# The Nile River Records Revisited:
## How good were Joseph's predictions?

*Michael Ghil*

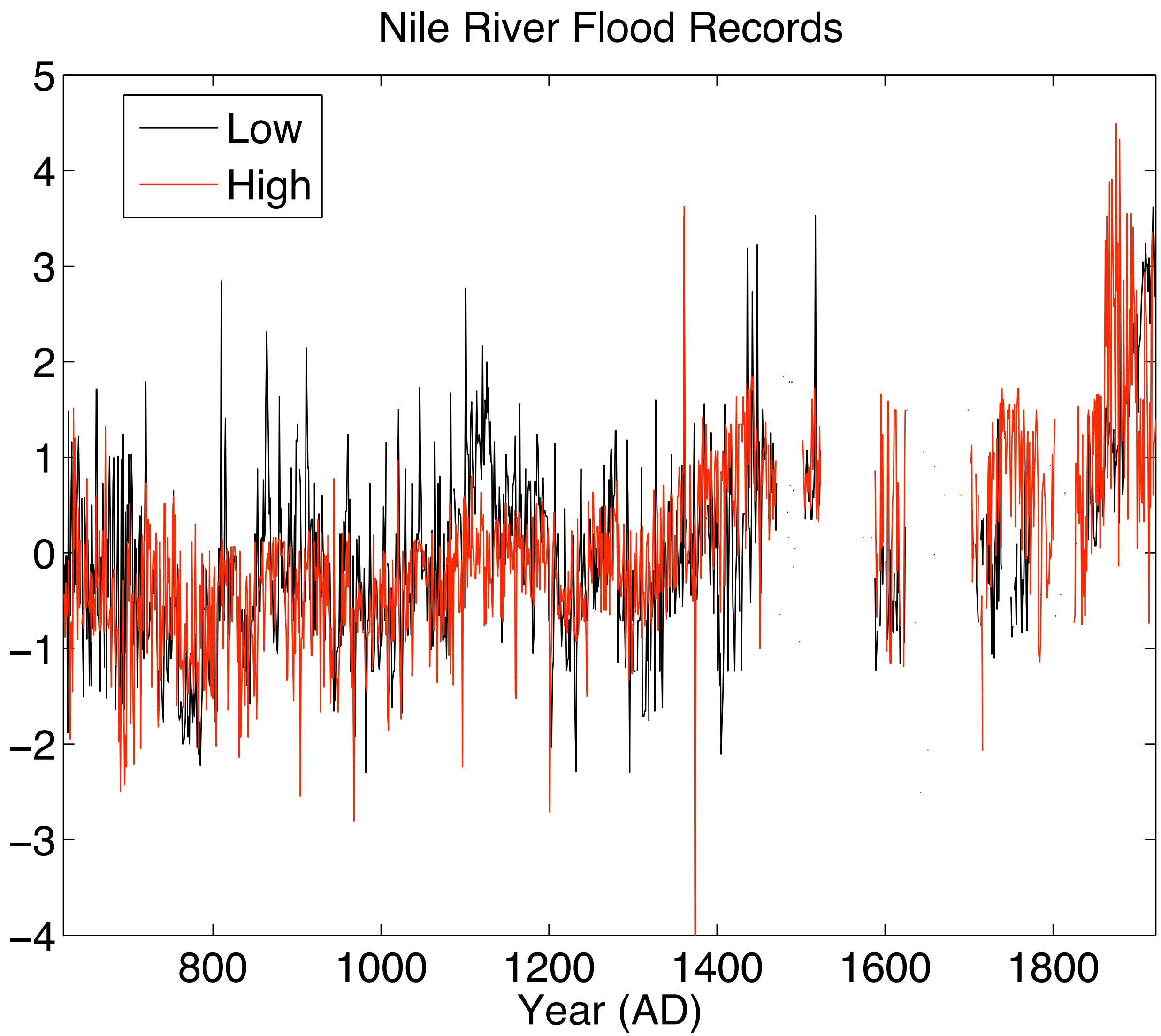
*Ecole Normale Supérieure, Paris,*
*and University of California, Los Angeles*

*Joint work with Yizhak Feliks, IIBR,*
*and Dmitri Kondrashov, UCLA*

# *Why are there data missing?*

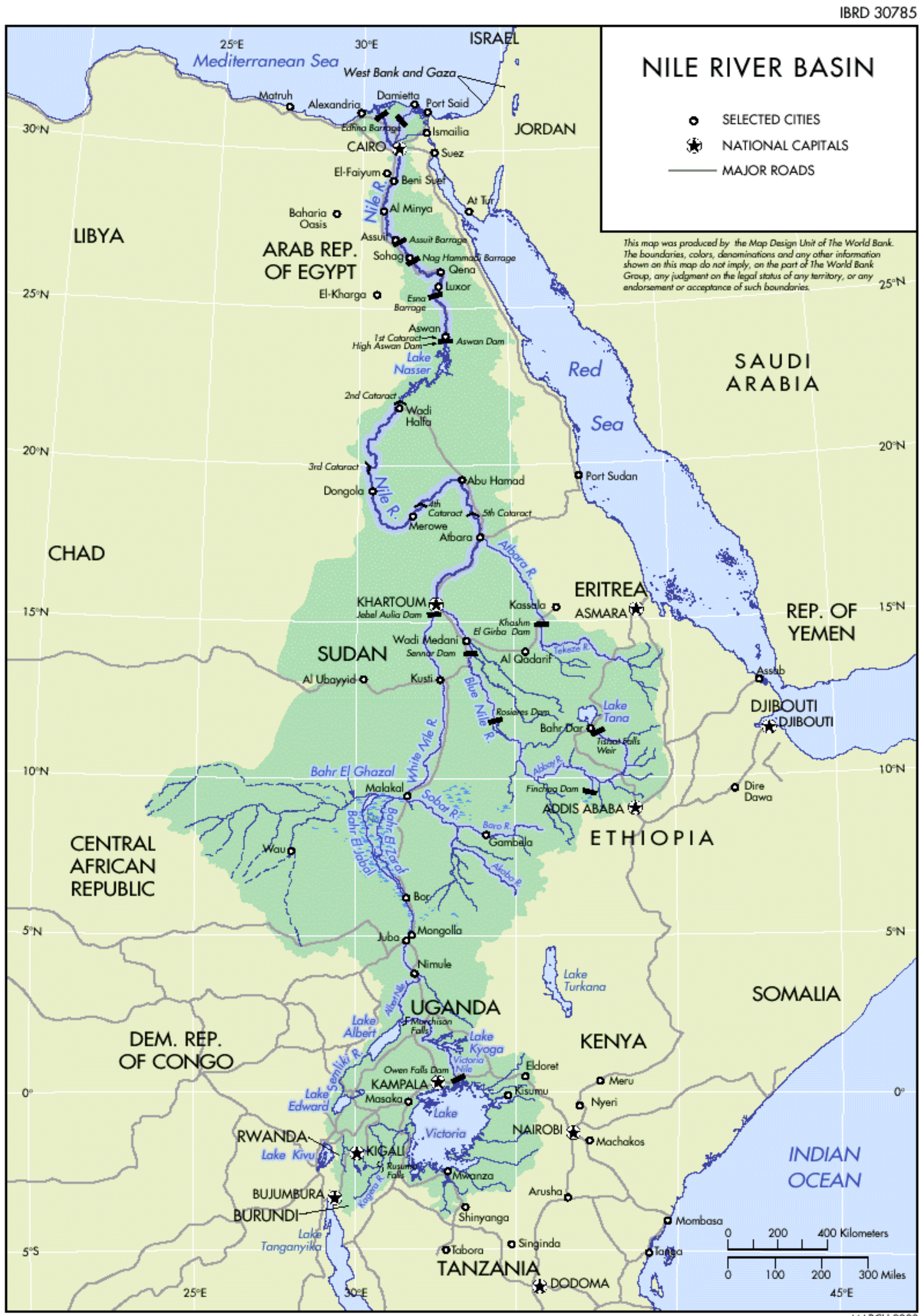

◦Byzantine-period mosaic from **Zippor**i, the capital of **Galilee** (1st century B.C. to 4th century A.D.); photo by **Yigal Feliks,** with *permission from the Israel Nature and Parks Protection Authority*)

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




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

# Nile River Basin Map




- The Blue Nile originates in the Ethiopian highlands, & contributes mainly to the floods.
- The White Nile originates from equatorial Africa, & is mainly responsible for the low level.
- Shorter-term, interannual fluctuations in the high-level record, however, do reflect also variations in the White Nile contributions (*i.e.*, in the low level)
- The accumulated precipitation of several preceding seasons affects the level of Lakes Victoria and Albert, at the sources of the White Nile.
- Hence, the time series of the annual difference between the maxima and minima is a better proxy record of the Nile floods than the high-level time series.

# *Motivation*

- Significant association between the rainfall in the catchment area of the Nile tributaries, the Indian monsoon, and ENSO [Walker, 1910; Quinn, 1992].
- A recent weakening of the close relationship between ENSO and the Indian monsoon [Kumar *et al*., 1999], however, might foreshadow a similar weakening of the East-Africa–ENSO correlation.
- North Atlantic influences over North Africa and the Middle East have been detected in several geologic [Felis *et al*., 2004] and historical [Mann, 2002] records, and may extend all the way into the Nile River's source area, further to the south.

*To ascertain whether this is so, we apply advanced methods for filling the gaps in the records and for their spectral analysis, with appropriate statistical confidence tests.*

# Singular Spectrum Analysis (SSA)

| Spatial EOFs | SSA |
|---|---|
| $\phi(\underline{x},t) = \sum a_k(t) e_k(\underline{x})$ | $X(t+\underline{s}) = \sum a_k(t) e_k(\underline{s})$ |
| PC EOF | |
| *x - space* | *s - lag* |

$$C_\phi(\underline{x},\underline{y}) = E\phi(\underline{x},\omega)\phi(\underline{y},\omega) \cong \frac{1}{T}\int_0^T \phi(\underline{x},t)\phi(\underline{y},t)dt$$

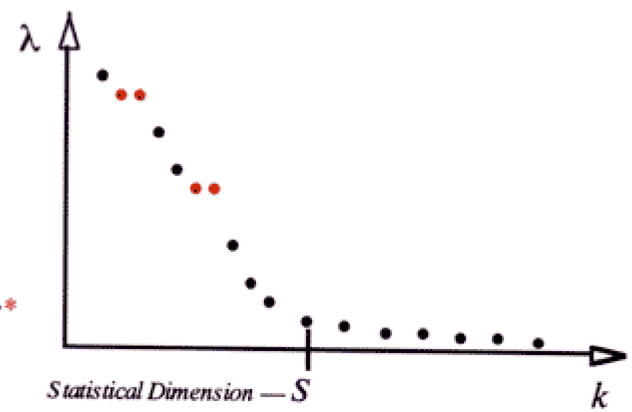
$$C_X(\underline{s}) = EX(\cdot,\omega)X(\underline{s},\omega) \cong \frac{1}{T}\int_0^T X(t)X(t+\underline{s})dt$$

$$C_\phi e_k(\underline{x}) = \lambda_k e_k(\underline{x})$$

$$C_X e_k(\underline{s}) = \lambda_k e_k(\underline{s})$$

*Pairs* $\Rightarrow$ *oscillations*

*(nonlinear) sine + cosine pair**




*Colebrook (1978), Weare and Nasstrom (1982), Broomhead and King (1986: BK), Fraedrich (1986)*

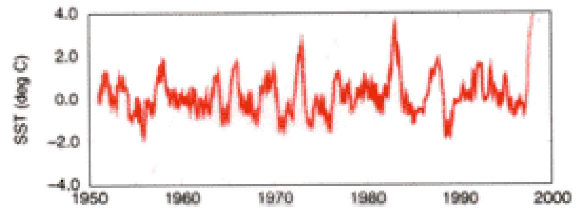
** Vautard and Ghil (1989: VG), Physica, 35D, 395–424*

*BK+VG: Analogy between Mañe-Takens embedding & Wiener-Khinchin theorem*

# Singular Spectrum Analysis

SSA decomposes geophysical time series:

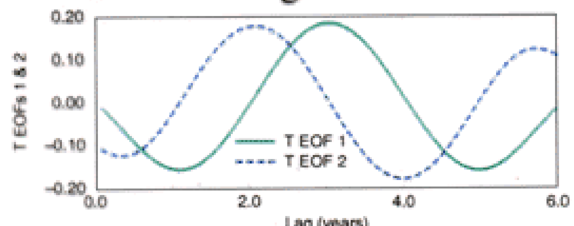
**Time series**




Niño-3
SST anomalies

into Temporal EOFs (T-EOFs) and
Temporal Principal Components (T-PCs),
based on the series' lag-covariance matrix:

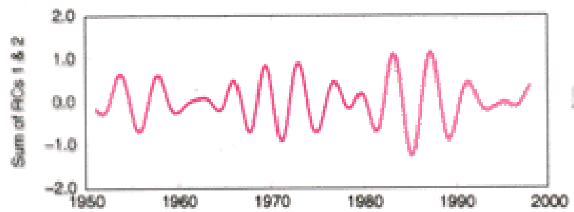
**T-EOFs**




Leading pair:
48-month period

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

**RCs**



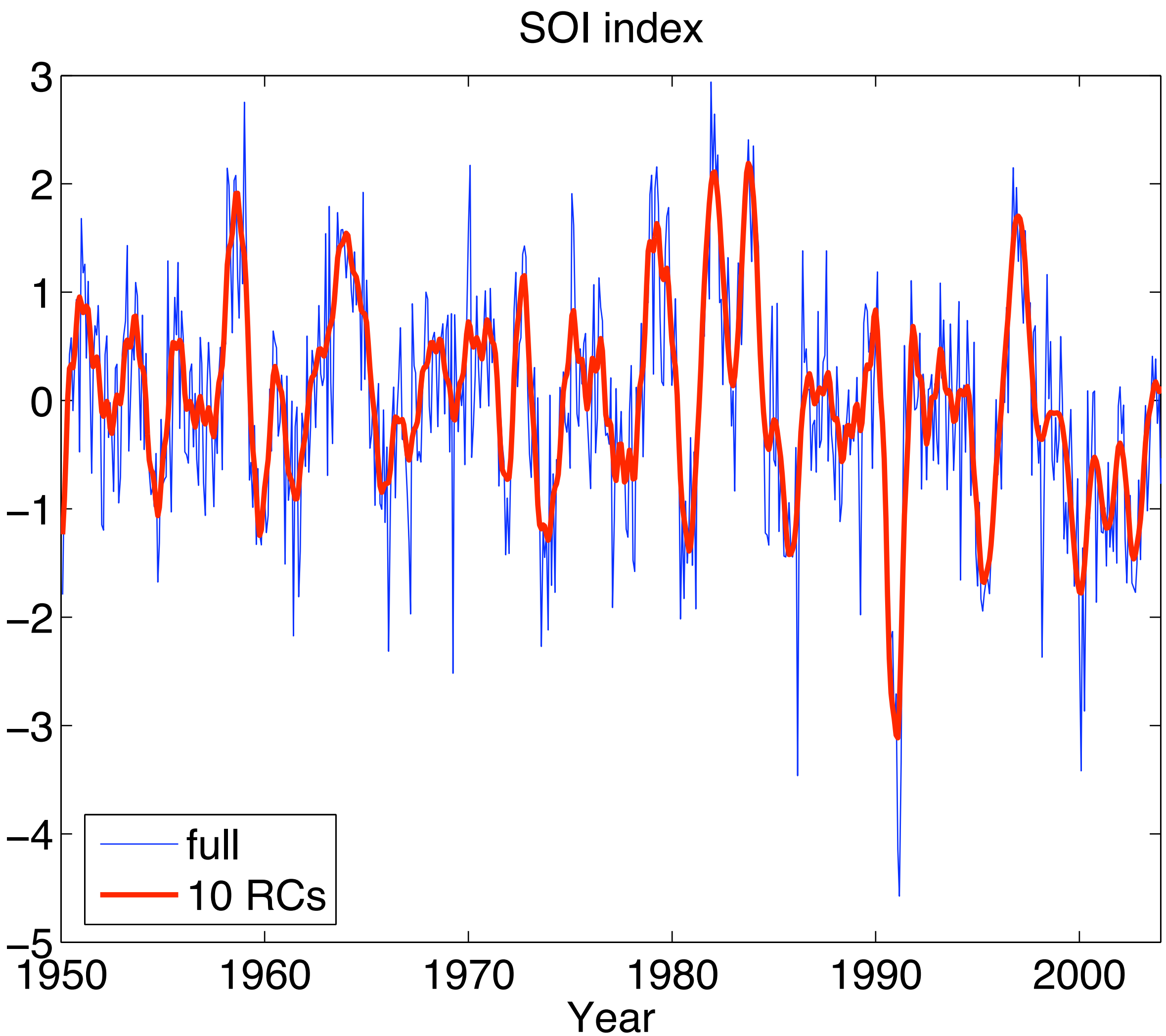

Leading pair

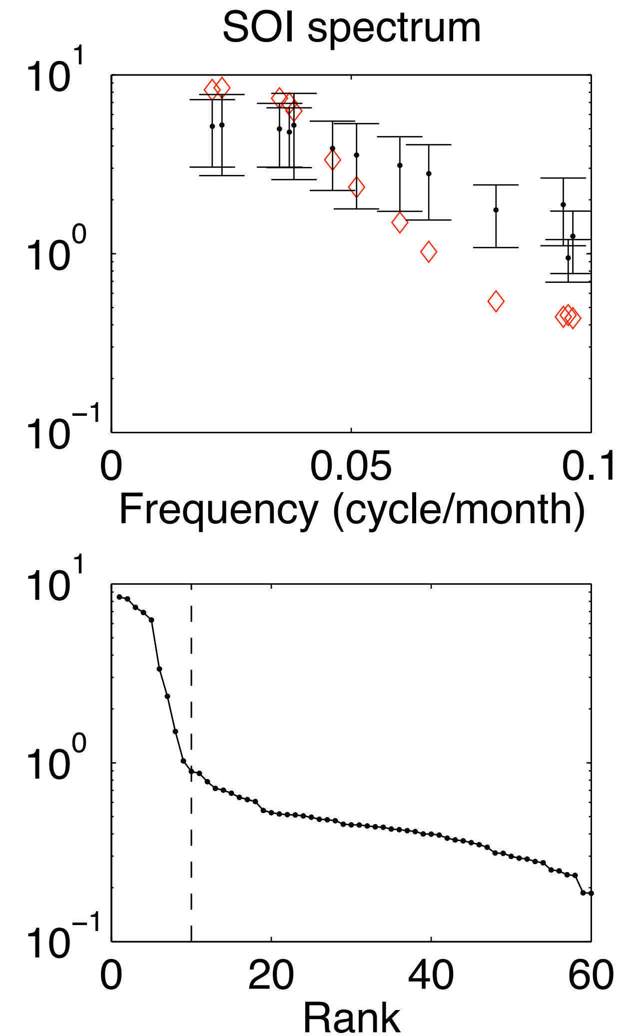
- SSA is good at isolating oscillatory behavior via paired eigen-elements.
- SSA tends to lump signals that are longer-term than the window into one or two trend components.

*Selected References*


Vautard & Ghil, 1989, *Physica D*; Dettinger *et al.*, 1995, *Eos, Trans. AGU.*

# SSA (& M-SSA),continued...




- Break in slope of SSA spectrum marks "**noise**" EOFs
- Formal Monte-Carlo test (*Allen and Smyth* 1994) identifies 4-yr and 2-yr ENSO oscillatory modes
- Size of window $M = 60$ is enough to "resolve" these modes in a monthly SOI time series

# *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 **to *fill in* gaps by applying iterative SSA (or M-SSA):**

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

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

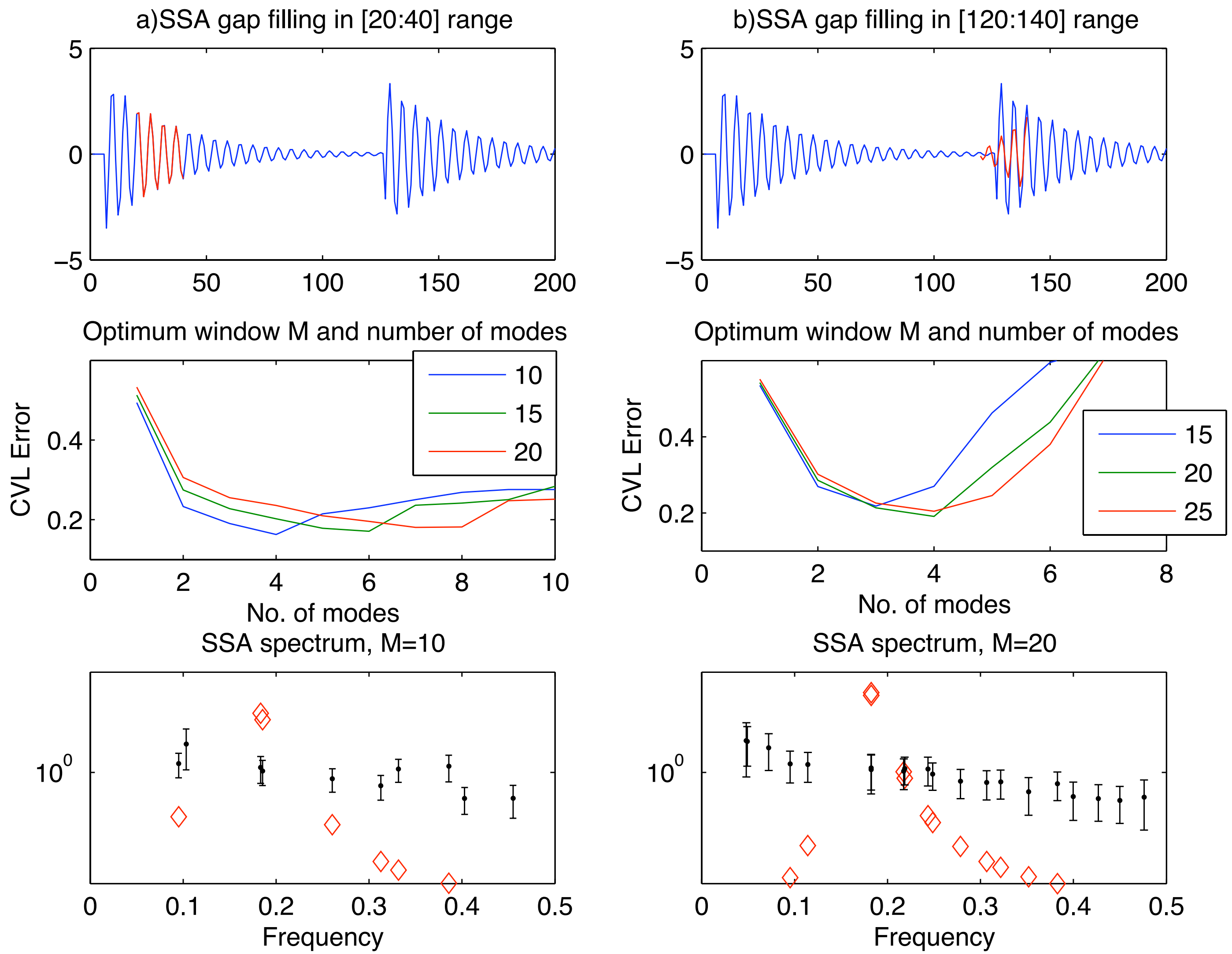
(1) for a given window width ***M***: center the original data by computing the unbiased value of the mean and set the missing-data values to zero.

(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.

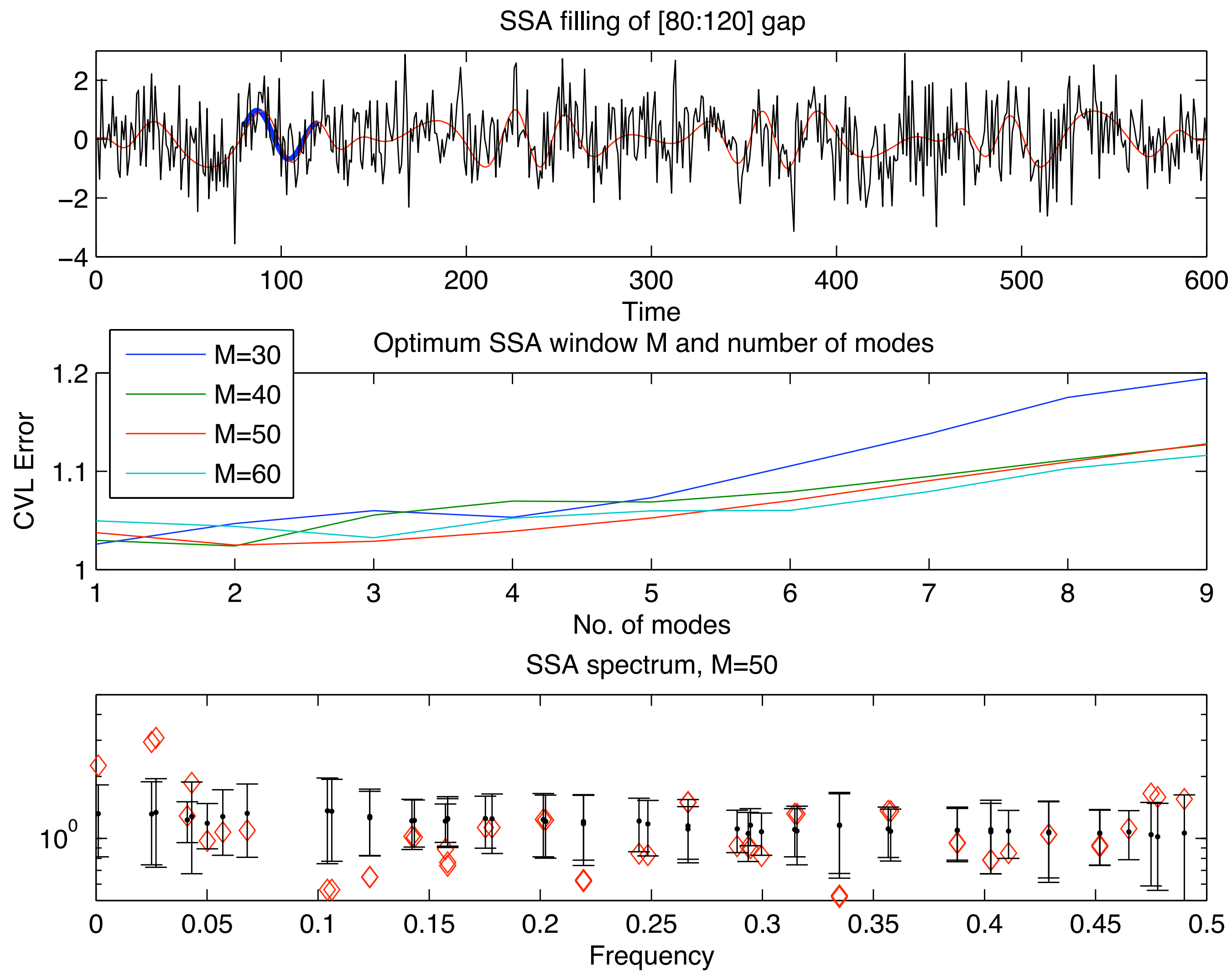
**(4) apply cross-validation** to optimize the ***window width M*** and ***number*** of dominant SSA (M-SSA) modes to fill the gaps: a portion of available data (selected at random) is flagged as missing and RMS error in reconstruction is computed.

# *Synthetic I: Gaps in Amplitude-Modulated Oscillatory Signal (with no noise)*



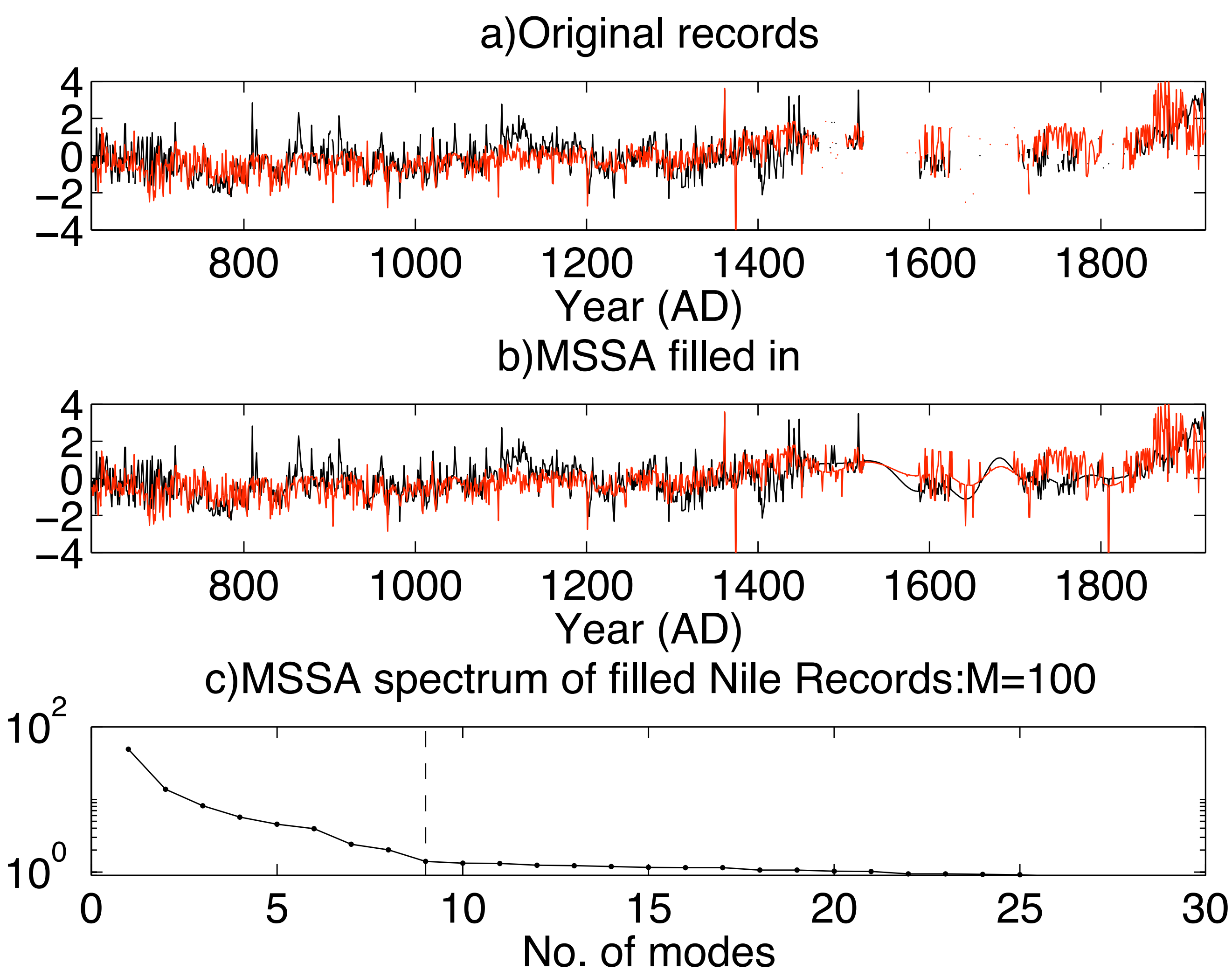

- Very good gap filling for smooth modulation; OK for sudden modulation.

# *Synthetic II: Gaps in Oscillatory Signal +Noise*



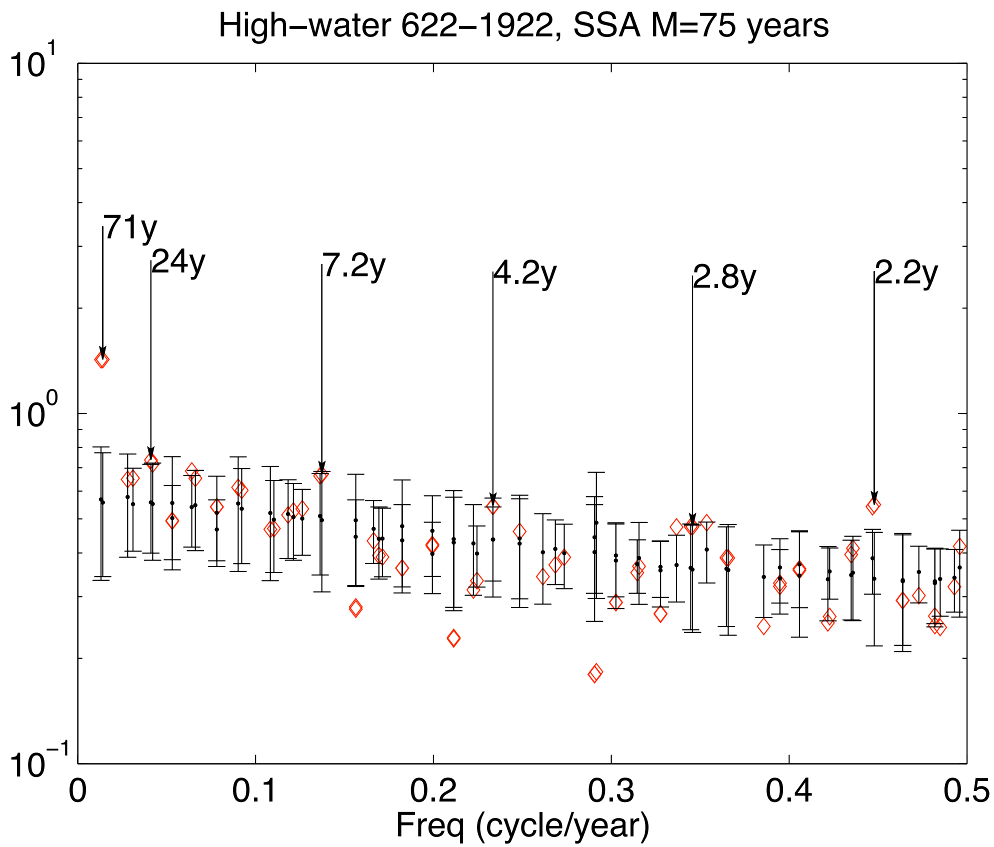

$$x(t) = sin(\frac{2\pi}{300}t) * cos(\frac{2\pi}{40}t + \frac{\pi}{2}sin\frac{2\pi}{120}t)$$

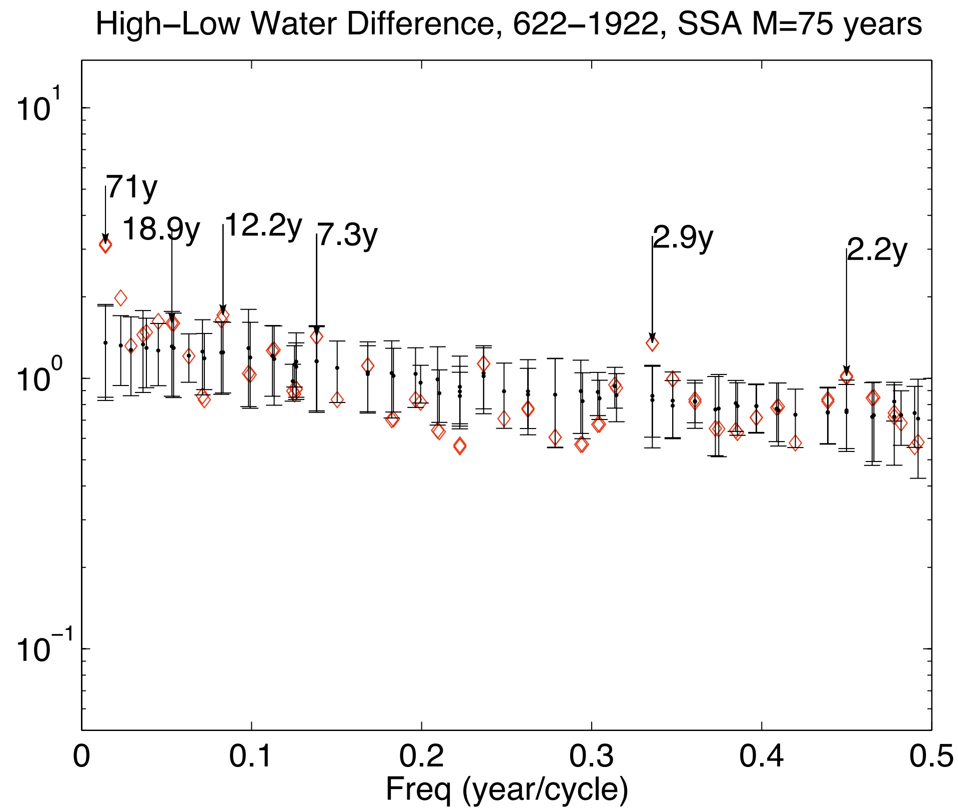
# *Nile River Records*




- 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.

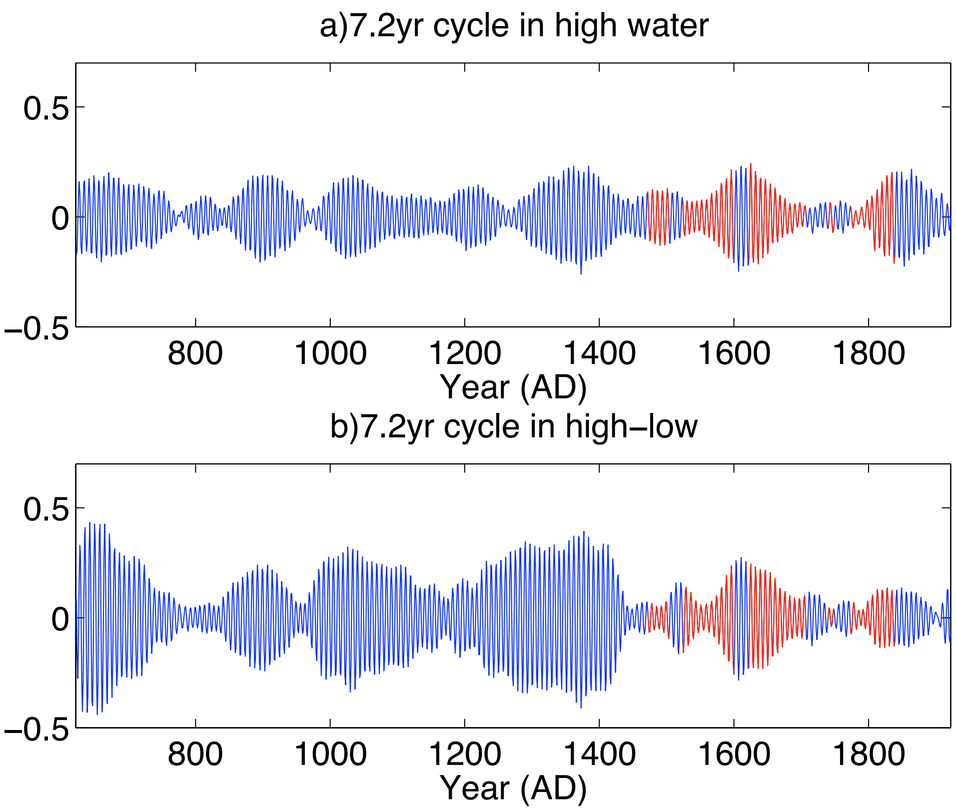
**Table 1a: Significant oscillatory modes in short records (A.D. 622–1470)**

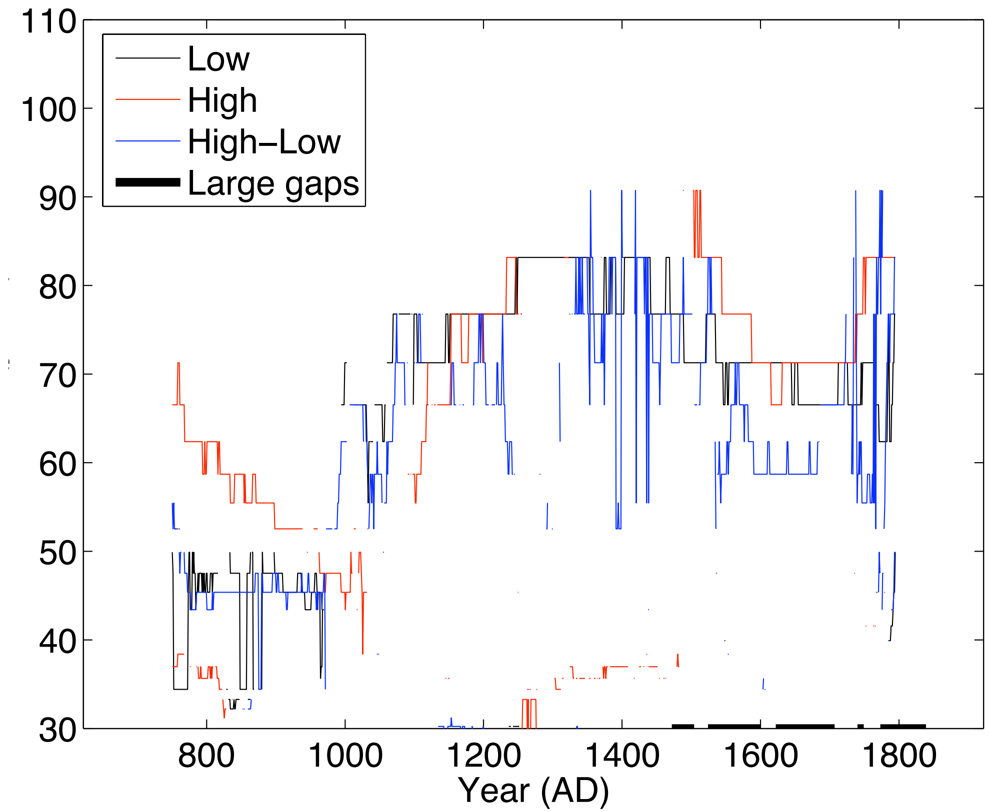
| Periods | Low | High | High-Low |
|---|---|---|---|
| 40–100yr | 64 (9.3%) | 64 (6.9%) | 64 (6.6%) |
| 20–40yr | | [32] | |
| 10–20yr | **12.2** (5.1%),<br>**18.0** (6.7%) | | **12.2** (4.7%),<br>**18.3** (5.0%) |
| 5–10yr | **6.2** (4.3%) | 7.2 **(4.4%)** | 7.3 **(4.4%)** |
| 0–5yr | 3.0 (2.9%),<br>2.2 (2.3%) | **3.6** (3.6%),<br>2.9 (3.4%),<br>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%),<br>**18.3** (4.2%) |
| 5–10yr | [6.2] | 7.3 (4.0%) | 7.3 (4.1%) |
| 0–5yr | 3.0 (4%),<br>2.2 (3.3%) | **4.2** (3.3%),<br>2.9 (3.3%),<br>2.2 (2.9%) | [4.2],<br>2.9 (3.6%),<br>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=3M*, with *M* = 85 yr.

# *Summary*

- **SSA (M-SSA)** can be used for filling **gaps in geophysical data sets**, by using **dominant spatio-temporal modes** of the data.
- The **7–8-yr period in Nile River records** equals that of a North Atlantic spectral peak [Ghil & Vautard, 1991; Moron *et al*., 1998; Wunsch, 1999].
- This peak may be due to intrinsic ocean variability [Speich *et al*., 1995; Dijkstra & Ghil, 2005; Simonnet *et al*., 2005].
- Our results suggest that the **climate of East Africa** has been subject to influences from the North Atlantic, besides those already documented from the Tropical Pacific.

Kondrashov, Feliks & M. Ghil (2005): Oscillatory modes of extended Nile River records (A.D. 622–1922), *Geophys. Res. Lett.*, in press.

*Free SSA-MTM Toolkit at http://www.atmos.ucla.edu/tcd/ssa*

# How good were Joseph's predictions?


Pretty good!