Gap-Filling of Solar Wind Data by Singular Spectrum Analysis

Dmitri Kondrashov

University of California, Los Angeles

Yuri Shprits

University of California, Los Angeles

Motivation

1. Observational data sets in the geosciences are often short, contain noise (errors) and are gappy: this is both an obstacle and an incentive. Continuous data is needed for modeling (boundary conditions), standard spectral estimation algorithms, etc.

2. Phenomena in the geosciences often have both regular (“cycles”) and irregular (“noise”) aspects. Regularities include (quasi-)periodicity $\Rightarrow$ singular spectrum analysis (SSA) $\Rightarrow$ Powerful method for spectral estimation, noise filtering and gap-filling (producing estimates of the missing points!)

2. Space Physics Application: Apply SSA to fill-in large gaps in historical solar wind and interplanetary magnetic field (IMF) data required for the state-of-the art empirical global magnetospheric magnetic field models. Important for radiation belt modeling.

dkondras@atmos.ucla.edu
Outline

1. Short brief on Singular Spectrum Analysis.

2. Illustrate SSA gap-filling on synthetic examples.


4. Verification of the algorithm on synthetic gaps in 2000-2001 continuous data.

5. First results of filled-in 1990-1991 gappy data; comparison with alternatives.

6. Conclusions.
SSA of Nino-3 index (El-Nino)

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

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

- SSA isolates oscillatory behavior via paired eigenelements.
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 Gap-filling algorithm

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 lag-covariance matrix, find leading $K$ EOFs

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

$$C_X = \frac{1}{N'}D^tD; \quad C_X E_k = \lambda_k E_k$$

3. Reconstruct missing points using $K$ EOFs

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

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

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


- Utilize both spatial and temporal correlations to iteratively compute maximum likelihood estimates of mean and lag-covariance matrix => can be applied to very gappy data.

- A few $K$ leading EOFs correspond to the “smooth” modes, while the rest is noise and can be discarded.

- Cross-validation provides error estimates and optimum SSA parameters.

Synthetic Gaps in Noisy Oscillatory Signal

\[ x(t) = \sin\left(\frac{2\pi}{300} t\right) \times \cos\left(\frac{2\pi}{40} t + \frac{\pi}{2} \sin\frac{2\pi}{120} t\right) \]
Synthetic Gaps in Multivariate data
Continuous inner-magnetosphere indices (Kp, Dst) are ground measured time-lagged magnetic disturbances caused by interaction of Earth’s magnetosphere with solar wind.

A measure of geomagnetic activity used to assess the severity of magnetic storms.

- Gaps in solar wind satellite data before the launch of the WIND spacecraft in 1994.

Ni B. et al., JGR, 2009

- Gaps of 1990-1991 are applied to hourly 2000-2001 data
- SSA Window M=25hr.
• Gaps of 1990-1991 are applied to hourly 2000-2001 data; SSA Window M=15hr.
Freeware ported to Sun, Dec, SGI, PC Linux, and Mac OS X

Graphics support for **IDL** and **Grace** (free)

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**.”

SSA Gap-filling.

For details and publications, please visit: **TCD**  [http://www.atmos.ucla.edu/tcd/ssa](http://www.atmos.ucla.edu/tcd/ssa)
Conclusions

Promising first results of applying SSA to estimate the missing data in gappy solar wind data by using inner-magnetospheric indices.

Future work:

- Consider longer datasets, varying temporal scales.
- Closer look at the SSA modes responsible for successful reconstruction.
- Include more solar wind parameters: solar wind speed ($V_{sw}$) and solar wind density ($N_{sw}$) and other components of magnetic field ($B$)
- Do systematic search for optimum combination of inner-magnetospheric indices


- SSA-MTM Toolkit, http://www.atmos.ucla.edu/tcd/ssa