Environmental Modelling & Software 48 (2013) 202–212

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

The application of a general time series model to floodplain fisheries in the Amazon

Ronny O. Vallejos [a,*], Nidia Noemi Fabré [b], Vandick da Silva Batista [b], Jonathan Acosta [a]

[a] Departamento de Matemática, Universidad Técnica Federico Santa María, Casilla 110-V, Valparaíso, Chile
[b] Universidade Federal de Alagoas, Instituto de Ciências Biológicas e da Saúde, Campus A.C. Simões, Av. Lourival Melo Mota, Tabuleiro do Martins, Maceió, AL 57072-970, Brazil

A R T I C L E I N F O

Article history:
Received 31 May 2012
Received in revised form
5 July 2013
Accepted 6 July 2013
Available online 1 August 2013

Keywords:
Time series
Common features
VAR model
Forecast
Manaus fishery

A B S T R A C T

Time series analysis is a common tool in environmental and ecological studies to construct models to explain and forecast serially correlated data. There are several statistical techniques that are used to deal with univariate and multivariate (more than one series) chronological patterns of fisheries data. In this paper, an additive stochastic model is proposed with explicative and predictive features to capture the main seasonal patterns and trends of a fisheries system in the Amazon. The model is constructed on the assumption that the multivariate response variable – vector containing fishery yield of eight periodic species and the total fishery yield – can be decomposed into three terms: an autoregression of the response vector, an exogenous environmental variable (river level), and a seasonal component (significant frequencies obtained by using spectral analysis and the periodogram indicating the regularity of periodic cycles in the natural and fisheries system). The estimation procedure is carried out via maximum likelihood estimation. The model explained, on average, 78% of the variability in yield of the study species. The model represents the optimal solution (minimum mean square mean error) among the class of all multivariate autoregressive processes with exogenous and seasonal variables. Predictions for one period ahead are provided to illustrate how the model works in practice.

1. Introduction

The use of statistical techniques for predictive purposes is common in ecology and fisheries (*e.g.* McCleary and Hassan, 2008; Van Steenbergen et al., 2012). However, although serial autocorrelation may be a common feature in nature, it has rarely been addressed in these studies leading to biased parameter estimates, misinterpretation of significance levels and, consequently, erroneous inferences concerning the magnitude and direction of effects (Ostrom, 1990). An important aspect of time series analyses relates to periodic variations caused by biological, physical or other environmental phenomena (Shumway and Stoffer, 2000). Dealing with these influences is of fundamental importance when making forecasts, particularly when there is some form of environmental cycle. Seasonal changes are common in natural systems, generating oscillations in populations and affecting community structure. Such seasonality drive changes in biomass and ecosystem structure, modifying the upward or downward use of energy and materials in the system (Holling, 1973; Winemiller, 1990).

The development of models, which summarize serial variations, is an important strategy to draw ecological inference (Austin, 2002). This is particularly the case in tropical ecosystems (Bayley, 1995). The analysis of time series based on frequencies, commonly known as spectral analysis, is particularly relevant in ecology where series are frequently characterized by seasonal components and cycles that cannot be easily estimated by simple inspection (Arnade et al., 2005). This includes the analysis of the spectral density associated with linear processes, a mathematical transformation of the correlation function. The estimation of the spectral density, called periodogram, can be used via hypothesis testing to determine whether a particular frequency is significant (allowing the inclusion of that frequency in the model). Spectral analysis can be combined with graphical tools and previous knowledge (usually fitting naive processes) to provide more sophisticated models. Such models are better able to describe some patterns of the series, or can be used to provide precise forecasting for unobserved future values of the process (Allende et al., 2002).

Many tropical aquatic ecosystems show cyclical patterns of flooding (Hamilton et al., 1996). Cycles periodically observed over

* Corresponding author. Departamento de Matemática, Universidad Técnica Federico Santa María, Avenida España 1680, Valparaíso, Chile. Tel.: +56 32 2654964; fax: +56 32 2654 563.
E-mail address: ronny.vallejos@usm.cl (R.O. Vallejos).

1364-8152/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.envsoft.2013.07.004

time have already been described in the Amazon region, and their correlation with flooded areas has been used to reconstruct regional inundation patterns (Sippel et al., 1998). Moreover, the ecological effects of regular flood cycles underlie the development of Junk's influential flood-pulse concept (Junk et al., 1989). Flood cycles have also been related to patterns of fish landings and primary production outputs which also follow cycles observed over time (*e.g.* Arnade et al., 2005; Gates, 2000; Waugh and Miller, 1970). If these cycles are not rigorously accounted for during analysis, they can introduce statistical noise into temporal and spatial distribution analysis of abundance. Consequently, fisheries managers may make decisions based on biased information.

Various types of mathematical or statistical models have been used for the analysis of fisheries data (multivariate regression models, exponential smoothing, and autoregressive integrated moving average –ARIMA– processes (Stergiou et al., 1997)). In the context of fisheries, ARIMA models are considered efficient tools for the analysis of the pilchard fishery (Koutroumanidis et al., 2006; Stergiou, 1989; Stergiou and Christou, 1996; Stergiou et al., 1997), and are most commonly used to make a small number (short-term) of future predictions (Stergiou, 1989; Lloret et al., 2000; Parsons and Colbourne, 2000).

Extensions of ARIMA models for series whose correlation function decays slowly as a function of the lag distance (long memory time series) have been used to study monthly river flows (Ooms and Franses, 2001). More recently, recursive algorithms that produce optimal estimates (*e.g.* the Kalman filter) have been used when there is large natural variability and measurement error in fisheries data (Peterman et al., 2003). Neurofuzzy networks have been developed to improve wastewater flow-rate forecasting (Fernandez et al., 2009). Multivariate time series have also been used to complement the univariate analyses, providing cross-correlation computations and information on stability, common features and causality (Engle and Granger, 1987; Engle and Kozichi, 1993). The vector autoregressive (VAR) model is one of the most common tools for describing multivariate characteristics and interactions of a time series vector, considering the serial correlation inside and across the series (Lutkepohl, 2005; Gan, 2006). Some of the models described above can be combined to analyze time series with more complex patterns (Ailliot and Monbet, 2012).

The model in this study assumes that the multivariate response variable – a vector containing fishery yield of eight species that migrate periodically and the total species yield – can be decomposed into three terms: an autoregression of the response variables, an exogenous environmental variable (river level), and a seasonal component (significant frequencies obtained by detecting the peaks of the periodogram indicating the regularity of periodic cycles in the natural and fisheries system). The model describes and predicts the yield of periodic fishes in tropical floodplains. These fishes represent species that maximizes age-specific fecundity at the expense of optimizing turnover time and juvenile survivorship. In this case, the maximization of the fecundity is driven by predictable seasonal environmental variation (Winemiller and Rose, 1992). Periodic environmental conditions lead to synchronize the reproduction. At least 77% of the fishery yield in the Central Amazon region is comprised by periodic strategy fishes (Barletta et al., 2010). Thus, yield of these fish species can be useful indicators of seasonal variations in the Amazonian rivers.

The following section provides a description of the study area, data set, and two statistical techniques: the notion of common features and multivariate autoregressive processes with seasonal and exogenous components. The Results Section includes a brief univariate analysis of the times series, the estimation of the most relevant frequencies, estimation of the parameters of the proposed model, residual analysis, forecasting, and cross-validation. The Discussion Section highlights important relationships of the methods used with fisheries-environmental data and also includes possible topics for further research.

2. Materials and methods

The following section is divided in four parts. First, we provide a description of the data set, study area, data collection, and the variables included in the study. Second, the common features framework is explained. This procedure is a decision rule to decide if a feature (trend, cycle or association) is common to a set of variables. Third, we explain the motivation of the predictive statistical model in an ecological context. Fourth, multivariate autoregressive processes (VAR) are introduced as a useful tool to obtain predictions. We then describe an additive VAR process (a combination of a multivariate autoregressive process, a seasonal variable, and an exogenous variable). Further mathematical details of the methods are provided in the supplementary information Appendix B.

2.1. Data sources

2.1.1. Study area

The study was carried out in an Amazon region and uses fisheries data from the Amazon basin (Fig. 1) which covers about 700,000 km^2 (Santos and Ferreira, 1999). Within the basin there is seasonal variation that occur periodically governed by the alternation of dry and wet seasons (although always in a warm, humid tropical climate). The basin is physically delimited to the north by the Guiana Shield, to the south by the Central-Brazil Shield, and to the west by the Sub-Andean foreland (Sioli, 1984). Geochemically, the basin can be divided into three sub-systems: Western, North-South, and Central Amazon (Fittkau et al., 1975). This last region contains a large floodplain area enriched by Andean sediments through the main channel and effluents (the Purus, the Juruá and the Japurá – Fig. 1) with nutrients that indirectly sustain the aquatic fauna. Commercial fishing landings in Central Amazon are centralized in Manaus (Batista and Petrere, 2003; Petrere, 1978). The Manaus artisanal fishery is typically based on fish caught from canoes using purse seine nets (Batista et al., 2004). The fish are then transferred to larger vessels which transport the catch to Manaus.

2.1.2. Data set

Landing data series were collected in the main fishing harbor of Central Amazon (Manaus city). The data were collected on a daily basis from January 1994 to December 2004 by the Federal University of Amazonas. Fishery yield time series for the years 1994–2004 were extracted from a central database. The following time series data were considered: 1. total monthly catch of commercial fishing boats (i.e. all fishing yields combined); 2. total monthly commercial catches of curimatá (*Prochilodus nigricans*), jaraqui-fina (*Semaprochilodus taeniurus*), jaraqui-grossa (*Semaprochilodus insignis*), matrinxã (*Brycon amazonicus*), pacu-comum (*Mylossoma duriventris*), sardinhas (*Triportheus* spp), and tambaqui (*Colossoma macropomum*). These species are the most important fish species landed in Manaus in terms of yield, totaling more than 70% of the total Central Amazon catches (Batista and Petrere, 2003; Barletta et al., 2010).

The daily river level series in Manaus was supplied by the Brazilian Geological Survey (CPRM) and was available for most days over the study period. However, as the lack of data for several days could bias the monthly average, we used average values for the 14th, 15th and 16th days each month. The independent variables used for the development of the models are listed in Table 1.

2.2. Common features in time series

Engle and Kozichi (1993) introduced a class of statistical tests for the hypothesis that some feature of a data set is common to several variables. Examples are serial correlation, trends, and seasonality. They define the relationship as follows: A feature that is present in each of a group of series is said to be common to those series if there exists a non-zero linear combination of the series that does not have the feature. To formalize these ideas the authors restricted the analysis to regression models so that the presence or absence of a common feature can be addressed considering a hypothesis testing procedure for a suitable parameter of interest. Engle and Kozichi (1993) used a regression model of the form

$$y_t = x_t\beta + z_t\gamma + \varepsilon_t, t = 1, 2, \ldots, T, \tag{1}$$

where $H_0{:}\gamma = 0$, represents no feature and $H_1{:}\gamma \neq 0$ represents the inclusion or presence of a feature. The parameter $\gamma = 0$ is necessary to assume some distributional assumptions for the vector $\{y, x, z\}$. This ensures that a Lagrange Multiplier (LM) statistic has a χ^2 limiting distribution. The reader is referred to Appendix A for the mathematical derivations associated to the common feature test.

In Section 3.1 the common feature procedure will be used to test if an annual cycle is common to each series. Also, the same test will be considered to test if the first and second order autocorrelations are significantly common to all series.