Gap filling in incomplete geophysical data sets

D. Kondrashov, M. Ghil1

Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles

Introduction

Data sets in the geosciences are often full of gaps. This is usually the case for geological and paleoclimatological proxy data from the remote past, as well as for historical records from the more recent past. In modern times, data points may be missing because of the way the measurements are obtained. We demonstrate here how Singular Spectrum Analysis (SSA) and multi-channel SSA can be applied to fill the missing data with smooth information from an iteratively inferred “signal” that represents coherent spatio-temporal structures, while the “noise” variance is discarded or reduced.

Gap filling by iterative SSA

1. Choose window $M$ and start with $K=1$: Flag fraction of dataset $X(t) (t = 1:N)$ as “missing” for cross-validation.
2. Update $X(t)$ mean and apply SSA embedding procedure for $N \times N-M+1$:
   \[
   D = \begin{pmatrix}
   X(1) & X(2) & \cdots & X(M) \\
   X(2) & X(3) & \cdots & X(M+1) \\
   \vdots & \vdots & \ddots & \vdots \\
   X(N-M) & X(N-M+1) & \cdots & X(N)
   \end{pmatrix}
   \]
3. Update covariance, find leading $K$ EOFs
   \[
   C_X = \frac{1}{N} D^T D; C_X E_k = \lambda_k E_k
   \]
4. Reconstruct missing points using $K$ EOFs
   \[
   \begin{aligned}
   A_k(t) &= \sum_{j=1}^{M} X(t + j - 1) E_k(j) \\
   B_k(t) &= \frac{1}{M} \sum_{j=1}^{M} \sum_{k=1}^{N} A_k(t + j - 1) E_k(j)
   \end{aligned}
   \]
5. Check cross-validation error: if convergence has occurred for the missing points, set $K = K + 1$; if not, go back to Step 2.

- For a window $M=1$, one utilizes both spatial and temporal correlations in a multivariate data set.
- The optimal number $K$ of “signal” EOFs is obtained through cross-validation.

SSA-MTM Toolkit

- Freeware for time series analysis.
- X11(Motif) GUI.
- Ported to Sun, Linux, and Mac OS X.
- Downloads at http://www.atmos.ucla.edu/tcd/ssa/

References


dkondras@atmos.ucla.edu

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