Estuarine, Coastal and Shelf Science 91 (2011) 411–422

Contents lists available at ScienceDirect

Estuarine, Coastal and Shelf Science

journal homepage: www.elsevier.com/locate/ecss

Exploring the long-term and interannual variability of biogeochemical variables in coastal areas by means of a data assimilation approach

Stefano Ciavatta [a,b,*], Roberto Pastres [c]

[a] Plymouth Marine Laboratory, Prospect Place, Plymouth PL1 3DH, United Kingdom
[b] EuroMediterranean Centre for Climate Change, Consorzio Venezia Ricerche, Via della Libertà 12, 30175 Venezia, Italy
[c] Department of Environmental Sciences, Informatics and Statistics, University Cà Foscari of Venice, Dorsoduro 2137, 30123 Venezia, Italy

ARTICLE INFO

Article history:
Received 21 March 2010
Accepted 28 November 2010
Available online 3 December 2010

Keywords:
time series analysis
Kalman filter
water quality monitoring
lagoon of Venice

ABSTRACT

Dynamic Harmonic Regression (DHR) models are applied here to the investigation of the interannual changes in the trend and seasonality of biogeochemical variables monitored in coastal areas. A DHR model can be regarded as a time-series component model, where the phases and amplitudes of the seasonal component, as well as the trend, are parameters that vary with time, reflecting relevant changes in the evolution of the biogeochemical variables. The model parameters and their confidence bounds are estimated by data assimilation algorithms, i.e. the Kalman filter and the Fixed Interval smoother. The DHR model structure is here identified by a preliminary spectral analysis and a subsequent minimization of the Bayesian Information Criterion, thus avoiding subjective choices of the frequencies in the seasonal component. The methodology was applied to the investigation of the long-term and interannual variability of ammonia, nitrate, orthophosphate and chlorophyll-a monitored monthly in the lagoon of Venice (Italy) during the years 1986–2008. It was found that the long-term evolutions of the biogeochemical variables were characterized by non-linear patterns and by statistically significant changes in the trend. The seasonal cycles of all the variables were characterized by a marked interannual variability. In particular, the changes in the seasonality of chlorophyll and nitrate were significantly related to the changes in the seasonality of water temperature at the study site and of nutrient concentrations in river discharges, respectively. These results indicate that the methodology could be a sound alternative to more traditional approaches for investigating the impacts of changes in environmental and anthropogenic forcings on the evolution of biogeochemical variables in coastal areas.

1. Introduction

In recent decades, the efforts in monitoring biogeochemical variables in coastal areas, e.g. chlorophyll and nutrient concentrations, have increased worldwide, fuelled by concerns about global warming (Smetacek and Cloern, 2008; Ducklow et al., 2009) or responding to environment protection legislation, such as the Water Framework Directive in Europe or the Ocean Act in USA, Australia and Canada (Borja et al., 2008).

Decadal time series of biogeochemical variables are currently available at several coastal sites such that appropriate time series analysis tools could be applied for estimating both multi-annual trends and systematic seasonal fluctuations.

Indeed, changes in the trend as well as interannual variations of the seasonal component could be related quantitatively to changes of the climatic forcings (see for example Villate et al., 2008) or changes of the anthropogenic pressures (see for example Guadayol et al., 2009 and Aravena et al., 2009), that may trigger changes and regime-shifts of the ecosystem (Folke et al., 2004; Viaroli et al., 2008; Zaldivar et al., 2008; Widdicombe et al., 2010). Therefore, univariate time series (TS) models that decompose the time series of monitoring data into long-term, seasonal and random components have proved to be valuable tools in the context of coastal-system investigations. Component TS models can support different analytical approaches such as ecosystem modelling (de Vries et al., 1998), transfer function models (Villate et al., 2008; Aravena et al., 2009), and wavelet analysis (Nezlin and Li, 2003; Kromkamp and Van Engeland, 2010).

Recent attempts at applying univariate models to time series of biogeochemical variables addressed their non-stationarity (Young et al., 1991), but focused separately either on the estimation of the non-linear trend or on the estimation of the seasonal

* Corresponding author. Plymouth Marine Laboratory, Prospect Place, Plymouth PL1 3DH, United Kingdom.
E-mail address: s.ciavatta@pml.ac.uk (S. Ciavatta).

0272-7714/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ecss.2010.11.006

component. For example, non-linear trends were estimated by means of moving averages of zooplankton observations (e.g. David et al., 2005), by fitting exponential and gamma models to nutrient concentrations (Pastres et al., 2004), or fitting second-order polynomials to chlorophyll concentrations (Kromkamp and Van Engeland, 2010). On the other hand, shifts in the seasonal patterns of a plankton time series were determined by Dowd et al. (2002, 2004) by means of a Fourier model, with a single frequency, whose time-varying phase and amplitude were estimated using the Kalman filter. The seasonality of the same time series was investigated by Ikeda et al. (2008) by means of functional data analysis with a Fourier basis. This method led a higher flexibility in the choice of the model structure, allowing the subjective inclusion of a second frequency in the seasonal component, which interannual variability was investigated by means of derivative calculation and curve registration. Nevertheless, the above methods were based on the assumption that the long-term trend of the plankton time series was not significant (Dowd et al., 2002, 2004; Ikeda et al., 2008).

Thus, the mentioned approaches to time series decomposition do not address the root of the problem, as pointed out by Ikeda et al. (2008), namely the simultaneous estimation of both the trend and the seasonal component in presence of non-linear changes of the mean level and interannual shifts of the periodical component. Equally relevant is to provide uncertainty measures of the estimates, in order to evaluate the statistical significance of the changes of the biogeochemical patterns in the coastal area (Beck, 1987).

Given the above, Dynamic Harmonic Regression models (DHR; Young et al., 1999) could represent a sound approach, since they are characterized by a very flexible structure that allows the decomposition of non-linear and non-stationary time series (Young et al., 1999). Therefore, DHR models can present several advantages with respect to classical analytical methods such as ARIMA or Census models (see for example the discussions in Young et al., 1999, and Pedregal and Trapero, 2007). The DHR parameters that define the trend, the amplitude and the phase of the seasonal component are regarded as time varying and they are estimated simultaneously, by means of data assimilation algorithms that process the data in sequence. This approach allows the estimate of the trajectory in time of the parameters and of the model output, as well as those of their standard errors, even with respect to missing data (Young et al., 1999). The trajectories of the parameters – and of their standard error – could provide insights into the dynamic of environmental systems, as shown in Young (1998), and were applied to detect statistically significant changes of the trend of air quality (Becker et al., 2006), and phase shifts of air temperature (Young, 2000).

In the framework of environmental studies, the DHR modelling approach has already been applied to non-stationary time series in hydrology (e.g. Keery et al., 2007; Chappel et al., 2009; Vogt et al., 2010), climate science (e.g. Young, 1998; Young, 2000; Taylor et al., 2007) and air quality studies (e.g. Romanowicz et al., 2006; Becker et al., 2006, 2008). Nevertheless, to the authors' knowledge, the potentiality and usefulness of this approach in the framework of coastal areas studies have not been explored as yet.

The objective of the present work is to demonstrate the potential advantages of applying Dynamic Harmonic Regression models to estimate the non-linear trends and the interannual variability of the seasonal cycles of highly noisy biogeochemical data collected in coastal areas. The identification of the most adequate DHR model is a key issue, which, however, has not been fully addressed as yet from the theoretical point of view (Pedregal and Trapero, 2007; Jiang et al., 2010). Therefore, in previous applications this problem was addressed by exploiting "a priori" a hypothesis (e.g. Vogt et al., 2010) or by using an arbitrary threshold of model performance (e.g. Jiang et al., 2010). In this work, we propose an operational procedure for the identification of the adequate DHR model, based on a preliminary spectral analysis (Young et al., 1999) and on the subsequent application of a Goodness-of-Fit criterion to a set of candidate models.

The method was successfully tested on twenty-year long time series of monthly, highly noisy (inherently variable) observations of chlorophyll-a, nitrogen, ammonia and orthophosphate monitored in the shallow-water lagoon of Venice, Italy, during the years 1986–2008. This case study explores the potential advantages offered by a peculiar feature of the method – i.e. the estimation of the trajectories of the parameters characterizing the trend, the seasonal component and their standard errors – for: i) detecting statistically significant changes of the trends of the biogeochemical variables, and ii) investigating the relationships between the interannual variability of the seasonal cycles and that of the environmental forcings.

2. Methods

2.1. The Dynamic Harmonic Regression model

The Dynamic Regression (DHR) model, described in detail in Young et al. (1999), is a non-stationary univariate time series model that can be represented in the following component form:

$$y_t = T_t + S_T + e_t \quad e_t \sim N\left(0, \sigma^2\right) \tag{1}$$

where y_t is the observed time series, and T_t, S_t, and e_t, represent the trend, seasonal and stochastic components, respectively. In Eq. (1) e_t is a normally distributed random sequence with zero mean and variance σ^2, and S_t has the form of a harmonic regression model or, equivalently, of a Fourier polynomial:

$$S_T = \sum_{i=1}^{R} \left(a_{i,t}\cos(\omega_i t) + b_{i,t}\sin(\omega_i t)\right) \tag{2a}$$

where $\omega_i = (2\pi i)/s$, $i = 1, 2, \ldots, R$ are the fundamental and harmonic frequencies of the sinusoidal term i, and s is the period of the fundamental cycle. The number R of the sinusoidal components needs to be opportunely estimated when applying the DHR model, as described in Section 2.2.

Equation (2a) can be rewritten in an equivalent form that puts in evidence the meaning of the parameters $a_{i,t}$ and $b_{i,t}$, $i = 1, \ldots, R$:

$$\begin{aligned} S_T &= \sum_{i=1}^{R} \left(\sqrt{a_{i,t}^2 + b_{i,t}^2}\cos\left(\omega_i t + \tan^{-1}(b_{i,t}/a_{i,t})\right)\right) \\ &= \sum_{i=1}^{R} \left(A_{i,t}\cos\left(\omega_i t + \phi_{i,t}\right)\right) \end{aligned} \tag{2b}$$

As one can see in Eq. (2b), the parameters $a_{i,t}$ and $b_{i,t}$ define the amplitude $A_{i,t} = \sqrt{a_{i,t}^2 + b_{i,t}^2}$ and the phase $\phi_{i,t} = \tan^{-1}(b_{i,t}/a_{i,t})$ of the sinusoidal term i.

The model in Eqs. (1) and (2) is different from a classical additive time series model (see for example the Census decomposition approach in David et al., 2005) because the parameters which define the trend, i.e. T_t itself, and the seasonal component S_t, i.e. $a_{i,t}$ and $b_{i,t}$, are modelled as Time Variable Parameters (TVPs), i.e. as stochastic variables. As a consequence, also the amplitudes $A_{i,t}$ and the phases $\phi_{i,t}$ of the harmonic components in Eq. (2b) can vary with time (Young et al., 1999).