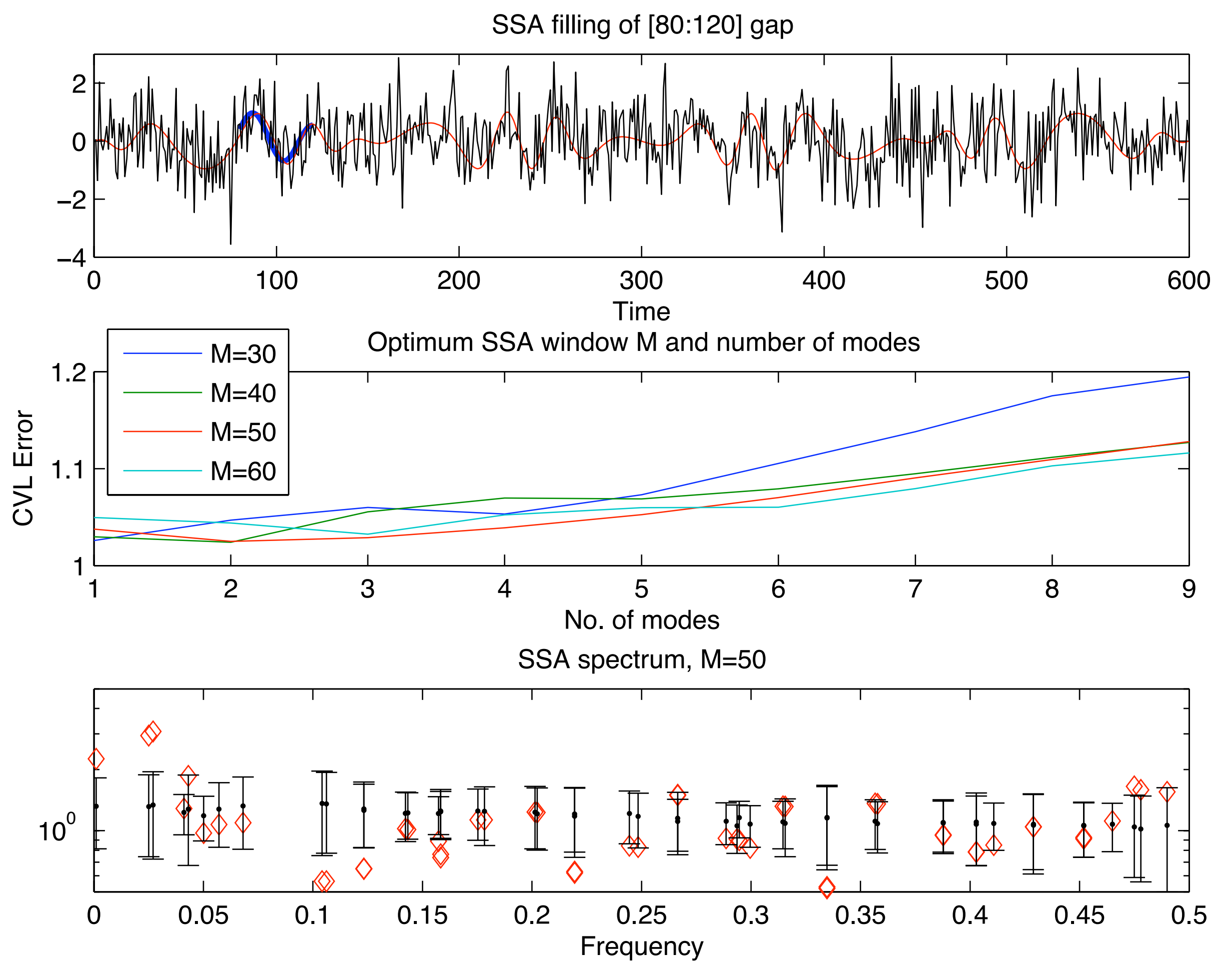
(4) apply cross-validation to optimize the value of M and the number of dominant SSA (M-SSA) modes K to fill the gaps: a portion of available data (selected at random) is flagged as missing and the RMS error in the reconstruction is computed.

Synthetic I: Gaps in Oscillatory Signal



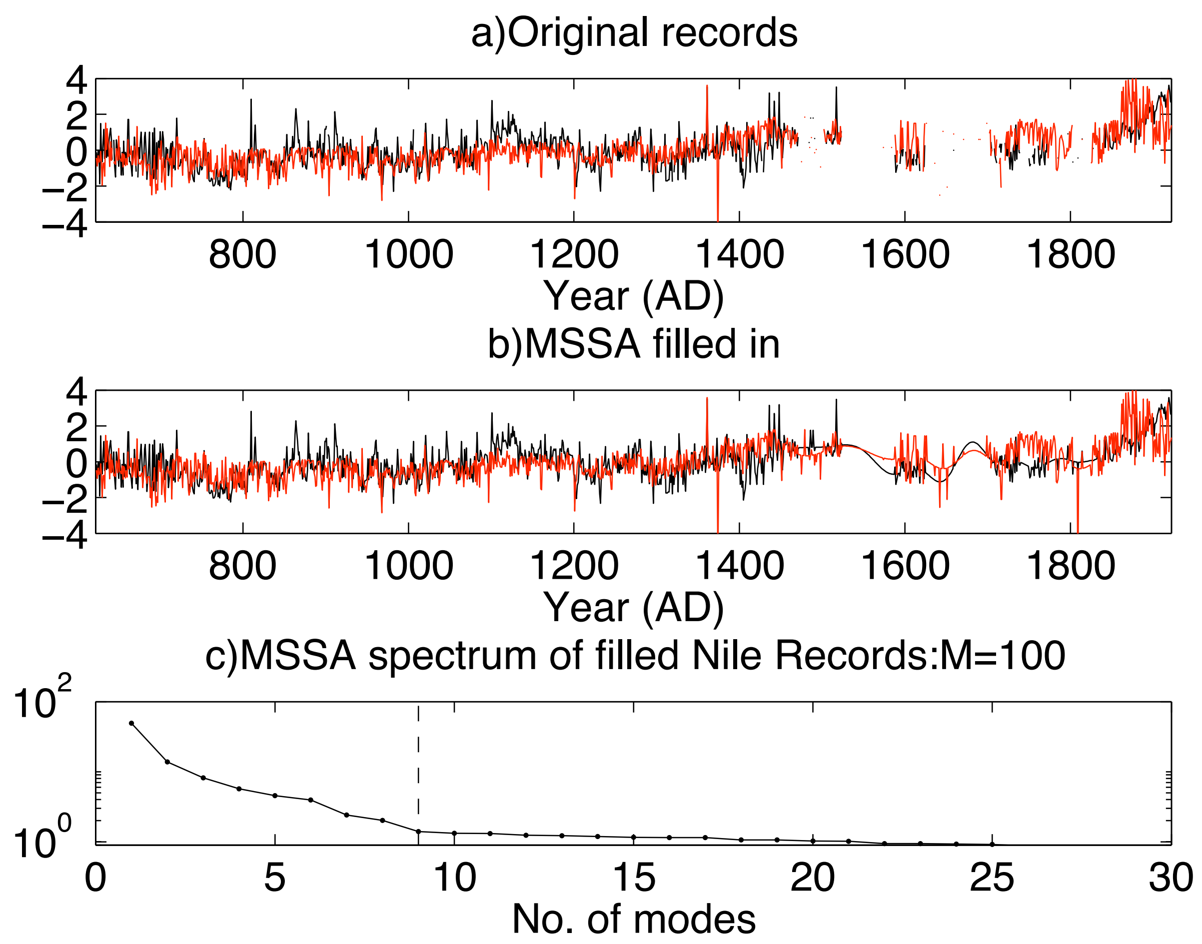
- Very good gap filling for smooth modulation; OK for sudden modulation.

Synthetic II: Gaps in Oscillatory Signal + Noise



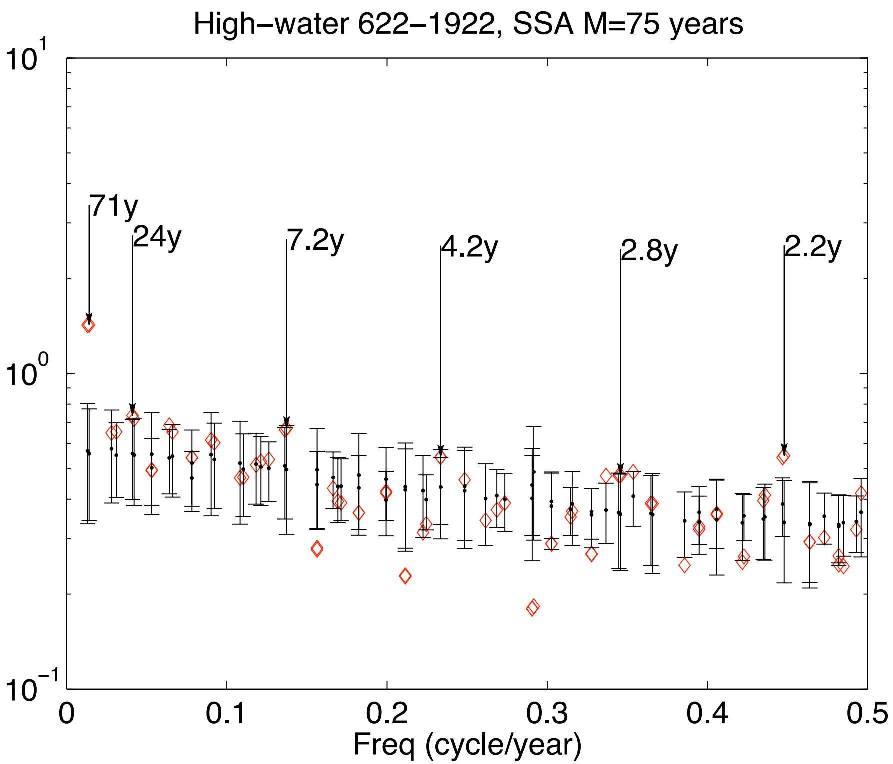
$$x(t) = sin(\frac{2\pi}{300}t) * cos(\frac{2\pi}{40}t + \frac{\pi}{2} sin\frac{2\pi}{120}t)$$

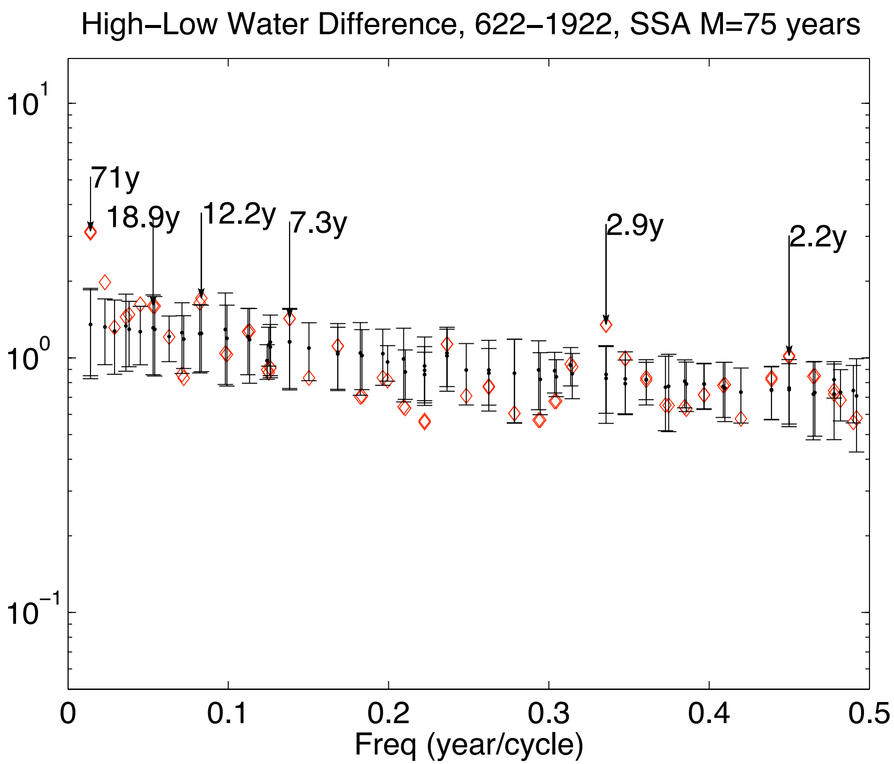
Nile River Records



- High level
- Low level

MC-SSA of Filled-in Records





SSA results for the extended Nile River records;

arrows mark highly significant peaks (at 95%), in both SSA and MTM.

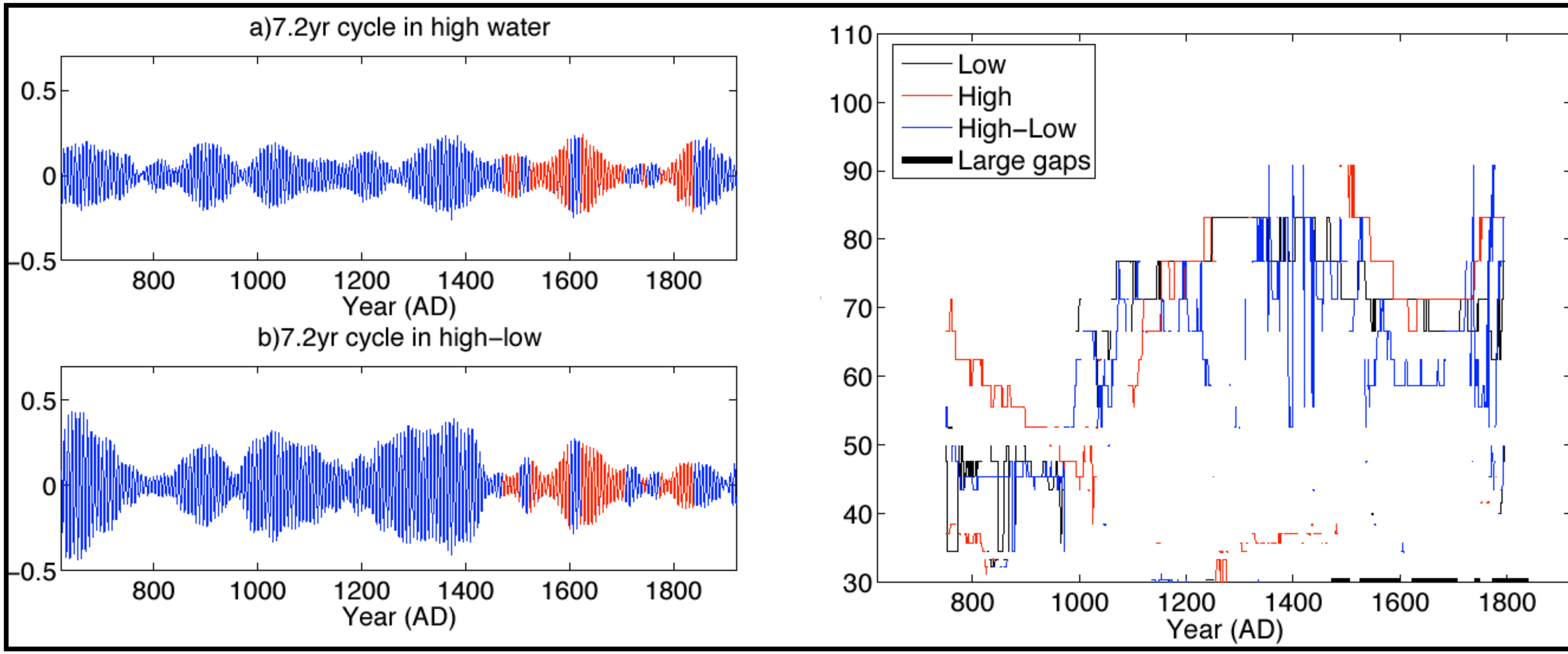
Table 1a: Significant oscillatory modes in short records (A.D. 622–1470)

Periods	Low	High	High-Low
40–100yr	**64** (9.3%)	**64** (6.9%)	**64** (6.6%)
20–40yr		[32]	
10–20yr	**12.2** (5.1%), **18.0** (6.7%)		**12.2** (4.7%), **18.3** (5.0%)
5–10yr	**6.2** (4.3%)	7.2 **(4.4%)**	7.3 **(4.4%)**
0–5yr	3.0 (2.9%), 2.2 (2.3%)	**3.6** (3.6%), 2.9 (3.4%), 2.3 (3.1%)	**2.9** (4.2%),

Table 1b: Significant oscillatory modes in extended records (A.D. 622–1922)

Periods	Low	High	High-Low
40–100yr	**64** (13%)	**85** (8.6%)	**64** (8.2%)
20–40yr		**23.2** (4.3%)	
10–20yr	[12], **19.7** (5.9%)		**12.2** (4.3%), **18.3** (4.2%)
5–10yr	[6.2]	7.3 (4.0%)	7.3 (4.1%)
0–5yr	3.0 (4%), 2.2 (3.3%)	**4.2** (3.3%), 2.9 (3.3%), 2.2 (2.9%)	[4.2], 2.9 (3.6%), 2.2 (2.6%)

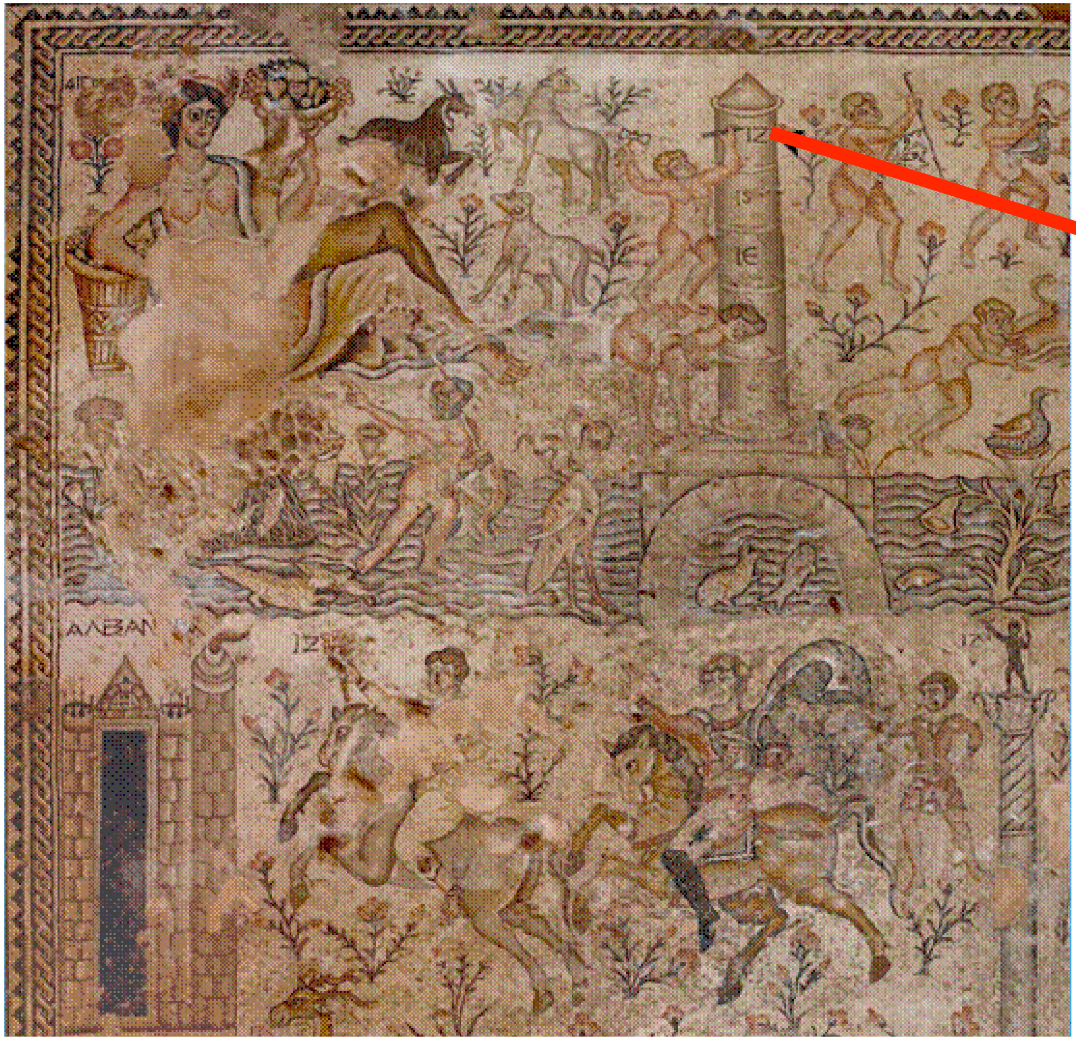
Significant Oscillatory Modes



SSA reconstruction of the 7.2-yr mode in the extended Nile River records: (a) high-water, and (b) difference. Normalized amplitude; reconstruction in the large gaps in red.

Instantaneous frequencies of the oscillatory pairs in the low-frequency range (40–100 yr). The plots are based on multi-scale SSA [Yiou *et al*., 2000]; local SSA performed in each window of width *W* = 3*M*, with *M* = 85 yr.

How good were Joseph's predictions?



Pretty good!