GEOPHYSICAL RESEARCH LETTERS, VOL. 30, NO. 14, 1727, doi:10.1029/2003GL017312, 2003

On the invariance of the first EOF of North Atlantic Sea Level Pressure to temporal filtering

E. D. da Costa

Laboratoire de Physique des Oceans, Brest, France

Received 12 March 2003; revised 4 June 2003; accepted 18 June 2003; published 16 July 2003.

[1] This article examines the frequency-dependence of the shape of the first empirical orthogonal function (EOF) of sea level pressure (SLP) over the North Atlantic region for time-scales in the range 8 months–45 years. It evidences that the time-sacle chosen for conducting the EOF analyses had no incidence on the main features of the first EOF of North Atlantic SLP. *INDEX TERMS:* 0399 Atmospheric Composition and Structure: General or miscellaneous; 1699 Global Change: General or miscellaneous; 3299 Mathematical Geophysics: General or miscellaneous; 3399 Meteorology and Atmospheric Dynamics: General or miscellaneous. **Citation:** da Costa, E. D., On the invariance of the first EOF of North Atlantic Sea Level Pressure to temporal filtering, *Geophys. Res. Lett.*, *30*(14), 1727, doi:10.1029/2003GL017312, 2003.

1. Introduction

[2] This study is mainly about the question: Is the shape of the dominant mode of variability in North Atlantic SLP dependent on the time–scale chosen to make the analysis? The first Empirical Orthogonal Function (EOF) and the first Principal Component (PC) of the North Atlantic surface pressure field are usually considered as representing, respectively, the space and time structure of the dominant mode of variability over the North Atlantic, i.e., the North Atlantic Oscillation (NAO) (e.g., *Hoerling et al.* [2001]). This interpretation ensues from the following observations: i) a good correlation of the NAO index time series with the first PC of the surface pressure field (e.g., *Hurrel* [1995] found a correlation coefficient of 0.91 between them for decadal time-scales), ii) the similarity between the first EOF and the spatial pattern issued from averaging the surface pressure fields that display a positive NAO index (e.g., Figure 2 of *Greatbatch*, [2000]), iii) the likeness between the first EOF and the teleconnections pattern for the North Atlantic surface pressure fields (e.g. *Wallace and Gutzler* [1981]). In virtue of the relationship between the first EOF of North Atlantic SLP and NAO, the study reported here suggests the dominance of the NAO pattern upon the variability at every time-scale from 8 months to 45 years.

2. Data

[3] We used the monthly mean-sea-level pressure data (MSLP) on a 5° latitude by 10° longitude grid–point basis from the Climatic Research Unit website. The sources of the original data are given in *Jones* [1987]. The data we analyzed had been recorded from 1 January 1873 to 31 December 2000. We extracted from this global MSLP database a set of data relative to a region located within [20°N, 65°N] in latitude and [100°W, 10°E] in longitude. This area of about 120 grid points rougly coincides with the North Atlantic region. Inside it, 20 grid–points time–series have missing values. The number of missing data by grid point remains less than 1.5% of the total length of the time series, that is, about 23 gaps in 128*12 = 1536 months. Each gap is filled with the climatological value of the MSLP for the corresponding month. However, for the calculations described in this paper, the seasonal cycle was removed by subtracting out the calendar monthly means.

[4] The data were not detrended. We found no significant effect of the trend in the MSLP data (not shown) on leading EOFs.

3. Methodology

[5] Let $X(p, t)$ denote the MSLP anomaly at the time t and grid point p. The whole data set can be represented either by a matrix $\mathbf{X}$, or by time series $X_p(t)$ or by 2D-fields $X_t(p)$, ($p = 1, 2, \ldots, P = 90$; $t = 1,2, \ldots, N = 1536$). Here, N and P denote the total number of months and grid points of the MSLP data set, respectively. We assume that the MSLP data set is a sample of a stationary normally-distributed random variable, $\mathbf{X}$, and that the mean $\mu(p)$ and variance $\sigma^2(p)$ of the random time series $X_p(t)$ can both be estimated from the sample. Hereafter we use the standardized variable $Z(p, t) \equiv (X(p, t) - \mu(p))/\sigma(p)$.

[6] Firstly, $Z(p, t)$ is filtered using 15 passband filters and then decomposed into EOFs by a PC analysis. Our aim is to show not only that the first EOF of the MSLP filtered data, denoted by E_W, is still independent of the filtering frequency band, W, but also that it remains alike the first EOF, E_{ref}, of the unfiltered data set. The spatial correlation between E_{ref} and E_W gives the similarity between them.

[7] A relevant issue is about the accidental finding of results. This question can partially be approached through statistical testing to establish the significance of the results. However, the correct use of statistical tests requires additional independent data, that is new independent realizations of the climatic system. The use of the same data to derive the result and conduct the test leads to the "Mexican Hat" problem described in *von Storch and Zwiers* [2000, pp. 106]. On the same page Storch and Zwiers suggest two alternatives to cope with this lack of data. The first one is to divide the observations into learning and validation data sets, assuming that the learning and validation periods behave independently. This, however, reduces the frequency range available for the analysis. Another alternative is to use a carefully constructed CGM to produce independent simulations of the phenomena to be analysed. This quite hard

0094-8276/03/2003GL017312$05.00

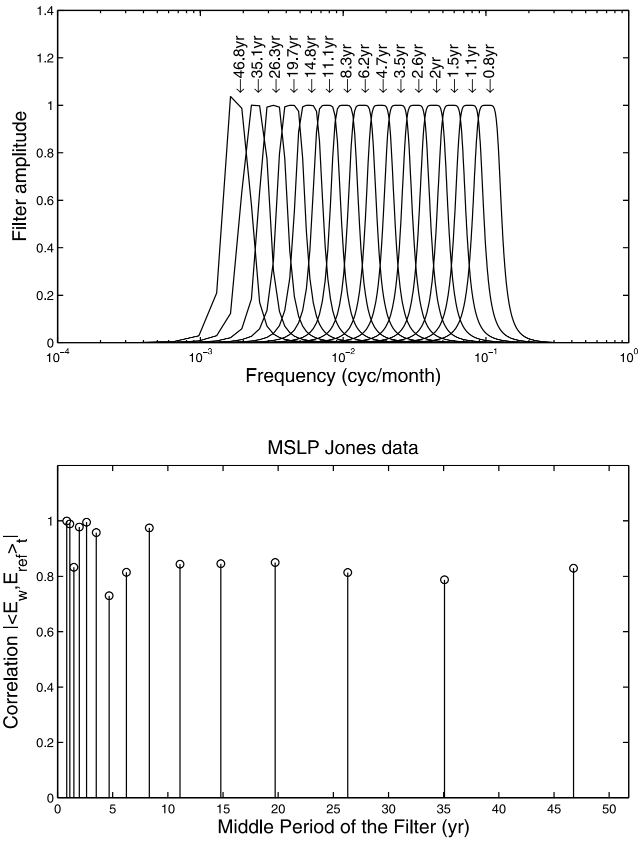

Figure 1. Top pannel: Shapes of the pass band filters; Bottom pannel: Absolute value of the space–correlation between the NAO pattern and the first EOF of the filtered data plotted against the middle period of the filter. Typical value of the space–correlation between MSLP fields is 0.15 and the variance of this value is 0.05.

procedure can be replaced by a statistical model approach, which gives independent data sets. Such a procedure was illustrated by *Smith* [1992] to distinguish red noise from non–linear determinism. He pointed out the availability of numerous statistical models to construct surrogate data series. Here our results are tested against both white and red noises.

4. PC Analysis of the Filtered Data

[8] The 15 filters used in this study are Butterworth ones, mainly because they give a maximally flat response in both the passband and stopband. The design of a Butterworth filter requires to specify the passband riple, *Rp*, the stopband attenuation, *Rs*, and the transition width ($Ws_i - Wp_i$), (i = 1, 2). For all filters we set N and *Rp* to 3 and to 3dB, respectively. As *Rp* and N are fixed, each Butterworth filter is only determined by the passband edge frequencies, Wp_i. The passband window length of each filter, W, varies as a function of its passband edge frequencies as $W = (Wp_1 + Wp_2)/3$. This leads to filters which have almost the same shape in the frequency domain (see Figure 1). Once the filters have been defined we proceed by filtering the MSLP anomalies, then further decomposed into EOFs through a PC analysis. Only the first EOF of each data set is retained (Figure 2), and its space correlation with the EOF of reference is calculated. The correlation is always above 0.7 (see Figure 1), which suggests that the main features of the first EOF of MSLP are not affected by the analysis time-scale.

5. Testing Against Noise

[9] In order to allow both autocorrelation in the synthetic time series and cross correlation between the individual time series, we followed *Parlange and Katz* [2000] who used a multivariate, first-order autoregressive [AR(1)] (space-time red noise) process to produce randomly standardized functions, $R_p(t)$. That is, the surrogates generator is given by:

$$R_p(t) = \Upsilon R_p(t-1) + W_p(t) \quad (1)$$

where Υ is a $P \times P$ matrix of lag-one time correlation coefficients and $W_p(t)$ is a time series drawn from a multivariate normal distribution with zero mean and $P \times P$ variance-covariance matrix Ξ.

[10] The matrices Υ and Ξ satisfy the equations $\Upsilon M_o = M_1$ and $\Xi = M_o - \Upsilon M_1^T$, where M_o and M_1 are the lag-zero and lag-one cross-covariance matrices with elements $m_o(p, q) = \langle Z_t(p), Z_t(q)\rangle_t$ and $m_1(p, q) = \langle Z_t(p), Z_{t-1}(q)\rangle_t$, with $p, q = 1, 2, \ldots, P$ and $\langle .,. \rangle_t$ the correlation operator. This technique reproduces not only the contemporaneous and lag-one cross-correlations, but also the first-order autocorrelations among the standardized random variables $Z_t(p)$.

[11] Using equation 1 we test for the significance of the time invariance of the E_{ref} pattern following the approach of

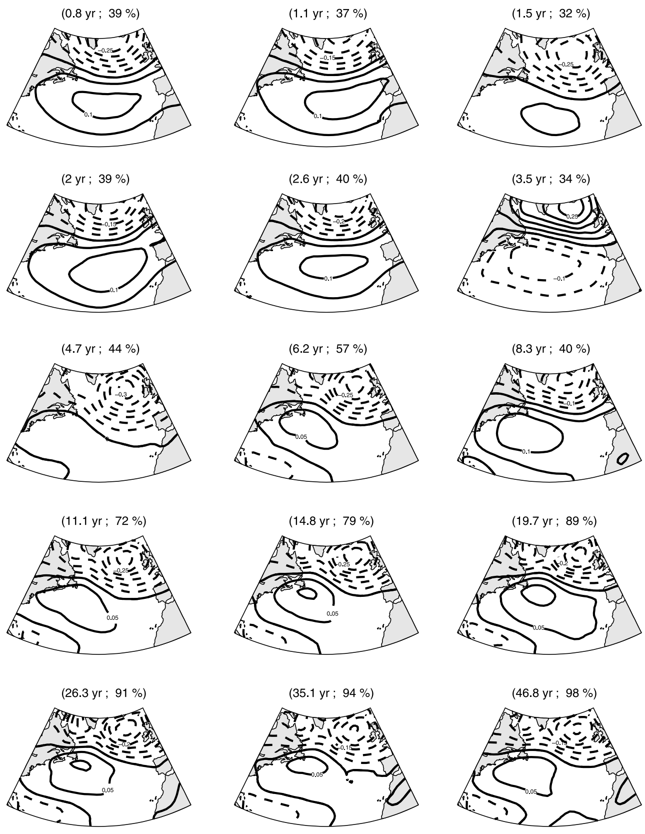

Figure 2. Patterns of the first EOF for the filtered data. The middle period of the filter and the explained variance are plotted at the top of the image.

Allen and Robertson [1996], a Monte Carlo surrogate data testing. According to the philosophy of surrogate data testing, we generate a large number of surrogates data segments $R(t, p)_s$, ($s = 1, 2, \ldots, 1000$ segments) with the same dimension as that of the original MSLP data. For each segment, the procedure of filtering and EOF decomposition are carried out in exactly the same way as for the original data. Thus, for each frequency band, we obtain 1000 values of the correlation coefficient between the firsts EOFs of the unfiltered and filtered surrogate data. Then, for each frequency band the distribution of correlation values is computed and the 97.5th percentile is compared to the correlations obtained by using the MSLP data over the same frequency band. Wherever this correlation is larger than the 97.5th percentile, we conclude that the time invariance of the E_{ref} is statistically more significant than expected from the hypothesis of data being produced by a red space-time noise. However, this never happened: the correlation coefficients obtained using the MSLP data felt always within the 25th and the 75th percentiles computed from surrogates. The time invariance of the first EOF can be reproduced by a red noise process with the same covariance matrix and lag-one autocorrelation as in the data.

[12] To further explore this result we conducted two additional tests. In the first one we used a red-white model, that is, a model producing data which are temporally alike red noise and spatially alike white noise (no cross-correlations between locations). This model can be obtained from the red space-time noise model by suppressing the second term on the r.h.s. of equation 1. This is equivalent to the Markov process used by *Allen and Smith* [1996]: $R_p(t) = \gamma_p R_p(t-1) + \alpha_p W_p(t)$, $t = 1, 2, \ldots, N$, where $W_p(t)$ is drawn from a Gaussian, unit-variance white noise, γ_p is given by the lag-one auto-covariance, c_{p1}, of the individual time series $Z_p(t)$, and α_p is given by $\sqrt{(1-c_{p1})^2}$.

[13] With this model the correlation coefficients obtained using the MSLP data always exceeded the 97.5th percentile value.

[14] The second test we used considers a white–red model producing data which are at each time spatially correlated with neighboring grid points, like in MSLP data, but on the other hand, they show no correlation either with the previous or the next observation time.

[15] The correlation coefficients obtained with this model felt always within the 25th and the 75th percentiles.

[16] These tests highlight that the key–factor is the correlation between locations for it contributes to the temporal invariance of the first EOF of North Atlantic MSLP. The role of temporal dependence is secondary.

6. Discussion

[17] The frequency–independence of EOFs shapes was first suggested by *North* [1984]. Even though his paper was mainly about the relationship between the modes of physical systems and their EOFs, in the last paragraph of section 3 author evidenced the existence of a large class of systems where EOFs are independent of the time–scale. In addition, he pointed out the benefits, in model identification studies, of finding "whether empirically derived EOFs taken from real data depend upon frequency or not".

[18] The frequency–independence of EOFs shapes can be also useful in EOF–based techniques for reconstruction of past geophysical fields. For instance, some of the statistical methods applied to reconstructing past variability of SST [e.g. *Smith et al*, 1996; *Kaplan et al.*, 1998] lie on the assumption that EOF shapes remain invariante during the whole reconstruction.

[19] The MSLP pattern shown in Figure 2, though similar in gross features has distinct differences, e.g. the low center around Iceland/Greenland seems to shift eastward with increasing filter window and the high center in the subtropics seems to shift westward. Differences among the patterns should be expected, however systematic shifts may suggest some hidden correlation. *North et al.* [1982] showed that the error in the estimation of a given EOFs results from contamination by the patterns of the neighboring EOFs. Moreover, the amplitude of the error is correlated with the number of independent samples used in computations. At low frequencies, the number of independent samples in the data set is reduced, and this decrease likely explains the observed shift. Another possibility is an ocean–atmosphere interaction at low–frequencies suggested by the similarity between EOF patterns for low–frequencies and the first EOF of Sea Surface Temperature (see Figure 2). However, the discussion of such a possibility requires additional investigations that are beyond the scope of this study.

[20] The NAO is mainly a winter time feature and one might presume that the invariant patterns might be much more robust in the winter time. We carried out the same analyses on winter–time data but the patterns remained similar to those of Figure 2 (not shown).

[21] Another point concerns the frequency range of our analyses. We used monthly data as the highest frequency input to our analyses; this choice failed the sampling of the frequencies at which most of the NAO variance occurs, namely from daily to weekly–time scales. However, as the pattern of the first EOF of North Atlantic SLP is known to coincide with the NAO pattern, we have focused on the low–frequency patterns.

[22] We also showed that a space–time red–noise process can reproduce the invariance of shape of the first EOF to temporal filtering. The random process to be successful must use correlations between locations computed from the observations. The source of such spatial correlations cannot be explained from a statistical point of view, they are related to deterministic mechanisms (e.g. the location of the mean subtropical and polar jets) and forcings (e.g. on an aqua–planet it would not be clear where NAO should have its center). This shows that NAO is not a random process, but can be statistically simulated by one where the spatial correlations are the observed ones.

7. Conclusions

[23] For time–scales ranging from 8 months to 45 years we showed the time invariance of the main features of the dominant pattern of variability in the North Atlantic sea level pressures.

[24] We also evidenced that a space–time red-noise process can produce this kind of time invariance. To be successful, the random process must use spatial correlations computed from the observations.

[25] **Acknowledgments.** The stay of E. D. da Costa at LPO was granted by Foundation for Science and Technology (Portugal)(SFRH/BPD/3565/2000). Two anonymous reviewers helped to improve the initial manuscript.

References

Allen, M., and A. Robertson, Distinguishing modulated oscillations from coloured noise in multivariate datasets, *Clim. Dyn.*, *12*, 775–784, 1996.

Allen, M., and L. Smith, Monte Carlo SSA: Detecting Irregular Oscillations in the Presence of Colored Noise, *J. Clim.*, *9*, 3373–3403, 1996.

Greatbatch, R. J., The North Atlantic Oscillation, *Stochastic Environmental Research and Risk*, 213–242, 2000.

Hoerling, M. P., J. W. Hurrell, and T. Xu, Tropical Origins for Recent North Atlantic Climate Change, *Science*, *292*, 90–92, 2001.

Hurrel, J. W., Decadal trends in the North Atlantic Oscillation regional temperatures and precipitation, *Science*, *269*, 676–679, 1995.

Jones, P. D., The early twentieth century Arctic High – fact or fiction?, *Climate Dynamics*, *1*, 63–75, 1987.

Kaplan, A., M. Cane, Y. Kushnir, A. Clement, M. Blumenthal, and B. Rajagopalan, Analyses of Global Sea Surface Temperature 1856–1991, *J. Geophys. Res.*, *102*, 27,835–27,860, 1998.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, Sampling errors in the estimation of empirical orthogonal functions, *Mon. Wea. Rev.*, *110*, 699–706, 1982.

North, G. R., Empirical Orthogonal Functions and Normal Modes, *J. Atmos Sci.*, *41*, 879–887, 1984.

Parlange, M. B., and R. W. Katz, An Extended Version of the Richardson Model for Simulating Daily Weather Variables, *J. of App. Meteo.*, *39*, 610–622, 2000.

Smith, L. A., Identification and prediction of low–dimensional dynamics, *Physica D*, *58*, 50–76, 1992.

Smith, T., R. W. Reynolds, R. E. Livezey, and D. C. Stokes, Reconstruction of Historical Sea Surface Temperatures Using Empirical Orthogonal Functions., *J. of Clim.*, *9*, 1403–1420, 1996.

von Stroch, H., and F. Zwiers, *Statistical Analysis in Climate Research*, Cambridge University Press, 484 pp., 2000.

Wallace, J. M., and D. S. Gutzler, Teleconnections in the geopotential height field during the northern hemisphere winter, *Mon. Wea. Rev.*, *109*, 784–812, 1981.

E. D. da Costa, Laboratoire de Physique des Oceans, Universite de Bretagne Occidentale – UFR Sciences et Techniques, 6 Avenue Le Gorgeau, BP 809, 29285, Brest Cedex, France. (dacosta@univ-brest.fr)