Journal of the Korean Statistical Society 46 (2017) 629–640

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Estimation and identification of periodic autoregressive models with one exogenous variable

Eugen Ursu [a,b,*], Jean-Christophe Pereau [a]

[a] GREThA UMR-CNRS 5113, Université de Bordeaux, France
[b] ECREB, West University of Timisoara, Romania

ARTICLE INFO

Article history:
Received 13 February 2017
Accepted 11 July 2017
Available online 7 August 2017

MSC 2010:
primary 62M10
secondary 90C59

Keywords:
Periodic time series
Autoregressive models
Genetic algorithms
Identification
Fisheries
Sea surface temperature
Catch-per-unit of effort

ABSTRACT

This paper analyzes the identification and estimation procedures for periodic autoregressive models with one exogenous variable (PARX). The identification of the optimal PARX model is based on the use of a genetic algorithm combined with the Bayes information criterion. The estimation of the parameters relies on the least squares method and their asymptotic properties are studied. Two simulation experiments are performed and indicate the success of the suggested method. A PARX model is used to study the relationship between the catch-per-unit-effort and the sea surface temperature as exogenous variable for the shrimp French Guiana fishery from January 1989 to December 2012.

1. Introduction

The use of periodic models appears to be well-suited to deal with many real life phenomena characterized by a seasonal behavior (Dudek, Hurd, & Wojtowicz, 2014; Lund, Shao, & Basawa, 2006; Tesfaye, Meerschaert, & Anderson, 2006). These models are increasingly used in the climatology or hydrology literature (Hipel & McLeod, 1994; Jones & Brelsford, 1967; Li & Lund, 2012; Lu, Lund, & Lee, 2010; Lund et al., 2007; Ursu & Pereau, 2016; Vecchia, 1985) but also in other disciplines like macroeconomics (Franses & Paap, 2004), engineering (Schlick, Duckwitz, & Schneider, 2013) and marine fisheries (Stoffer, 1986). In many applications of time series analysis, the variable of interest may be affected by other variables, called exogenous or unmodeled variables, which are determined outside the system of interest. In the case study of the paper, climate change and global warming in particular through its effect on the sea surface temperature appears as an exogenous major driver on inter-annual fluctuations in fish abundance (Brander, 2007; Cheung et al., 2009).

Autoregressive models with exogenous variables (ARX) have been extensively used in the econometric literature. The dependent variable is assumed to depend on its past values and the present and lagged values of exogenous variables. These models are also called conditional or partial models (Lütkepohl, 2005), distributed lag models (Reinsel, 1997) or transfer function models (Wei, 2006). The introduction of ARX in a var model (VARX) with systematically varying coefficients in a state-space model form has been studied by Lütkepohl (2005). Paroli and Spezia (2008) introduced a periodic component in an ARX model. Extending ARX models to periodic autoregressive (PAR) models with exogenous variables (PARX) in a Bayesian

* Corresponding author.
E-mail address: eugen.ursu@u-bordeaux.fr (E. Ursu).

http://dx.doi.org/10.1016/j.jkss.2017.07.001
1226-3192/

framework has been proposed by Andel (1987, 1989). Maçaira, Oliveira, Ferreira, de Almeida, and Souza (2017) proposed a PAR with exogenous variables to generate scenarios for hydrological inflows for some Brazilian reservoirs and showed that the inclusion of exogenous variables decreases the error measure by 3%. This notation PARX should not be confused with other acronyms like Poisson AutoRegression with eXogenous covariates (Agosto, Cavaliere, Kristensen, & Rahbek, 2016; Angelini & Angelis, 2016) or Polynomial Autoregressive Regression with eXogenous variables (Wojciechowski, 2001).

The main challenge with PARX models relies on the large number of parameters to estimate. Sometimes there exist several sets of model parameters that give reasonable results. A way to reduce the parameter space in PAR models consists in introducing restrictions on the parameters (Ursu & Duchesne, 2009; Ursu & Turkman, 2012). A similar problem occurs in time series analysis with a large number of models which needs to be compared and estimated. Genetic Algorithms (GA) appear to be an useful tool to investigate the space of solutions and to select the combination of parameters that corresponds to the best model (Baragona & Battaglia, 2009). Baragona, Battaglia, and Cucina (2004), Wu and Chang (2002) developed a method which exploits GA with Akaike's Information Criterion (AIC). GA coupled with Minimum Description Length (MDL) have recently been used in Song and Bondon (2013) and Yau, Tang, and Lee (2015). GA in Bayesian context was also used by Jeong and Kim (2013) to locate change points for an autoregressive model. To deal with PARX models, this paper uses an automatic procedure based on GA combined with Bayesian Information Criterion (BIC).

Our contributions to the PARX literature are twofold. A first contribution is to provide theoretical results on the least squares estimators and derives their asymptotic properties in PARX models. Contrary to Maçaira et al. (2017), the estimation and identification of the model are analyzed taking into account constraints on the parameters. Our paper also considers a global optimization strategy relying on genetic algorithms and not a sequential procedure which consists in estimating first a PAR model and then adding exogenous variables. A second contribution refers to the applied fisheries literature concerning the stock assessment of fish species populations which is often performed on the basis of the well-known "Virtual Population Analysis" method (Lassen & Medley, 2001; Sparre & Venema, 1998). This age-structured method aims at providing to the fishery manager information on the recruit abundance, the spawning stock biomass as well as the fishing mortality. However this method appears to be stringent in terms of data collection (Haddon, 2011). PARX models can be more easier to implement and provide short-term forecasts to managers concerning the catch-per-unit-effort for fish species population.

The rest of the article is organized as follows. In Section 2, the PARX model is introduced and a brief review of the estimation techniques is presented. Section 3 develops a genetic algorithm to conduct the identification process and Section 4 reports some Monte Carlo simulation results. In Section 5, an application to the shrimp French Guiana fishery is presented. Section 6 concludes.

2. Periodic models with one exogenous variable

The dependent variable is denoted by Y_t, the exogenous variable by X_t and the error process by ϵ_t. Y_t and X_t are assumed to be periodic stationary processes according the following definition.

A stochastic process W_t is periodic stationary if

$$E(W_{n+s}) = E(W_n) \quad \text{and} \quad cov(W_{n+s}, W_{m+s}) = cov(W_n, W_m),$$

for all integers n and m, where s stands for the period. Periodic series are also called periodically correlated (Gladyshev, 1961) or cyclostationary (Lund & Basawa, 2000).

Based on Andel (1989), we consider the following periodic autoregressive model with one exogenous variable:

$$Y_{ns+\nu} = \sum_{k=1}^{p(\nu)} \phi_k(\nu) Y_{ns+\nu-k} + \sum_{j=0}^{m(\nu)} \theta_j(\nu) X_{ns+\nu-j} + \epsilon_{ns+\nu}, \tag{1}$$

for $n = 0, 1, \ldots, N-1$ and $\nu = 1, 2, \ldots, s$. The autoregressive model order at season ν for Y_t is given by $p(\nu)$, while the terms $\phi_k(\nu)$, $k = 1, \ldots, p(\nu)$, represent the autoregressive model coefficients during season ν. Concerning the exogenous variable X_t, the autoregressive order at season ν is given by $m(\nu)$ while the terms $\theta_j(\nu)$, $j = 0, \ldots, m(\nu)$ are the autoregressive model coefficients during season ν. The mean of Y_t is equal to zero in each of the s seasons, that is $E(Y_{ns+\nu}) = 0$, $\nu = 1, 2, \ldots, s$. Without loss of generality, the mean of process X_t is assumed to be zero. The error process $\epsilon = \{\epsilon_t, t \in \mathbb{Z}\}$ corresponds to a periodic white noise, with $E(\epsilon_t) = 0$ and $\text{var}(\epsilon_{ns+\nu}) = \sigma^2(\nu) > 0$, $\nu = 1, \ldots, s$. In the following we assumed that X_t and ϵ_t are independent processes, although most results can be obtained under less restrictive conditions. Note that if $s = 1$, then Eq. (1) reduces to a classical autoregressive model with exogenous variables (ARX). As in Andel (1989), the model can be extended to several exogenous variables.

2.1. Unconstrained least squares estimators

To estimate the parameters of the model, we consider the time series data $Y_{ns+\nu}$, $n = 0, 1, \ldots, N-1$, $\nu = 1, \ldots, s$ with sample size Ns. Let $\mathbf{z}(\nu) = \left(Y_\nu, Y_{s+\nu}, \ldots, Y_{(N-1)s+\nu}\right)^\top$ and $\mathbf{e}(\nu) = \left(\epsilon_\nu, \epsilon_{s+\nu}, \ldots, \epsilon_{(N-1)s+\nu}\right)^\top$ be $(N \times 1)$ vectors with $\top$ the

transpose operator and $\mathbf{Z}(\nu) = [\mathbf{Y}(\nu)\ \mathbf{X}(\nu)]$ the $N \times (p(\nu) + 1 + m(\nu))$ a random matrix defined as

$$\mathbf{Y}(\nu) = \begin{bmatrix} Y_{\nu-1} & Y_{\nu-2} & \dots & Y_{\nu-p(\nu)} \\ Y_{s+\nu-1} & Y_{s+\nu-2} & \dots & Y_{s+\nu-p(\nu)} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{(N-1)s+\nu-1} & Y_{(N-1)s+\nu-2} & \dots & Y_{(N-1)s+\nu-p(\nu)} \end{bmatrix}, \quad \text{and}$$

$$\mathbf{X}(\nu) = \begin{bmatrix} X_{\nu} & X_{\nu-1} & \dots & X_{\nu-m(\nu)} \\ X_{s+\nu} & X_{s+\nu-1} & \dots & X_{s+\nu-m(\nu)} \\ \vdots & \vdots & \ddots & \vdots \\ X_{(N-1)s+\nu} & X_{(N-1)s+\nu-1} & \dots & X_{(N-1)s+\nu-m(\nu)} \end{bmatrix}.$$

Let

$$\boldsymbol{\beta}(\nu) = (\boldsymbol{\phi}(\nu), \boldsymbol{\theta}(\nu))^{\top},$$

be the parameter vector, where

$$\boldsymbol{\phi}(\nu) = \left(\phi_1(\nu), \dots, \phi_{p(\nu)}(\nu)\right)^{\top} \quad \text{and} \quad \boldsymbol{\theta}(\nu) = \left(\theta_0(\nu), \theta_1(\nu), \dots, \theta_{m(\nu)}(\nu)\right)^{\top}.$$

Since Eq. (1) is a linear model, it may be written in a regression model form:

$$\mathbf{z}(\nu) = \mathbf{Z}(\nu)\boldsymbol{\beta}(\nu) + \mathbf{e}(\nu), \quad \nu = 1, \dots, s. \tag{2}$$

The covariance matrix of the random vector $\mathbf{e}(\nu)$ is $\sigma^2(\nu)\mathbf{I}_N$ with $\mathbf{I}_N$ the identity matrix of size N. The least squares estimate of $\boldsymbol{\beta}(\nu)$ is obtained by minimizing

$$S(\boldsymbol{\beta}) = \sum_{\nu=1}^{s} \mathbf{e}^{\top}(\nu)\mathbf{e}(\nu) = \sum_{n=0}^{N-1}\sum_{\nu=1}^{s}\left(Y_{ns+\nu} - \sum_{k=1}^{p(\nu)}\phi_k(\nu)Y_{ns+\nu-k} - \sum_{j=0}^{m(\nu)}\theta_j(\nu)X_{ns+\nu-j}\right)^2. \tag{3}$$

Hence, the least-squares estimators for $\hat{\boldsymbol{\beta}}(\nu) = \left(\hat{\boldsymbol{\phi}}(\nu), \hat{\boldsymbol{\theta}}(\nu)\right)^{\top}$ for any fixed ν are solution of the following $(p(\nu)+m(\nu)+1)$ equations obtained by taking derivatives of Eq. (3):

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \phi_k(\nu)} = 2\sum_{n=0}^{N-1}\epsilon_{ns+\nu}\frac{\partial \epsilon_{ns+\nu}}{\partial \phi_k(\nu)} = -2\sum_{n=0}^{N-1}Y_{ns+\nu-k}\epsilon_{ns+\nu},$$

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \theta_j(\nu)} = 2\sum_{n=0}^{N-1}\epsilon_{ns+\nu}\frac{\partial \epsilon_{ns+\nu}}{\partial \theta_j(\nu)} = -2\sum_{n=0}^{N-1}X_{ns+\nu-j}\epsilon_{ns+\nu}.$$

Setting the derivatives equal to zero, we obtain the following system for a given season ν:

$$\mathbf{Z}^{\top}(\nu)\mathbf{e}(\nu) = 0.$$

Since $\mathbf{e}(\nu) = \mathbf{z}(\nu) - \mathbf{Z}(\nu)\boldsymbol{\beta}(\nu)$ we have

$$\mathbf{Z}^{\top}(\nu)\mathbf{z}(\nu) = \mathbf{Z}^{\top}(\nu)\mathbf{Z}(\nu)\boldsymbol{\beta}(\nu),$$

which gives

$$\hat{\boldsymbol{\beta}}(\nu) = \{\mathbf{Z}^{\top}(\nu)\mathbf{Z}(\nu)\}^{-1}\mathbf{Z}^{\top}(\nu)\mathbf{z}(\nu). \tag{4}$$

The asymptotic properties of the least squares estimators in the unrestricted case are stated in the following theorem. The symbols '$\xrightarrow{d}$' and '$\xrightarrow{p}$' stand for convergence in distribution and probability, respectively, and $\mathcal{N}_d$ denotes a d-dimensional normal distribution.

Theorem 1. *Consider that a process generated by* Eq. (1) *is causal. Assume that* $\{\epsilon_t\}$ *is mean zero independent periodic white noise with* $var(\epsilon_{ns+\nu}) = \sigma^2(\nu)$ *and* $E[\epsilon_{ns+\nu}^4] < \infty$ *for all seasons* ν. *Then*

$$\sqrt{N}\left(\hat{\boldsymbol{\beta}}(\nu) - \boldsymbol{\beta}(\nu)\right) \xrightarrow{d} \mathcal{N}_{p(\nu)+m(\nu)+1}\left(\mathbf{0}, \sigma^2(\nu)\boldsymbol{\Omega}^{-1}(\nu)\right), \quad \text{as } N \to \infty. \tag{5}$$

$\boldsymbol{\Omega}(\nu)$ *represents the covariance matrix of the* $\{p(\nu) + m(\nu) + 1\} \times 1$ *random vector* $\mathbf{Z}_n(\nu)$, *where* $\mathbf{Z}_n(\nu)$ *stands for the nth line of the matrix* $\mathbf{Z}(\nu)$. *Moreover,* $\sqrt{N}\left(\hat{\boldsymbol{\beta}}(\nu) - \boldsymbol{\beta}(\nu)\right)$ *are asymptotically independent across different seasons* ν.

Proof. Technical details are similar to those used in Basawa and Lund (2001). It is easy to check that $\mathbf{Z}_n^\top(\nu)\epsilon_{ns+\nu}$ is a martingale difference sequence given the independence assumption of the error term $\epsilon_{ns+\nu}$. In addition, the unconditional covariance matrix of $\epsilon(ns+\nu)\mathbf{Z}_n^\top(\nu)$ is given by $\sigma^2(\nu)\mathbf{\Omega}(\nu)$.

Using the law of large numbers and a central limit theorem for martingale difference sequences, it follows that:

$$N^{-1}\sum_{n=0}^{N-1}\epsilon_{ns+\nu}^2\mathbf{Z}_n^\top(\nu)\mathbf{Z}_n(\nu)\xrightarrow{p}\sigma^2(\nu)\mathbf{\Omega}(\nu)$$

$$\sqrt{N}\sum_{n=0}^{N-1}\epsilon_{ns+\nu}\mathbf{Z}_n^\top(\nu)\xrightarrow{d}\mathcal{N}_{p(\nu)+m(\nu)+1}\left(\mathbf{0},\sigma^2(\nu)\mathbf{\Omega}(\nu)\right).$$

Note that we have actually proved that

$$\{N^{-1}\mathbf{Z}^\top(\nu)\mathbf{Z}(\nu)\}^{-1}\xrightarrow{p}\mathbf{\Omega}^{-1}(\nu)\quad\text{and}\quad N^{-1/2}\mathbf{Z}^\top(\nu)\mathbf{e}(\nu)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\sigma^2(\nu)\mathbf{\Omega}(\nu)\right).$$

Using Eq. (4), we obtain:

$$\hat{\boldsymbol{\beta}}(\nu)=\left[\{\mathbf{Z}^\top(\nu)\mathbf{Z}(\nu)\}^{-1}\mathbf{Z}^\top(\nu)\right]\mathbf{Z}(\nu)\boldsymbol{\beta}(\nu)+\mathbf{e}(\nu)=\boldsymbol{\beta}(\nu)+\{\mathbf{Z}^\top(\nu)\mathbf{Z}(\nu)\}^{-1}\mathbf{Z}^\top(\nu)\mathbf{e}(\nu)$$

from where

$$\sqrt{N}\{\hat{\boldsymbol{\beta}}(\nu)-\boldsymbol{\beta}(\nu)\}=\{N^{-1}\mathbf{Z}^\top(\nu)\mathbf{Z}(\nu)\}^{-1}N^{-1/2}\mathbf{Z}^\top(\nu)\mathbf{e}(\nu).$$

Slutsky's theorem yields Eq. (5). The joint asymptotic normality of $N^{1/2}\{\hat{\boldsymbol{\beta}}(1)-\boldsymbol{\beta}(1),\ldots,\hat{\boldsymbol{\beta}}(s)-\boldsymbol{\beta}(s)\}$ follows using the same kind of manipulations as those for a single season ν. From this, the asymptotic independence between $N^{1/2}\{\hat{\boldsymbol{\beta}}(\nu)-\boldsymbol{\beta}(\nu)\}$ and $N^{1/2}\{\hat{\boldsymbol{\beta}}(\nu')-\boldsymbol{\beta}(\nu')\}$, $\nu\neq\nu'$ is easily deduced. □

2.2. Least squares estimation with linear constraints on the parameters

Assume that linear constraints for the $(p(\nu)+1+m(\nu))\times 1$ vector $\boldsymbol{\beta}(\nu)$ are given by

$$\boldsymbol{\beta}(\nu)=\mathbf{R}(\nu)\boldsymbol{\xi}(\nu)+\mathbf{b}(\nu),\tag{6}$$

where $\mathbf{R}(\nu)$ stands for a known $(p(\nu)+1+m(\nu))\times K(\nu)$ matrix of order $K(\nu)$, $\boldsymbol{\xi}(\nu)$ a unrestricted $K(\nu)\times 1$ vector of unknown parameters and $\mathbf{b}(\nu)$ a $(p(\nu)+1+m(\nu))\times 1$ vector of known constants.

The unconstrained case described in Section 2.1 is obtained by assuming that $\mathbf{R}(\nu)=\mathbf{I}_{p(\nu)+1+m(\nu)}$, where $\mathbf{I}_{p(\nu)+1+m(\nu)}$ stands for the $(p(\nu)+1+m(\nu))\times(p(\nu)+1+m(\nu))$ identity matrix and $\mathbf{b}(\nu)=\mathbf{0}$, $\nu=1,2,\ldots,s$. In general, matrices $\mathbf{R}(\nu)$ and vectors $\mathbf{b}(\nu)$ allow for linear constraints on the parameters of the same season ν. When constrained parameters are introduced, it follows that

$$\mathbf{e}(\nu)=\mathbf{z}(\nu)-\mathbf{Z}(\nu)\Big(\mathbf{R}(\nu)\boldsymbol{\xi}(\nu)+\mathbf{b}(\nu)\Big).$$

The least square calculation of $\boldsymbol{\xi}(\nu)$ yields

$$\hat{\boldsymbol{\xi}}(\nu)=\left[\mathbf{R}^\top(\nu)\mathbf{Z}^\top(\nu)\mathbf{Z}(\nu)\mathbf{R}(\nu)\right]^{-1}\mathbf{R}^\top(\nu)\mathbf{Z}^\top(\nu)\left[\mathbf{z}(\nu)-\mathbf{Z}(\nu)\mathbf{b}(\nu)\right].$$

Furthermore, $\hat{\boldsymbol{\xi}}(\nu)$ follows asymptotically a multivariate normal distribution.

Duchesne and Lafaye de Micheaux (2013) consider a more general PAR framework with structured parametrization and propose an estimation method based on maximum likelihood. However applying their procedure would not changed our results because the identification procedure we develop requires to impose nullity restrictions on the value of coefficients.

3. Applying genetic algorithm to the identification problem

A crucial issue in the estimation of a PARX model relies on the identification of the model. Basically, fitting a PARX model can be seen as fitting an ARX model for each season separately. The PARX model depends on a set of parameters which needs to be properly selected in order to maximize some selection criteria. For PAR models, McLeod (1994) suggests to use the BIC selection criteria separately for each of the seasonal components:

$$BIC(\nu)=\log\hat{\sigma}^2(\nu)+\frac{\log(N)}{N}(p(\nu)+m(\nu)),\tag{7}$$

where $\hat{\sigma}(\nu)$ stands for the least squares estimators of $\sigma(\nu)$ and $p(\nu)+m(\nu)$ is the total number of autoregressive parameters of the dependent and exogenous variables in the season ν. The same criteria may be used in the identification method for PARX models. Another point to consider is that the method based on periodic autocorrelation function used by Noakes, McLeod, and Hipel (1985) to identify a PAR model cannot be applied to a PARX model. These models may have different patterns and

the choice between several models can be difficult. Let us remark that the identification of a periodic model is not as simple as it is sometimes the case for non-periodic models.

The order of AR operators also varies across the seasons. Since the number of parameters in periodic time series can be large, situations with potentially linear constraints on the parameters of a given season and zero-valued parameters are considered. When s, $p(\nu)$ and $m(\nu)$ $(\nu = 1, 2, \ldots, s)$ increase and when some parameters are constrained to zero, the number of models to be investigated becomes huge. The most common values taken by the period s are 4 for quarterly data, 12 for monthly data or 48 for quarter-monthly data. Computational time is critical for the model selection since there are $s \times 2^{p(\nu)+m(\nu)+1}$ potential sub-models. This is a notable weakness of the periodic models to deal with. For that reason, genetic algorithms appear as an efficient tool to examine the space of solutions, to select the combination of parameters that corresponds to the best model and to reduce computational time for finding the optimum. The chromosome can be expressed in binary digits to represent different orders and zero-valued parameters. The identification is made for each of the separate periods. One gene for each possible lag is reserved, filling it with 1 if the parameter is free and with 0 if the parameter is constrained to zero.

A maximum search order has to be selected for the autoregressive part of the dependent variable (P) and for the exogenous variable (M). The maximal length of the chromosome is equal to $P + 1 + M$ with $0 \leq p(\nu) \leq P$ and $0 \leq m(\nu) \leq M$. For instance, if $s = 12$, $\nu = 1$ and $P = M = 7$ the chromosome

101000011010000

corresponds to the following ARX model

$$Y_{12n+1} - \phi_1(1)Y_{12n} - \phi_3(1)Y_{12n-2} = \theta_0 X_{12n+1} + \theta_1 X_{12n} + \theta_3 X_{12n-2} + \epsilon_{12n+1}.$$

An alternative coding and its advantages are described in Minerva and Poli (2001).

The algorithm follows the standard steps of GA, starting with an initial population of chromosomes generated at random. A large number of individuals (population size N_p) affects the computational time of GA but ensures that the GA investigates the space of solutions more efficiently.

The aim of the fitness function is to numerically represent the performance of each chromosome. Essentially, the fitness function is linked to the BIC criteria given in Eq. (7) which is one of the most popular identification criteria in time series. A frequent problem with this choice is that the BIC criteria has to be minimized and therefore cannot be implemented directly since the fitness function needs to be maximized. It turns out that due to the negativity of the BIC, the fitness function is defined as follows in Baragona et al. (2004)

$$f_j(\nu) = \exp(-BIC_j(\nu)), \tag{8}$$

where $BIC_j(\nu)$ stands for the $BIC(\nu)$ value for the jth chromosome in the current population.

The selection process determines which chromosomes will survive or not in the next generation, according to their fitness values. Several selection methods are used in the literature (Goldberg, 1989). Our paper uses the tournament selection method which is based on running several tournaments among a group of chromosomes chosen at random from the current population. The winner of each tournament (the one with the best fitness) is selected to be part of a new population, that replaces the previous one. Crossover enables the pairs of parents (assuming that N_p is an even integer) to exchange individual characteristics between chromosomes. Several crossover techniques (one-point, two-points, etc.) are described in Mitchell (1996). Each chromosome is mutated with a very low probability and this reduces the chances to be trapped on a local optimum of the space of solutions. We adopt an elitist generation, meaning that the best chromosomes (ranked by their fitness values) remain unchanged for the next generation. The procedure is stopped after a fixed number N_g of generations.

To emphasize the interest of GA methods over a complete enumeration of all the $2^{P+1+M} \times s$ possible models, our framework requires to evaluate the fitness of $N_p \times N_g \times s$ models. Moreover the number of possible models increases rapidly for PARX models with more than one exogenous variable or in the case of multivariate PARX models but even in these cases, the same method defined in Sections 2 and 3 can be implemented.

4. Simulation experiments

Monte-Carlo simulations have been performed to analyze the estimation and identification methods described in Sections 2 and 3. Their aims are to evaluate the capacity of our procedure to identify and estimate the parameters of a PARX model generated by Eq. (1).

We begin by generating data from a PARX process for $s = 4$ periods. Parameters are displayed in Table 1. The stochastic process $\boldsymbol{\epsilon} = \{\boldsymbol{\epsilon}_t, t \in \mathbb{Z}\}$ corresponds to a zero mean periodic white noise with the error variance given in the last column of the table. We use a random gaussian noise for the exogenous variable X_t.

Firstly, we assume that the GA structure (i.e. the values of P and M) for each period is known and we aim at evaluating the performances of the estimated parameters. For each period, the values of the autoregressive part for the dependent and exogenous variables are known. These values are different according the season as well as the periodic variances. Concerning the PARX process given in Table 1, 1000 time series of length $N = 50$ (by period) have been generated. The empirical mean and the standard deviation of each parameter estimator are presented in Table 2. The coefficients appear to be well estimated even with a small sample size like $N = 50$.

Table 1
The parameters of the PARX model with 4 periods; the periodic variances are given in the last column.

Season	Parameters						
	ϕ_1	ϕ_2	ϕ_3	θ_0	θ_1	θ_2	σ^2
$\nu = 1$		0.50			0.20	0.60	1.00
$\nu = 2$	0.42			−0.70	0.50		1.60
$\nu = 3$	−0.80	0.20	0.35	1.00			2.00
$\nu = 4$	−0.30			1.40	0.30		2.50

Table 2
Mean values of the estimated parameters after 1000 replications of the PARX model given in Table 1 (standard deviation are in parentheses).

Season	Parameters estimates						
	ϕ_1	ϕ_2	ϕ_3	θ_0	θ_1	θ_2	σ^2
$\nu = 1$		0.506			0.205	0.589	0.927
		(0.073)			(0.138)	(0.157)	(0.191)
$\nu = 2$	0.419			−0.689	0.503		1.482
	(0.109)			(0.178)	(0.187)		(0.314)
$\nu = 3$	−0.793	0.196	0.351	0.973			1.806
	(0.145)	(0.144)	(0.095)	(0.199)			(0.382)
$\nu = 4$	−0.301			1.387	0.331		2.309
	(0.114)			(0.219)	(0.240)		(0.478)

Table 3
Identification of PARX model with 4 periods.

Season	Number of observations											
	$N = 100$			$N = 200$			$N = 400$			$N = 600$		
	q_1	q_2	q_3	q_1	q_2	q_3	q_1	q_2	q_3	q_1	q_2	q_3
$\nu = 1$	0.28	0.46	0.26	0.52	0.33	0.15	0.77	0.16	0.07	0.82	0.14	0.04
$\nu = 2$	0.56	0.29	0.15	0.77	0.19	0.04	0.83	0.15	0.02	0.90	0.09	0.01
$\nu = 3$	0.32	0.35	0.33	0.55	0.25	0.20	0.78	0.12	0.10	0.83	0.12	0.05
$\nu = 4$	0.21	0.43	0.36	0.41	0.43	0.16	0.70	0.24	0.06	0.82	0.15	0.03

Table 4
The parameters of the PARX model with 12 periods; the periodic variances are given in the last column.

Season	Parameters							
	ϕ_1	ϕ_2	ϕ_3	θ_0	θ_1	θ_2	θ_3	σ^2
$\nu = 1$	0.30	0.50		0.40	1.00	1.20		1.00
$\nu = 2$	0.42			1.20				1.60
$\nu = 3$	0.65			0.80			2.30	2.00
$\nu = 4$	−0.30			2.00	2.00	−0.90	0.50	2.50
$\nu = 5$	0.70	−0.35	0.80					3.00
$\nu = 6$	0.40	−0.50		0.70	1.20			0.90
$\nu = 7$	0.70			−0.70	−0.75			1.70
$\nu = 8$	−0.60			0.90	0.60			2.40
$\nu = 9$	0.80		0.80	−1.40	1.00			1.50
$\nu = 10$	0.90			1.60		0.92	0.55	3.20
$\nu = 11$				0.9				2.80
$\nu = 12$	0.72			−2.00	0.35			0.70

Secondly, we consider that the GA structure is unknown. In our simulations, we set the maximum order for autoregressive processes at $P = 7$ and for the exogenous variable at $M = 7$, yielding $2^{15} = 32\,768$ potential models by period. 500 replications series with N observations by period have been generated. The empirical frequencies when the true model is detected are given in Table 3. Three statistics are used to summarize the results as in Gaetan (2000): q_1 is the proportion of simulations in which the true model is detected, q_2 is the proportion of simulations in which the detected chromosome differs from the true string in one point (it means that at most a non-existent lag was included or an existent lag was excluded) and q_3 is the proportion of the remaining cases, namely the proportion of simulations in which the detected chromosome differs from the true string in at least two points.

In a second simulation experiment, we consider a PARX model with $s = 12$ periods. Here, the objective is to evaluate the ability of our procedure to identify the "right" model. The parameters of the model and the variances of the periodic white noise are displayed in Table 4. The selection algorithm is applied to 500 independent simulations (N observations by season). The chromosome length was set to 15 ($M = P = 7$). The empirical frequencies are given in Table 5.

Table 5
Identification of PARX model with 12 periods.

Season	Number of observations											
	$N = 100$			$N = 200$			$N = 400$			$N = 600$		
	q_1	q_2	q_3	q_1	q_2	q_3	q_1	q_2	q_3	q_1	q_2	q_3
$\nu = 1$	0.40	0.41	0.19	0.72	0.24	0.04	0.88	0.10	0.02	0.89	0.10	0.01
$\nu = 2$	0.64	0.25	0.09	0.77	0.20	0.03	0.84	0.14	0.02	0.87	0.12	0.01
$\nu = 3$	0.68	0.25	0.07	0.78	0.18	0.04	0.85	0.14	0.01	0.87	0.12	0.01
$\nu = 4$	0.59	0.26	0.15	0.75	0.19	0.06	0.89	0.10	0.01	0.88	0.10	0.02
$\nu = 5$	0.63	0.24	0.13	0.77	0.18	0.05	0.84	0.15	0.01	0.87	0.12	0.01
$\nu = 6$	0.71	0.21	0.08	0.81	0.16	0.03	0.88	0.11	0.01	0.89	0.10	0.01
$\nu = 7$	0.64	0.26	0.10	0.80	0.16	0.04	0.85	0.13	0.02	0.89	0.09	0.02
$\nu = 8$	0.61	0.25	0.14	0.74	0.21	0.05	0.86	0.13	0.01	0.90	0.09	0.01
$\nu = 9$	0.68	0.25	0.07	0.78	0.19	0.03	0.86	0.13	0.01	0.88	0.11	0.01
$\nu = 10$	0.57	0.27	0.16	0.75	0.20	0.05	0.86	0.13	0.01	0.90	0.10	0.00
$\nu = 11$	0.61	0.29	0.10	0.76	0.20	0.04	0.84	0.15	0.01	0.87	0.12	0.01
$\nu = 12$	0.66	0.27	0.07	0.78	0.19	0.03	0.89	0.10	0.01	0.87	0.12	0.01

Concerning the parameters used for the GA optimization, we assume that the size of GA population is set to $N_p = 50$; the uniform crossover with probability of crossover is equal to 0.8; the mutation probability is equal to 0.01 and the algorithm stops after $N_g = 30$ generations.

The performance of the genetic algorithm for a given period is satisfactory with a high rate of correct identification. It should be noted that frequencies of identification of wrong lags (at most a non-existent lag was included or an existent lag was excluded) are small. In empirical applications we suggest to run the GA algorithm several times by period and choose the model with the minimum *BIC* and/or the minimum number of parameters. This procedure may eventually prevent the algorithm to choose a "wrong" model.

5. Numerical illustration

To highlight the interest of PARX model, we illustrate the main findings of the paper with the shrimp French Guiana fishery. This case-study has been analyzed in Sanz, Diop, Blanchard, and Lampert (2016). Two shrimp species are mainly exploited in this fishery, the brown and the pink shrimps (respectively, Farfantepenaeus subtilis and Farfantepenaeus brasiliensis). The F. subtilis represents more than 85% of shrimp landings. We denoted by C the total catch of this shrimp in tons for the whole French Guiana fleet. This catch C is the product of the catchability coefficient q, the fishing effort measured by the number of days at sea E and the abundance of the fish population B. Based on the Schaeffer relation $C = qEB$, the catch-per-unit-effort (CPUE) is equal to ratio C/E. CPUE is the catch extracted from one unit of fishing effort. Changes in CPUE are assumed to correspond to a proxy of proportional changes in the abundance of the fish population modulo the catchability coefficient (Clark, 2000; Dunn, Harley, Doonan, & Bull, 2000). We use the data collected by IFREMER (French institute of research for the exploitation of the sea) on C and E between January 1989 to December 2012 to get the CPUE. The monthly sea surface temperature (SST) refers to an indicator of environmental conditions for the shrimp fisheries. It is provided by the National Climatic Data Center (NCDC) database. This monthly one-degree global SST climatology was constructed using the analysis carried out by the Climate Prediction Center (CPC/NOAA). The area coverage of the data is 2°9°N and 47°59°W. As mentioned in the introduction, SST appears to be a major driver in fish abundance changes through its impact on the intrinsic growth rate and the carrying capacity of the species. Sanz et al. (2016) estimate a Schaeffer relation and show that SST appears to be significant among a set of other environmental variables.

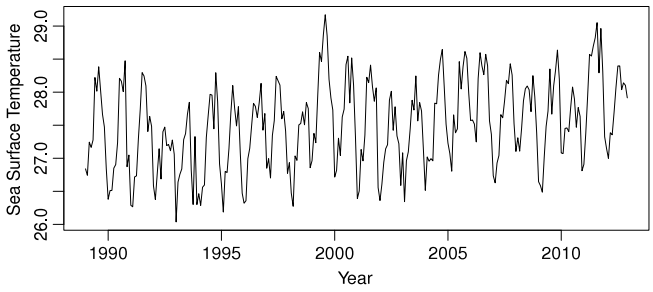
CPUE and SST refer respectively to the dependent Y_t and exogenous X_t variables. Time series Y_t and X_t have 288 monthly observations. In the SST series, two missing values have been replaced by the mean of all other values. The plots of the two series are depicted in Fig. 1 and exhibit a trend and a seasonal component for the two series. As we analyze monthly data, the period $s = 12$ has been naturally selected.

In order to stabilize the variance and make the data more normal distribution-like, we use a Box–Cox transformation for both variables. The trend was removed by using the first difference of the transformed data. The detrended series are depicted in Fig. 2. Results indicate that they are stationary.

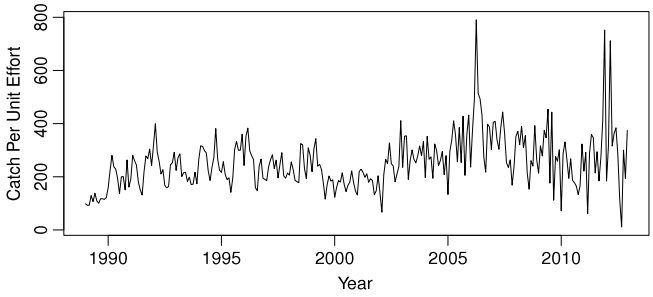
Vector autoregressive (VAR) models are often used to model population time series data. However instead of considering season as categorical variable as in Zhou, Fujiwara, and Grant (2016), our initial model consists in a 2-dimensional vector periodic autoregressive (PVAR) given by:

$$\mathbf{Y}_{ns+\nu} = \sum_{k=1}^{p(\nu)} \mathbf{\Phi}_k(\nu)\mathbf{Y}_{ns+\nu-k} + \boldsymbol{\epsilon}_{ns+\nu}, \tag{9}$$

where $\mathbf{Y}_t = (Y_t, X_t)^\top$ is a 2×1 vector and $\mathbf{\Phi}_k(\nu) = \left(\Phi_{k,ij}(\nu)\right)_{i,j=1,2}, k = 1, \ldots, p(\nu)$, are the autoregressive model coefficients during season ν. For a detailed presentation of PVAR models see Ursu and Duchesne (2009).

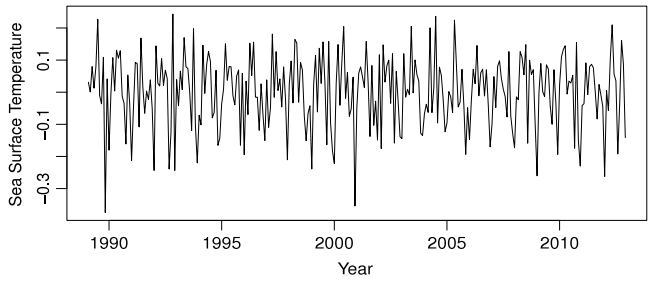


(a) Sea surface temperature.

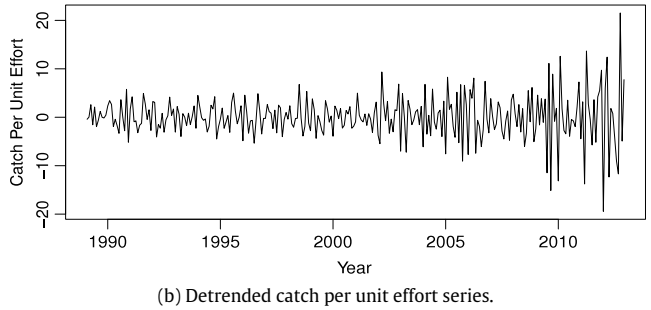


(b) Catch per unit effort.

Fig. 1. Plot of average monthly sea surface temperature and catch per unit effort between 1989 and 2012.



(a) Detrended sea surface temperature.



(b) Detrended catch per unit effort series.

Fig. 2. Plot of the first difference of transformed time series.

Periodic vector AR models of orders $p(\nu) = 1, \ldots, 5$ have been fitted by the least squares method. The least squares estimators show that the PVAR coefficients in the (2, 1) positions of the $\Phi_k(\nu)$ matrix denoted by $\Phi_{k,21}(\nu)$ are all close to zero and are very small relative to their respective standard errors (not presented here). This suggests that the input variable X_t is not influenced by the variable Y_t, and, hence, that X_t is exogenous for Y_t. Estimation results suggest the relevance of a PARX model given in Eq. (1) which relates the SST to the CPUE. Such a relation has been analyzed by Stoffer (1986) in a spatial context.

Table 6
Fitting a PARX model to the shrimp data.

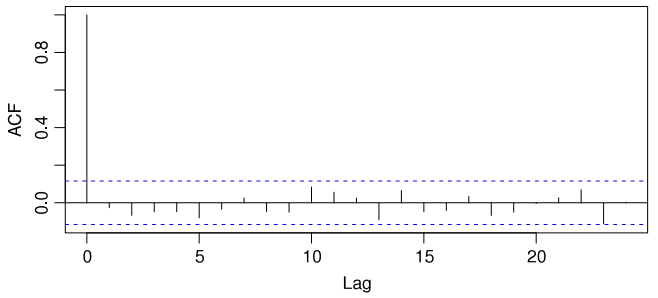
P	M	BIC	nb. of parameters	nb. of possible models
12	7	599.78	77	12582912
11	8	620.43	78	12582912
11	9	610.12	81	25165824
12	8	595.39	71	25165824

Table 7
Parameters estimates and their standard errors (in parentheses) for the PARX model. The residual standard deviation is given in the last column. Period 1 corresponds to February.

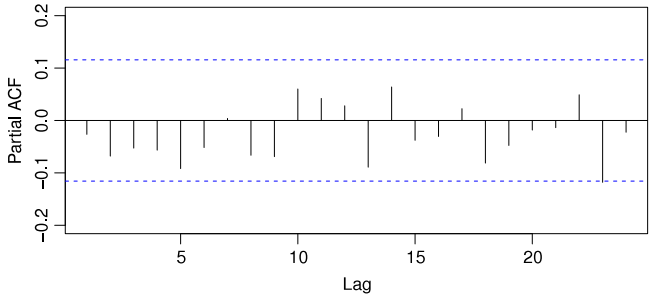
Parameter estimates	Season											
	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{\phi}_1$	−0.49 (0.09)		−0.75 (0.09)	−0.43 (0.05)		−0.89 (0.21)		−0.83 (0.07)	−0.62 (0.14)	−0.83 (0.11)		−1.23 (0.16)
$\hat{\phi}_2$				−0.36 (0.06)				−0.29 (0.08)	−0.40 (0.13)	−0.19 (0.11)		−0.72 (0.19)
$\hat{\phi}_3$			−0.27	−0.57 (0.08)	(0.07)		0.43 (0.18)	−0.67 (0.11)		0.44 (0.11)		−0.52 (0.20)
$\hat{\phi}_4$				−0.28 (0.05)				−0.88 (0.10)				
$\hat{\phi}_5$							−0.21 (0.10)	−0.67 (0.07)				−0.35 (0.16)
$\hat{\phi}_6$			−0.46 (0.11)					−0.74 (0.06)		−0.27 (0.11)	0.55 (0.22)	
$\hat{\phi}_7$											0.70 (0.21)	
$\hat{\phi}_8$			−0.57 (0.11)								0.29 (0.11)	
$\hat{\phi}_9$								−0.48 (0.06)				
$\hat{\phi}_{10}$		0.64 (0.26)						−1.21 (0.07)				
$\hat{\phi}_{11}$		0.74 (0.16)					0.52 (0.13)	−0.73 (0.09)	0.74 (0.19)	0.63 (0.10)	−0.45 (0.09)	
$\hat{\phi}_{12}$				−0.22 (0.08)			0.33 (0.15)	−0.51 (0.07)		0.46 (0.13)		
$\hat{\theta}_0$	17.31 (7.47)									7.44 (2.32)		
$\hat{\theta}_1$		−20.27 (9.21)	12.14 (4.63)	−10.63 (2.44)				16.09 (2.41)				
$\hat{\theta}_2$					13.27 (4.49)				−17.00 (6.84)			
$\hat{\theta}_3$			12.52 (2.83)				−18.42 (4.43)	26.29 (2.79)		−11.38 (3.92)		
$\hat{\theta}_4$								−9.52 (2.77)				−12.92 (6.64)
$\hat{\theta}_5$		−10.02 (6.44)		−4.29 (1.58)		17.87 (7.12)	28.42 (5.35)	−17.89 (3.24)		−9.15 (3.71)		
$\hat{\theta}_6$						19.43 (3.44)		−7.56 (3.86)		19.61 (3.20)		−13.79 (6.48)
$\hat{\theta}_7$				8.14 (2.08)				−24.01 (2.97)				
$\hat{\theta}_8$			−27.30 (6.05)									
$\hat{\sigma}$	6.96	9.39	3.33	0.82	4.19	5.93	4.59	0.52	6.75	1.61	5.96	7.69

GA is run for 50 iterations with a size of the population equal to $N_p = 40$, a probability of crossover of $P_c = 0.8$, a mutation probability $P_m = 0.01$ and with one elite individual. Several maximal values for AR order and for exogenous variables have been considered. The order of $P = 12$ seems to be a reasonable choice and implies that for each observation, lags from a whole year are included.

Table 6 shows the results. We choose the last model $P = 12, M = 8$ for our case-study which implies a good balance in the trade-off value of BIC and number of parameters. Let us remark that the number of parameters decreases from 252 in the saturated model (in which all the parameters are estimated) to 71 in our model. Estimates of the parameters and their associated standard errors are given in Table 7.

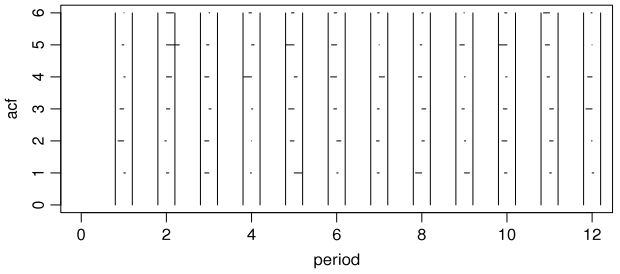


(a) Autocorrelation function for residuals.

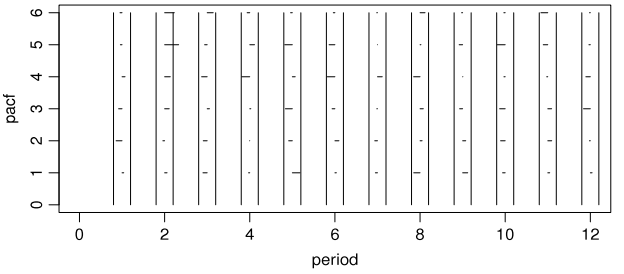


(b) Partial autocorrelation function for residuals.

Fig. 3. The ACF and PACF plots of the model residuals, showing the bounds $\pm 1.96/\sqrt{Ns}$.



(a) Periodic ACF of residuals.



(b) Periodic PACF of residuals.

Fig. 4. The periodic ACF and PACF plots of the model residuals.

The analysis of the residuals for the estimated model $P = 12, M = 8$ is needed to test its relevance. The stationarity of the residuals allows us to apply the standard 95% confidence limits (that is $1.96\sqrt{Ns}$). The usual sample autocorrelation (ACF) and partial autocorrelation (PACF) plots show no significant autocorrelation (Fig. 3). We examine also the periodic ACF (Fig. 4(a)) and PACF (Fig. 4(b)) as in Hipel and McLeod (1994). The residuals for each season fall within the confidence limits (that is $1.96\sqrt{N}$), meaning that the seasonal residuals are white. The normal probability plot of the residuals shows no obvious violation of the normality assumption (Fig. 5).

This numerical example shows that our identification and estimation methods perform well for an analysis of the relation between the CPUE and SST as in Stoffer (1986). To show the interest of PARX model, results have to be compared with the ones we could obtained with PAR models for each variable and in particular for the variable of interest. The results of the PAR model for the CPUE variable give a $BIC = 686.1$ for 41 estimates parameters. The comparison between the PAR and the PARX models shows an improvement of the BIC criteria with a value of $BIC = 595.39$ at the expense of more parameters to estimate, 71. We also compared our selection method based on GA with other methods like the stepwise regression. Using Matlab stepwise routine, the value for the BIC criteria is $BIC = 690.17$ and the number of estimated parameters is 24. This example puts forward the trade-off between the value of the BIC and the number of estimated parameters. However the performance of the automatic procedure based on GA with respect to other methods allows to investigate a large range of models and to select the best one as in the case-study.

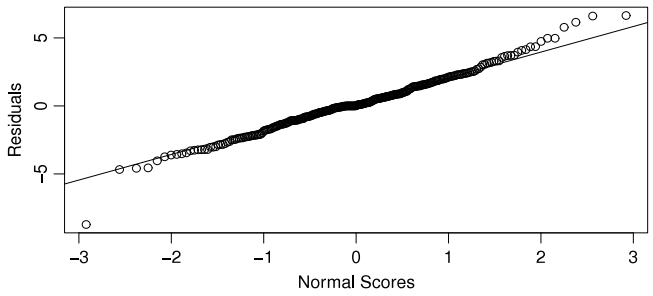


Fig. 5. Normal probability plot of residuals.

This paper shows that the PARX model gives better results than the PAR model in both simulation experiments and the real case-study. It also shows that the procedure of variables selection is more efficient with GA method than other methods like stepwise as pointed out by Hocking (1976).

6. Conclusions

This paper studies the properties of identification and estimation procedure for periodic autoregressive models with one exogenous variable (PARX). This procedure is based on genetic algorithm combined with Bayesian Information Criterion which is used to investigate the space of solutions and to select the set of parameters that corresponds to the best model. Results of several Monte-Carlo simulation experiments show that PARX models with different number of seasons can be correctly identified by the proposed procedure.

Using time series data from the shrimp French Guiana fishery, a PARX model relying the catch-per-unit-effort as dependent variable to the sea surface temperature as exogenous variable has been estimated. Results show that the PARX model performs better that the PAR model applied only for the dependent variable. The introduction of the exogenous variable improves the Bayesian Information Criterion at the expense of more parameters to estimate. However the trade-off remains in favor of the use of PARX model since the main advantage of our procedure is a sharp fall of the number of parameters to estimate with respect to the saturated model. It turns out that this method appears to be easier to implement for providing short term forecasts to fisheries managers.

Considering PARX models with several exogenous variables and multivariate PARX models should be interesting to analyze. In both cases the number of models to investigate becomes so huge that the use of genetic algorithms is needed to overcome the computational challenge and the difficulty for researchers and decision-makers to identify the right model. In a marine fishery context for instance, genetic algorithms could be used to select a subset of exogenous variables among the sea surface temperature, the wind velocity, the nutrient flows from rivers, rain precipitation or the southern oscillation index in the Pacific Ocean concerning the El Niño-La Niña phenomena that explain fish recruitment and harvest (Sanz et al., 2016). Genetic algorithms could also be useful to analyze space–time models by considering multivariate PARX models dealing with different fishing areas and different environmental conditions.

Acknowledgments

The authors thank L. Baulier and F. Blanchard from the Ifremer in French Guiana for their comments and providing data for the case-study. This work has been carried out with the financial support of the CNRS and from the French National Research Agency (ANR) in the frame of the Investments for the future Programme, within the Cluster of Excellence COTE (ANR-10-LABX-45).

References

Agosto, A., Cavaliere, G., Kristensen, D., & Rahbek, A. (2016). Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX). *Journal of Empirical Finance*, *38*, 640–663.

Andel, J. (1987). Periodic autoregression with exogenous variables and equal variances. In Grossmann, Mogyoródi, Vincze, & Wertz (Eds.), *Proc. 5th pannonian symp.* (pp. 237–245). Budapest: Akadémiai Kiadó.

Andel, J. (1989). Periodic autoregression with exogenous variables and periodic variances. *Aplikace Matematiky*, *34*, 387–395.

Angelini, G., & Angelis, L. (2016). PARX model for football matches predictions. *Working paper* (pp. 1–26). University of Bologna.

Baragona, R., & Battaglia, F. (2009). Evolutionary computing in statistical data analysis. In A. Abraham, A. E. Hassanien, P. Siarry, & A. Engelbrecht (Eds.), *Global optimization: Vol. 3. Fundations of computational intelligence* (pp. 347–386). Springer.

Baragona, R., Battaglia, F., & Cucina, D. (2004). Fitting piecewise linear threshold autoregressive models by means of genetic algorithm. *Computational Statistics & Data Analysis*, *47*, 277–295.

Basawa, I. V., & Lund, R. B. (2001). Large sample properties of parameter estimates for periodic ARMA models. *Journal of Time Series Analysis*, *22*, 651–663.

Brander, K. (2007). Global fish production and climate change. *Proceedings of the National Academy of Sciences*, *104*(50), 19709–19714.

Cheung, W., Lam, V. W., Sarmiento, J. L., Kearney, K., Watson, R., & Pauly, D. (2009). Projecting global marine biodiversity impacts under climate change scenarios. *Fish and Fisheries*, *10*(3), 235–251.
Clark, C. (2000). *The worldwide crisis in fisheries: economic models and human behavior*. Cambridge University Press.
Duchesne, P., & Lafaye de Micheaux, P. (2013). Distributions for residual autocovariances in parsimonious periodic vector autoregressive models with applications. *Journal of Time Series Analysis*, *34*, 496–507.
Dudek, A. E., Hurd, H., & Wojtowicz, W. (2014). PARMA models with applications in R. In F. Chaari, J. Leskow, A. Napolitano, R. Zimroz, A. Wylomanska, & A. Dudek (Eds.), *Cyclostationarity : theory and methods* (pp. 131–153). Springer.
Dunn, A., Harley, S., Doonan, I., & Bull, B. (2000). Calculation and interpretation of catch-per-unit-effort (CPUE) indices, *New Zealand fisheries assessment report* (pp. 1–44).
Franses, P. H., & Paap, R. (2004). *Periodic time series models*. Oxford University Press.
Gaetan, C. (2000). Subset ARMA model identification using genetic algorithms. *Journal of Time Series Analysis*, *21*, 559–570.
Gladyshev, E. G. (1961). Periodically correlated random sequences. *Soviet Mathematics*, *2*, 385–388.
Goldberg, D. E. (1989). *Genetic algorithms in search, optimization & machine learning*. Addison-Wesley.
Haddon, M. (2011). *Modelling and quantitative methods in fisheries*. (2nd ed.). Taylor and Francis Group, LLC.
Hipel, K. W., & McLeod, A. I. (1994). *Time series modelling of water resources and environmental systems*. Amsterdam: Elsevier.
Hocking, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics*, *32*, 1–49.
Jeong, C., & Kim, J. (2013). Bayesian multiple structural change-points estimation in time series models with genetic algorithm. *Journal of the Korean Statistical Society*, *42*, 459–468.
Jones, R. H., & Brelsford, W. (1967). Time series with periodic structure. *Biometrika*, *54*, 403–408.
Lassen, H., & Medley, P. (2001). Virtual population analysis - A practical manual for stock assessment. *FAO fisheries technical paper 400*.
Li, S., & Lund, R. (2012). Multiple changepoint detection via genetic algorithms. *Journal of Climate*, *25*, 674–686.
Lu, Q., Lund, R., & Lee, T. C. M. (2010). An MDL approach to the climate segmentation problem. *The Annals of Applied Statistics*, *4*, 299–319.
Lund, R. B., & Basawa, I. V. (2000). Recursive prediction and likelihood evaluation for periodic ARMA models. *Journal of Time Series Analysis*, *21*, 75–93.
Lund, R. B., Shao, Q., & Basawa, I. V. (2006). Parsimonious periodic time series modeling. *Australian & New Zealand Journal of Statistics*, *48*, 33–47.
Lund, R. B., Wang, X. L., Lu, Q., Reeves, J., Gallagher, C., & Feng, Y. (2007). Changepoint detection in periodic and autocorrelated time series. *Journal of Climate*, *20*, 5178–5190.
Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin: Springer.
Maçaira, P. M., Oliveira, F. L. C., Ferreira, P. G. C., de Almeida, F. V. N., & Souza, R. C. (2017). Introducing a causal PAR(p) model to evaluate the influence of climate variables in reservoir inflows : a brazilian case. *Pesquisa Operacional*, *37*(1), 107–128.
McLeod, A. I. (1994). Diagnostic checking periodic autoregression models with applications. *Journal of Time Series Analysis*, *15*, 221–233.
Minerva, T., & Poli, I. (2001). Building ARMA modles with genetic algorithms. In E. J. W. Boers, et al. (Eds.), *LNCS*: *Vol. 2037*. *Applications of evolutionary computing* (pp. 335–342). Springer.
Mitchell, M. (1996). *An introduction to genetic algorithms*. MIT Press.
Noakes, D. J., McLeod, A. I., & Hipel, K. W. (1985). Forecasting monthly riverflow time series. *International Journal of Forecasting*, *1*, 179–190.
Paroli, R., & Spezia, L. (2008). Bayesian inference in non-homogeneous Markov mixtures of periodic autoregressions with state-dependent exogenous variables. *Computational Statistics & Data Analysis*, *52*, 2311–2330.
Reinsel, G. C. (1997). *Elements of multivariate time series analysis*. (2nd ed.). Springer.
Sanz, N., Diop, B., Blanchard, F., & Lampert, L. (2016). On the influence of environmental factors on harvest: the French Guiana shrimp fishery paradox. *Environmental Economics and Policy Studies*. http://dx.doi.org/10.1007/s10018-016-0153-6. Online First Articles.
Schlick, C. M., Duckwitz, S., & Schneider, S. (2013). Project dynamics and emergent complexity. *Computational and Mathematical Organization Theory*, *19*, 480–515.
Song, L., & Bondon, P. (2013). Structural changes estimation for strongly-dependent processes. *Journal of Statistical Computation and Simulation*, *83*, 1783–1806.
Sparre, P., & Venema, S. (1998). Introduction to tropical sh stock assessment. Part 1 manual. FAO. Fish. Tech. Pap. 306/1.
Stoffer, D. (1986). Estimation and identification of space-time ARMAX Models in the presence of missing data. *Journal of the American Statistical Association*, *81*, 762–772.
Tesfaye, Y. G., Meerschaert, M. M., & Anderson, P. L. (2006). Identification of periodic autoregressive moving average models and their application to the modeling of river flows. *Water Resources Research*, *42*, 1–11.
Ursu, E., & Duchesne, P. (2009). On modelling and diagnostic checking of vector periodic autoregressive time series models. *Journal of Time Series Analysis*, *30*, 70–96.
Ursu, E., & Pereau, J. C. (2016). Application of periodic autoregressive process to the modeling of the Garonne river flows. *Stochastic Environmental Research and Risk Assessment*, *30*, 1785–1795.
Ursu, E., & Turkman, K. F. (2012). Periodic autoregressive model identification using genetic algorithm. *Journal of Time Series Analysis*, *33*, 398–405.
Vecchia, A. V. (1985). Periodic autoregressive-moving average (PARMA) modeling with applications to water resources. *Water Resources Bulletin*, *21*, 721–730.
Wei, W. S. W. (2006). *Time series analysis: univariate and multivariate methods*. (2nd ed.). Pearson.
Wojciechowski, E. (2001). *Identification of PARX models with hydrological and hydrogeological applications. Thesis*. Stevens Institute of Technology. 300 North Zeeb Road, Ann Arbor, MI 48106-1346.
Wu, B., & Chang, C.-L. (2002). Using genetic algorithms to parameters (d,r) estimation for threshold autoregressive models. *Computational Statistics & Data Analysis*, *38*, 315–330.
Yau, C. Y., Tang, C. M., & Lee, T. C. M (2015). Estimation of multiple-regime threshold autoregressive models with structural breaks. *Journal of the American Statistical Association*, *511*, 1175–1186.
Zhou, C., Fujiwara, M., & Grant, W. E. (2016). Finding regulation among seemingly unregulated populations: a practical framework for analyzing multivariate population time series for their interactions. *Environmental and Ecological Statistics*, *23*, 181–204.