Singular Spectrum Analysis with Missing Data

D. Kondrashov (1), and M. Ghil (1,2)
(1) Department of Atmospheric Sciences and
Institute of Geophysics and Planetary Physics
University of California, Los Angeles
(2) Département Terre-Atmosphère-Océan and
Laboratoire de Météorologie Dynamique,
Ecole Normale Supérieure, Paris, France
dkondras@ucla.edu

We study the performance of Singular-Spectrum Analysis (SSA) with missing data. The standard SSA algorithm computes eigenvalues and eigenvectors of a given time series' lag-covariance matrix. The name of the algorithm refers to the singular values of the so-called trajectory matrix, whose squares are the eigenvalues of the lag-covariance matrix, while the eigenvectors of the latter are referred to often as temporal empirical orthogonal functions (T-EOFs).

When data are missing, we apply an iterative procedure that finds each of the T-EOFs, one by one; these are then used to update the estimate of the missing data by a standard SSA reconstruction. A cross-validation procedure finds the optimum number of T-EOFs by including randomly chosen points of the existing record into the set of missing data, and computing the interpolation error of the reconstructed time series for each of the T-EOFs included in the reconstruction. The minimum of the interpolation error averaged over all the validations corresponds to the optimum number of T-EOFs that capture the useful signal and thus separate it from the uncorrelated white noise in the underlying time series. The algorithm is demonstrated on synthetic data sets, the well-known Niño-3 sea-surface temperature index, and the Nile river flood level record from 622 A.D. (1 A.H.) to 1921.