



Singular Spectrum Analysis with Missing Data and Oscillatory Climate Modes in Extended Nile River Record (A.D. 622-1922)



D. Kondrashov, Y. Feliks², and M. Ghil³

Department of Atmospheric Sciences and Institute of Geophysics and Planetary Physics, University of California Los Angeles

²Mathematics Department, Israel Institute of Biological Research, Nes-Ziona, Israel

³Additional affiliation: Département Terre-Atmosphère-Océan and Laboratoire de Météorologie Dynamique, Ecole Normale Supérieure, F-75231 Paris Cedex 05, France.

dkondras@atmos.ucla.edu; http://www.atmos.ucla.edu/tcd

Introduction and Motivation

The time series of the Nile flood levels (Popper, 1951) represents a remarkable human historical record of climate variability for over more than 1000 years, that reflects water intake from the Blue and the White Nile in Ethiopia and equatorial Africa. Climate researchers have studied extensively the resulting multi-century, annually resolved records and have demonstrated the significant association between the rainfall in the catchment area of the Nile tributaries, the Indian monsoon and ENSO [Walker, 1910; Quinn, 1992; Nicholson and Entekhabi, 1986; Ropelewski and Halpert, 1987]. 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. In any case, some regularities in the Nile River records cannot be explained by the ENSO connection. In particular, North Atlantic influences over North Africa and the Middle East have been detected in several geologic [Felis et al., 2004] and historical [Mann, 2002] time series, and may extend all the way into the Nile River's source area, further to the south.

The traditional view is that it is the Blue Nile, which originates in the Ethiopian highlands, that contributes mainly to the floods, while the White Nile, originating from equatorial Africa, is mainly responsible for the low level [Said, 1993; Hurst, 1952]. The short-term, interannual fluctuations in the high-level record, however, do reflect also variations in the contributions from the White Nile (i.e., in the low level) [Hassan, 1981]. Moreover, on longer time scales, not only the rains of the given season determine the level of Lakes Victoria and Albert, at the sources of the White Nile, but also the accumulated precipitation of several preceding seasons [Hurst, 1952]. Examining both the high- and low-water records (see Fig. 1 in De Putter et al. [1998] and Figs 2a,b in this work), it is clear that the variance of the minimum level is much larger than that of the maximum; furthermore, correlations between ENSO activity and equatorial Africa are quite comparable to those between ENSO and the Ethiopian highlands (see Fig. 6.1 in Quinn [1992]). We thus conclude that the time series of the annual difference between the maxima and minima represents better the Nile floods than the high-level time series.

We revisit therefore the detection and attribution of regularities in the 1300-yr long Nile River records. Our work provides five important new results on the interannual, interdecadal and centennial evolution of climate in northeast Africa over the last millennium-and-a-half: (i) a complete 1300-yr record of Nile River floods with annual resolution that permits one to study the evolution of the regularities over the most recent 450 years (A.D. 1471-1922); (ii) more robust statistical-significance tests; (iii) an analysis of the net flood record, which we argue is better represented by the time series of the difference between the high- and low-water records; (iv) sharper and more reliable determination of climatic-regime transitions; and last, but certainly not least, (v) evidence for a novel source of interannual climatic variability for tropical East Africa, namely changes in the North Atlantic ocean circulation. We address these points by investigating both low- and high-water Nile-river records, and their difference. A modified singular spectrum analysis algorithm (SSA) is applied to self-consistently fill missing points with the leading oscillatory modes of the time series.

Spectral Methods

We apply advanced, well-documented spectral methods [Ghil and Vautard, 1991; Allen and Smith, 1996; Mann and Lees, 1996; Ghil et al. 2002] to estimate the regularities in our data sets. Our significance tests are carried out with respect to red, rather than white noise; the former is a more appropriate null hypothesis for the Nile River records, which — like most climatic and other geophysical time series — have larger power at lower frequencies. The main methods we used are the Monte Carlo version [Allen and Smith, 1996] of SSA and the robust MTM version for both line and background estimation [Mann and Lees, 1996], as provided in version 4.2 of the SSA-MTM Toolkit [Ghil et al., 2002] (see also the freeware toolkit documentation at <http://www.atmos.ucla.edu/tcd/ssa/>).

SSA is a data-adaptive, non-parametric spectral method based on diagonalizing the lag-covariance matrix of a time series [Coblebrook, 1978; Fraedrich, 1986; Vautard and Ghil, 1989]. The eigenvectors of this matrix are the empirical orthogonal functions (EOFs) and projection onto them yields the corresponding principal components. The entire time series or parts of it that correspond to trends, oscillatory modes or noise can be reconstructed by using linear combinations of these principal components and EOFs, which provide the reconstructed components (RCs) [Ghil and Vautard, 1991; Ghil et al., 2002].

Missing Data Filling by Iterative SSA/MSSA.

Analyzing the full extent of the available water-level records, with the missing points filled in, allows for greater accuracy and better significance testing in the spectral analysis. It also improves our knowledge on the evolution of the oscillatory modes over the entire 13 centuries of record and especially over the last 450 years.

Schoellhamer [2001] first suggested using SSA [Broomhead and King, 1986; Vautard et al., 1992; Ghil et al., 2002] to obtain spectral estimates from records with a large fraction of missing data. The method applied here differs from his in two ways: (i) we iteratively produce estimates of missing data points, which are then used to compute a self-consistent lag-covariance matrix; and (ii) we use cross-validation to optimize the window width and number of dominant SSA modes to fill the gaps. For a univariate record, we first center the original data by computing the unbiased value of the mean and set the missing-data values to zero. At each subsequent iteration step, new estimates of the record, its lag-covariance matrix, EOFs and principal components are obtained [Ghil et al., 2002].

For many climatic and geophysical records, a few leading EOFs correspond to the record's dominant oscillatory modes, while the rest is noise [Ghil et al., 2002]. Using this idea, we start the inner-loop iteration with a first EOF, and repeat the SSA algorithm on the new time series, in which the RC of that EOF alone was used to replace the missing points. Each new reconstruction of the entire record is tested against the previous reconstruction until a tolerance is met. Next, outer-loop iterations are performed with two leading EOFs, and so on. The optimum number of EOFs is found by cross-validation: a portion of the available data is marked as missing, and the estimation error is computed for each EOF being added to the reconstruction. In practice, this error starts to increase once noise-related EOFs are added. The full procedure and the tests on synthetic data will appear elsewhere. For a univariate record, our procedure uses only temporal correlations in the data to fill in the missing points. For a multivariate record, multi-channel SSA imputation takes advantage of both spatial and temporal correlations. This provides a substantial improvement in cases like the Rodah Island records, where data points are often missing in one time series, but not the other.

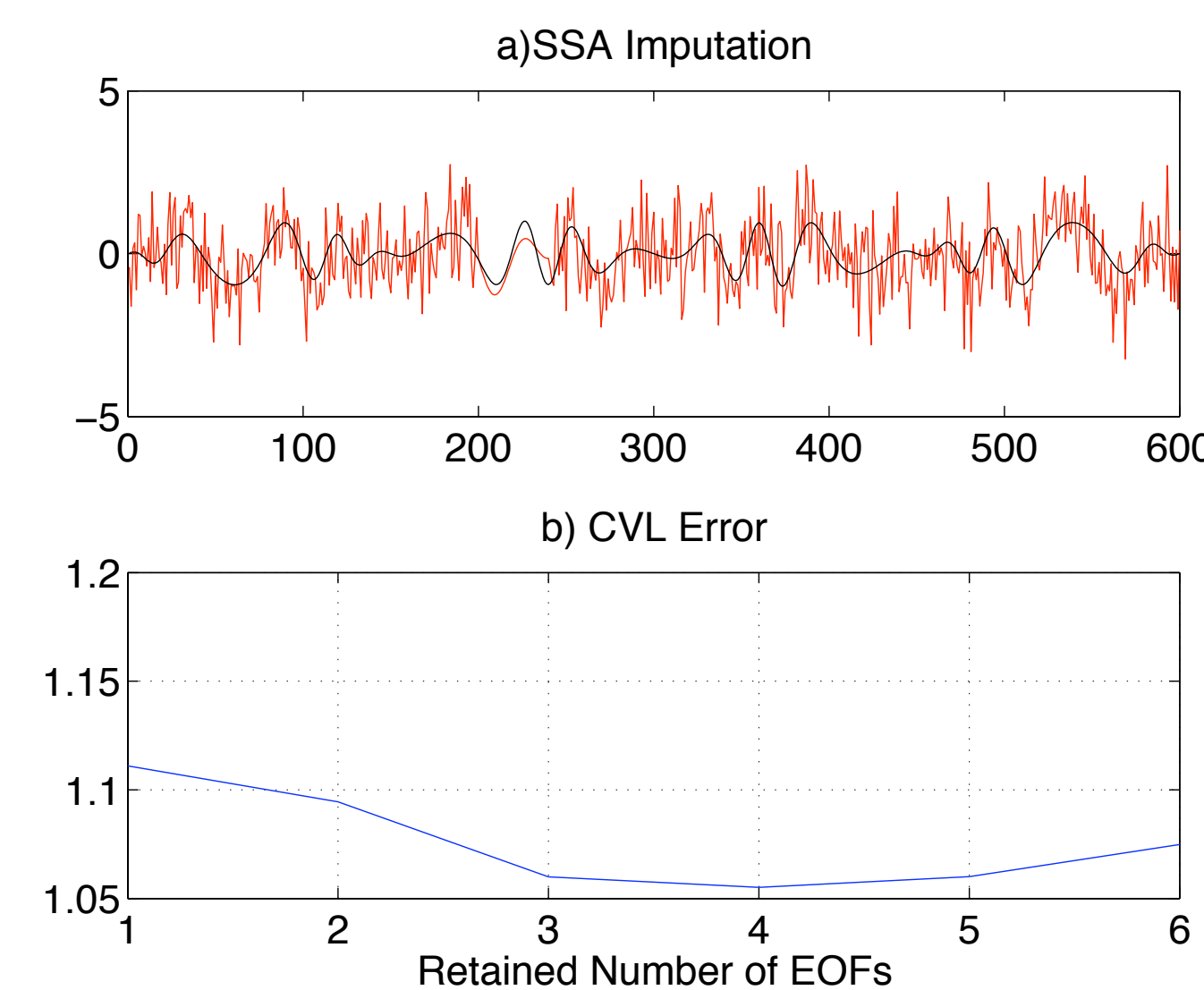


Figure 1. SSA (M=40) imputation of a gap (200<t<240) in amplitude- and phase-modulated signal, contaminated by white noise. Four leading EOFs are optimum number to use in imputation, based on CVL error.

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

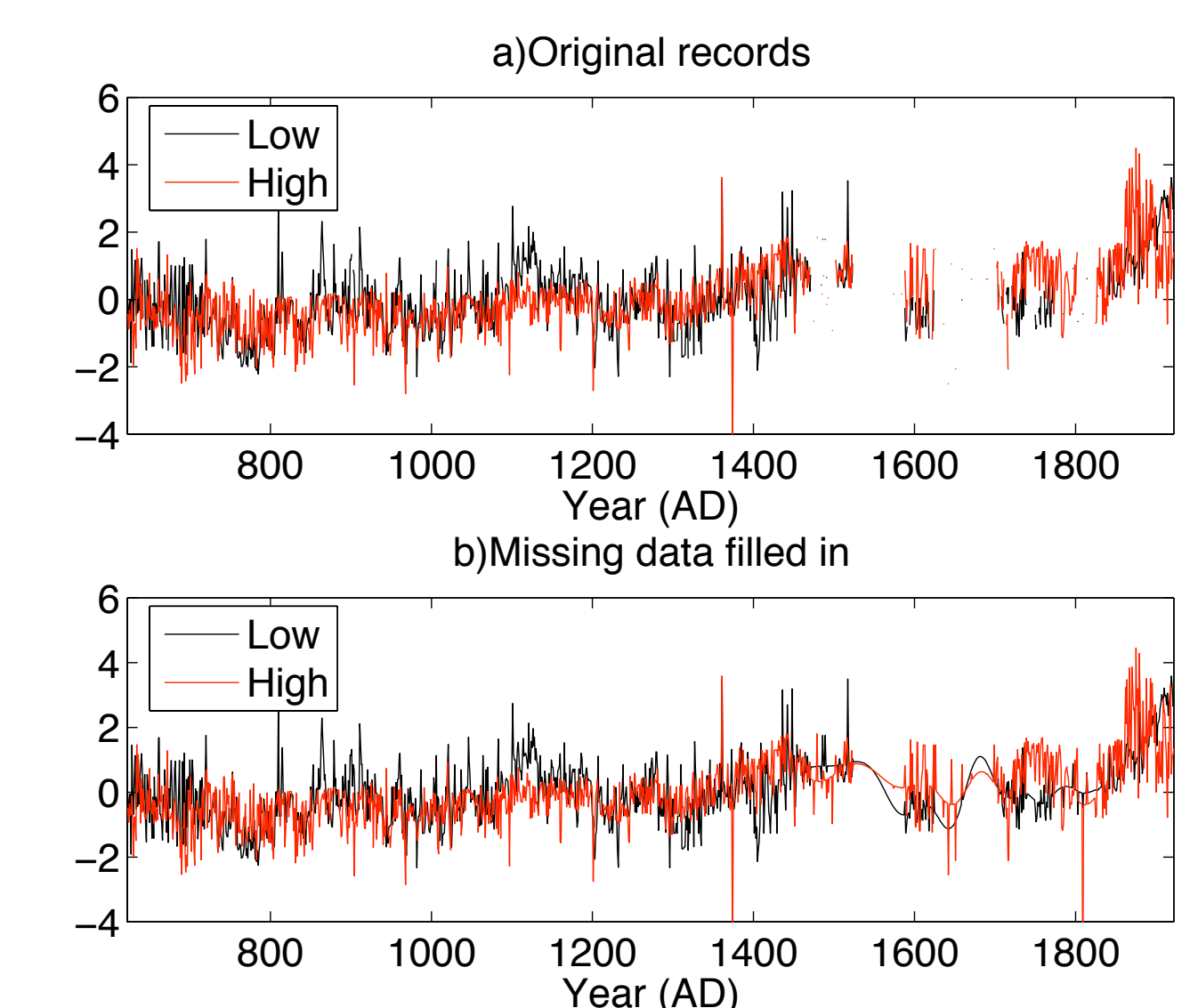


Figure 2. We apply MSSA to fill up missing gaps in both low- and high-water records (black and red lines, respectively). Upper panel is original data and low panel is data with missing points imputed by MSSA using M=100 years and two channels (low- and high-water). Eight leading EOFs are optimum for imputation based on CVL error (not shown).

Spectral Analysis

The extended low-water, high-water and difference records, like the short ones, have a very low-frequency, trend-plus-oscillatory component, captured by the two leading EOFs of an SSA analysis with a 100-yr window. The nonlinear, data-adaptive upward trend that continues over the entire extended record might be due to siltation [Popper, 1951]. The very low-frequency oscillation has the same 256-yr period as in the short records.

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%), 18.3 (6.7%)		12.2 (4.7%), 18.3 (5.0%)
5-10yr	6.2 (4.3%)	7.2 (4.4%)	7.3 (4.4%)
0-5yr	3.0 (2.9%), 2.2 (2.3%)	3.6 (3.6%), 2.9 (3.4%), 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%)	71(8.6%)	71(8.2%)
20-40yr		23.2 (4.3%)	
10-20yr	[12], 19.7 (5.9%)		12.2 (4.3%), 18.3 (4.2%)
5-10yr	[6.2]	7.3 (4.0%)	7.3 (4.1%)
0-5yr	3.0 (4%), 2.2 (3.3%)	4.2 (3.3%), 2.9 (3.3%), 2.2 (2.9%)	[4.2], 2.9 (3.6%), 2.2 (2.6%)

Table 1. Results for detrended time series. The main entries give the periods in years: bold entries without brackets indicate that the mode is significant at the 95% level against a null hypothesis of red noise, in both SSA and MTM results; entries in square brackets are significant at this level only in the MTM analysis. Entries in parentheses provide the percentage of variance captured by the mode with the given period.

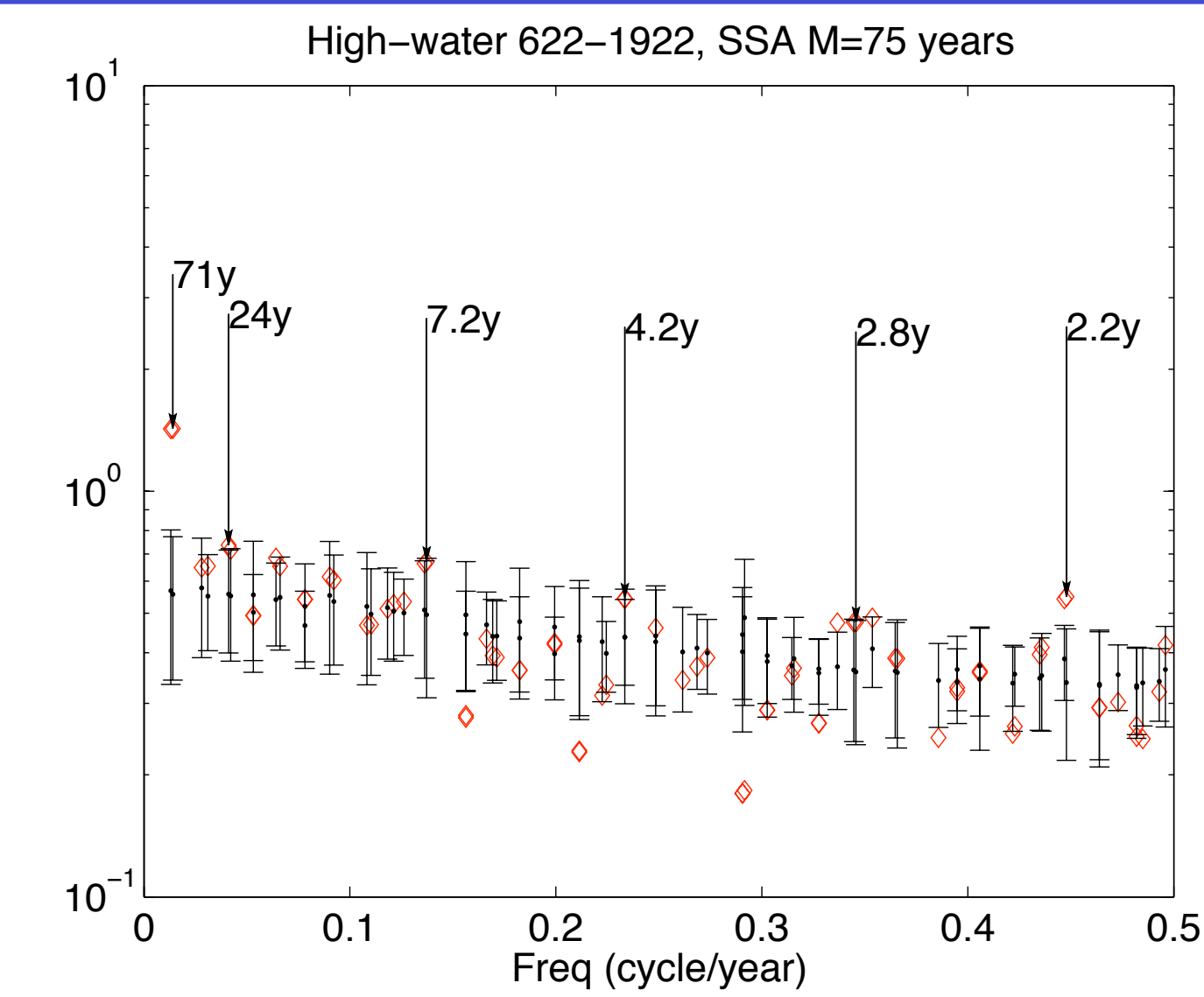


Figure 3. SSA results for the extended Nile time series, arrows mark peaks that are highly significant in both (at 95%) and MTM.

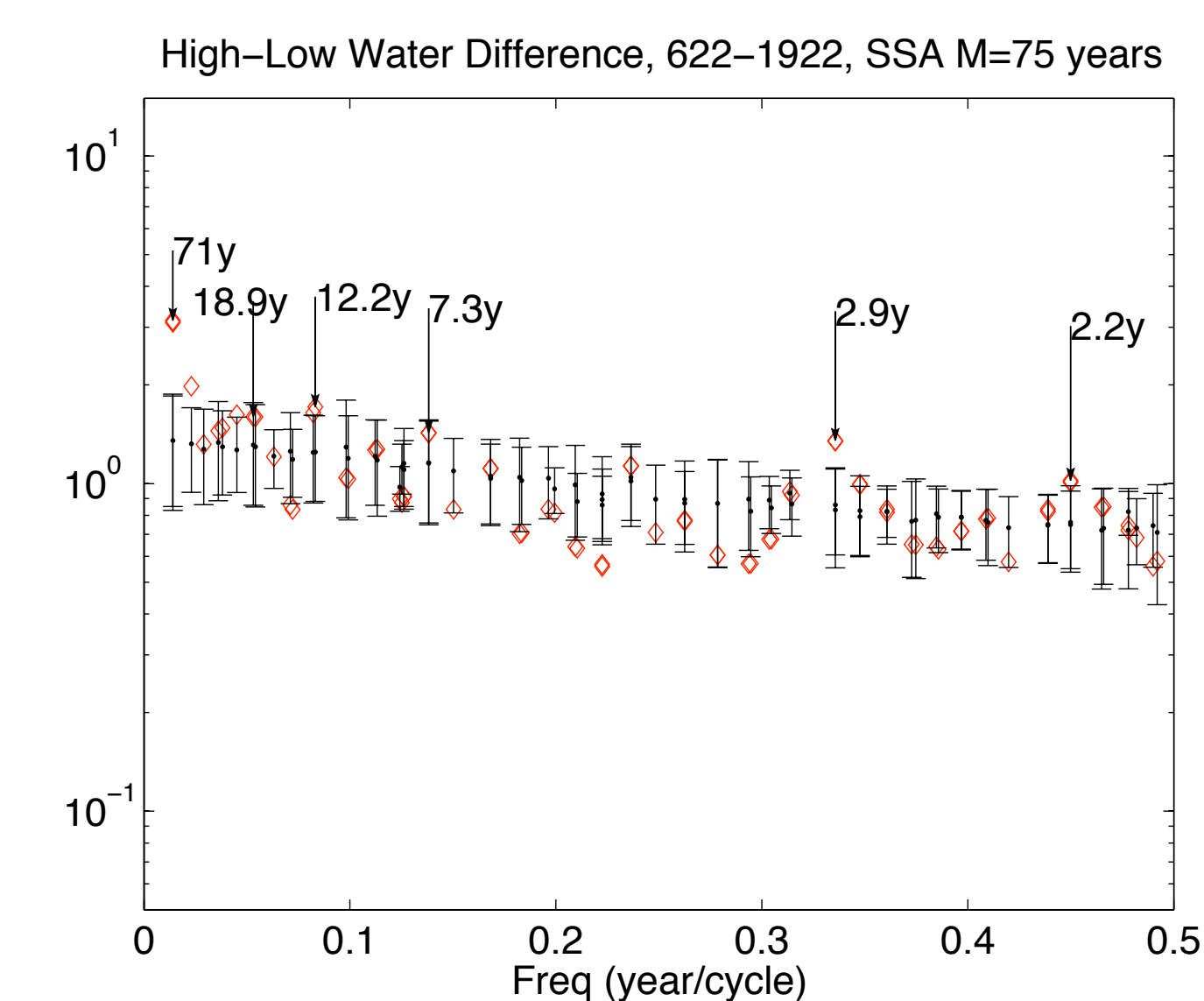


Figure 4. SSA reconstruction of the 7.2-yr mode in the extended Nile River records: (a) high-water, and (b) difference. The amplitude has been normalized by one standard deviation of the detrended time series; Reconstruction in the large gaps is indicated in red.

Discussion and Conclusions

The 7-yr mode in the high-water and difference records, both short and extended, is quite robust and does not seem to have been previously discussed, aside from a parenthetical reference to an 8.1-yr peak in the high-water record of [De Putter et al., 1998]. Since the difference record is representative of net flood conditions, the presence of this peak in it lends further weight to its interpretation. A 7-8-yr peak exists in instrumental records of North Atlantic sea surface temperatures [Meron et al., 1998] and sea level pressure [Da Costa et al., 2002]. Felis et al. [2004] have shown 5-6-yr variability to persist in isotopic coral records from the northernmost Red Sea (Gulf of Aqaba) during the last interglacial, 120kyr ago. Mann [2002], based on the global temperature reconstructions of [Mann et al., 1998] has found 8-9-yr variability in an index of Middle-and-Near East temperature, back to 1750 or 1650 A. D.

Our clear and persistent 7-yr peak for the extended Nile River records (A.D. 622-1922) demonstrates the extent of North-Atlantic influences over the last millennium-and-a-half and into Eastern Africa's tropical regions. The spatio-temporal pattern associated with this peak is well simulated by a whole hierarchy of ocean models for the North Atlantic's wind-driven circulation, with increasingly realistic physics, resolution and geometry [Dijkstra and Ghil, 2004; Simonnet et al., 2004]. Atmospheric circulation patterns over the North Atlantic have well-known downwind effects, including over the Mediterranean and North Africa [Felis et al., 2004; Mann, 2002; Matthews, 2004; Hurrell, 1995]. The effects over the source region of the Nile River, however, are new. It is tempting, moreover, to identify, the 7-yr peak in the Nile River records with the cycle mentioned in Joseph's biblical story. Such an identification, though, has to be taken with the grain of salt: seven is a sacred number in the Hebrew tradition and its use in the story may designate merely a near-regularity of several years, not a specific, exact periodicity.

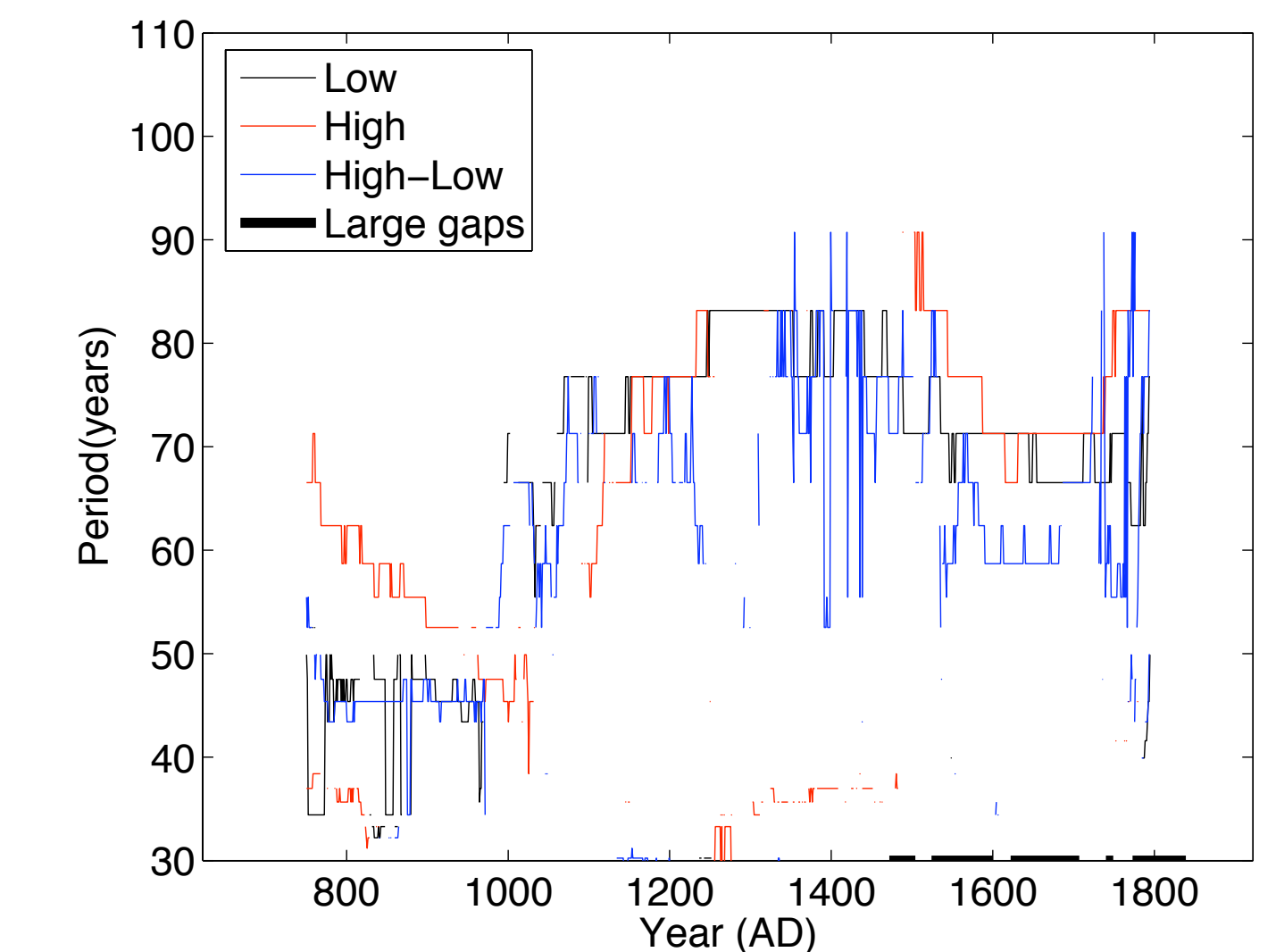


Figure 5. Temporal evolution of the instantaneous frequencies of the oscillatory pairs that lie in the low-frequency range (40-100 yr), for the low-water, high-water, and difference records.

The plots are based on multi-scale SSA [Yiou et al., 2000] results with a sliding window of W = 255 yr and M = 85 yr; local SSA analysis is performed in each window of width W, with M lags, and W = 3M.

Previous studies of various climate proxy and instrumental records reveal oscillatory climate modes that are consistent with those identified in our analysis, i.e. a 250-yr oscillation [Mann et al., 1995; Vaquero et al., 1997], as well as 50-70-yr [Mann et al., 1995; Schlesinger and Ramankutty, 1994] and 15-35-yr periods [Ghil and Vautard, 1991; Moron et al., 1998]. Ghil et al. [2002] discussed in detail the physical interpretation of interannual and interdecadal climate modes, in terms of the internal variability of the mid-latitude and global ocean circulation, wind-driven and thermohaline [Dijkstra and Ghil, 2004; Simonnet et al., 2004]. The periodicities shared by the present analysis of Nile River records and global or North Atlantic modes of ocean or coupled ocean-atmosphere variability [Ghil and Vautard, 1991; Dijkstra and Ghil, 2004; Simonnet et al., 2004] do not demonstrate a causal relationship. Our results suggest, though, quite strongly that the climate of East Africa has been subject to influences from the North Atlantic, besides those already documented from the Tropical Pacific [Walker, 1910; Quinn, 1992; Kumar et al. 1999; Nicholson and Entekhabi, 1986] and those of possibly astronomical origin [Hameed, 1984; De Putter et al., 1984; Mann et al., 1995; Vaquero et al., 1997]. Moreover, the fairly sharp shifts in the amplitude and period of the interannual and interdecadal modes over the last millennium-and-a-half support concerns about the possible effect of climate shifts in the not-so-distant future [Alley et al., 2003].

References

Allen, M. R., and L. A. Smith (1996), Monte Carlo SSA: Detecting irregular oscillations in the presence of coloured noise. *J. Clim.*, 9, 3373-3404.

Alley, R. B. et al. (2003), Abrupt climate change. *Science*, 299, 2005-2010.

Broomhead, D. S., and G. P. King (1986), Extracting qualitative dynamics from experimental data. *Physica D*, 20, 217-236.

Coblebrook, J. M., (1978), Continuous plankton records: zooplankton and environment, North-East Atlantic and North Sea, 1948-1975. *Oceanol. Acta*, 1, 9-23.

Da Costa, E. D., and A. C. de Verdière (2002), The 7.7-year North Atlantic oscillation. *Q. J. R. Meteorol. Soc.*, 128A, 797-817.

De Putter, T., M. F. Loure, and G. Wansard (1998), Decadal periodicities of Nile River historical discharge (A.D. 622-1470) and climatic implications. *Geophys. Res. Lett.*, 25, 3193-3196.

Deser, C., and M. L. Blackmon (1993), Surface climate variations over the North Atlantic ocean during winter 1900-1989. *J. Clim.*, 6, 1743-1753.

Dijkstra, H. A., and M. Ghil (2004), Low-frequency variability of the large-scale ocean circulation: A dynamical systems approach. *Rev. Geophys.*, conditionally accepted.

Felis, T. et al. (2004), Increased seasonality in Middle East temperatures during the last interglacial period. *Nature*, 429, 164-168.

Fraedrich, K. (1986), Estimating the dimensions of weather and climatic attractors. *J. Atmos. Sci.*, 43, 419-432.

Fraedrich, K., J. Jiang, F. W. Gerstengarbe, and P. C. Werner (1997), Multiscale detection of abrupt climate changes: Application to River Nile flood levels. *Int. J. Climatol.*, 17, 1301-1315.

Hameed, S. (1984), Fourier analysis of Nile flood levels. *Geophys. Res. Lett.*, 1, 843-845.

Hassan, F. A. (1981), Historical Nile floods and their implications for climatic change. *Science*, 212, 1142-1145.

Hurst, H. E. (1952), *The Nile*, 326 pp., Constable, London.

Hurrell, J., (1995) Decadal Trends in the North Atlantic Oscillation: Regional Temperatures and Precipitation. *Science*, 269, 676-679.

Ghaleb, K. O. (1951), Le Mikyas ou Nilomètre de File de Rodah. *Mémoires de l'Institut d'Égypte*, 54, 182 pp. Cairo.

Ghil, M., and R. Vautard (1991), Interdecadal oscillations and the warming trend in global temperature time series. *Nature*, 350, 324-327.

Ghil, M., et al. (2002), Advanced spectral methods for climatic time series. *Rev. Geophys.* 40(1), 3.1-3.41, doi: 10.1029/2000GR00092 (2002).

Kumar, K. K., B. Rajagopalan, and M. A. Cane (1999), On the weakening relationship between the Indian monsoon and ENSO. *Science*, 284, 2156-2159.

Matthews, A. J. (2004), Intraseasonal variability over Tropical Africa during northern summer. *J. Climate*, 17, 2427-2440.

Mann, M. (2002), Large-scale climate variability and connections with the Middle East in past centuries. *Climatic Change*, 55, 287-314.

Mann, M., J. Park, and R. S. Bradley (1995), Global interdecadal and century-scale climate oscillations during the past five centuries. *Nature*, 378, 266-270.

Mann, M., and J. M. Lees (1996), Robust estimation of background noise and signal detection in climatic time series. *Clim. Change*, 33, 409-445.

Mann, M., R. S. Bradley, and M. K. Hughes (1998), Global-scale patterns and climate forcing over the past six centuries. *Nature*, 392, 779-787.

Moron, V., R. Vautard, and M. Ghil (1998), Trends, interdecadal and interannual oscillations in global sea-surface temperatures. *Clim. Dyn.*, 14, 545-569.

Nicholson, S. E., and D. Entekhabi (1986), The quasi-periodic behavior of rainfall variability in Africa and its relationship to the Southern Oscillation. *J. Clim. Appl. Meteor.*, 24, 331-348.

Popper, W. (1951), *The Cairo Nilometer*, 269 pp., University of California Press, Berkeley/Los Angeles.

Quinn, W. H. (1992), A study of Southern Oscillation-related climatic activity for A.D. 622-1900 incorporating Nile River flood data, in *El Niño: Historical and Paleoclimatic Aspects of the Southern Oscillation*, edited by H.F. Diaz and V. Markgraf, pp. 119-149, Cambridge Univ. Press, Cambridge.

Ropelewski, C. F., and M. S. Halpert, (1987), Global and regional scale precipitation associated with El Niño Southern Oscillation. *Mon. Wea. Rev.* 115, 985-996.

Said, R. (1993), *The River Nile*, Geology, Hydrology and Utilization, 320 pp., Pergamon Press.

Schlesinger, M. E., and N. Ramankutty (1994), An oscillation in the global climate system of period 65-70 years. *Nature*, 367, 723-726.

Schoellhamer, D. (2001), Singular spectrum analysis for time series with missing data. *Geophys. Res. Lett.*, 16, 3187-3190.

Simonnet, E., Ghil, M. and Dijkstra, H. A. (2004), Homoclinic bifurcations in the quasi-geostrophic double-gyre circulation. *J. Mar. Res.*, submitted.

Thomson, J. D. (1982), Spectrum estimation and harmonic analysis. *Proc. IEEE*, 70, 1055-1096.

Toussoun, O. (1925), Mémoire sur l'histoire du Nil, Mémoires de l'Institut d'Égypte, 18, pp. 366-404, Cairo.

Vaquero, J. M., M. C. Gallego, and J. A. Garcia (1997), A 250-year cycle in naked-eye observations of sunspots. *Geophys. Res. Lett.*, 29, doi:10.1029/2002GL014782, 2002.

Vautard, R., and M. Ghil (1989), Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D*, 35, 395-424.

Vautard, R., P. Yiou, and M. Ghil (1992), Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D*, 58, 95-126.

Walker, G. T. (1910), Correlation in seasonal variations of weather, II, Indian Meteorol. Mem., 21, 1-21.

Yiou, P., D. Sornette, and M. Ghil (2000), Data-adaptive wavelets and multi-scale SSA. *Physica D*, 142, 254-290, oscillations in global sea-surface temperatures. *Clim. Dyn.*, 14, 545-569.