Smooth Time Series - Applications

Dmitri Kondrashov

University of California, Los Angeles

Joint work with M. Ghil, and many others

http://www.atmos.ucla.edu/tcd

Spectral Analysis

S(f) ~ f $^{-p}$ ***+ 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)

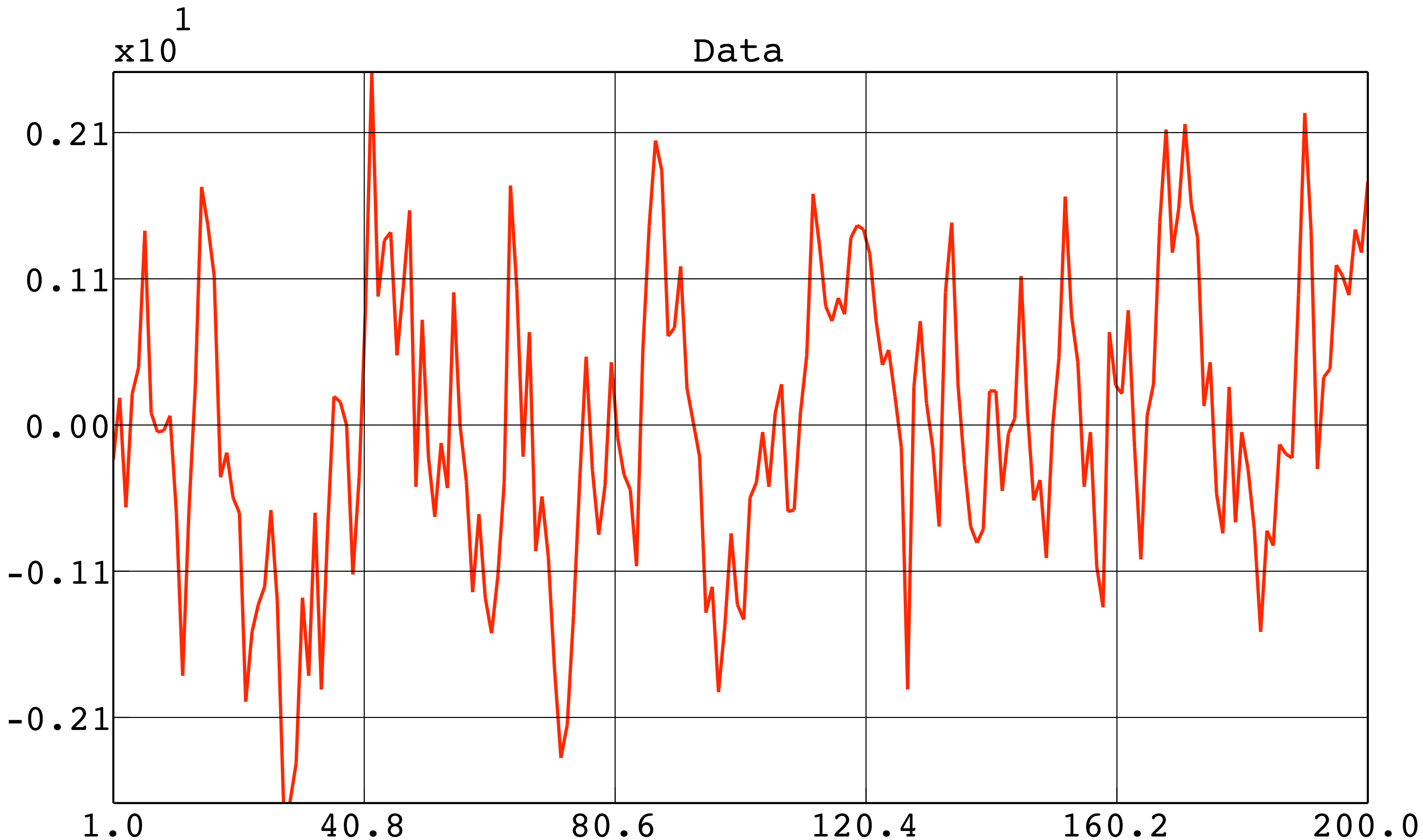
"White" noise, *p* = 0

Main challenge for ***short and noisy*** geophysical time series to distinguish between **poles** and (**red) noise.**

$$\ddot{x} = -\omega^2 x \;\; vs. \;\; \dot{x} = -\lambda x$$

Tradeoff for spectral methods: **resolution (smoothing, windowing) vs. spurious peaks & power leakage.**

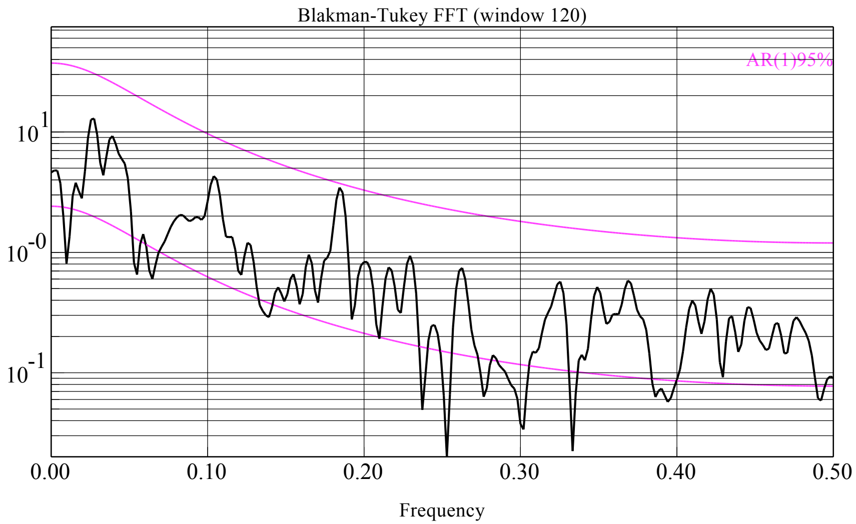
Synthetic example (Task & Prize!!!)

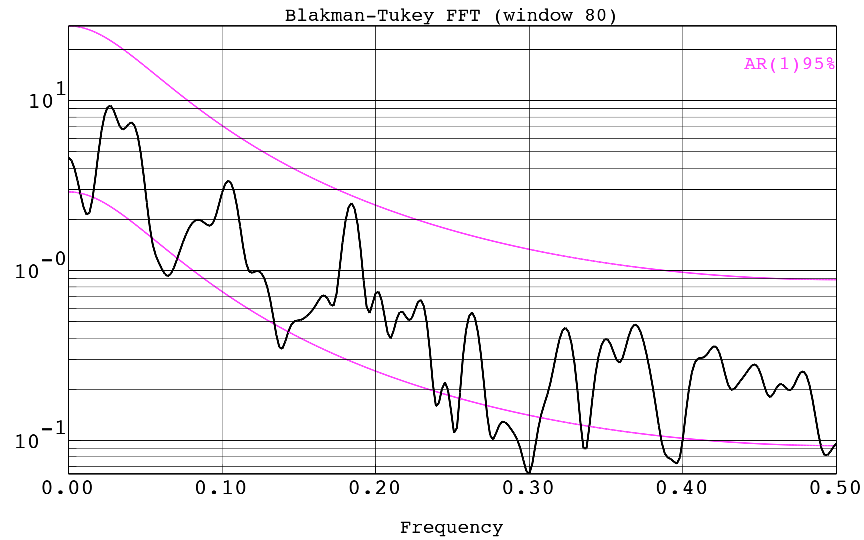


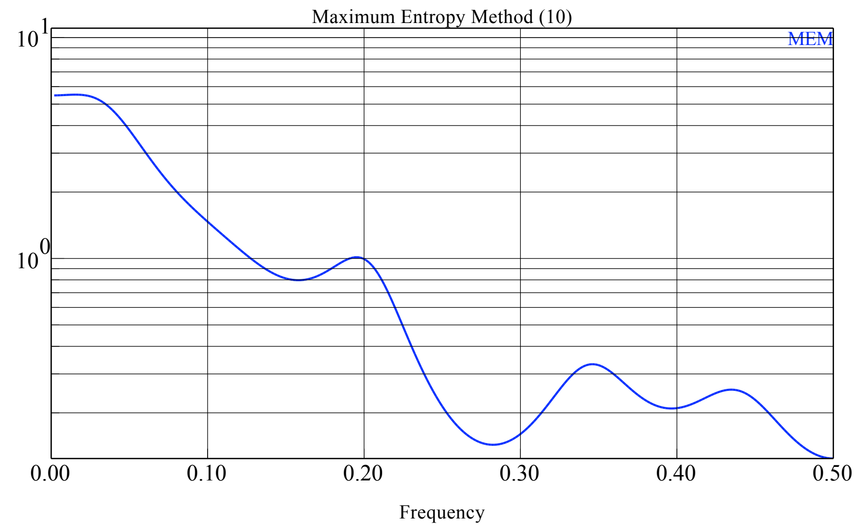
Q: ***Is there a periodicity and what is its period?***

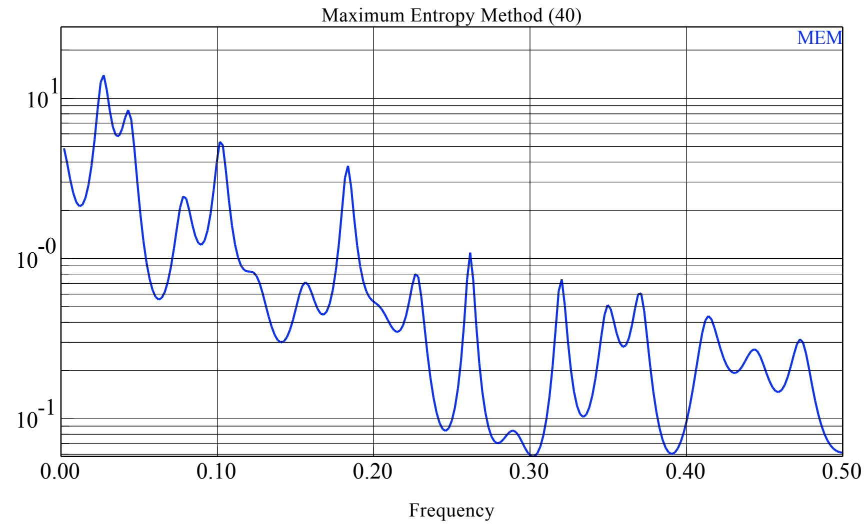
A: ***What is the underlying noise "null hypothesis"?***

Classical Spectral Methods









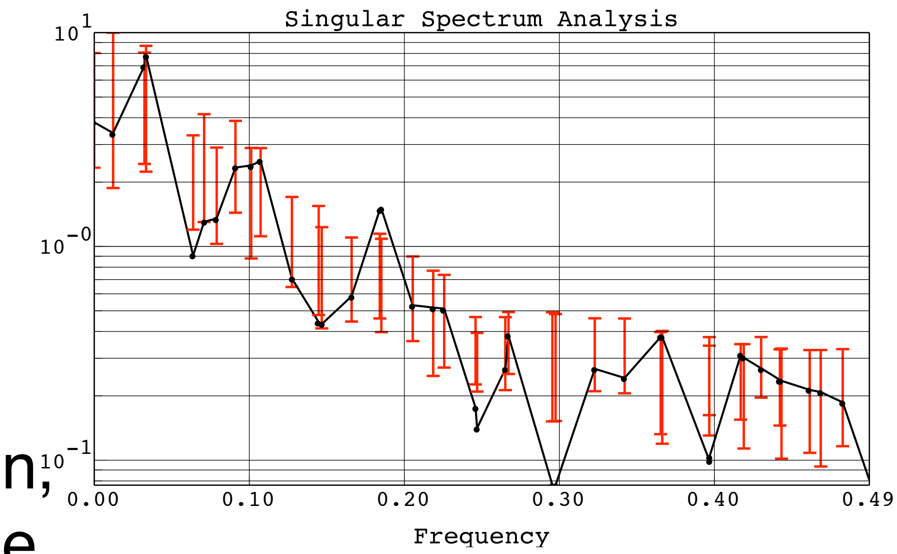
Advanced Spectral Methods

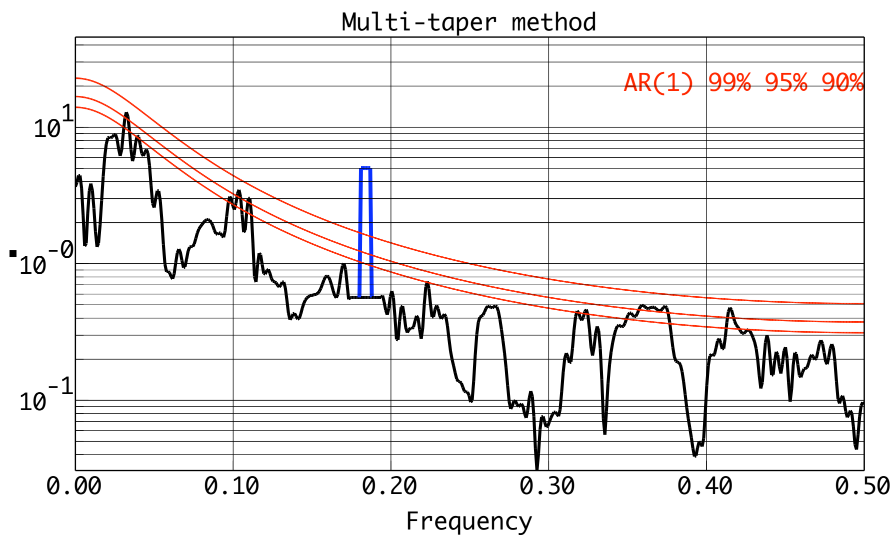
- **Singular spectrum analysis (SSA)** and **Multi-taper method (MTM)**.

- detection of periodic signals: phase and amplitude modulation, intermittent behavior, large noise.

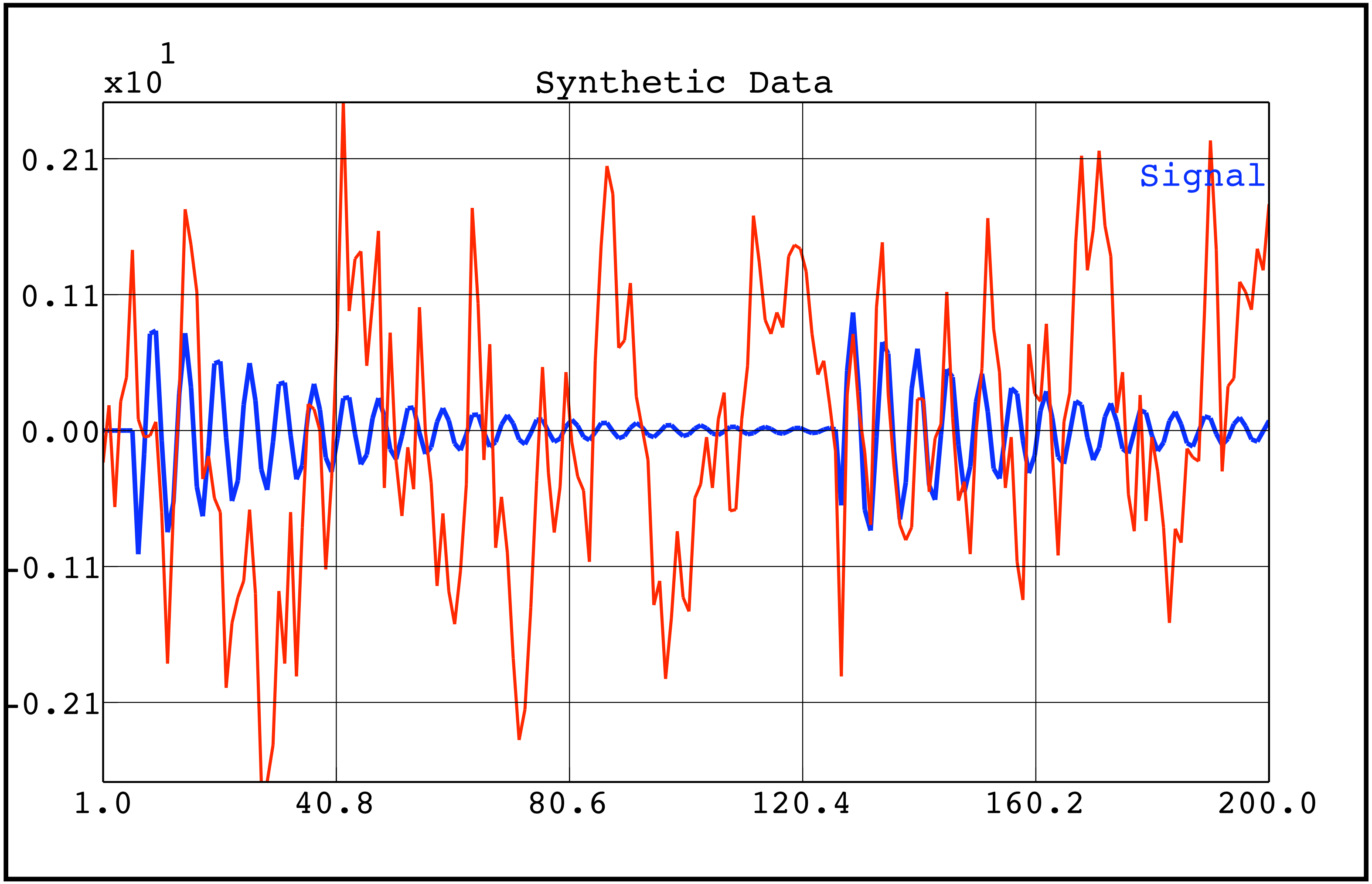
- use **data-adaptive** orthogonal basis in **frequency domain** (MTM) and **time domain** (SSA).

- significance tests for spectral peaks.





Right answer!



SSA Power Spectra & Reconstruction

- **A. Transform pair**:

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

For given window ***M***, e_k's are **adaptive filters** (empirical orthogonal functions)

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

the a_k's are **filtered time series**, principal components in time domain.

B. Power spectra

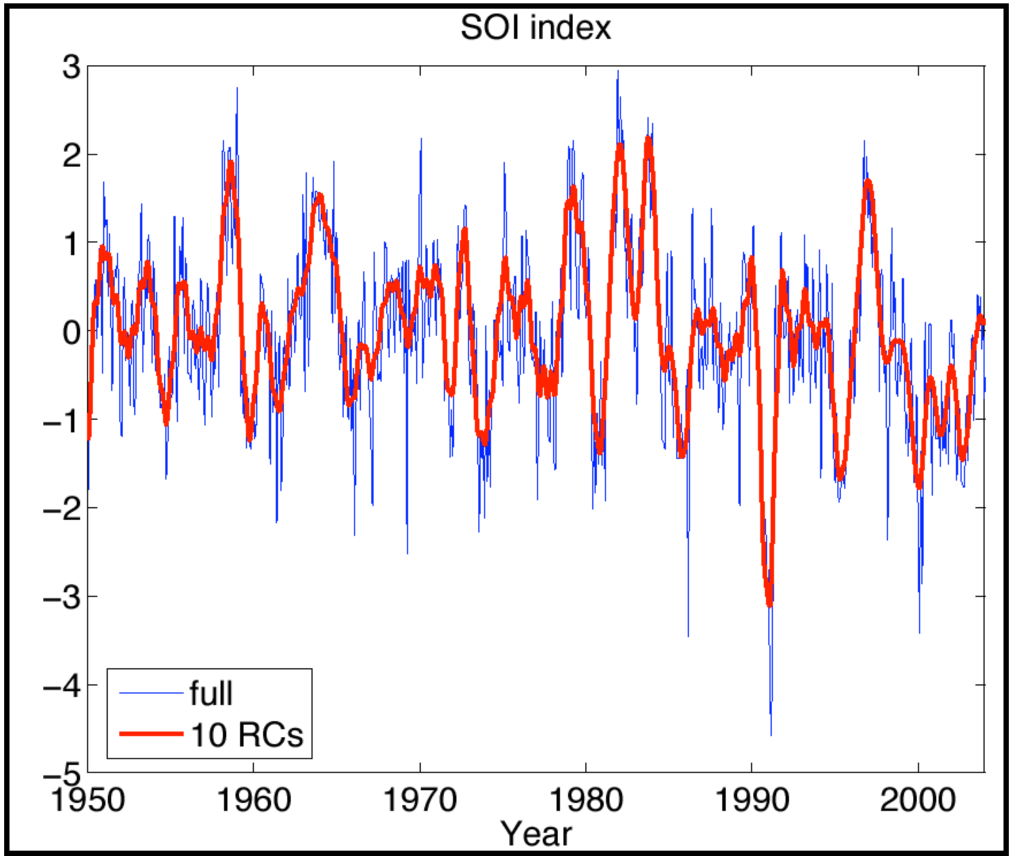
$$S_X(f) = \sum_{k=1}^{M} S_k(f); \quad S_k(f) = \hat{R}_k(s); \quad R_k(s) \approx \frac{1}{T} \int_0^T a_k(t) a_k(t+s) dt$$

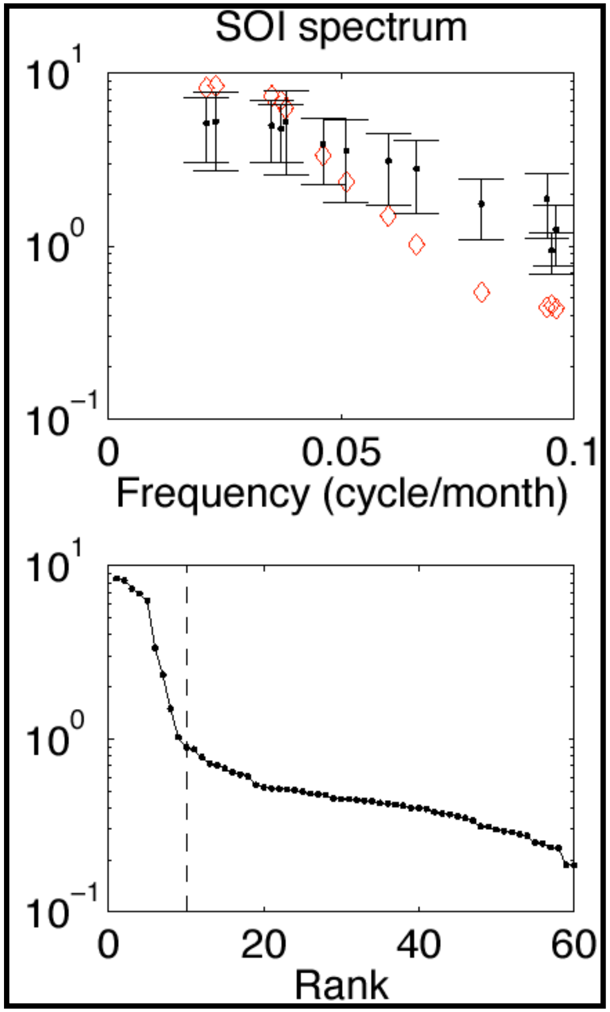
C. Reconstruction

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

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

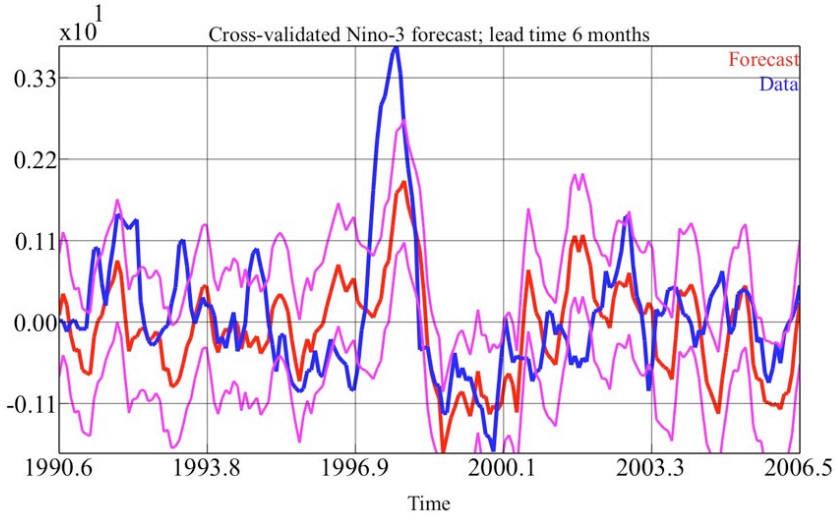
SSA of Southern Oscillation Index (ENSO)

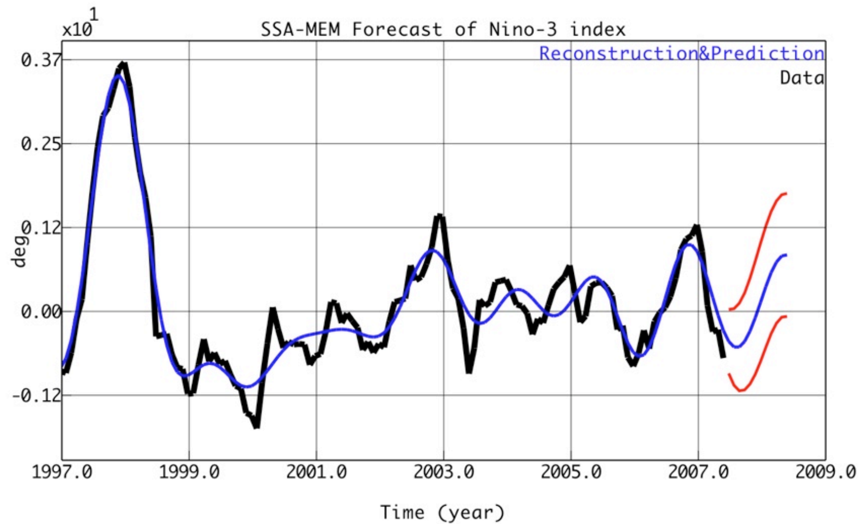




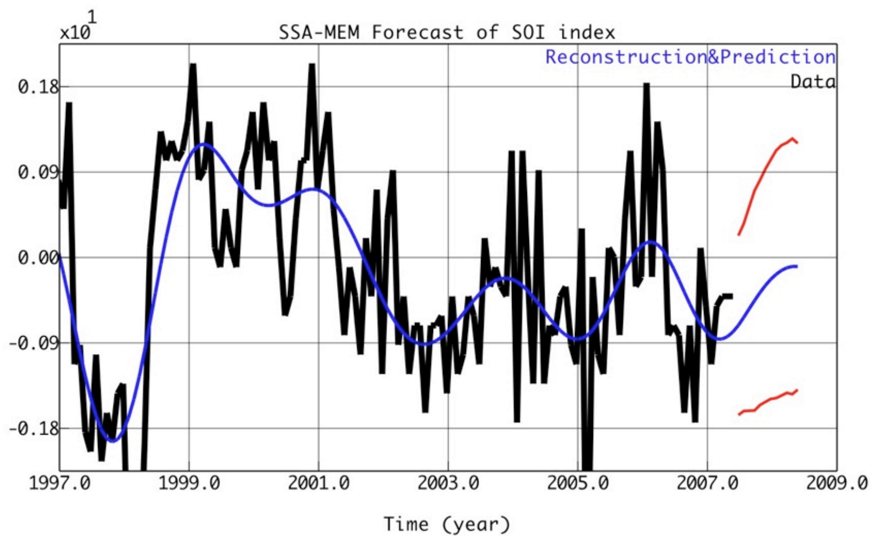
- **Powerful noise filter**: Break in slope of SSA spectrum distinguishes **"significant"** from **"noise"** EOFs
- Formal Monte-Carlo test identifies 4-yr and 2-yr ENSO oscillatory modes (**SSA pairs**). A window size of M = 60 is enough to "resolve" these modes in a monthly SOI time series.

SSA Forecast (ENSO)

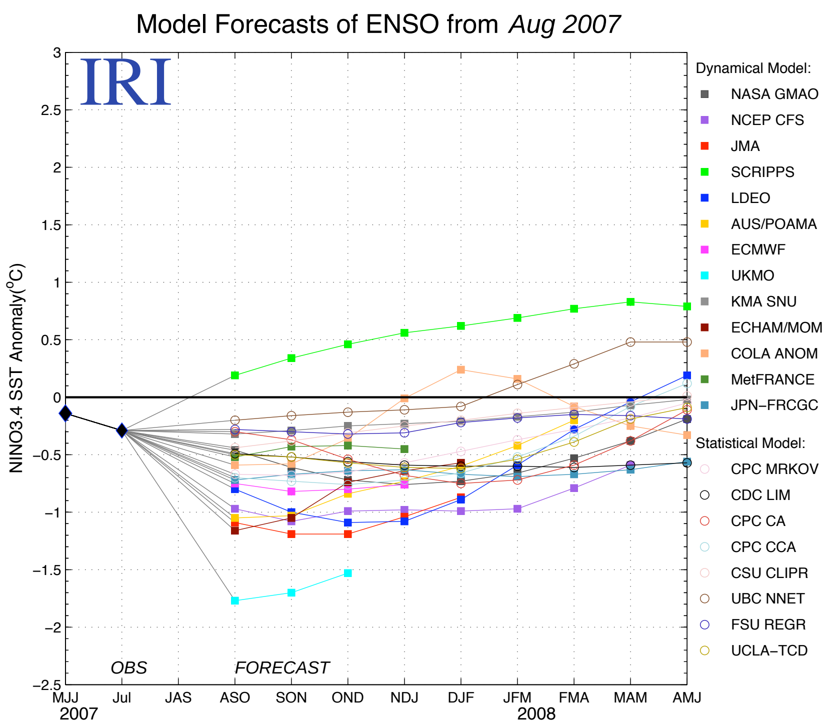


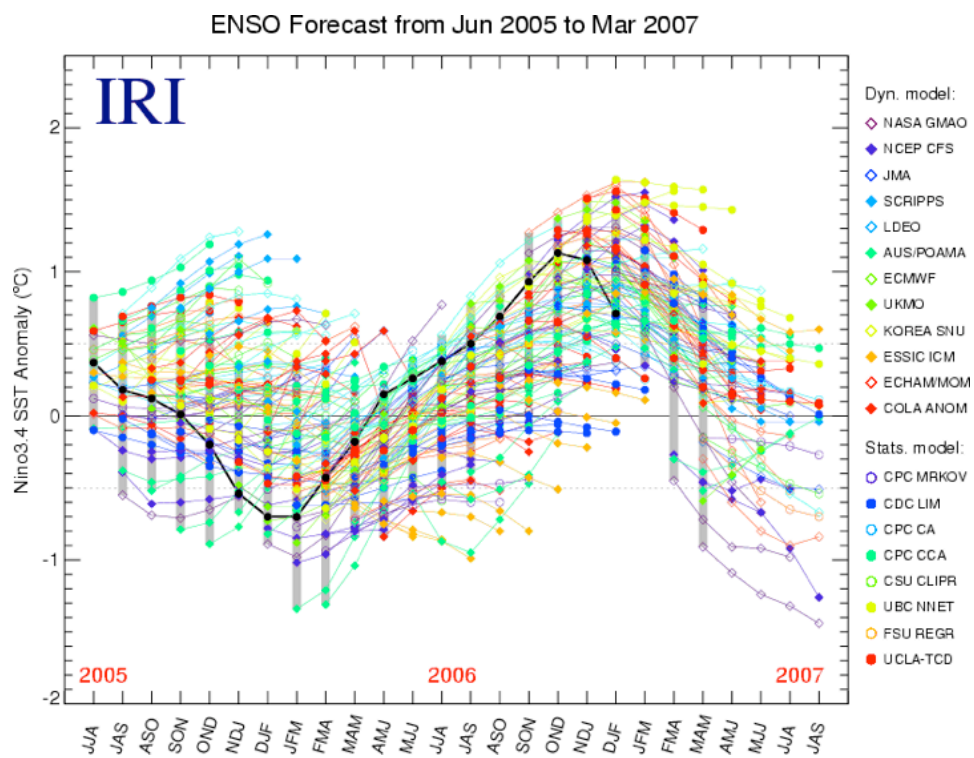


- Filter "**signal**" and forecast with AR(M) model.
- Cross-validation to find optimum number of "**signal**" components and error bars of the forecast.
- Correlations are both advantage and limitations of empirical models.



Real-time Plume of Climate Forecasts (ENSO)

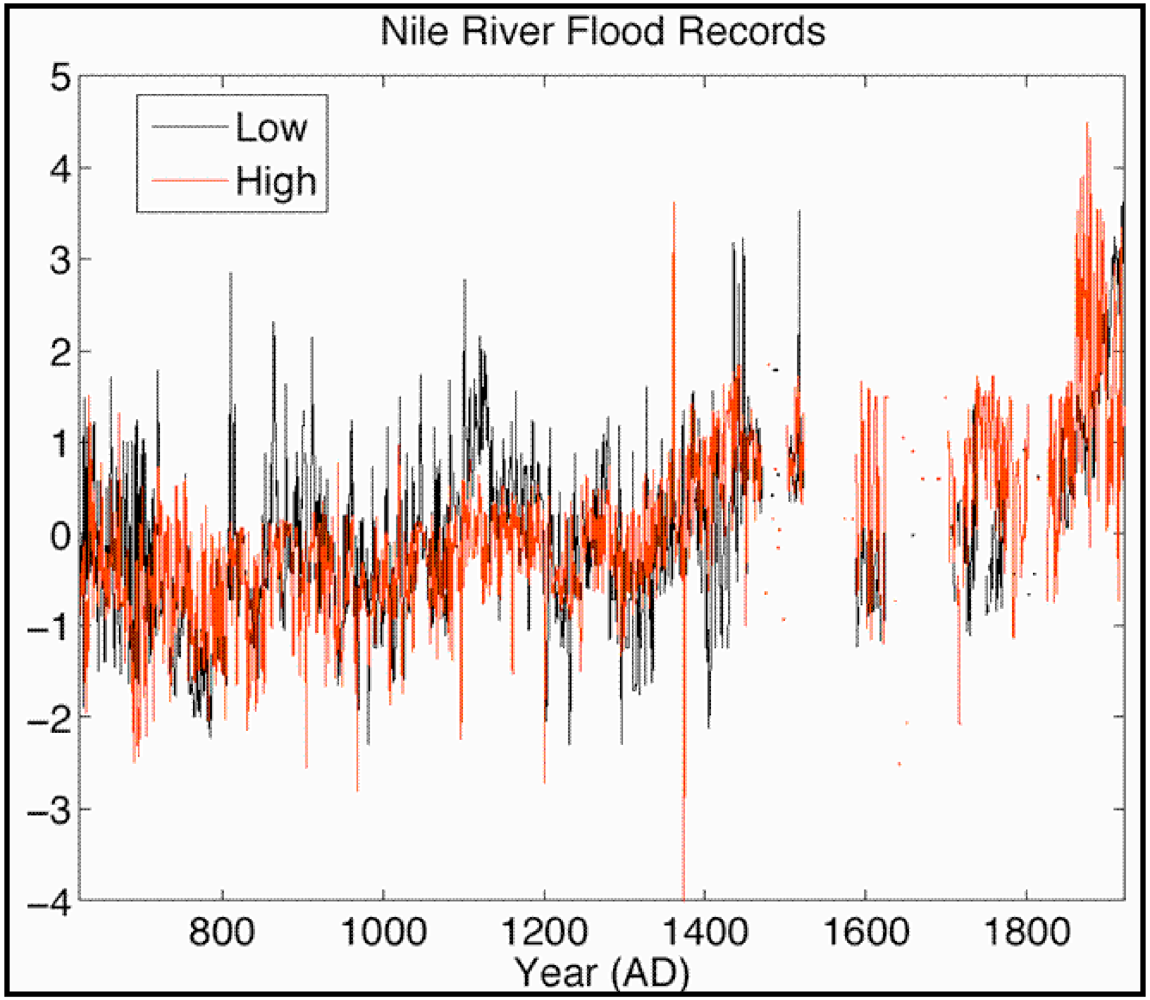




- IRI Multi-model forecast of Nino-34 index
- UCLA-TCD: Kondrashov, D., S. Kravtsov, A. W. Robertson, and M. Ghil (2005), A hierarchy of data-based ENSO models, J. Climate, 18, 4425–4444. 495

Dealing with Missing Data

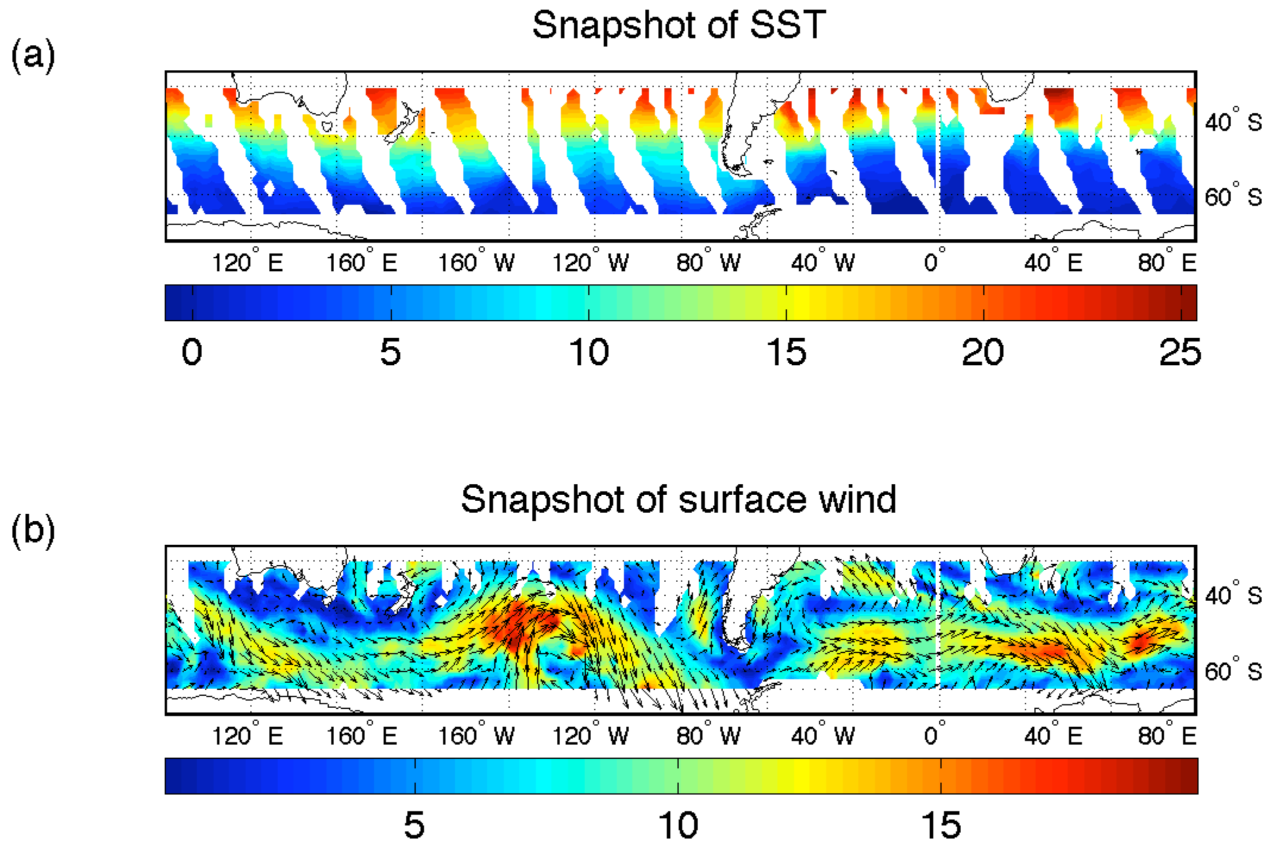
Historical records are full of “gaps”....



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

... nowdays on Earth...

- SST (AMSR-E), daily 2x2, June 2002 – February 2007: 38.2% of missing points
- Wind (QuikSCAT), daily 2x2, July 1999 -- February 2007:17.2% of missing points



- Gaps: satellite coverage, precipitation and clouds.

SSA gap-filling

1. Choose window ***M*** and set ***K***=1. Flag fraction of dataset ***X(t)(t=1:N)*** as "missing" for cross-validation.

2. Update mean and covariance, find leading ***K*** EOFs

$$\mathbf{D} = \begin{pmatrix} X(1) & X(2) & . & . & X(M) \\ X(2) & X(3) & . & . & X(M+1) \\ . & . & . & . & . \\ X(N'-1) & . & . & . & X(N-1) \\ X(N') & X(N'+1) & . & . & X(N) \end{pmatrix}$$

$$\mathbf{C}_X = \frac{1}{N'}\mathbf{D}^{\mathrm{t}}\mathbf{D}; \mathbf{C}_X E_k = \lambda_k E_k$$

3. Reconstruct missing points using ***K*** EOFs

$$A_k(t) = \sum_{j=1}^{M} X(t+j-1)E_k(j)$$

$$R_{\mathcal{K}}(t) = \frac{1}{M_t}\sum_{k\in\mathcal{K}}\sum_{j=L_t}^{U_t} A_k(t-j+1)E_k(j);$$

4. If convergence for missing points, ***K = K +1.*** Check cross-validation error, and Go to Step 2 if necessary.

Utilize (spatio) temporal correlations to iteratively compute self-consistent lag-covariance matrix => can be applied to very gappy data.

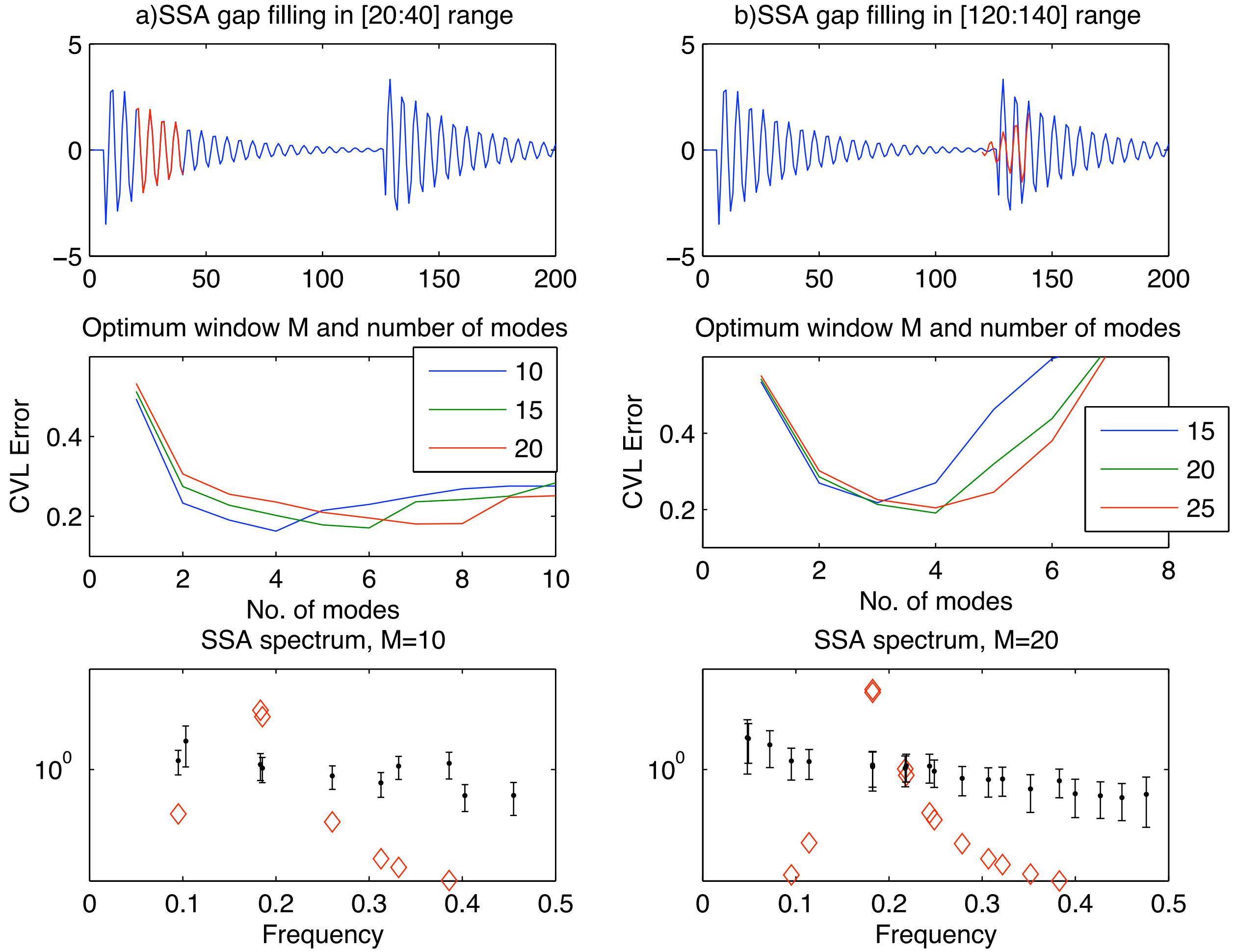
Follows expectation maximization (EM) procedure for finding maximum likelihood estimates of mean and covariance matrix.

A few ***K*** leading EOFs correspond to the "smooth" modes, while the rest is noise.

Provides both spectral analysis and estimates of missing data.

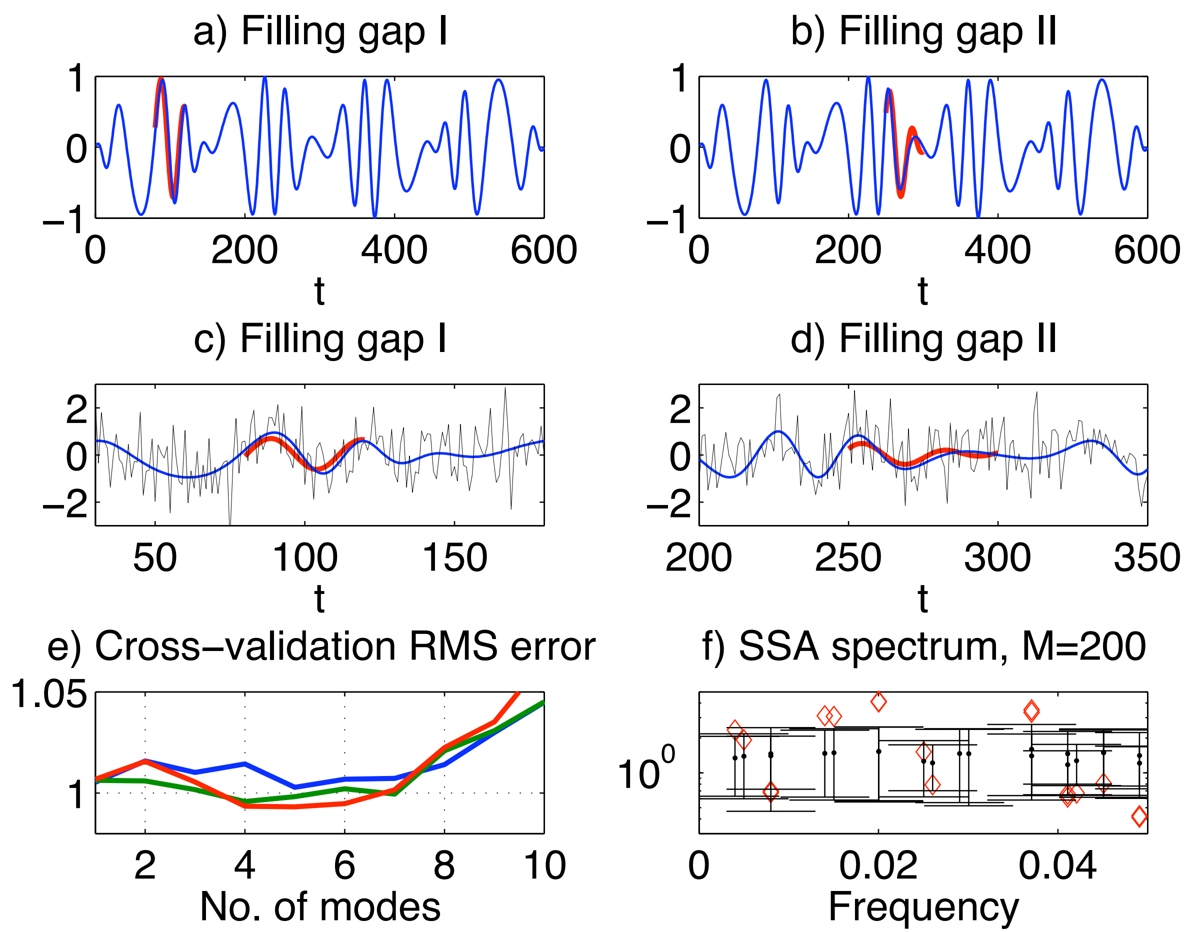
- D. Kondrashov and M. Ghil, 2006: Spatio-temporal filling of missing points in geophysical data sets, Nonl. Proc. Geophys., 13, 151-159.

Synthetic I: Gaps in Oscillatory Signal



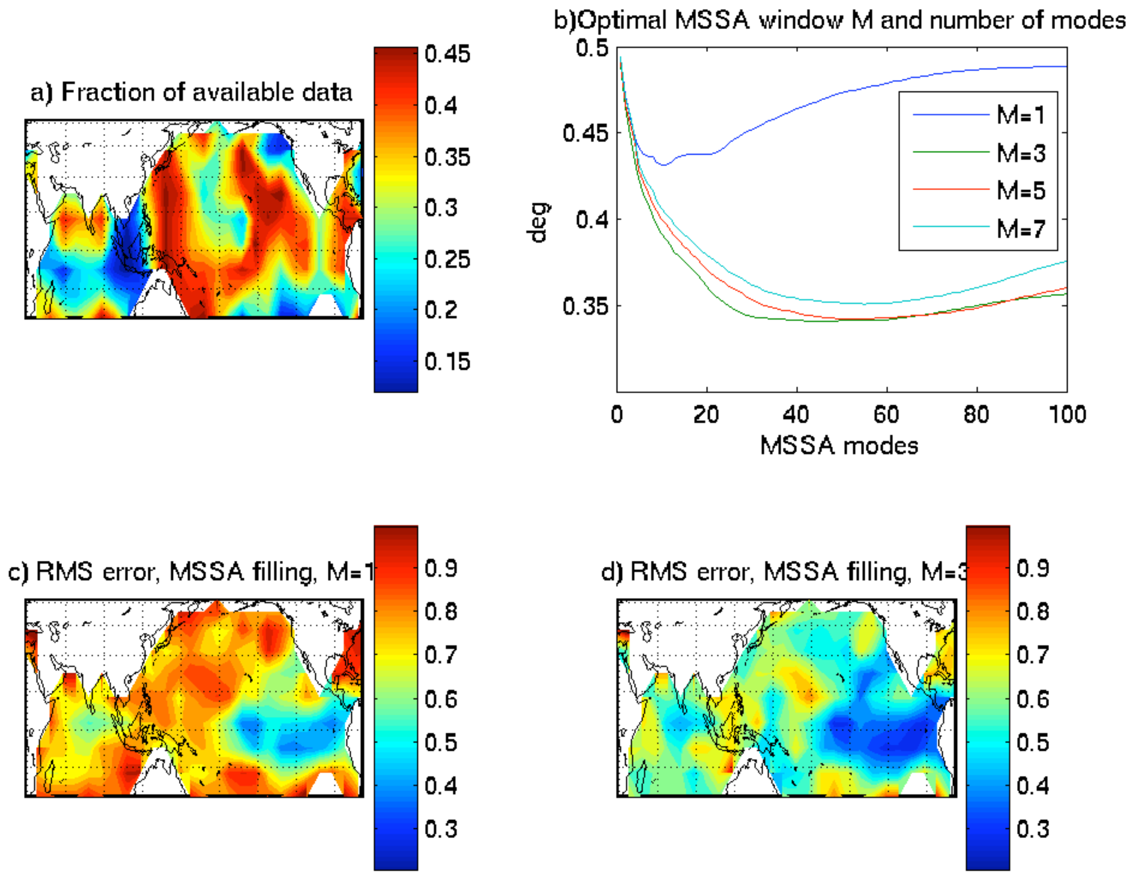
- Very good gap filling for smooth modulation; OK for sudden modulation.

Synthetic II: Gaps in Oscillatory Signal + Noise



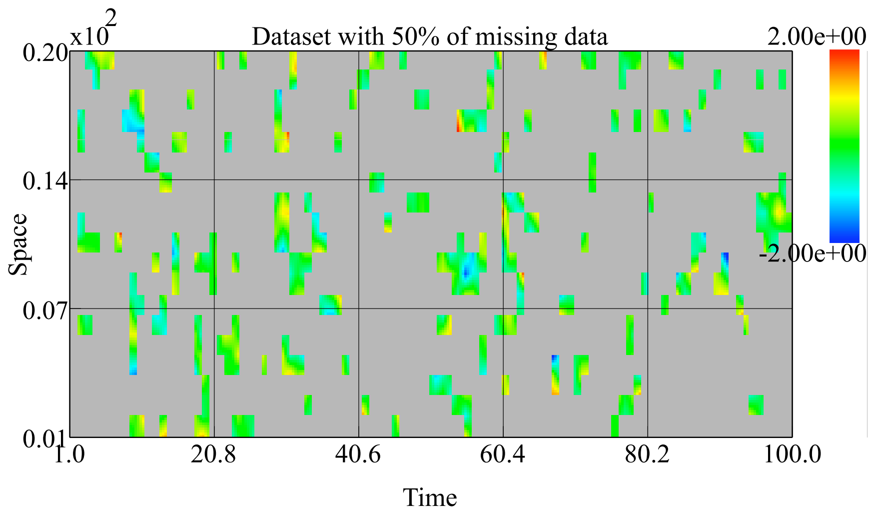
$$x(t) = sin(\frac{2\pi}{300}t) * cos(\frac{2\pi}{40}t + \frac{\pi}{2}sin\frac{2\pi}{120}t)$$

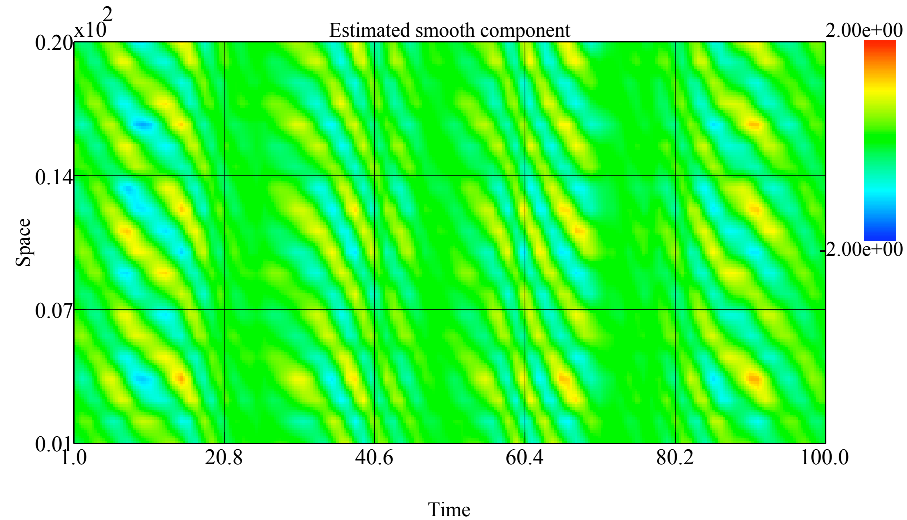
Synthetic III: Synthetic gaps in SST data

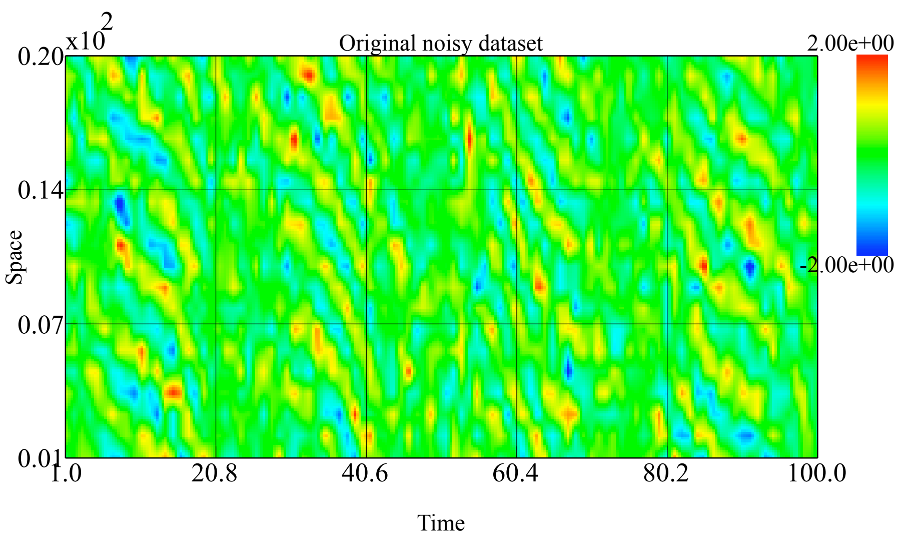


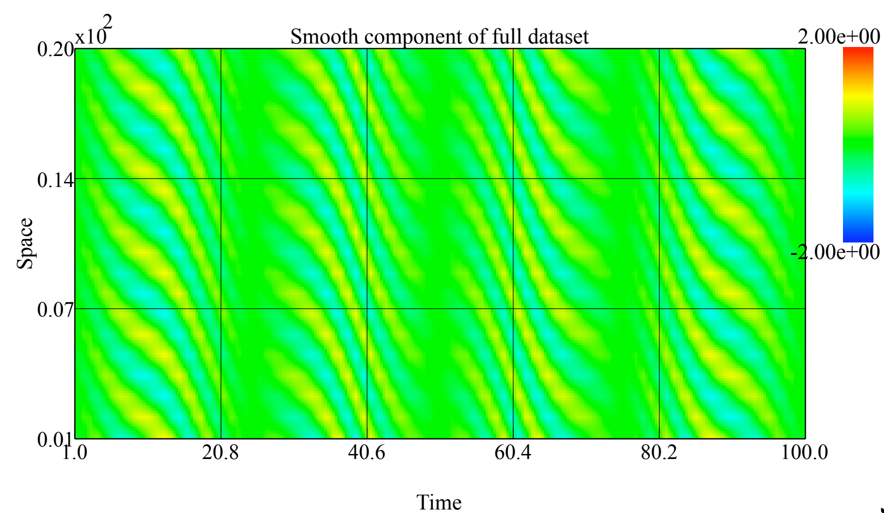
- 1950-2004 IRI monthly SST dataset (10°x10°, 660x237 grid points)
- see improvement with ***MSSA***; ***"random"*** pattern favors small ***M***!
- Error is smaller in CE Pacific Sector where "signal" (ENSO) mode is dominant!

Synthetic IV: multivariate example (Prize!!!)

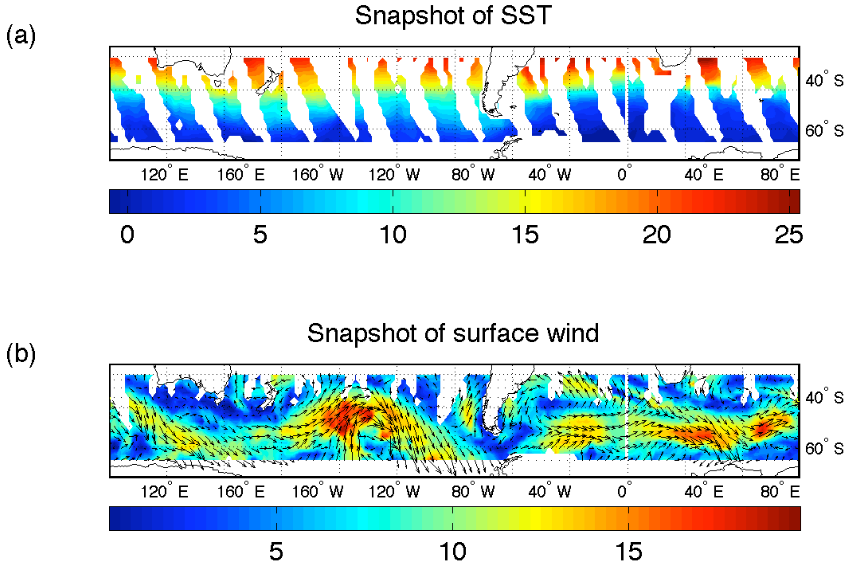


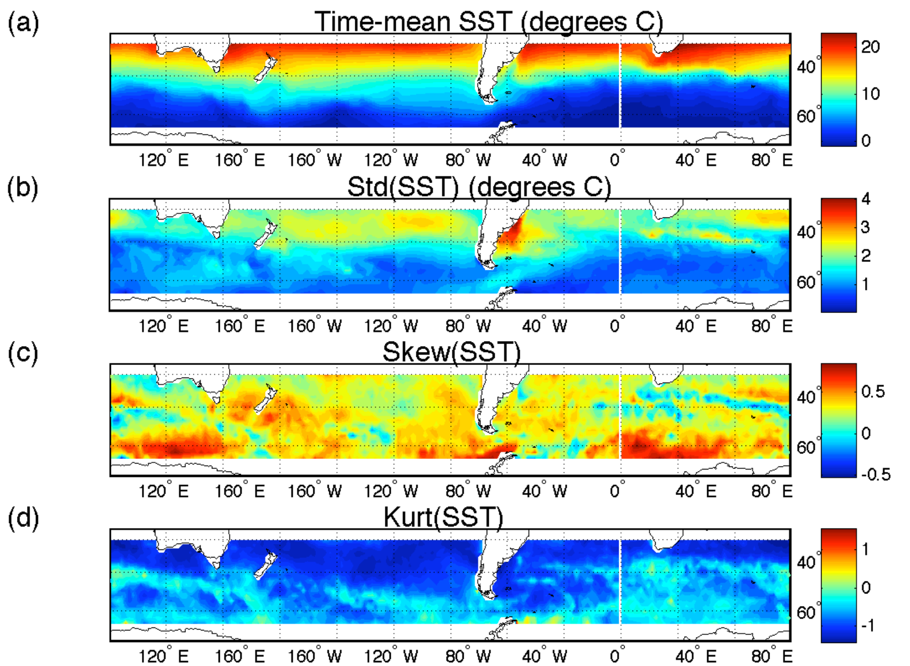


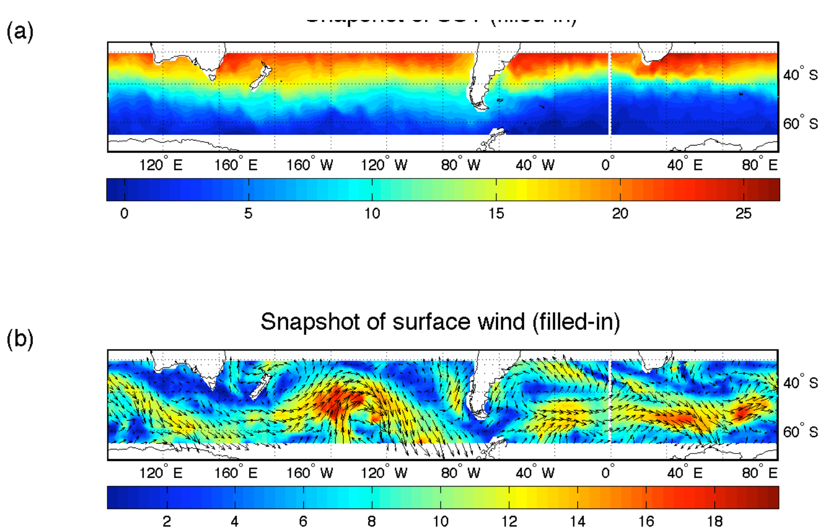




Filled-in Southern Ocean data

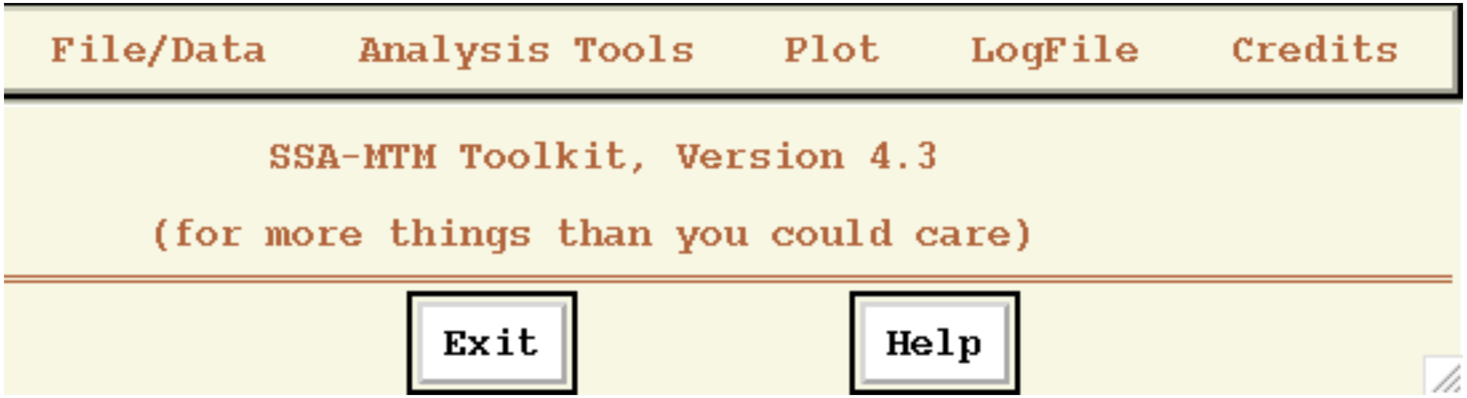






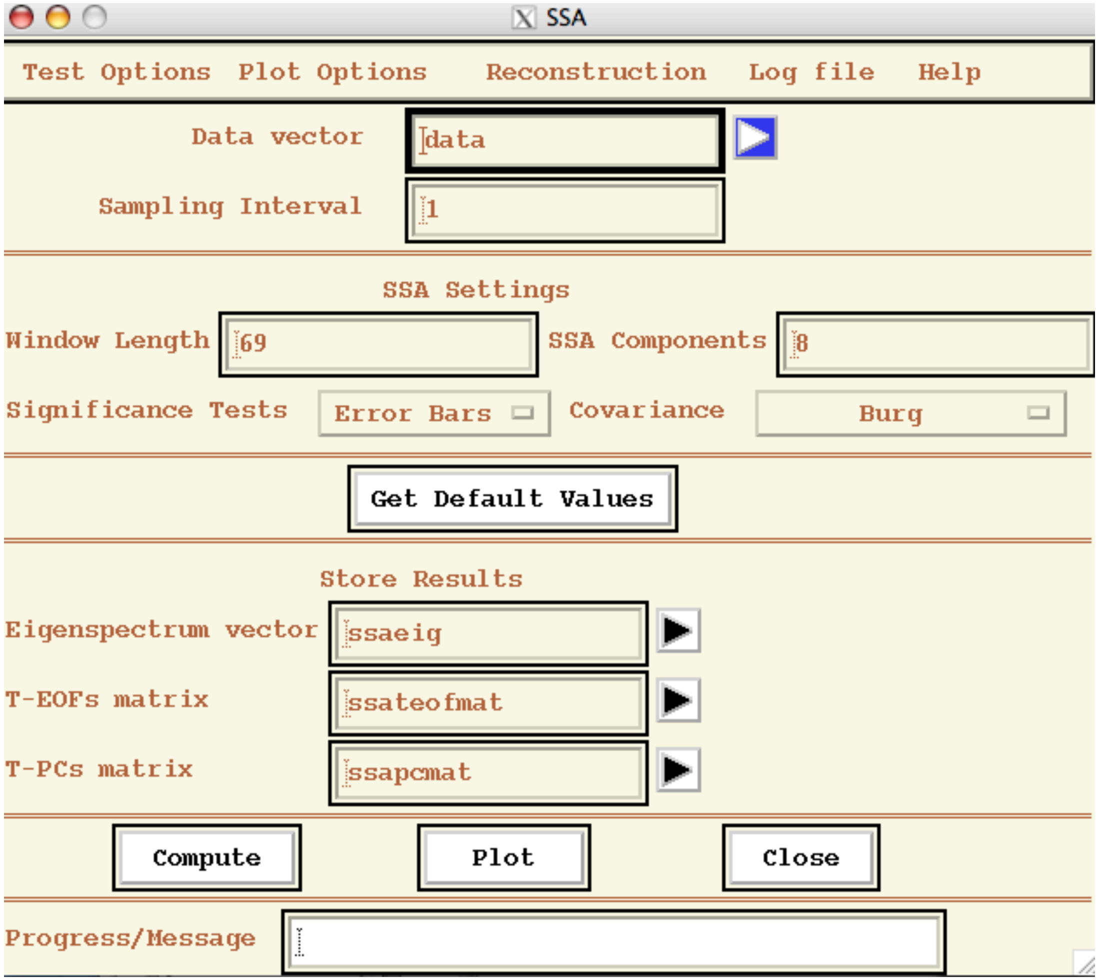
- Gap-filling needs to respect physical limits

SSA-MTM Toolkit



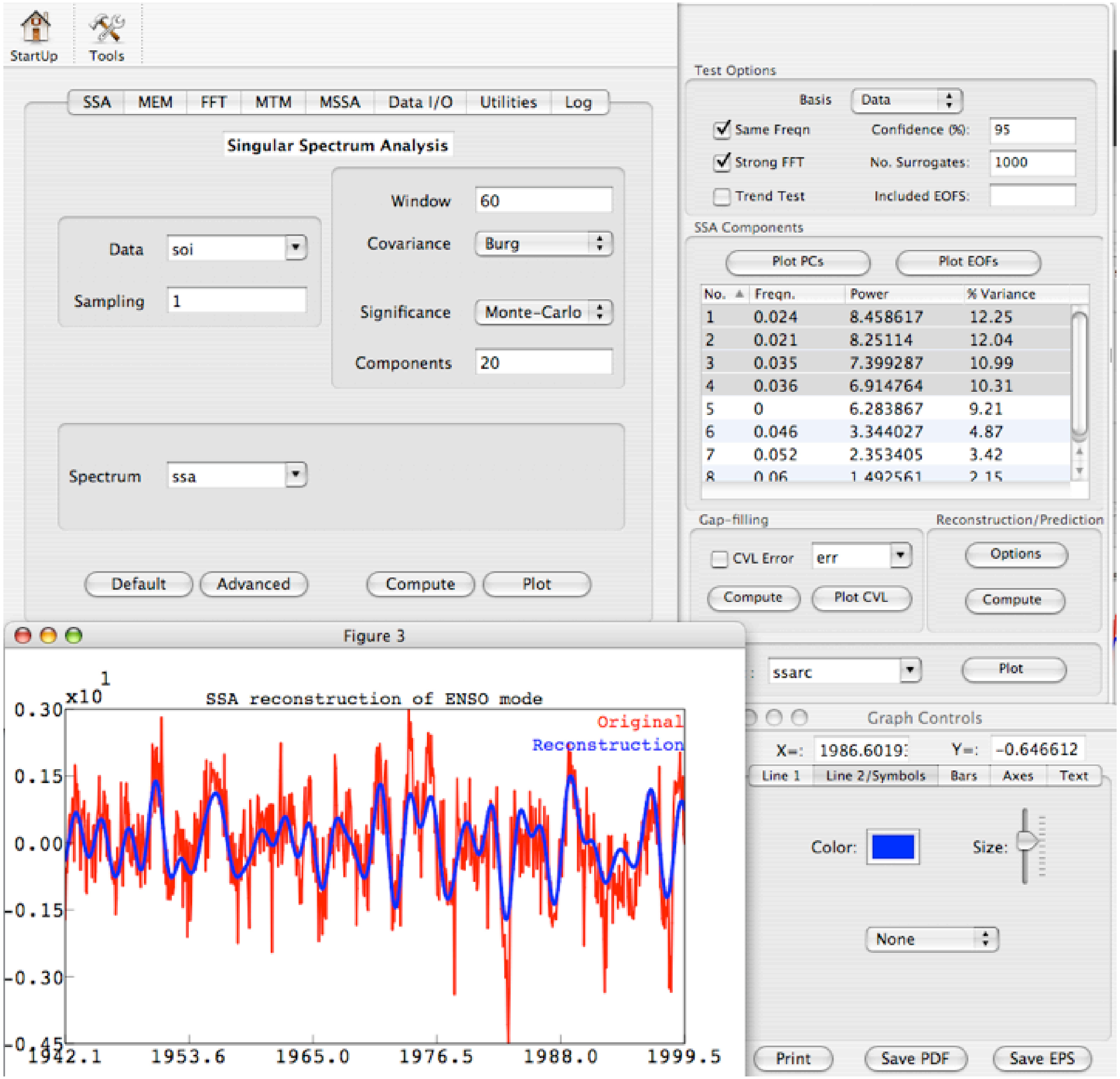
- Freeware ported to Sun, Dec, SGI, Linux, and Mac OS X: self-contained binary (~2-5Mb) depending on the Unix platform.
- http://www.atmos.ucla.edu/tcd/ssa/
- Needs external graphics package: Grace (free, default) is a part of standard Linux installation, may need compiling for other OS; IDL ($$)
- Includes ***Blackman-Tukey FFT***, ***Maximum Entropy Method***, ***Multi-Taper Method (MTM)***, ***SSA and M-SSA***.
- Spectral estimation, decomposition, reconstruction, gap-filling
- Significance tests of "**oscillatory modes**" vs. "**noise.**"

SSA-MTM Toolkit (cont'd)



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

Mac OS X: kSpectra Toolkit



- Project files
- SSA Forecasts
- Automated tasks
- Built-in plots
- Animations (QuickTime)
- Automation (Automator)
- www.spectraworks.com

END

- *Ghil M., R. M. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A. Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou, 2002: "Advanced spectral methods for climatic time series," Rev. Geophys., 40(1), pp. 3.1-3.41, 10.1029/2000RG000092.*
- *D. Kondrashov and M. Ghil, 2006: Spatio-temporal filling of missing points in geophysical data sets,*
  Nonl. Proc. Geophys., 13, 151-159.
- *more at http://www.atmos.ucla.edu/tcd/ssa.*

- *Computer Lab: SOI (ENSO), "small signal", gap-filling, multivariate example (time permitting)*