A TEMPORAL EXTENSION TO TRADITIONAL EMPIRICAL ORTHOGONAL FUNCTION ANALYSIS

ALLAN AASBJERG NIELSEN AND KLAUS BAGGESEN HILGER

Informatics and Mathematical Modelling, Technical University of Denmark, Building 321
DK-2800 Kgs. Lyngby, Denmark, Telephone +45 4525 3425, Telefax +45 4588 1397
{aa,kbh}@imm.dtu.dk, www.imm.dtu.dk/~{aa,kbh}

OLE BALTAZAR ANDERSEN AND PER KNUDSEN

National Survey and Cadastre, Rentemestervej 8
DK-2400 Copenhagen NV, Denmark, Telephone +45 3587 5050, Telefax +45 3587 5051
{oa,pk}@kms.dk, research.kms.dk/~{oa,pk}

This paper describes the application of temporal maximum autocorrelation factor analysis to global monthly mean values of 1996-1997 sea surface temperature (SST) and sea surface height (SSH) data. This type of analysis can be considered as an extension of traditional empirical orthogonal function (EOF) analysis, which provides a non-temporal analysis of one variable over time. The temporal extension proves its strength in separating the signals at different periods in an analysis of relevant oceanographic properties related to one of the largest El Niño events ever recorded.

1 Introduction

Empirical orthogonal functions (EOF) [5,10] analysis is often used in oceanography and other geophysical sciences to analyse temporal sequences of scalar (image) data. This paper deals with an extension of the EOF method, by applying maximum autocorrelation factor (MAF) analysis to two temporal sequences of scalar image data, namely global sea surface temperature (SST) and global sea surface height (SSH) one at a time. This type of analysis is typically carried out to maximise spatial autocorrelation. Here we maximise temporal autocorrelation. The data used to illustrate the analysis carried out come from the Ocean Pathfinder programmes set up by NASA/NOAA [7,9]. The data chosen represent relevant oceanographic properties related to one of the largest El Niño events ever recorded.

El Niño is a large-scale warm ocean event in the Pacific off the coast of Peru and Ecuador caused by eastward drifting toward the west coast of South America of the pool of warm waters normally residing in the western part of the Pacific. This event is not local but may influence weather conditions world-wide. This disruption of normal climate may have drastic socio-economic consequences [6]. The strength and frequency of El Niño events may be influenced by human activities.

2 Method

EOF analysis is a name often used in geophysical data processing for principal component analysis (PCA) [4,1]. Often the usual PCA assumption on variables with mean zero is replaced by an assumption of temporal means of zero. PCA finds linear combinations (called principal components, PCs) of the original data that maximise variance. PC1 contains maximum variance. Higher order PCs contain maximum variance subject to orthogonality or uncorrelatedness with lower order PCs. Maximum autocorrelation factor (MAF) analysis [11,2] finds linear combinations of the zero mean original variables that maximise autocorrelation between neighbouring observations. MAF1 is a linear combination of the original data that have maximum autocorrelation. Higher order MAFs have maximum autocorrelation subject to orthogonality or uncorrelatedness with lower order MAFs. Often the spatial autocorrelation is maximised, here we maximise the temporal autocorrelation. The temporal MAF (T-mode MAF or T-MAF) analysis is carried out first on the SST and then on the SSH data.

Linear combinations $a^T X$ with maximum (auto)covariance between a (zero mean) signal $X(t)$ and a temporally shifted version of the same signal $X(t+\Delta)$ are obtained by finding coefficients a that maximise $a^T \Gamma(\Delta) a$ where $\Gamma(\Delta)$ is the covariance between $X(t)$ and $X(t+\Delta)$. Introducing Σ as the dispersion matrix of the original variables $X(t)$ and Σ_Δ as the dispersion matrix of the difference between the original and the temporally shifted variables $X(t)-X(t+\Delta)$, we get for the autocovariance $a^T(\Sigma - 1/2\,\Sigma_\Delta)a$ and hence for the autocorrelation $1 - a^T \Sigma_\Delta a / 2a^T \Sigma a$. Therefore the desired linear combinations are obtained by solving the generalised eigenvalue problem

$$\Sigma a = \lambda \Sigma_\Delta a$$

where a are (conjugate eigen-) vectors containing the desired weights to construct the variables. λ is the eigenvalue and $1 - 1/(2\lambda)$ is the autocorrelation.

The results from the analysis depend on Δ; here Δ is one month.

3 Data and results

The data used are global monthly mean values of 1996-1997 SST data from the NOAA/NASA Oceans Pathfinder AVHRR SST database [9] and global monthly mean values of 1996-1997 SSH data from the NASA/GSFC Ocean Altimeter Pathfinder database [7]. The SSH data are interpolated point observations from the TOPEX/Poseidon radar altimeter mission. The SST data come as 360 rows by 720

columns half degree data starting at 180° longitude, the SSH data come as 179 rows by 360 columns one degree data starting at 0° longitude. The SST data have been resampled to the SSH grid. The AVHRR instrument is influenced by cloud coverage whereas the radar altimeter provides uninterrupted data. Statistics for the SST analysis are calculated where SST have non-missing values for all 24 months and similarly for SSH.

Figure 1 shows the first six SST T-MAFs. The T-MAFs are stretched linearly from mean ± three standard deviations, the pseudo-colour scale goes from blue (minimum) over cyan, green, yellow to red (maximum). Figure 2 shows the first six SSH T-MAFs (same stretch and colouring).

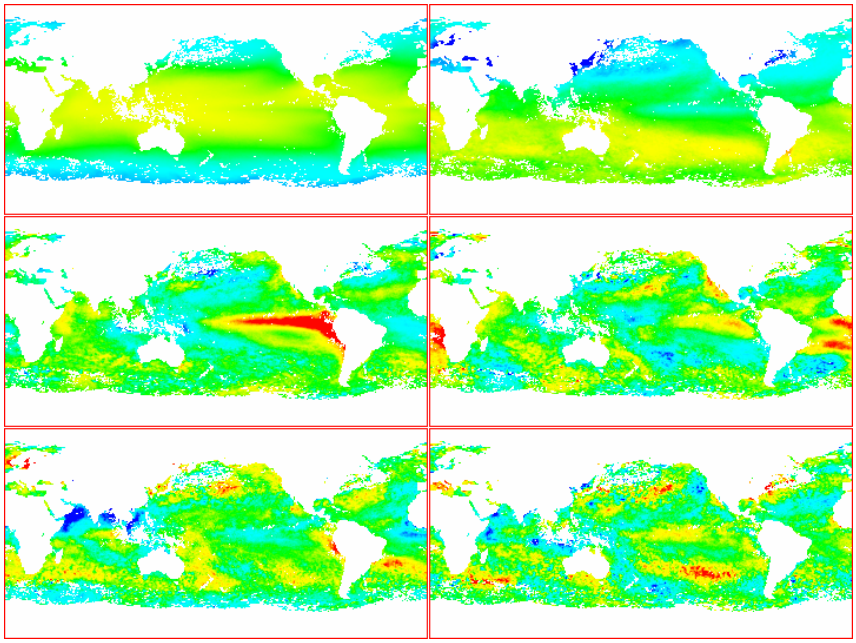

Figure 1. SST T-MAFs 1-6 row-wise.

Figure 3 shows the SST T-MAF autocorrelations (in black), and the correlations between the original data and the first six SST T-MAFs for both SST and SSH (correlations with T-MAFs 1-6 are shown in red, green, blue, cyan, magenta and yellow, respectively).

Figure 4 shows the SSH T-MAF autocorrelations (in black), and the correlations between the original data and the first six SST T-MAFs for both SST and SSH (correlations with T-MAFs 1-6 are shown in red, green, blue, cyan, magenta and yellow, respectively).

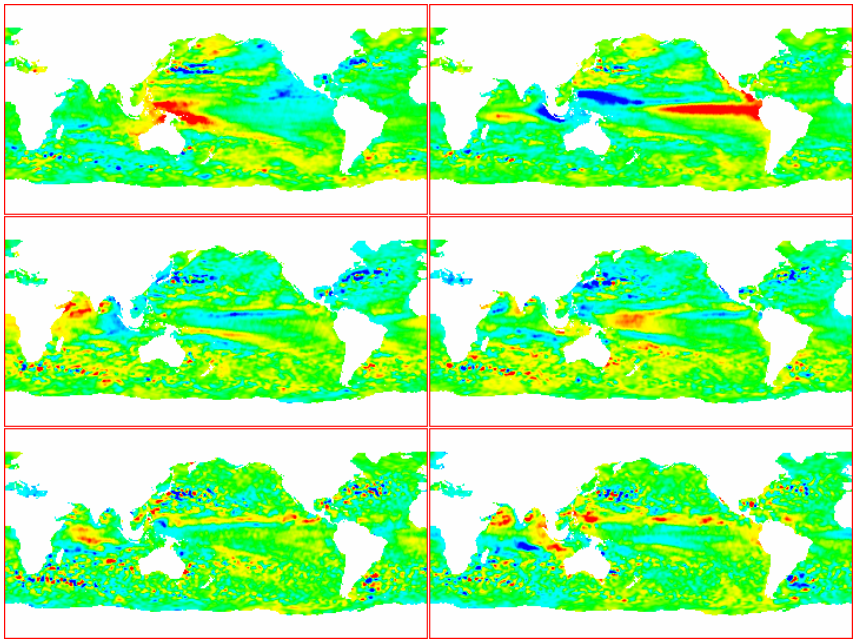

Figure 2. SSH T-MAFs 1-6 row-wise.

4 Discussion and conclusions

By maximizing the temporal autocorrelation we generally isolate features with the longest period in the first MAFs. This means that signals related to the mean and possible trend of the data will be captured first. Thereafter signals with shorter periods like annual signals and intra-annual variations will be captured.

Figure 3 shows that the first seven SST T-MAFs (especially T-MAFs 1-4) are highly autocorrelated temporally. SST T-MAF1 (shown in red) is highly (positively) correlated with all 24 months of SST data with a small semi-annual oscillation. Accordingly, the spatial structure of SST T-MAF1 shows the overall mean of the SST field having high temperatures in the tropics and cold temperatures away from the Equator. The SST T-MAF1 is nearly uncorrelated with all 24 months of SSH shown in indices 25-48. This corresponds well with the fact that the mean of the SSH field has been removed prior to the analysis.

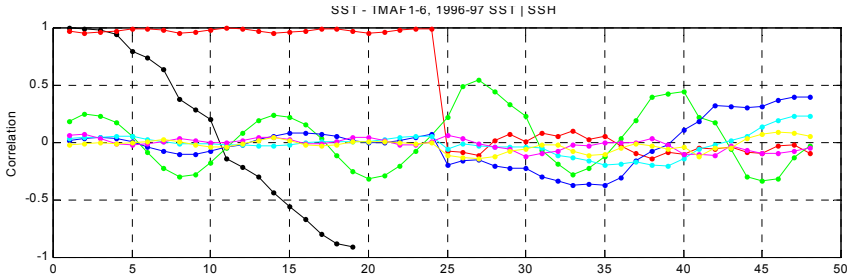

Figure 3. SST T-MAF analysis: autocorrelations (in black), and correlations between original SST (indices 1-24) and SSH variables (indices 25-48), and SST T-MAFs 1-6 (in red, green, blue, cyan, magenta and yellow, respectively).

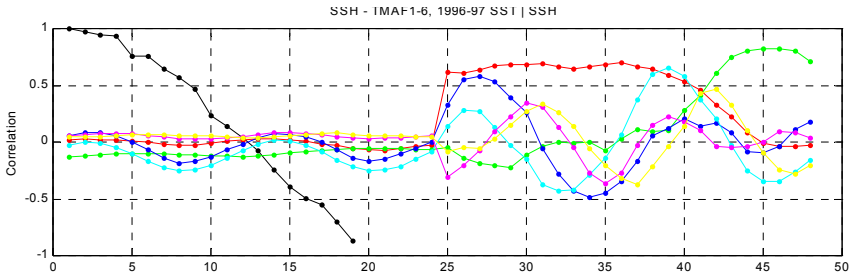

Figure 4. SSH T-MAF analysis: autocorrelations (in black), and correlations between original SST (indices 1-24) and SSH variables (indices 25-48), and SSH T-MAFs 1-6 (in red, green, blue, cyan, magenta and yellow, respectively).

Correlations between SST T-MAF2 and the 24 months of SST data (in green) exhibit strong annual oscillations [5] with positive correlations in the Northern Hemisphere's winter months and negative correlations in the summer months. The correlations with the 24 months of SSH are also high, but note that the extreme correlations occur one month later for the SSH than for the SST (e.g.

, minimum in month 8 and 20 for SST, but month 9 and 21 for SSH). The spatial structure of SST T-MAF2 shown in Figure 1 shows highs in the Southern Hemisphere and lows in the Northern Hemisphere and corresponds to an annual heating and cooling of the ocean. The apparent phase lag of one month between the highs of the SST and the SSH is explained by the fact that the SST represent the instant temperature of the sea surface, whereas the steric expansion causing the sea level to rise is a more integrated effect.

For both the SST T-MAF3 and the SST T-MAF4 (blue and cyan curves) the correlations also show some oscillations, but not on annual periods. This correlation pattern is furthermore dramatically disturbed in the second half of 1997.

This dramatic change in correlation is in accordance with the El Niño build-up during 1997. The connection to the El Nino is also captured by the corresponding spatial structure of SST T-MAF3. This exhibits a conspicuous high in the equatorial Pacific off the South American west coast. The spatial structure of the SST T-MAF4 is also El Nino related in the tropical Pacific, but some very interesting signals in the south Atlantic Ocean are also found.

The first seven to nine SSH T-MAFs (again especially T-MAFs 1-4) are highly autocorrelated temporally. SSH T-MAF1 is nearly uncorrelated with the SST, but highly (positively) correlated with the SSH data in 1996 and well into the first half of 1997. During the spring of 1997 this correlation drops to around zero where it stays for the rest of 1997. The corresponding spatial pattern of SSH T-MAF1 shows the warm and therefore high SSH "waters normally residing in the western part of the Pacific" mentioned in the introduction. By and large, SSH T-MAF2 exhibits an opposite correlation pattern: going from small oscillations close to zero to being very high in the second half of 1997. Accordingly, the spatial pattern of SSH T-MAF2 shows very high values in the equatorial Pacific off the South American west coast and also highs along the Mexican and south west U. S. coasts. This, again, complies with the El Niño phenomenon.

Generally all correlation structures between SSH T-MAFs and the 24 months of SSH data show large differences between 1996 and 1997. For T-MAF3 amplitudes are high and the annual oscillation is very conspicuous in 1996 with much smaller correlation values and oscillations in 1997. The spatial pattern of SSH T-MAF3 may show El Niño related edge effects with lows immediately to the north of the Equator and highs immediately to the south of the Equator in the Pacific.

For a joint analysis of SST and SSH data by means of canonical correlations analysis (CCA), see [8]. For a non-linear CCA of the same data, see [3].

In spite of the very short time span of the data and the associated risk of over-interpreting the results, simultaneous inspection of spatial patterns of the T-MAFs and the correlations between the original and transformed variables from the analysis gives good indications of a large anomaly in both the SST and SSH fields off the South American west coast taking place in the second half of 1997. This is in good agreement with established oceanographic knowledge on the build-up of one of the largest El Niño events on record.

Future analysis should include simultaneous maximisation of temporal and spatial autocorrelation. Also analysis of longer time series should be carried out in order to establish whether 1997 (and 1998) represent anomalous events in terms of global SST and SSH.

Acknowledgement

The Pathfinder SST data provision at JPL is due to J. Vazques, R. Sumagaysay and co-workers. The Pathfinder SSH data provision at GSFC is due to V. Zlotnicki and co-workers and the authors would like to thank B. Beckley for providing the

data.. This work is done as a part of the GEOSONAR project funded by the Danish National Research Councils under the Earth Observation Programme. GEOSONAR is headed by Dr. Per Knudsen, the National Survey and Cadastre, Denmark.

References

1. Anderson T. W., *An Introduction to Multivariate Statistical Analysis*, second edition, John Wiley, 1984.
2. Green A. A., Berman M., Switzer P. and Craig M. D., A transformation for ordering multispectral data in terms of image quality with implications for noise removal, *IEEE Transactions on Geoscience and Remote Sensing*, 26, 65-74, 1988.
3. Hilger K. B., Nielsen A. A., Andersen O. A. and Knudsen P., An ACE-based nonlinear extension to traditional empirical orthogonal function analysis, *MultiTemp 2001*.
4. Hotelling H., Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology*, 24, 417-441, 1933.
5. Knudsen P., Andersen O.B. and Knudsen T., ATSR sea surface temperature data in a global analysis with TOPEX/Poseidon altimetry, *Geophysical Research Letters,* 23(8), 821-824, 1996.
6. NASA Facts, *El Niño*, The Earth Science Enterprise Series, NF-211, 1999.
7. NASA/GSCF, http://neptune.gsfc.nasa.gov/ocean.html, Goddard Space Flight Center, National Aeronautics and Space Administration, Greenbelt, Maryland, USA.
8. Nielsen A. A., Hilger K. B., Andersen, O. A. and Knudsen P., A bivariate extension to traditional empirical orthogonal function analysis, *MultiTemp 2001*.
9. PO.DAAC, http://podaac.jpl.nasa.gov, Jet Propulsion Laboratory, National Aeronautics and Space Administration, Pasadena, California, USA.
10. Preisendorfer R. W., *Principal Component Analysis in Meteorology and Oceanography*, posthumously compiled and edited by C. D. Mobley. Developments in Atmospheric Science, 17, Elsevier, 1988.
11. Switzer P. and Green A. A., Min/max autocorrelation factors for multivariate spatial imagery, Technical Report, Department of Statistics, Stanford University, 1984.