A change oriented extension of EOF analysis applied to the 1996–1997 AVHRR sea surface temperature data

Allan Aasbjerg Nielsen [a,*], Knut Conradsen [a], Ole Baltazar Andersen [b,1]

[a] *Informatics and Mathematical Modelling, Technical University of Denmark, Building 321, DK-2800 Kgs. Lyngby, Denmark*
[b] *Danish National Survey and Cadastre, Rentemestervej 8, DK-2400 Copenhagen NV, Denmark*

Abstract

This paper describes the application of orthogonal transformations to detect multivariate change in the monthly mean sea surface temperature (SST) as given by the NOAA/NASA Oceans Pathfinder data. The transforms applied include multivariate alteration detection (MAD) variates based on canonical correlation analysis, and maximum autocorrelation factors (MAFs). The method described can be considered as an extension to empirical orthogonal function analysis that is specially tailored for change detection in spatial data since it first maximises differences in the data between two points in time and then maximises autocorrelation between neighbouring observations. The results show that the large scale ocean events associated with the El Niño/Southern Oscillation related changes are concentrated in the first SST MAF/MAD mode and the two first sea surface height MAF/MAD modes. The MAD/MAF analysis also revealed a spatially correlated structure in the Western Mediterranean Sea that turned out to be related to a strong semi-annual variation in the SST for 1997 which was difficult to resolve from a traditional principal component analysis.
© 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Sea surface temperature; Sea surface height; El Niño; Change detection; Spatial autocorrelation

1. Introduction

This paper deals with detection of non-trivial change in multivariate, bi-temporal data, i.e., data from two points in time. The term ''non-trivial change'' here means non-affine change between two points in time, i.e., change due to an additive shift in mean level (offset) or a multiplicative change is not detected.

The method applied, which is called multivariate alteration detection (MAD) (Nielsen, 1994, 1999; Nielsen and Conradsen, 1997; Nielsen et al., 1998), is based on the established multivariate statistical technique canonical correlation analysis (CCA) (Hotelling, 1936; Cooley and Lohnes, 1971; Anderson, 1984) and post-processing by the maximum autocorrelation factors (MAF) transformation (Switzer and Green, 1984; Switzer and Ingebritsen, 1986; Green et al., 1988). Here, the method is applied to detect change in the 1996–1997 NOAA/

NASA AVHRR Oceans Pathfinder sea surface temperature (SST). The data from 1996 are considered as 12 variables from one point in time and the data from 1997 are considered as 12 variables from another point in time.

Analysis of multitemporal oceanographic data is often performed by means of principal component (PC) analysis or empirical orthogonal functions (EOFs) as described in (Preisendorfer, 1988). Since the method applied here first maximises differences in the data between two points in time and then maximises autocorrelation between neighbouring observations the analysis carried out can be considered as a spatial change or difference focused extension to the usual PC or EOF analysis.

PC analysis concentrates information on different uncorrelated modes with high variance in the time-varying signal analysed whereas the combined MAF/MAD analysis finds modes of high degree of change between two points in time with high spatial autocorrelation. As opposed to the traditional PC analysis the applied MAF/MAD analysis thus concentrates specifically on finding regions with a high degree of change and high autocorrelation between neighbouring observations simultaneously.

---

[*] Corresponding author. Tel.: +45-4525-3425; fax: +45-4588-1397.

E-mail addresses: aa@imm.dtu.dk (A.A. Nielsen), oa@kms.dk (O.B. Andersen).

URLs: http://www.imm.dtu.dk/~aa, http://www.research.kms.dk/~oa.

[1] Tel.: +45-3587-5255; fax: +45-3587-5052.

The drawback of PC or EOF is mostly related to the lack of physical basis for the decomposition potentially making interpretation difficult. However, it can also be considered as an advantage that there is no prejudice as to what results to expect. The MAF/MAD transformation implicitly assumes that high spatial autocorrelation is useful. This is a very relevant assumption since the monthly mean SST is dominated by atmospheric scale signals (Leeuwenburgh, 2001).

It must be noted, however, that the regions of largest variance in SST (captured by the PC analysis) are normally also associated with spatially correlated signals on regional and basin-wide scales. Consequently, the test case will yield roughly similar results when investigating basin-wide structures like El Niño/Southern Oscillation (ENSO).

2. Methods

When analysing changes in images taken at different points in time it is customary to analyse the difference between two images, possibly after some normalisation. The idea is of course that areas with no or little change have zero or low absolute values and areas with large changes have large absolute values in the difference image. Below vectors are written in bold, scalars are not. If $X(r)$ is a multivariate observation at location r at one point in time and $Y(r)$ is an observation of the same variables at the same location r at another point in time

$$X(r) = [X_1(r) \cdots X_k(r)]^T \quad \text{and}$$

$$Y(r) = [Y_1(r) \cdots Y_k(r)]^T \quad (1)$$

(without loss of generality we assume that $E\{X(r)\} = E\{Y(r)\} = 0$) where k is the number of variables, then a simple change detection transformation is

$$X - Y = [X_1 - Y_1 \cdots X_k - Y_k]^T \quad (2)$$

(for ease of notation we drop the location index r).

To concentrate information on change, linear transformations of the data that optimise some difference criterion can be applied. Consider linear combinations

$$a^T X = a_1 X_1 + \cdots + a_p X_p \quad (3)$$

$$b^T Y = b_1 Y_1 + \cdots + b_q Y_q \quad (4)$$

($p \leqslant q$) and the difference between them $a^T X - b^T Y$. This measure in principle also accounts for situations where the variables at the two points in time are not the same or equal in number. In this case one must be cautious when interpreting the multivariate difference as multivariate change.

To maximise a measure of change let us maximise the variance of $a^T X - b^T Y$. A multiplication of a and b with a constant c will multiply the variance with c^2. Therefore

we must make choices concerning a and b, and natural choices are requesting unit variance of $a^T X$ and $b^T Y$. The criterion then is: maximise $\text{Var}\{a^T X - b^T Y\}$ with $\text{Var}\{a^T X\} = \text{Var}\{b^T Y\} = 1$. With this choice we have

$$\begin{aligned} \text{Var}\{a^T X - b^T Y\} &= \text{Var}\{a^T X\} + \text{Var}\{b^T Y\} \\ &\quad - 2\text{Cov}\{a^T X, b^T Y\} \\ &= 2(1 - \text{Corr}\{a^T X, b^T Y\}) \end{aligned} \quad (5)$$

We shall request that $a^T X$ and $b^T Y$ are positively correlated. Therefore, determining the difference between linear combinations with maximum variance corresponds to determining linear combinations with minimum (non-negative) correlation. Determination of linear combinations with extreme correlations brings the theory of CCA to mind.

2.1. Canonical correlations analysis

CCA investigates the relationship between two groups of variables. It finds two sets of linear combinations of the original variables, one for each group. The first two linear combinations are the ones with the largest correlation. This correlation is called the first canonical correlation and the two linear combinations are called the first canonical variates. The second two linear combinations are the ones with the largest correlation subject to the condition that they are orthogonal to the first canonical variates. This correlation is called the second canonical correlation and the two linear combinations are called the second canonical variates. Higher order canonical correlations and canonical variates are defined similarly.

This technique was first described in (Hotelling, 1936) and a treatment is given in most textbooks on multivariate statistics; good references are (Cooley and Lohnes, 1971; Anderson, 1984). For a description of CCA extended to more than two groups of data, see (Kettenring, 1971; Nielsen, 1994; Nielsen, 2002).

2.2. The MAD transformation

In accordance with the above we define the MAD transformation as

$$\begin{bmatrix} X \\ Y \end{bmatrix} \rightarrow \begin{bmatrix} a_p^T X - b_p^T Y \\ \vdots \\ a_1^T X - b_1^T Y \end{bmatrix} \quad (6)$$

where a_i, and b_i are the defining coefficients from a standard CCA. X and Y and are vectors with $E\{X\} = E\{Y\} = 0$. The dispersion matrix of the MAD variates is

$$D\{a^T X - b^T Y\} = 2(I - R) \quad (7)$$

where I is the $p \times p$ unit matrix and R is a $p \times p$ matrix containing the sorted canonical correlations on the diagonal and zeros off the diagonal.

The MAD transformation has the very important property that if we consider linear combinations of two sets of p respectively q ($p \leqslant q$) variables that are positively correlated then the pth difference shows maximum variance among such variables. The $(p - j)$th difference shows maximum variance subject to the constraint that this difference is uncorrelated with the previous j ones. In this way we may sequentially extract uncorrelated difference images where each new image shows maximum difference (change) under the constraint of being uncorrelated with the previous ones. The MAD variates are invariant to affine transformations (including linear scaling), which means that they are not sensitive to for example offset in the data and to linear radiometric and atmospheric correction schemes.

2.3. The MAF transformation

To find maximum change areas with high spatial autocorrelation a MAF post-processing of the MAD variates is suggested. The MAF transformation can be considered as a spatial extension of PC analysis in which the new variates maximise autocorrelation between neighbouring pixels rather than variance (as with PCs). Also the MAF transformation is invariant to affine transformations (including linear scaling). The MAF transformation is described in (Switzer and Green, 1984; Switzer and Ingebritsen, 1986; Green et al., 1988).

3. Data

The data used are global monthly mean values of 1996–1997 SST data from the NOAA/NASA Oceans Pathfinder AVHRR SST database.

Global 1996–1997 monthly mean values of sea surface height (SSH) anomalies are from the NASA/GSFC Ocean Altimeter Pathfinder database. These are interpolated TOPEX/Poseidon radar altimeter point observations.

In the analysis of these data statistics for the transformations applied are calculated only where non-missing values are available for all 24 months.
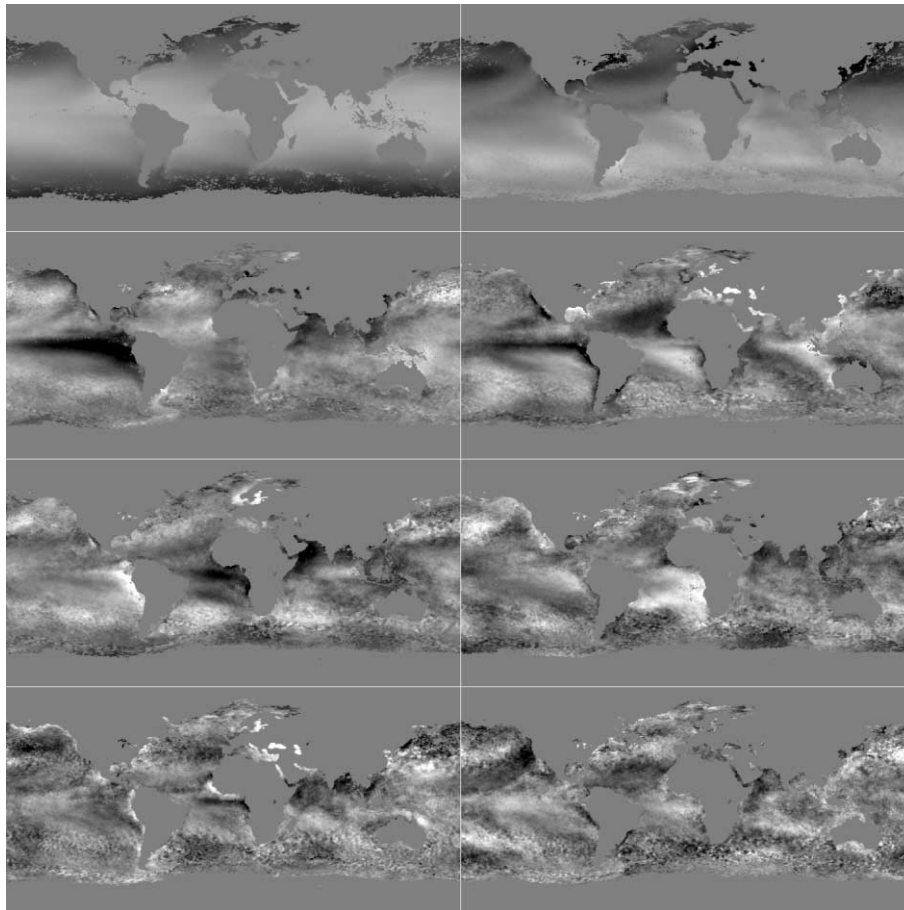

Fig. 1. Eight first PCs of 1996–1997 AVHRR monthly mean SST.

4. Results

Regional and global scales SST analysis, using EOF and related orthogonal decompositions, have been described in numerous studies, e.g. (Weare et al., 1976; Hsiung and Newell, 1983; Hu et al., 1994). Also, EOF analysis using two years of T/P SSH data have been presented by (Hendricks et al., 1996). Coupled pattern analysis between SST and SSH have been presented by (Leuliette and Wahr, 1999; Nielsen et al., 2001; Hilger et al., 2001).

Fig. 1 shows the first eight of 24 SST PC modes. PC 1 clearly corresponds to the mean value of the SST being highest at the Equator. The correlation with each monthly mean SST varies around 0.95 (Fig. 2) with a semi-annual period. The second PC mode corresponds to the annual variation in sea level. PC modes 3 and 4 represent the largest inter-annual variations mostly associated with the ENSO event in the Pacific Ocean.

The following PC modes 5–7 are relatively difficult to interpret. The correlations of the PC 5–7 all show a clear semi-annual pattern but the spatial structure in the Pacific Ocean resembles the structure of the Intertropical Convergence Zone and the South Pacific Convergence Zone. The MAF/MAD analysis does a much better job in separating these signals. The remaining PC modes 9–24 are not shown here as they are noisy with small correlations to the original monthly mean SST.

Fig. 3 shows the first eight of 12 SST MAF/MAD modes. Fig. 4 shows the correlations between the the monthly mean SST and the first eight MAF/MADs. These correlations are calculated only where MAF/MAD mode 1 has values greater than 1.5 or smaller than −1.5. In the analysis the 1996 data are considered as 12 variables (X in Eq. (6)) and the 1997 data are considered as 12 variables (Y in Eq. (6)).

By considering the 1996 and 1997 monthly mean SST as two sets of 12 variables any identical signals in the two years get suppressed. Consequently, the annual signal, which corresponds to the largest variation in SST will get suppressed along with the semi-annual signal. Only changes in the annual and semi-annual signal will appear in the MAF/MADs along with other inter-annual changes such as the ENSO event. Transforming the MADs with MAF creates change variables which have maximum autocorrelation between neighbouring pixels. Features of large spatial autocorrelation like the ENSO signal will consequently concentrate in the first MAF/MADs as seen in Fig. 3.
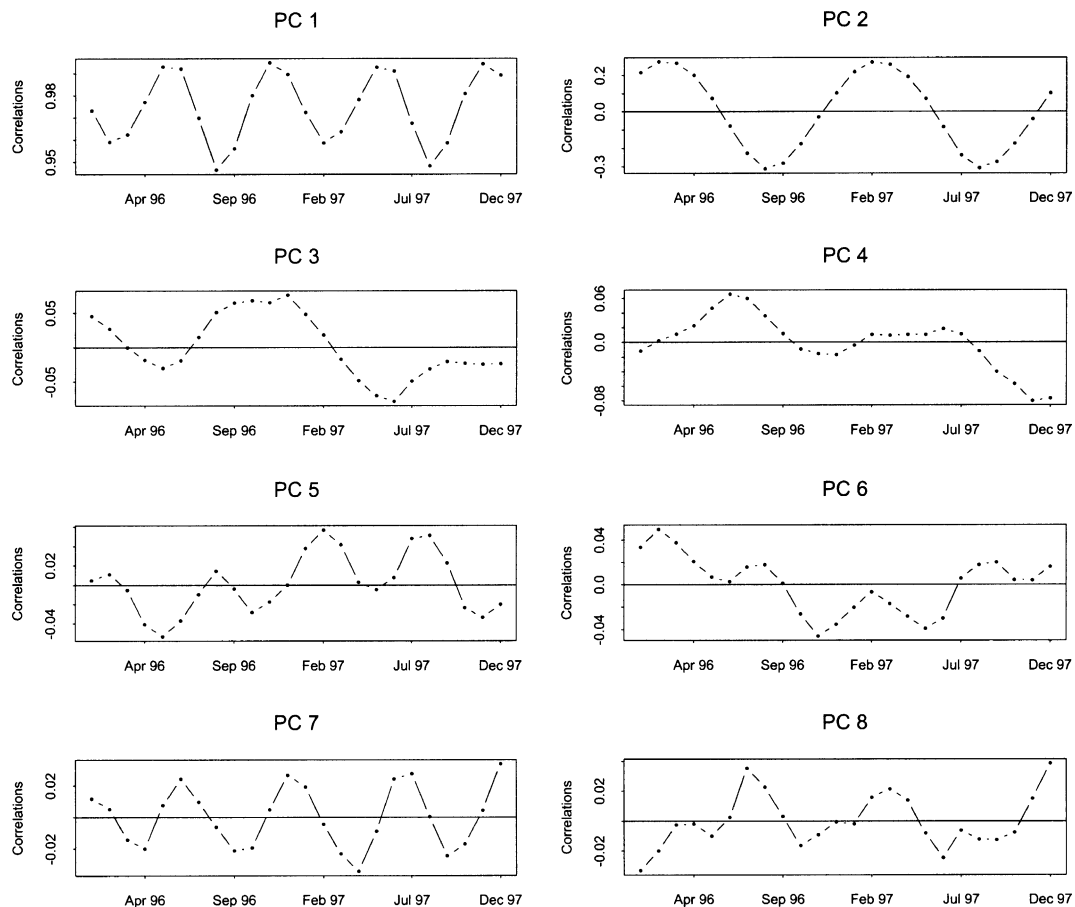

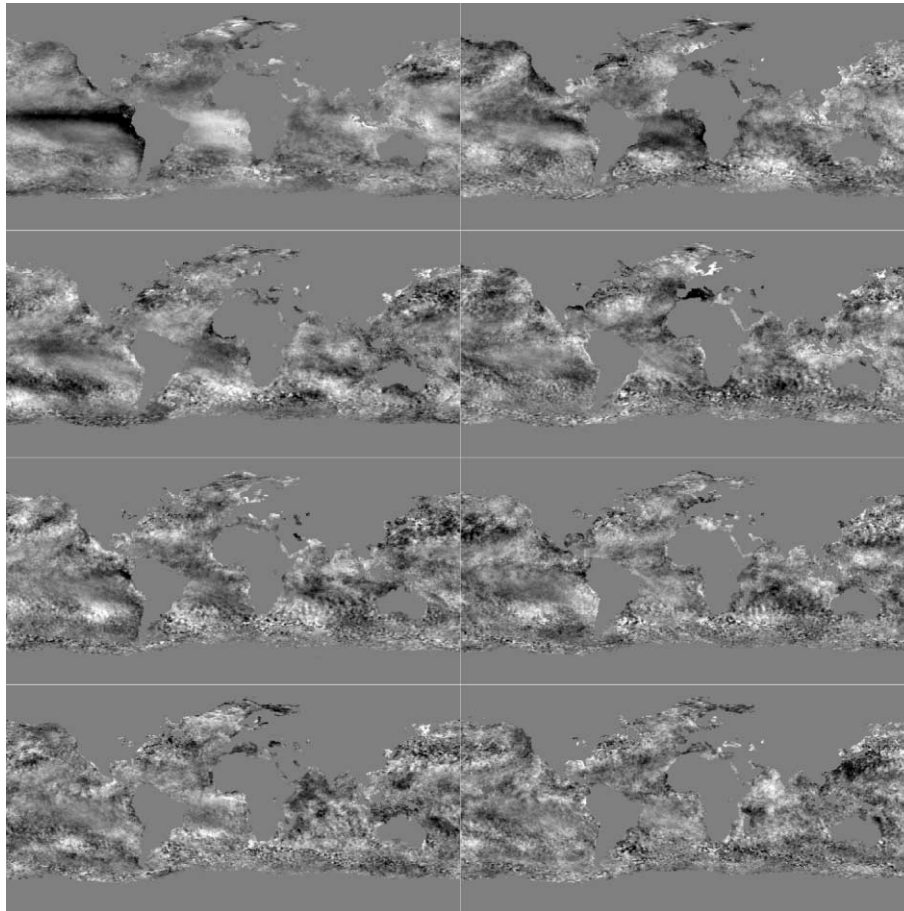Fig. 2. Correlations between 1996–1997 AVHRR monthly mean SST and PCs 1–8.

Fig. 3. Eight first MAF/MADs of 1996–1997 AVHRR monthly mean SST.

The correlations between the first MAF/MAD mode and the monthly mean SST (Fig. 4) is clearly dominated by the the El Niño event in the Pacific Ocean. In the North Atlantic Ocean the MAF/MAD modes 1 and 2 both exhibit a structure similar to that observed by Andersen et al. (2002).

Most other MAF/MADs are seen to have spatial structure in the Pacific that resembles the structure of the Intertropical Convergence Zone and the South Pacific Convergence Zone, and consequently being mostly related with ENSO dynamics. MAF/MADs 9–12 are not shown here as they are noisy with small correlations to the original monthly mean SST.

A closer investigation is carried out below, for the Central Pacific Ocean and the Western Mediterranean Sea to demonstrate the differences and similarities between the PC and the MAF/MAD analyses of monthly mean SST data.

4.1. El Niño/Southern Oscillation

In 1997, one of the largest ENSO events on record in the Pacific Ocean occurred, with the onset in early 1997 leading to a peak in both SST and sea surface height in the eastern Pacific Ocean towards the end of 1997, following the characterisation in (Rasmusson and Carpenter, 1982). The features associated with the ENSO event are very large basin-wide coherent structures, and as such exhibiting large autocorrelation. The MAF/MAD analysis therefore concentrates these in the lowest order MAF/MAD modes as seen in Fig. 3. On the other hand the ENSO event is also associated with the largest variance in the SST, and will also be clearly isolated using a PC or EOF analysis as well. Referring to Fig. 1 (PC) and Fig. 3 (MAF/MAD) there is an obvious similarity between PC 3 and PC 4 and, MAF/MAD 1 and 2. However, the MAF/MAD analysis focuses the ENSO related events in the lowest order MAF/MADs.

In Fig. 5 an El Niño mask has been introduced and the correlations with the original monthly mean SST are calculated under this mask. The mask is limited by 15°S–15°N and 150–280°E. Introducing such a region increases the correlations between the MAF/MADs and the monthly mean SST dramatically. The shape of the correlation curves for MAF/MADs 1 and 2 in Fig. 4 are
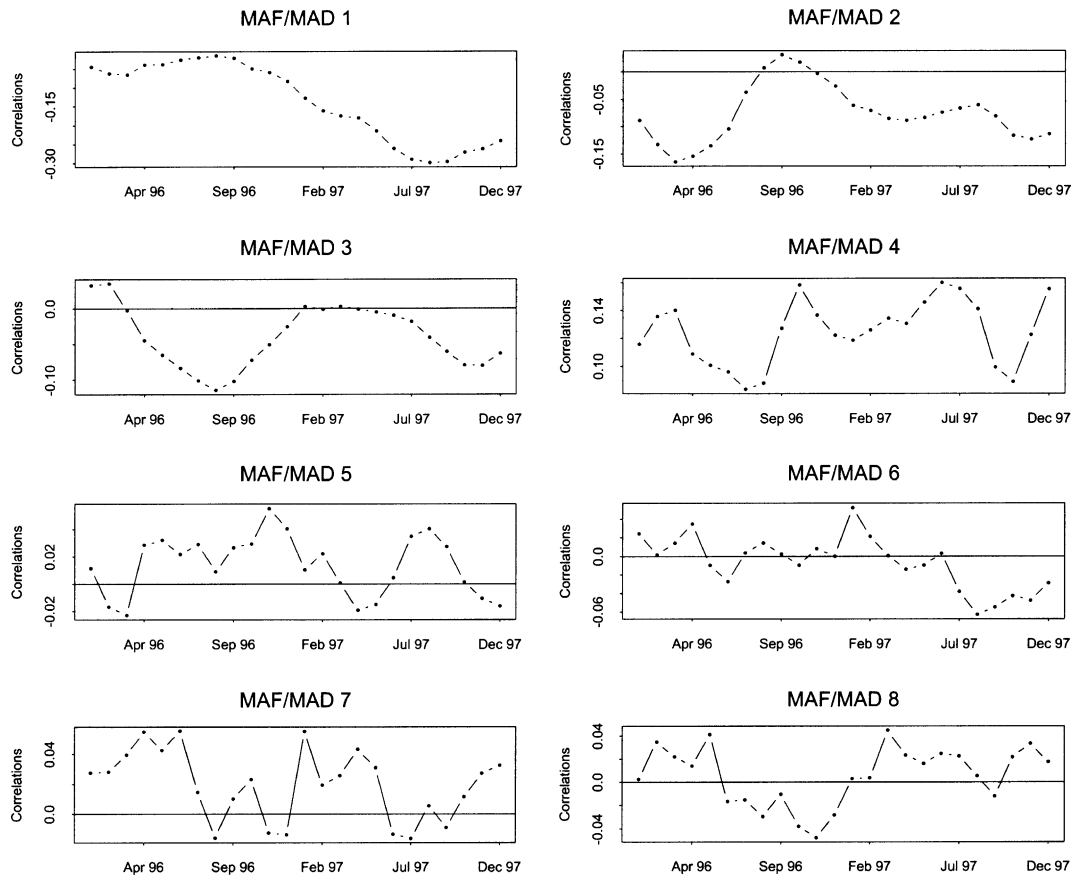
Fig. 4. Correlations between 1996–1997 AVHRR monthly mean SST and MAF/MADs 1–8 (calculated only where the absolute value of MAF/MAD 1 is greater than 1.5).
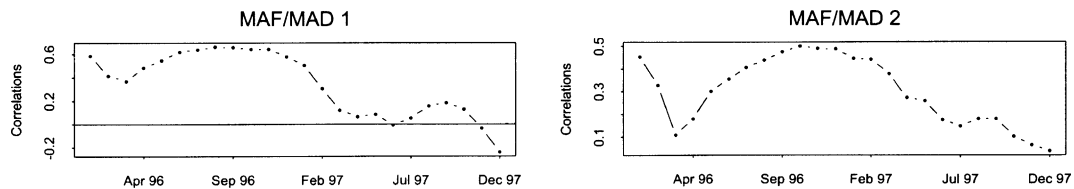

Fig. 5. Correlations between 1996–1997 AVHRR monthly mean SST and MAF/MADs 1–2 calculated under El Niño mask.

similar those in Fig. 5 showing how this region dominates the global correlation pattern.

In the east, months after the initial relaxation of the trade winds, the thermocline deepens in the eastern part of the Pacific. This causes the SST to increase in the eastern part while remaining virtually constant in the western part of the Pacific. On the other hand, the SSH, which measures the integrated effect of heating of the entire water column, increases dramatically in the eastern parts of the Pacific Ocean and decreases in the western part of the Pacific. Detailed comparison with the TOPEX/Poseidon monthly mean SSH effect is presented in Figs. 6 and 7 where the monthly mean SSH anomalies have been subjected to the same MAF/MAD analysis as the SST data. The correlations in Fig. 7 have

been calculated under the El Niño mask as defined above. Whereas MAF/MAD 1 shows the prolonged signal in the Central Equatorial Pacific, that is a precursor to the El Niño event (Hendricks et al., 1996), the MAF/MAD 2 shows a strong east–west oscillation that is characteristic of the ENSO event. The correlation plots in Fig. 7 reveal the timing offset between these modes. The SSH MAF/MAD mode 1 decreases from around October/November 1996, the increase in sea level in the eastern part of the Pacific (MAF/MAD 2) is not initiated until March 1997. Closer inspection of the MAF/MAD of the T/P SSH reveals track-like structures associated with the location of the T/P ground tracks. This "trackiness" should be of the order of a few centimetres but is enhanced in the current investigation
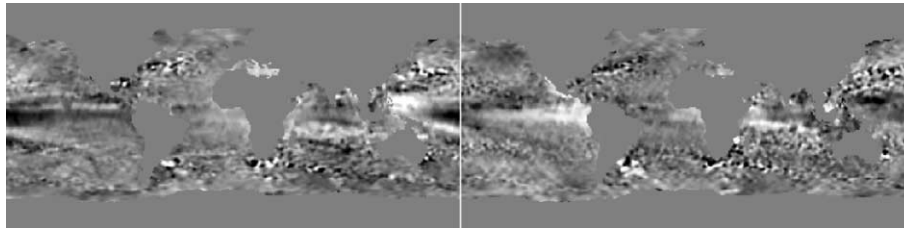
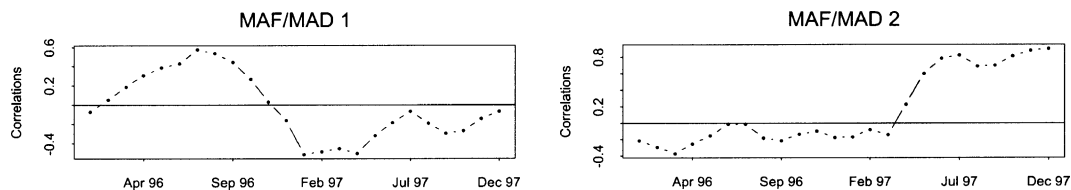Fig. 6. Two first MAF/MADs of 1996–1997 T/P monthly mean SSH.

Fig. 7. Correlations between 1996–1997 T/P monthly mean SSH and MAF/MADs 1–2 calculated under El Niño mask.
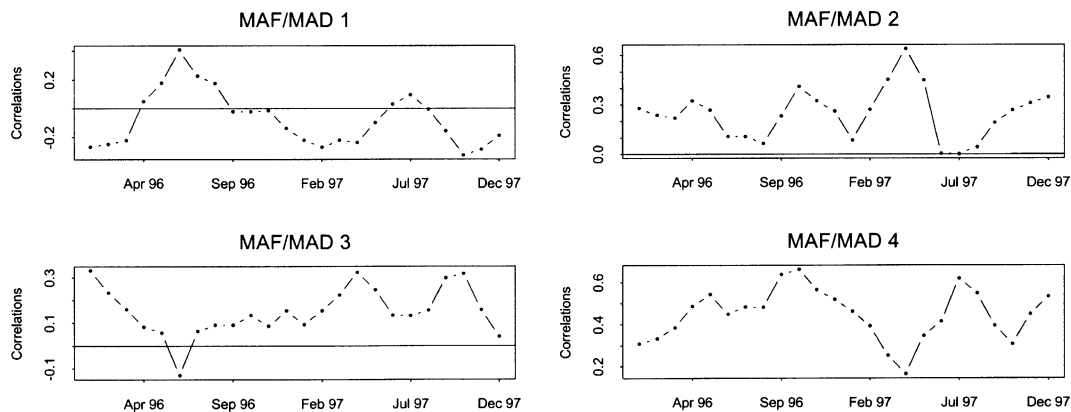

Fig. 8. Correlations between 1996–1997 AVHRR monthly mean SST and MAF/MADs 1–4 calculated under Western Mediterranean mask.

because the differences between T/P observations from 1996 and 1997 are investigated.

4.2. Western Mediterranean

Inspecting the lowest order MAF/MAD modes in Fig. 3 reveals that particularly MAF/MAD 4 and to some extent also MAF/MAD 2 exhibit a clear spatially correlated signal in the Western Mediterranean Sea. No clear conclusions can be drawn from inspecting the global correlations with the monthly mean SST (Fig. 4). All the first six PC modes (Fig. 1) show some signal in the Western Mediterranean but the MAF/MAD analysis focuses this spatially correlated signal in a few components. To investigate the signal further, the correlations were calculated under a Western Mediterranean mask limited by 30–47°N and 0–15°E, and shown in Fig. 8. MAF/MAD 4 shows a semi-annual signal in 1997 peaking at a correlation above 0.6 in July. A semi-annual signal peaking at a correlation of 0.6 in April is

also seen in MAF/MAD 2 in 1997. However, for both MAF/MADs no clear semi-annual variation is seen for 1996. Comparing this with a wavelet analysis of the averaged SST time series for the Western Mediterranean Sea (not shown) leads to the conclusion, that the semi-annual signal was particularly strong in the Western Mediterranean in 1997.

5. Conclusions

By design the MAD/MAF method isolates changes with the highest autocorrelation, often the most important changes, in the lowest order modes. In the MAF/MAD method we focus on change and autocorrelation, and in the case shown we have 12 MAF/MAD modes as opposed to 24 PC or EOF modes. The proposed method seems to be superior for change detection, since it focuses on patterns of large difference and high autocorrelations as opposed to PC or EOF, which

focuses on patterns of high variance. The largest variance in SST (captured by PC analysis) is, however, often also associated with large spatially correlated signals on regional and basin-wide scales, and as such the PC and MAF/MAD methods will both capture signals like the El Niño. The current investigation also showed that this was the case, but where El Niño was concentrated in modes 1 and 2 of the MAF/MAD analysis, the PC analysis isolated this in the higher order modes 3 and 4. The investigation of the El Niño was furthermore extended to include the TOPEX/Poseidon monthly mean SSH to demonstrate how the MAF/MAD method was able to capture the prolonged precursor to the El Niño in one mode with a timing offset to the strong east–west oscillation characteristic of the ENSO event.

The PC modes 5–7 were relatively hard to interpret because there is no physical basis for the decomposition, and these all showed temporal patterns of correlation with a semi-annual period, whereas the spatial pattern more assembled ENSO like structures. This was not the case with the MAF/MAD analysis. The MAF/MAD analysis was also able to identify a spatially correlated signal in the western Mediterennean Sea which showed up to be a stronger than usual semi-annual variation in 1997.

References

Andersen, O.B., Knudsen, P., Beckley, B., 2002. Monitoring sea level and sea surface temperature trends using ESA satellites. Physics and Chemistry of the Earth 27, 1413–1417.

Anderson, T.W., 1984. An Introduction to Multivariate Statistical Analysis, second ed. John Wiley, New York.

Cooley, W.W., Lohnes, P.R., 1971. Multivariate Data Analysis. John Wiley and Sons, New York.

Green, A.A., Berman, M., Switzer, P., Craig, M.D., 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. IEEE Transactions on Geoscience and Remote Sensing 26 (1), 65–74.

Hendricks, J.R., Leven, R.R., Born, G.H., Koblinsky, C.J., 1996. Empirical orthogonal function analysis of global TOPEX/Poseidon altimeter data and implications for detection for global sea level rise. Journal of Geophysical Research 101 (C6), 14131–14145.

Hilger, K.B., Nielsen, A.A., Andersen, O.B., Knudsen, P., 2001. An ACE-based nonlinear extension to traditional empirical orthogonal function analysis. In: MultiTemp 2001, Trento, Italy.

Hotelling, H., 1936. Relations between two sets of variates. Biometrika XXVIII, 321–377.

Hsiung, J., Newell, R.E., 1983. The principal nonseasonal modes of variation of global sea surface temperature. Journal of Physical Oceanography 13, 1957–1967.

Hu, W., Newell, R.E., Wu, Z.X., 1994. Modes of variability of global sea surface temperature, free atmosphere temperature and oceanic surface energy flux. Climate Dynamics 10, 377–393.

Kettenring, J.R., 1971. Canonical analysis of several sets of variables. Biometrika 58, 433–451.

Leeuwenburgh, O., 2001. Combined analysis of sea surface height and temperature for mapping and climate studies. Ph.D. Thesis, Department of Geophysics, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark. National survey and Cadastre, Technical Report 16.

Leuliette, E.W., Wahr, J.M., 1999. Coupled pattern analysis of sea surface temperature and TOPEX/Poseidon sea surface height. Journal of Physical Oceanography 29, 599–611.

Nielsen, A.A., 1994. Analysis of regularly and irregularly sampled spatial, multivariate, multi-temporal data. Ph.D. Thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby. Available from <http://www.imm.dtu.dk/~aa/phd/.

Nielsen, A.A., 1999. Multi-channel remote sensing data and orthogonal transformations for change detection. In: Kanellopoulos, I., Wilkinson, G.G., Moons, T. (Eds.), Machine Vision and Advanced Image Processing in Remote Sensing. Springer, Berlin.

Nielsen, A.A., 2002. Multiset canonical correlations analysis and multispectral, truly multi-temporal remote sensing data. IEEE Transactions on Image Processing 11 (3), 293–305.

Nielsen, A.A., Conradsen, K., 1997. Multivariate alteration detection (MAD) in multi-spectral, bi-temporal image data: A new approach to change detection studies. Technical Report 1997-11, Department of Mathematical Modelling, Technical University of Denmark. Available from <http://www.imm.dtu.dk/~aa/tech-rep-1997-11/>.

Nielsen, A.A., Conradsen, K., Simpson, J.J., 1998. Multivariate alteration detection (MAD) and MAF post-processing in multispectral, bi-temporal image data: New approaches to change detection studies. Remote Sensing of Environment 64, 1–19.

Nielsen, A.A., Hilger, K.B., Andersen, O.B., Knudsen, P., 2001. A bivariate extension to traditional empirical orthogonal function analysis. In: MultiTemp 2001, Trento, Italy.

Preisendorfer, R.W., 1988. Principal component analysis in meteorology and oceanography. In: Mobley, C.D. (Ed.), Developments in Atmospheric Science, vol. 17. Elsevier, Amsterdam.

Rasmusson, E.M., Carpenter, T.H., 1982. Variations in tropical sea surface temperature and surface wind field associated with the Southern Ocean/El Niño. Monthly Weather Review 110, 354–384.

Switzer, P., Green, A.A., 1984. Min/max autocorrelation factors for multivariate spatial imagery. Technical Report 6, Department of Statistics, Stanford University.

Switzer, P., Ingebritsen, S.E., 1986. Ordering of time-difference data from multispectral imagery. Remote Sensing of Environment 20, 85–94.

Weare, B.C., Navato, A.R., Newell, R.E., 1976. Empirical orthogonal analysis of Pacific sea surface temperatures. Journal of Physical Oceanography 6, 671–678.