Statistical Analysis of Historical Atlantic SST Data: Methodology and Application*

A. Kaplan, Y. Kushnir, M. A. Cane, and M. B. Blumenthal
Lamont Doherty Earth Observatory of Columbia University, Palisades, NY 10964

March 8, 1996

Historical archives of marine observations present the user with two major problems: observational errors, and spatial and temporal gaps. It is these problems that statistical data analysis is trying to address. Here we develop a computationally feasible way to produce a climate data analysis, coherent in space and time, and demonstrate its application.

The technique is shown here in application to Atlantic monthly SST anomalies taken from the UK Meteorological Office archives. These data are included in the Global Ocean Surface Temperature Atlas (GOSTA, Bottomley et al., 1990; Folland and Parker, 1995). They represent monthly averaged SST anomalies from ship observations binned in 5 by 5 degree boxes for the period 1856–1991. The actual number of observations inside each gridbox (along with intrabox spatial variability which we estimate from COADS dataset) are used in order to create realistic observational error estimates.

Our method combines the idea of variational analysis with a reduced space approach. The analyzed field is presented as a time dependent linear combination of a few spatial structures, namely the field leading Empirical Orthogonal Functions (EOFs), and inserted into a properly formed quadratic functional whose minimum is expected to give an optimal estimate of the field. This estimate is gap free, and its error is constrained. More precisely: let T_n denote the SST field (rearranged as a one-dimensional vector of length M) for the month $n = 1, \ldots, N$, and let superscripts o and a denote the monthly averaged observed and analyzed anomalies, respectively. Then express the field as a linear combination of a subset of EOFs: $T_n^a = Ea_n$ (E is a $M \times L$ matrix of EOF patterns used, and a_n is a L-vector of their amplitudes for the

*Appeared in Atlantic Climate Change Program (ACCP) Notes, vol III, No.1, April 1996, 4–7.

month n). The functional we minimize is

$$S[a_1, a_2, \ldots, a_N] = \sum_{n=1}^{N} (H_n E a_n - T_n^o)' R_n^{-1} (H_n E a_n - T_n^o) + \sum_{n=1}^{N-1} (E a_{n+1} - A_n E a_n)' Q_n^{-1} (E a_{n+1} - A_n E a_n).$$

Here R_n is the matrix of observational errors, and H_n is the matrix that transforms an analyzed field into the format of observations (here it designates those gridboxes where month n observations are actually available). The matrices A_n and Q_n determine a linear, first-order autoregressive model (in a multivariate sense) for monthly transitions of SST anomalies. A_n is the transition matrix (in EOF space), and Q_n the model error-covariance matrix estimate (in data space), respectively. Formally, minimizing the expression S achieves an optimization where the multi-dimensional "distance" between the analysis and the data, on one hand, and the previous analysis, on the other hand, is minimal (e.g, Sasaki, 1970).

Equating A_n to 0 and correspondingly replacing Q_n by the anomaly covariance matrix C_n in the formula for the functional S, eliminates the connection between estimates for consecutive months. As a result, analyzed fields for each month minimize the functional separately, not constrained by time continuity, but constrained to have energy distribution in the EOF modes similar to that prescribed by C_n. This in essence is the well known optimal interpolation (OI) procedure discussed by Gandin (1963). If only the first quadratic term in S is retained (i.e., the covariance constraint is removed), the estimate becomes a weighted least squares fit to observations, in EOF space, as used e.g. by Shriver and O'Brien (1995), and Smith et al. (1996).

We compute the EOFs using a covariance matrix calculated from the 1951-1991 observations (when the data coverage was best and data collection methods were reliable). The covariance matrix is smoothed by Shapiro filter (Shapiro, 1971) of the second order before the EOF calculation to reduce sampling errors. The OI variant of the analysis is then used to derive an initial estimates of the SST field (T_n^{OI}), and to calculate seasonally varying A_n matrices for the first order autoregressive model for the period 1956-1981. All A_n models are represented by $L \times L$ matrices, but, to avoid overfitting, each of them is restricted not to have non-zero elements outside of diagonal with any index exceeding certain $L_1 < L$. The A_n models is tested separately for the years where they were calculated and for the months in years not used in their calculation (1946–55 and 1982–91). The conclusion was that with

$L_1 = 5$ and $L = 30$ we can resolve essential part of the variance without introducing artificial skill. The minimization of the functional S was implemented as a "sweep" procedure, which, because of the statistical meaning of R_n and Q_n matrices, can be interpreted here as an optimal smoother (OS, Rauch et al., 1965). This approach not only provides the analyzed field itself, but also produces a spatial error estimate for each monthly field.

Figures 1 and 2 compare the present analysis to the NMC OI analysis (Reynolds and Smith, 1994). The OI analysis is a relatively high resolution monthly analysis of ship observations and satellite estimates of SST. In Figure 1 an arbitrary month is presented. The figure also attests to the robustness of the OS analysis when part of the data are withheld in its application. Figure 2 presents the spatial correlation and rms difference between the NMC OI and the present OS analysis. Figure 3 displays the area averaged analyzed SST anomalies north of 35^oN as a function of time since the middle 1850's. The multidecadal fluctuation displayed by this series was studied in Kushnir (1994). Figure 4 shows the difference between area averaged analyzed SST anomalies in the tropical North Atlantic (5^oN to 28^oN), and area averaged analyzed anomalies in the tropical South Atlantic (20^oS to $5^o N$). This index was suggested by Servain (1991) as a measure of the SST gradient across the equatorial Atlantic, an important part of the mechanism governing multi-year fluctuations in precipitation over the continents bordering the region.

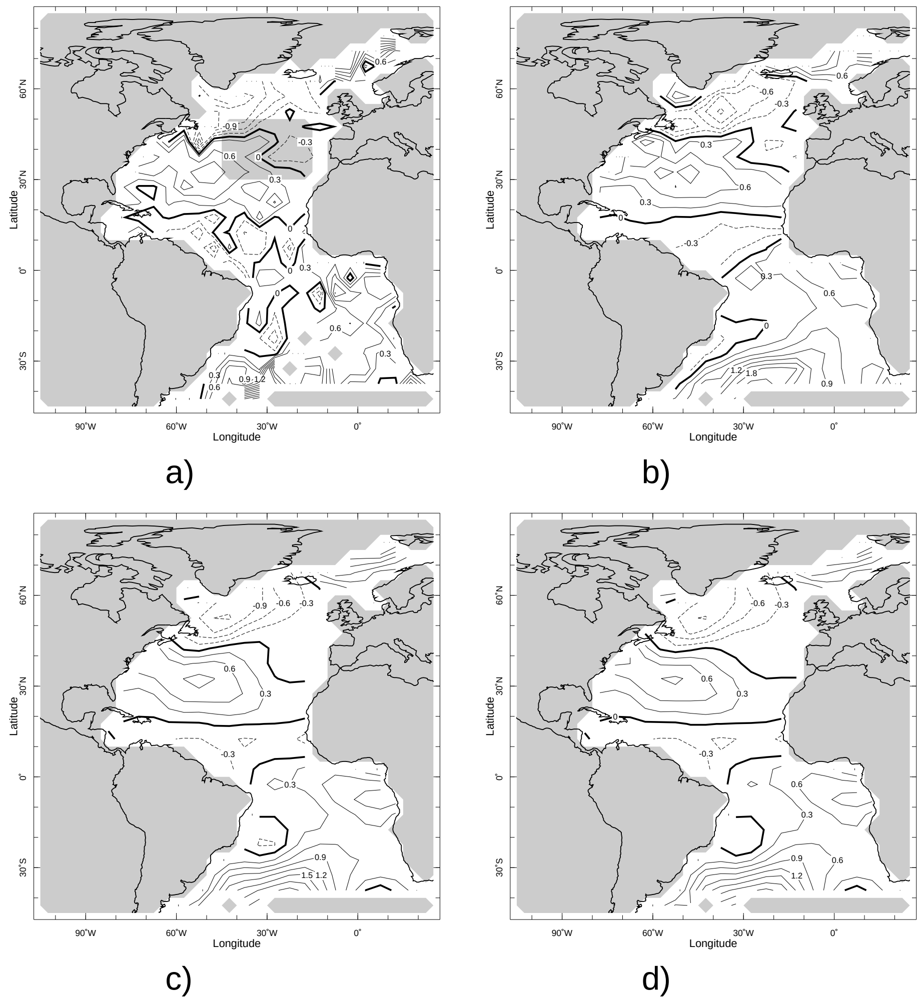

Figure 1: Verification of optimal smoother (OS) analysis for March 1991: *a)* observed GOSTA SST, *b)* NMC OI analysis, *c)* analysis by OS when all the data shown on *a* are used, *d)* same as *c*, but when all observational data from $30^o \times 20^o$ rectangular area (shown by shading underneath of SST isolines) are withheld for all months, demonstrating the robustness of OS estimate. Contour interval is 0.3^o C, negative contours are dashed.

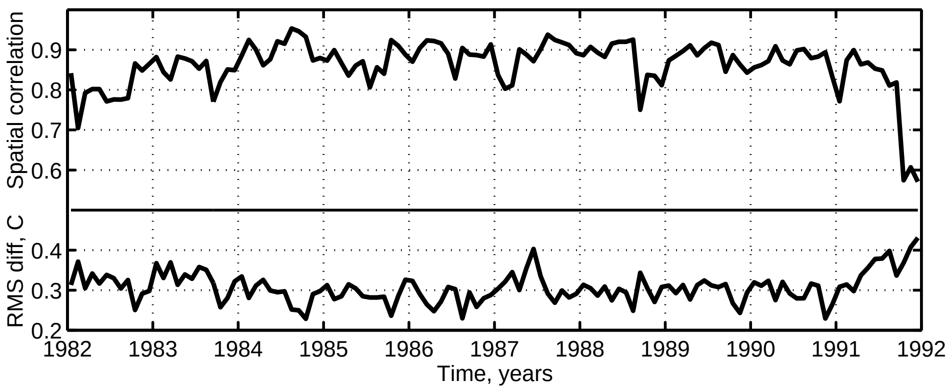

Figure 2: Spatial correlation and RMS difference between OS and NMC OI analysis shown as a function of time from January 1982 to December 1991

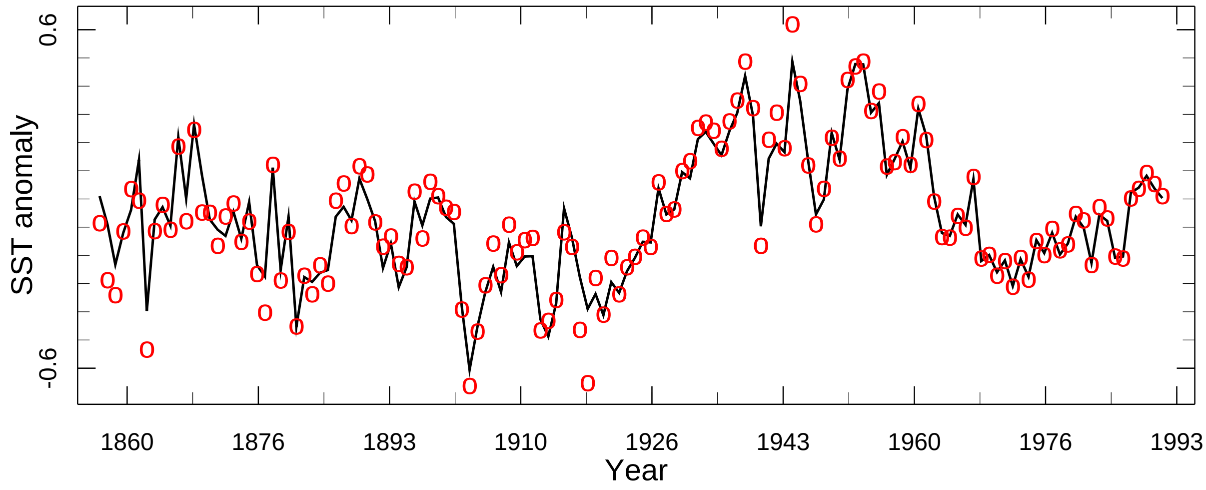

Figure 3: Area averaged annual SST anomalies, ^{o}C, in North Atlantic north of 35^{o}N, from 1856 to 1991. The solid curve is based on the OS analysis. The circles represent an average of all avialable data in the region.

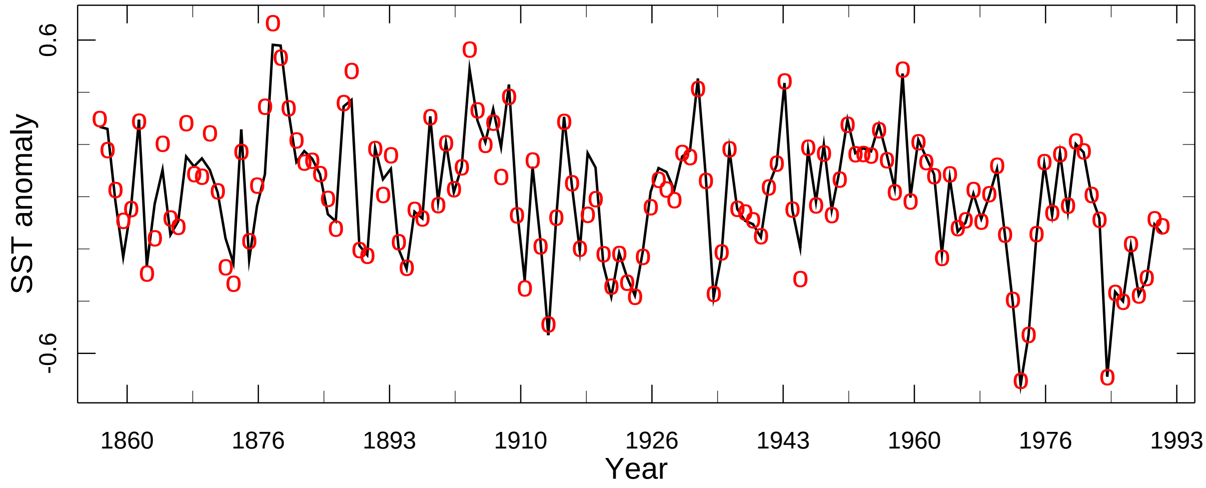

Figure 4: Same as Figure 3 but for the difference between the average over 5^{o}N to 28^{o}N and the average over 20^{o}S to 5^{o}N.

References

Bottomley, M., C. K. Folland, J. Hsiung, N. R.E. and D. E. Parker, 1990: *Global Ocean Surface Temperature Atlas.* Vol. Meteorological Office Bracknell UK and Dept. of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, 20 pp., 313 plates.

Folland, C. K. and D. E. Parker, 1995: Correction of instrumental biases in historical sea surface temperature data. *Quart. J. Roy. Meteor. Soc.*, **121**, 319-367.

Gandin, L. S., 1963: *Objective Analysis of Meteorological Fields.* Gidrometeorologicheskoye Izdatel'stvo, Leningrad. Translated from Russian by Israel Program for Scientific Translations, Jerusalem, 1965, 242 pp.

Kushnir, Y., 1994: Interdecadal variations in North Atlantic sea surface temperature and associated atmospheric conditions. *J. Climate*, **7**, 141-157.

Rauch, H.E., F. Tung, and C.T. Striebel, 1965: Maximum likelihood estimates of linear dynamic systems. *AIAA J.*, **3**, 1445–1450.

Reynolds, R.W., and T.M. Smith, 1994: Improved global sea surface temperature analysis using optimum interpolation. *J. Climate*,**7**, 929–948.

Sasaki, Y., 1970: Some basic formalisms in numerical variational analysis. *Monthly Weather Review*, **98**, 875-883.

Servain, J., 1991: Simple climatic indices for the tropical Atlantic Ocean and some applications. *J. Geophys. Res.*, **96**, 15,137-15,146.

Shapiro, R., 1971: The use of linear filtering as a parameterization of atmospheric diffusion. *J. Atmos. Sci.*, **28**, 523–531.

Shriver, J.F. and J. J. O'Brien, 1993: Low-frequency variability of the equatorial Pacific ocean using a new pseudostress dataset: 1930-1989. *J. Climate*, **8**, 2762-2786.

Smith, T. M., R. W. Reynolds, R. E. Livezey, and D.C. Stokes, 1996: reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Climate*, **9**, (in press).