



**HAL**  
open science

## Multivariate autoregressive modelling of sea level time series from TOPEX/Poseidon satellite altimetry

S. M. Barbosa, M. E. Silva, M. J. Fernandes

► **To cite this version:**

S. M. Barbosa, M. E. Silva, M. J. Fernandes. Multivariate autoregressive modelling of sea level time series from TOPEX/Poseidon satellite altimetry. *Nonlinear Processes in Geophysics*, 2006, 13 (2), pp.177-184. hal-00302723

**HAL Id: hal-00302723**

**<https://hal.science/hal-00302723>**

Submitted on 18 Jun 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Multivariate autoregressive modelling of sea level time series from TOPEX/Poseidon satellite altimetry

S. M. Barbosa, M. E. Silva, and M. J. Fernandes

Department of Applied Mathematics, Faculty of Science, University of Porto, Portugal

Received: 16 January 2006 – Revised: 30 March 2006 – Accepted: 18 April 2006 – Published: 20 June 2006

**Abstract.** This work addresses the autoregressive modelling of sea level time series from TOPEX/Poseidon satellite altimetry mission. Datasets from remote sensing applications are typically very large and correlated both in time and space. Multivariate analysis methods are useful tools to summarise and extract information from such large space-time datasets. Multivariate autoregressive analysis is a generalisation of Principal Oscillation Pattern (POP) analysis, widely used in the geosciences for the extraction of dynamical modes by eigen-decomposition of a first order autoregressive model fitted to the multivariate dataset of observations. The extension of the POP methodology to autoregressions of higher order, although increasing the difficulties in estimation, allows one to model a larger class of complex systems. Here, sea level variability in the North Atlantic is modelled by a third order multivariate autoregressive model estimated by stepwise least squares. Eigen-decomposition of the fitted model yields physically-interpretable seasonal modes. The leading autoregressive mode is an annual oscillation and exhibits a very homogeneous spatial structure in terms of amplitude reflecting the large scale coherent behaviour of the annual pattern in the Northern hemisphere. The phase structure reflects the seesaw pattern between the western and eastern regions in the tropical North Atlantic associated with the trade winds regime. The second mode is close to a semi-annual oscillation. Multivariate autoregressive models provide a useful framework for the description of time-varying fields while enclosing a predictive potential.

## 1 Introduction

Geophysical systems often exhibit complex variability patterns over a wide range of spatial and temporal scales. Observation of the Earth by remote sensing techniques is yielding huge space-time datasets of geophysical observations. It is a challenging task to summarise and extract information from

*Correspondence to:* S. M. Barbosa  
(susana.barbosa@fc.up.pt)

such large datasets and characterise the dynamical properties of the associated geophysical time series.

This work focuses on the autoregressive modelling of time series of sea surface heights from TOPEX/Poseidon satellite altimetry mission. Sea level is a key indicator of climate change and an important observational constraint on global climate models. Furthermore, sea level changes have considerable environmental, social and economical impacts. The height of the sea surface relative to a geocentric reference ellipsoid is measured from space through radar altimetry. TOPEX/Poseidon (T/P) mission achieved an unprecedented accuracy, yielding a huge, high quality, space-time dataset of precise sea level measurements.

A stochastic space-time process  $\{X_{s,t}\}$  is defined as a collection of random variables indexed by parameters  $[s, t] \in \mathbb{R}^2$  where  $t$  indicates time and  $s$  a location. Observed values of  $\{X_{s,t}\}$  from  $m$  spatial locations ( $s=1, \dots, m$ ) at  $N$  times ( $t=1, \dots, N$ ) constitute a realisation of the stochastic process  $\{X_{s,t}\}$ . The term field is used hereafter to designate a particular realisation and is denoted by

$$X_t = [X_{1,t} X_{2,t} \dots X_{m,t}]^T \in \mathbb{R}^{m \times 1}, t = 1, \dots, N. \quad (1)$$

Statistical space-time methods play a key role in the analysis of time-varying fields. Given an observed field  $X_t$ , dimensionality reduction, spatio-temporal description and extraction of dominant variability modes are often primary goals. In the case of systems dominated by nonlinear interactions, nonlinear methods for dimensionality reduction are required, such as the Isomap procedure (Tenenbaum et al., 2000; Gámez et al., 2004), based on the replacement of Euclidean by geodesic distances, or nonlinear principal component analyses (Hsieh, 2001; Hsieh and Wu, 2002). Modelling of nonlinear systems requires methods extending the linear empirical models to a nonlinear setting such as in Kravtsov et al. (2005) and Kondrashov et al. (2005). However, linearity is assumed henceforth as a first approximation to the dynamics of the system.

Principal Component Analysis (PCA), or Empirical Orthogonal Function (EOF) analysis is a useful multivariate

analysis method for extraction of the dominant variability patterns from an observed field (Preisendorfer, 1988; Von Storch and Zwiers, 1999; Jolliffe, 2002). PCA yields dominant spatial structures, in terms of maximal variance explained and the corresponding temporal evolution for these structures; the observed field is decomposed into modes of variability, each of which is the product of a spatial pattern and a time-varying amplitude. Although principal components (PCs) are often interpretable, dynamical modes are not necessarily uncorrelated and/or yielding maximal variance. Furthermore, PCA depends on the size and shape of the spatial domain. Thus interpretation of PCs as physical/dynamical modes must be always done with caution.

Principal Oscillation Pattern (POP) analysis (Hasselmann, 1988; Von Storch et al., 1995; Von Storch and Zwiers, 1999) yields dynamical modes from a spatio-temporal dataset through the analysis of a stochastic model fitted to the observations. POP analysis assumes that the observed field has a temporal autoregressive structure of order one, i.e. follows a  $m$ -variate  $AR(1)$  model

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t \quad (2)$$

where  $\mathbf{A} \in \mathbb{R}^{m \times m}$  is the matrix of real autoregressive coefficients and  $\boldsymbol{\varepsilon}_t$  is a temporally uncorrelated noise vector with mean  $\mathbf{0}$  and covariance matrix  $\Sigma \in \mathbb{R}^{m \times m}$ . A first order autoregressive model is the discrete-equivalent of a first order ordinary differential equation and thus corresponds, depending on the autoregressive parameters, to a combination of stochastically forced relaxators and oscillators. If a first order multivariate autoregressive model provides an adequate fit to the observed field, the dynamical characteristics of the system can be empirically inferred from the fitted model, assuming the space-time characteristics of the model to be representative of the full system. The eigen-decomposition of the autoregressive model describing the temporal evolution of the field yields the dominant modes of variability from the multivariate dataset in terms of relaxation and oscillation modes. From eigen-decomposition of the matrix  $\mathbf{A}$  of autoregressive coefficients,  $\mathbf{A} = \mathbf{P}\mathbf{L}\mathbf{P}^{-1}$ , the state of the system at any time  $t$  can be expressed as

$$\mathbf{X}_t = \sum_{k=1}^m u_t^k \mathbf{p}_k \quad (3)$$

where the eigenvectors  $\mathbf{p}_k$  (the columns of matrix  $\mathbf{P}$ ) are the principal oscillation patterns (POP) or empirical normal modes, and the time series  $u_t^k$  are the coefficient time series, computed from the adjoint patterns  $\mathbf{p}_k^a$  (eigenvectors of  $\mathbf{A}^T$ ) as  $u_t^k = \mathbf{p}_k^{aT} \mathbf{X}_t$ . In the absence of noise the eigenvectors of matrix  $\mathbf{A}$  are the system normal modes, representing the natural modes of variability of the evolving field. Complex eigenvalues  $\lambda_j = r_j e^{i w_j}$  are associated under stationary conditions ( $|\lambda_j| \leq 1$ ,  $j = 1, \dots, m$ ) to damped oscillations with characteristic damping rate  $r_j$  and frequency  $w_j$  while real eigenvalues describe damped, non-oscillatory, patterns.

POP analysis yields a space-time description of dominant variability modes and can also be used for prediction, but the applicability of POP analysis is restricted to fields for which an  $AR(1)$  model provides an adequate fit to the observations. However, the methodology can be generalised, allowing to model a larger class of systems, by extending the analysis to autoregressive models of arbitrary order  $p$ ,  $AR(p)$  (Lütkepohl, 1993; Neumaier and Schneider, 2001). Multivariate autoregressive models ( $m-AR(p)$ ), also called vector autoregressive models (VAR), are the most used models for multiple time series and are being increasingly used in geophysical applications (e.g. Maharaj and Wheeler, 2005; Rashid and Simmonds, 2005).

In this study sea level variability in the North Atlantic is examined through multivariate autoregressive modelling. The paper is organised as follows. Multivariate autoregressive analysis is summarised in Sect. 2. The particular application of multivariate autoregressive modelling to the analysis of sea level observations from satellite altimetry is described in Sect. 3. Concluding remarks are presented in Sect. 4.

## 2 Multivariate autoregressive analysis

### 2.1 Multivariate autoregressive modelling

A  $m$ -variate autoregressive model of order  $p$  ( $m-AR(p)$ ) is defined as

$$\mathbf{X}_t = \mathbf{A}_1 \mathbf{X}_{t-1} + \mathbf{A}_2 \mathbf{X}_{t-2} + \dots + \mathbf{A}_p \mathbf{X}_{t-p} + \boldsymbol{\varepsilon}_t \quad (4)$$

where  $\mathbf{A}_i \in \mathbb{R}^{m \times m}$   $i = 1 \dots p$  are the matrices of autoregressive coefficients and  $\boldsymbol{\varepsilon}_t \in \mathbb{R}^m$  is a temporally uncorrelated noise vector with mean  $\mathbf{0}$  and covariance matrix  $\Sigma \in \mathbb{R}^{m \times m}$ .

The fit of a  $m-AR(p)$  model to a time-varying field involves the selection of the model order  $p$  and the estimation of the model parameters  $\mathbf{A}_1, \mathbf{A}_2 \dots \mathbf{A}_p$ ,  $\Sigma$  from a spatio-temporal dataset. The model order can be selected by minimising an order selection criterion such as the Schwartz Bayesian Criterion (SBC) (Schwarz, 1978) reflecting the trade-off between over-fitting and over-simplification. The SBC is superior in terms of consistency to the Akaike Information Criterion (AIC) and the Final Prediction Error Criterion (FPE) as discussed in Lütkepohl (1993). A  $m-AR(p)$  model can be cast in the form of a regression model and the parameters  $\mathbf{A}_1, \mathbf{A}_2 \dots \mathbf{A}_p$  estimated by least squares (Lütkepohl, 1993, chapter 3). The parameters for a model with order  $p$  optimising an order selection criterion such as SBC can be estimated in a computationally efficient way through stepwise least squares i.e. by stepwise down-dating the least squares estimates from a QR factorisation of the data matrix for a model with a pre-set maximum order (Neumaier and Schneider, 2001). Although the least squares approach does not have many of the desirable properties of an exact maximum likelihood algorithm, such as allowance for

missing values, the least squares estimator for a  $m-AR(p)$  model is consistent and asymptotically normal (Lütkepohl, 1993) and is recommended, provided that the model to be fitted is stable (Schneider and Griffies, 1999). Difficulties in the estimation of a  $m-AR(p)$  are often associated with the large number of parameters involved. For large fields the number of spatial degrees of freedom can be reduced by considering the leading components from a PCA analysis thereby reducing dimensionality while retaining most of the variance in the original field. Although PCA is a multivariate analysis method for independent observations, and thus should not be used for time series, non-independence does not have a serious effect when the main objective of the analysis is descriptive rather than inferential (Jolliffe, 2002). A positive by-product of estimating a multivariate autoregressive model in PCA space is the exclusion of noisy components from the analysis and diagonalisation of the error covariance matrix.

### 2.2 Multivariate autoregressive modes

A  $m-AR(p)$  model can be written as a first order model  $m-AR(1)$ ,

$$\tilde{\mathbf{X}}_t = \tilde{\mathbf{A}}\tilde{\mathbf{X}}_{t-1} + \tilde{\boldsymbol{\varepsilon}}_t \quad (5)$$

where  $\tilde{\mathbf{X}}_t = [\mathbf{X}_t, \mathbf{X}_{t-1} \dots \mathbf{X}_{t-p+1}]^T \in \mathbb{R}^{m \times p}$ ,  $\tilde{\boldsymbol{\varepsilon}}_t = [\boldsymbol{\varepsilon}_t \ 0 \dots 0]^T$  with covariance matrix  $\tilde{\boldsymbol{\Sigma}} = \begin{bmatrix} C & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{mp \times mp}$  and  $\tilde{\mathbf{A}} \in \mathbb{R}^{mp \times mp}$  is the augmented covariance matrix given by

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & & \mathbf{A}_p \\ \mathbf{I} & 0 & & 0 \\ 0 & \mathbf{I} & & \\ \vdots & & \ddots & \vdots \\ 0 & 0 & & 0 \end{bmatrix}. \quad (6)$$

Since a  $m-AR(p)$  model can be formulated as a first order autoregressive model, higher order multivariate autoregressive models can be used, analogously to POP analysis, to infer the dynamical characteristics of a system through decomposition into eigenmodes with characteristic oscillation periods and damping times.

Eigen-decomposition of the augmented coefficient matrix  $\tilde{\mathbf{A}}$  as  $\tilde{\mathbf{A}} = \tilde{\mathbf{P}}\mathbf{L}\tilde{\mathbf{P}}^{-1}$ , where  $\tilde{\mathbf{P}} = [\tilde{\mathbf{p}}_1 \tilde{\mathbf{p}}_2 \dots \tilde{\mathbf{p}}_{mp}]$  is the matrix of eigenvectors and  $\mathbf{L}$  is the diagonal matrix of the eigenvalues  $\lambda_k, k=1, \dots, mp$  yields  $mp$  m-dimensional eigenmodes (Neumaier and Schneider, 2001). Complex eigenvalues  $\lambda_j = r_j e^{iw_j}$  correspond to damped oscillatory modes with period  $T_j = 2\pi/w_j$  and characteristic damping rate  $r_j$  or e-folding time  $-1/\ln(r_j)$  (the time interval in which a exponentially decaying quantity decreases to  $1/e$  of its previous value). Real eigenvalues correspond to damped, non-oscillatory modes.

### 3 Multivariate autoregressive analysis of sea level series

Multivariate autoregressive analysis is illustrated for the time series of sea level observations from TOPEX altimeter.

#### 3.1 Data

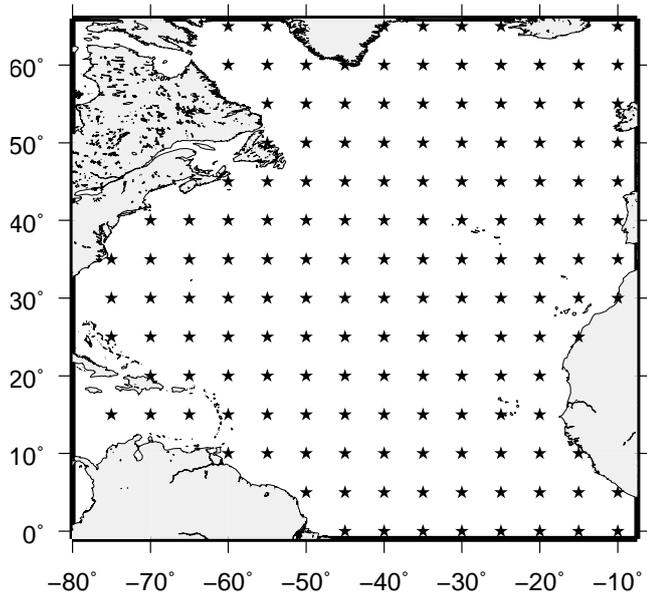
The analysed altimetry data are measurements from TOPEX altimeter for the North Atlantic. The dataset covers nearly 12 years from September 1992 to March 2005 (cycles 1 to 460) at approximately 10-days intervals (9.9156 days, corresponding to the satellite repeat period). Corrected sea surface heights are derived from the Merged Geophysical Data Records (MGDR) products (AVISO, 1996) by applying standard instrumental and geophysical corrections and editing procedures, including the Inverse Barometer (IB) correction (Dorandeu and Le Traon, 1999), the smoothing of TOPEX dual frequency ionospheric correction (Fernandes and Antunes, 2003), the cycle dependent drift effect in the wet tropospheric correction derived from the onboard TOPEX Microwave Radiometer (Scharroo et al., 2004), the sea state bias (SSB) correction (Chambers et al., 2003) and a residual SSB correction of  $-3$  mm applied to cycles 236 and greater (Berwin, 2003). Tides are removed using the NAO99b model (Matsumoto et al., 2000). Sea level anomalies are derived by subtracting to corrected sea surface heights the GSFC00.1 mean sea surface model (Wang, 2001).

For each cycle, a regular 5 degree grid of altimetry data is obtained from the along-track measurements. In order to avoid spatial aliasing and eliminate redundant data spatial blocks with size equal to the grid spacing are considered and the along-track observations within every non-empty block are replaced by a median position and value. Regular grids of sea level anomalies are then obtained using the adjustable tension continuous curvature surface gridding algorithm of Smith and Wessel (1990), with a tension value of 0.35, which is the value suggested by experience for topography data. The resulting dataset comprises 153 time series of sea level anomalies at the grid nodes (Fig. 1) at approximately 10-days intervals ( $m=153$  locations,  $N=460$  observations). The typical standard deviation for the time series of sea level anomalies is below 5 cm.

#### 3.2 Multivariate autoregressive modelling

Multivariate autoregressive modelling of the sea level field is carried out in three steps: i) data preprocessing for error and dimensionality reduction, ii) model estimation and order selection and iii) computation of autoregressive modes from eigen-decomposition of the estimated matrix of coefficients.

The R software environment (R Development Core Team, 2005) has been used in the analysis. The R-package mAr for multivariate Autoregressive analysis has been developed, implementing in R the stepwise least squares estimation for a multivariate autoregressive model based on the algorithm of



**Fig. 1.** Study area (North Atlantic) and time series locations (\*).

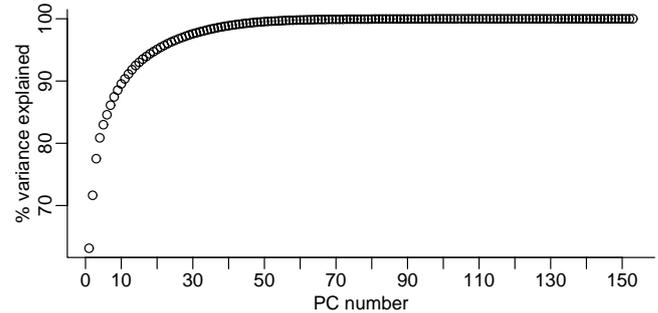
Neumaier and Schneider (2001) and the code of Schneider and Neumaier (2001). The mAr package has been released under GPL (General Public License) Version 2 and is available from the CRAN repository (<http://www.cran.r-project.org>).

### 3.2.1 Data preprocessing

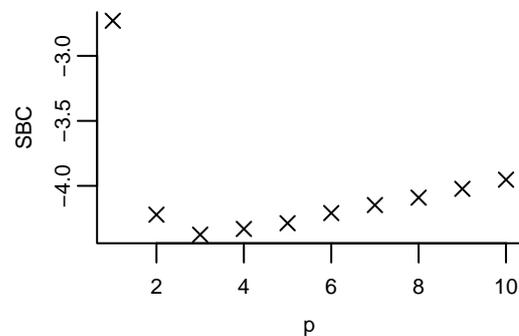
Altimeter records are short and include substantial high-frequency variability associated with mesoscale circulation. Denoising based on the discrete wavelet transform is an efficient way to remove high frequency variability from the satellite altimetry time series while preserving non-smooth features (Barbosa et al., 2005). Variability at scales shorter than 20 days is reduced through nonlinear thresholding in the wavelet domain using the universal level-dependent threshold rule (Donoho and Johnstone, 1995) yielding an overall reduction in variability of sea level anomalies of 1 cm.

The gridded dataset of denoised sea level anomalies is standardised by subtracting the mean and dividing each time series by its standard deviation. The normalisation of the sea level series allows one to obtain spatial patterns without possible domination by gridpoints with larger variances. Although avoiding the problem of spatial variability being driven by the most energetic gridpoints, normalisation yields coefficients corresponding to standardised sea level, which are therefore less easy to interpret directly.

PCA is carried out on the normalised field for dimensionality reduction. The large number of spatial degrees of freedom is reduced by considering a truncated PCA version of the original sea level field. A subspace of  $k=8$  principal com-



**Fig. 2.** Fraction of variance accounted by the principal components (PC).



**Fig. 3.** Schwartz Bayesian Criterion (SBC) as a function of model order ( $p$ ).

ponents (PCs) is considered, explaining more than 85% of the overall variance of the sea level field (Fig. 2).

### 3.2.2 Model estimation

A multivariate autoregressive model is estimated for the denoised and normalised sea level field. The estimation of the parameters in the model is carried out in the PCA subspace by least squares. The Schwartz Bayesian Criterion (SBC) is used for selection of the order  $p$  of the model (Fig. 3). The plot shows that a considerable improvement in model fit (as measured by SBC) is obtained by considering rather than a first order (corresponding to the assumption in POP analysis) higher order models. The SBC criterion suggests a model of order  $p=3$ .

Since the estimated model can be written as a system of  $k=8$  multiple regressions, inference on the fitted model can be carried out from the univariate form of the equations. Each equation is a linear regression with  $k \times 3=24$  variables for  $N-p$  observations. The statistics (residual sum of squares, R square and F test) describing how well each univariate equation fits the data, assuming normally-distributed errors, are presented in Table 1. The results indicate a very good agreement between the model and the observations.

**Table 1.** Statistics of model fit for each variable: residual sum of squares (RSS), R-square, F statistic and corresponding p-value.

Variable	RSS	R-square	F	p-value
PC1	08.55	0.9998	93 093	$<10^{-16}$
PC2	02.72	0.9995	39 440	$<10^{-16}$
PC3	11.58	0.9972	6397	$<10^{-16}$
PC4	14.28	0.9939	2949	$<10^{-16}$
PC5	06.59	0.9955	4015	$<10^{-16}$
PC6	07.81	0.9930	2553	$<10^{-16}$
PC7	12.40	0.9884	1541	$<10^{-16}$
PC8	04.87	0.9947	3403	$<10^{-16}$

**Table 2.** Eigenmodes for the sea level field estimated from the  $m-AR(3)$  model.

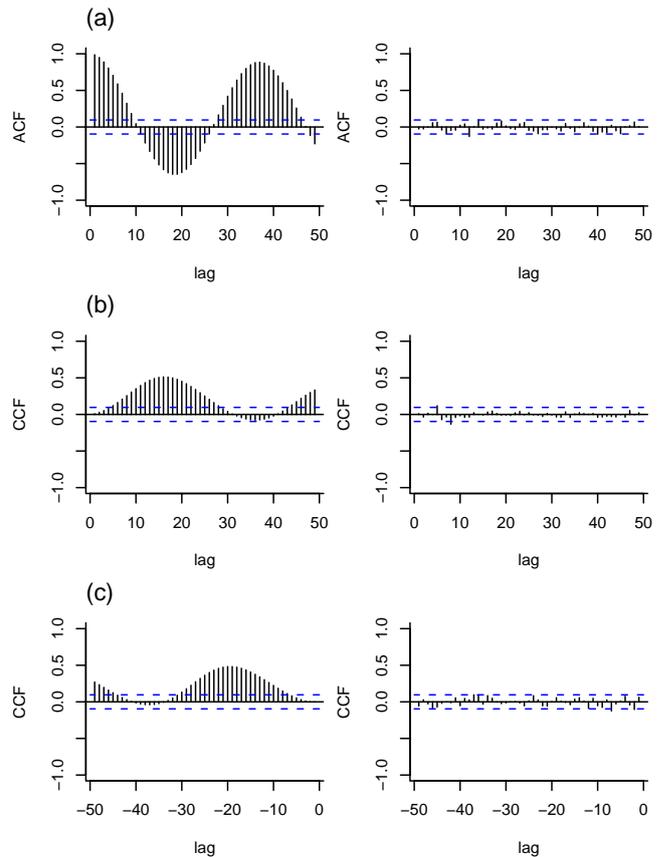
Period (years)	e-folding time (years)
0.97	9.70
0.52	0.94

However such agreement can be misleading due to the danger of overfitting in which case the model is not a parsimonious representation of the multivariate dataset. A trade-off between parsimony and agreement with observations is unavoidable and to a large extent dependent on the objectives of the analysis. If the model is used for forecasting for example, a “worst” fit may yield better predictions; on the other hand if the model is used to describe and understand the measured system a closer agreement between the model and the dataset may be preferable.

Since a multivariate autoregressive process can be interpreted as a filter that transforms the observations into white noise series, uncorrelatedness of the residuals is a primary criterion for checking the adequacy of an estimated model. The autocorrelation function for the leading component (Fig. 4a) indicates that the model yields a uncorrelated residual series. Figures 4b and c show that the estimated model is able to describe the relationship between the first two leading components of the observed field yielding uncorrelated residual series. These results suggest that the fitted model is an adequate representation of the sea level field.

### 3.2.3 Eigenmodes

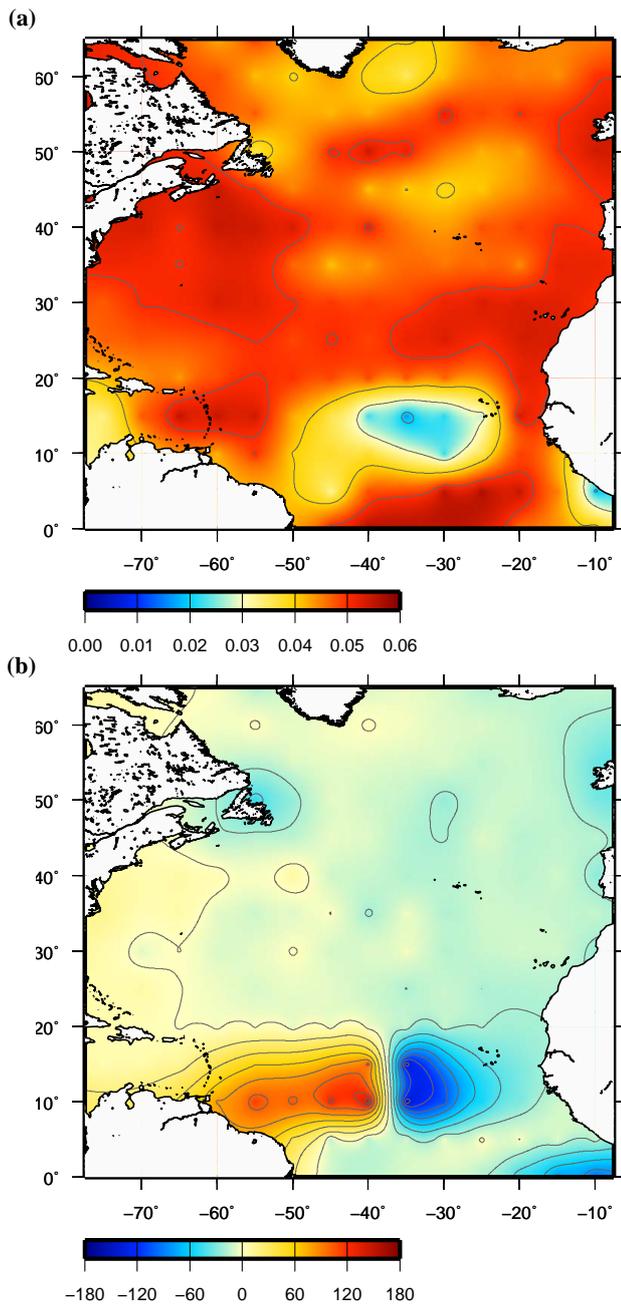
Eigenmodes are computed from the eigen-decomposition of the coefficients matrix from the  $m-AR(3)$  model. The estimated model is stable since for all the eigenvalues  $|\lambda_i| < 1 \forall i$ . Only sustained modes, as measured by the corresponding e-folding time (e-folding  $>$  period), are considered (Table 2).



**Fig. 4.** Correlation of leadings PCs (left) and corresponding residuals from the estimated  $m-AR(3)$  model (right): (a) autocorrelation, PC1; (b) cross-correlation, PC1 & PC2 (positive lags); (c) cross-correlation, PC1 & PC2 (negative lags). Horizontal dashed lines represent 95% confidence levels for correlation of white noise realisations. Lags are spaced by 10 days.

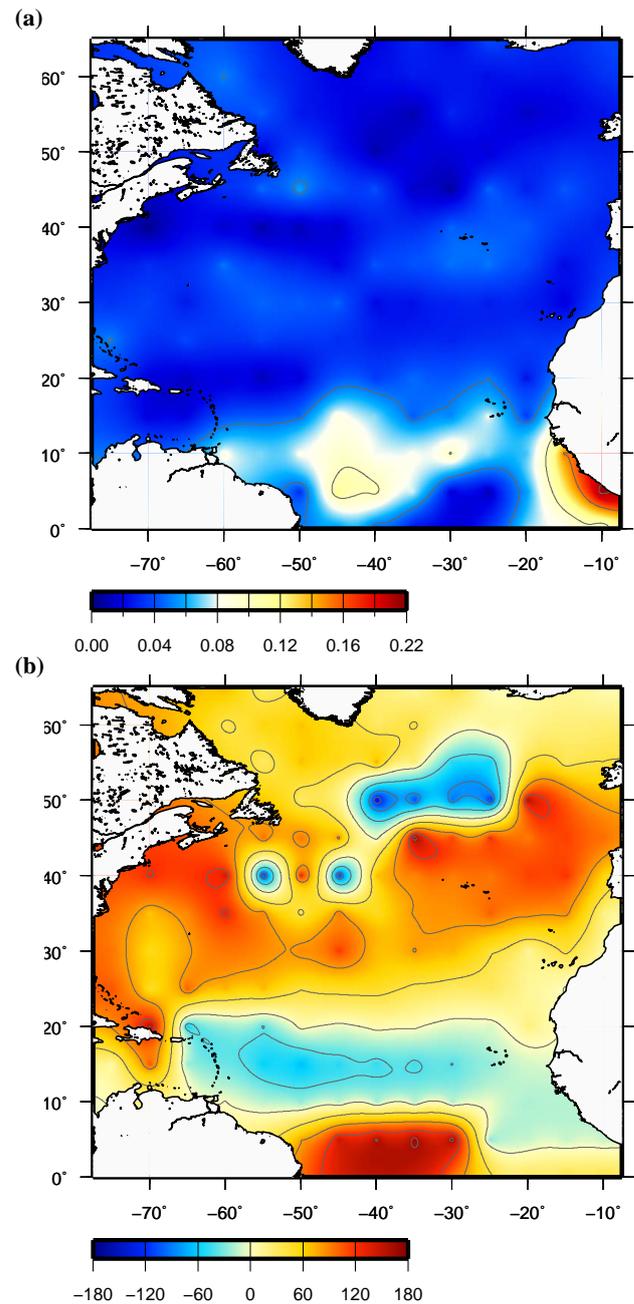
The leading, least damped mode, associated with the largest eigenvalue, is an annual oscillation. The second mode is close to a semi-annual oscillation. These are natural modes, since the seasonal signal is usually the dominant signal in sea level time series. The structure of the autoregressive modes is represented by the amplitude ( $A$ ) and phase ( $\psi$ ) obtained from the real ( $p^r$ ) and imaginary ( $p^i$ ) components of the associated eigenvectors  $\mathbf{p} = p^r + i p^i$  as  $A = \sqrt{p_r^2 + p_i^2}$  and  $\psi = \arctan(p_i / p_r)$ .

The annual mode (Fig. 5) exhibits, as expected, a very homogeneous spatial structure reflecting the large-scale coherent behaviour of the annual cycle of sea level. The amplitude values are very similar over the North Atlantic, only slightly lower in the tropical Atlantic near 15 degrees latitude, in agreement with the known weakening of the annual cycle and increase dominance of the semi-annual cycle at these locations. The phase pattern is fairly homogeneous



**Fig. 5.** First autoregressive mode: (a) Amplitude; (b) Phase.

throughout the North Atlantic except for the seesaw in the tropical Atlantic, reflecting opposing phases on the west and east sides. Sea level in this region is closely related with the winds regime. The relaxation of trade winds at the beginning of the year (February–March) causes a decrease in sea level on the west side and an increase on the east side; the onset of trade winds in May–June is responsible for a decrease in the sea level in the east and an increase in the west (Schouten et al., 2005). A difference in phase is also visible along the coast of Newfoundland and Ireland.



**Fig. 6.** Second autoregressive mode: (a) Amplitude; (b) Phase.

The semi-annual mode (Fig. 6) exhibits higher amplitudes on the eastern boundary near the equator and in the intertropical convergence zone (ITCZ) reflecting the dominance of the semi-annual signal in the tropical Atlantic. The phase of the semi-annual mode exhibits a complex spatial pattern. In the tropical Atlantic the equatorial region (particularly near the western boundary) and the ITCZ exhibit contrasting phases possibly associated with dynamic influences through mechanical and thermal forcings on the semi-annual sea level signal.

#### 4 Conclusions

Remote sensing of the Earth is yielding large datasets of geophysical observations in both time and space. Finding patterns from an observed field is a challenging task, even more challenging when one comes to interpretation. The use of autoregressive models as linear approximations to the dynamics of a field has a strong physical motivation, since autoregressions can be interpreted as discretised versions of ordinary differential equations.

Multivariate autoregressive models constitute an useful extension to POP analysis. By considering higher order models, a larger class of systems can be modelled. Multivariate autoregressive models allow the extraction of dominant modes of variability from an observed field. Similarly to POP analysis, if a multivariate autoregressive model of arbitrary order  $p$  provides an adequate fit to the observations of an observed field, the dynamical characteristics of the field can be empirically inferred from the fitted model through eigen-decomposition, yielding the dominant modes of variability from the multivariate dataset in terms of relaxations and oscillations. A drawback of the approach is the large number of parameters that need to be estimated, leading to difficulties in estimation and stability. Estimation is thus carried out in reduced PCA space for dimensionality reduction. When fitting an autoregression to an observed field, the estimation method and the number of principal components to retain when considering the field in PCA reduced space are important issues that must be handled on an application-specific basis.

Here satellite altimetry observations in the North Atlantic have been modelled by a multivariate autoregression of third order. The estimated model is according to SBC a significantly better representation of the observed sea level field than a first order model. Eigendecomposition of the fitted autoregressive model yielded physical interpretable modes, constituting a useful space-time description of North Atlantic sea level variability. Although only seasonal modes have been inferred by autoregressive modelling, this is a limitation of the dataset (strongly constrained by the short length of the satellite altimetry series) rather than of the approach itself.

Autoregressive models have a predictive potential. Thus the estimation of a multivariate autoregressive model can be applied not only to produce a space-time description in terms of dominant variability modes, as in this work, but also to prediction. Although the linear stationary nature of autoregressive models implies the decay of the oscillatory solutions, which is an undesirable property in a forecasting framework, this can be handled by assuming a constant unitary amplitude for prediction from an estimated model. In the sea level application discussed here, forecasting is hampered by the short length of the available time series. With the expected increasing length of sea level series as satellite

missions extend in time, forecasting sea level from an autoregressive model should become feasible.

*Acknowledgements.* This work has been supported by program POCTI through the Centro de Investigação em Ciências Geo-espaciais (CICGE) of the Faculty of Science, University of Porto.

Edited by: M. Thiel

Reviewed by: H. Rust and three other referees

#### References

- AVISO: AVISO User Handbook for Merged TOPEX/POSEIDON Products, AVI-NT-02-101-CN ed. 3.0, 1996.
- Barbosa, S. M., Fernandes, M. J., and Silva, M. E.: Space-time analysis of sea level in the North Atlantic from TOPEX/Poseidon satellite altimetry, International Association of Geodesy Symposia, 129, 248–253, Springer, 2005.
- Berwin, R.: Topex/Poseidon Sea Surface Height Anomaly Product. User's Reference Manual, NASA JPL Physical Oceanography DAAC, Pasadena, CA., 2003.
- Chambers, D. P., Hayes, S. A., Ries, J. C., and Urban, T. J.: New Topex sea state bias models and their effect on global mean sea level, *J. Geophys. Res.*, 108, 3305–3311, 2003.
- Donoho, D. L. and Johnstone, I. M.: Adapting to unknown smoothness via wavelet shrinkage, *J. Amer. Stat. Assoc.*, 90, 1200–1224, 1995.
- Dorandeu, J. and Le Traon, P. Y.: Effects of global mean atmospheric pressure variations on mean sea level changes from TOPEX/Poseidon, *J. Atmos. Oceanic Technol.*, 16, 1279–1283, 1999.
- Fernandes, M. J. and Antunes, M. A.: Eight years of radar altimetry in the North-East Atlantic, *Proceedings 3 Assembleia Luso-Espanhola de Geodesia e Geofísica*, 1, 226–230, 2003.
- Gámez, A. J. G., Zhou, C. S., Timmermann, A., and Kurths, J.: Nonlinear dimensionality reduction in climate data, *Nonlin. Processes Geophys.*, 11, 393–398, 2004, <http://www.nonlin-processes-geophys.net/11/393/2004/>.
- Hasselmann, K.: PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns, *J. Geophys. Res.*, 93, 11 015–11 021, 1988.
- Hsieh, W. W.: Nonlinear principal component analysis by neural networks, *Tellus A*, 53, 599–615, 2001.
- Hsieh, W. W. and Wu, A.: Nonlinear multichannel singular spectrum analysis of the tropical Pacific climate variability using a neural network approach, *J. Geophys. Res.*, 107, doi:10.1029/2001JC000957, 2002.
- Jolliffe, I. T.: *Principal Component Analysis*, Springer, 2002.
- Kondrashov, D., Kravtsov, S., Robertson, A. W., and Ghil, M.: A hierarchy of data-based ENSO models, *J. Climate*, 18, 4425–4444, 2005.
- Kravtsov, S., Kondrashov, D., and Ghil, M.: Multi-level regression modeling of nonlinear processes: derivation and applications to climatic variability, *J. Climate*, 18, 4404–4424, 2005.
- Lütkepohl, H.: *Introduction to Multiple Time Series Analysis*, Springer, 1993.
- Maharaj, E. A. and Wheeler, M. C.: Forecasting an index of the Madden-Oscillation, *Int. J. Climatol.*, 25, 1611–1618, 2005.

- Matsumoto, K., Takanezawa, T., and Masatsugu, O.: Ocean tide models developed by assimilating Topex/Poseidon altimeter data into hydrodynamical model: a global model and a regional model around Japan., *J. Oceanography*, 56, 567–581, 2000.
- Neumaier, A. and Schneider, T.: Estimation of parameters and eigenmodes of multivariate autoregressive models, *ACM – Transactions on mathematical software*, 27, 27–57, 2001.
- Preisendorfer, R. W.: *Principal Component Analysis in Meteorology and Oceanography*, Elsevier, 1988.
- R Development Core Team: *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org>, ISBN 3-900051-07-0, 2005.
- Rashid, H. A. and Simmonds, I.: Southern Hemisphere annular mode variability and the role of optimal nonmodal growth, *J. Atmos. Sci.*, 62, 1947–1961, 2005.
- Scharroo, R., Lillibridge, J. L., Smith, W. H. F., and Schrama, E. J. O.: Cross-calibration and long term monitoring of the microwave radiometers of ERS, TOPEX, GFO, Jason and Envisat, *Marine Geodesy*, 27, 279–297, 2004.
- Schneider, T. and Griffies, S. M.: A conceptual framework for predictability studies, *J. Climate*, 12, 3133–3155, 1999.
- Schneider, T. and Neumaier, A.: ARfit – A MATLAB package for the estimation of parameters and eigenmodes of multivariate autoregressive models, *ACM Transactions on Mathematical Software*, 27, 58–65, 2001.
- Schouten, M. W., Matano, R. P., and Strub, T. P.: A description of the seasonal cycle of the equatorial Atlantic from altimeter data, *Deep Sea Res. I*, 52, 477–493, 2005.
- Schwarz, G.: Estimating the dimension of a model, *Annals of Statistics*, 6, 461–464, 1978.
- Smith, W. H. F. and Wessel, P.: Gridding with continuous curvature splines in tension, *Geophysics*, 55, 293–305, 1990.
- Tenenbaum, J. B., de Silva, V., and Langford, J. C.: A global geometric framework for nonlinear dimensionality reduction, *Science*, 290, 2319–2323, 2000.
- Von Storch, H. and Zwiers, F. W.: *Statistical Analysis in Climate Research*, Cambridge University Press, 1999.
- Von Storch, H., Burger, G., Schnur, R., and von Stoch, J.: Principal Oscillation Patterns: a Review, *J. Climate*, 8, 377–399, 1995.
- Wang, Y. M.: Mean Sea Surface, gravity anomaly, and vertical gravity gradient from satellite altimeter data, *J. Geophys. Res.*, 106, 31 167–31 174, 2001.