A GENERALIZATION OF PERIODIC AUTOREGRESSIVE MODELS FOR SEASONAL TIME SERIES

Francesco Battaglia
Domenico Cucina
Manuel Rizzo

Dipartimento di Scienze Statistiche
Rapporto Tecnico N. 2 2018

SAPIENZA
UNIVERSITÀ DI ROMA

ISSN: 2279-798X

A generalization of periodic autoregressive models for seasonal time series

Francesco Battaglia*, Domenico Cucina**, Manuel Rizzo*

*Department of Statistical Sciences, University of Rome La Sapienza, Piazzale Aldo Moro 5, I-00100 Roma, Italy

**Department of Economics and Statistics, University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, Italy

Abstract

Many nonstationary time series exhibit changes in the trend and seasonality structure, that may be modeled by splitting the time axis into different regimes. We propose multi-regime models where, inside each regime, the trend is linear and seasonality is explained by a Periodic Autoregressive model. In addition, for achieving parsimony, we allow season grouping, i.e. seasons may consist of one, two, or more consecutive observations. Since the set of possible solutions is very large, the choice of number of regimes, change times and order and structure of the Autoregressive models is obtained by means of a Genetic Algorithm, and the evaluation of each possible solution is left to an identification criterion such as AIC, BIC or MDL. The comparison and performance of the proposed method are illustrated by a real data analysis. The results suggest that the proposed procedure is useful for analyzing complex phenomena with structural breaks, changes in trend and evolving seasonality.

Keywords: Genetic algorithms, Structural break, Regime change

1. Introduction

Many seasonal time series exhibit an autocorrelation structure which depends not only on the time between observations but also on the season of the year. Moreover, the time series of observations within a given season is usually second order stationary [9]. In order to model appropriately these and similar types of time series, Periodic AutoRegressive models (PAR) can

be employed. These models are appropriate, for instance, for describing time series drawn from different areas such as economy, hydrology, climatology and signal processing ([7, 9, 19]). When fitting a PAR model to periodic time series a separate AR model for each season of the year is estimated.

In this study we consider a generalization of PAR models with linear trend in two directions. First, the model may follow different regimes in time, and regime changes may occur at any time. The regime changes may affect the linear trend, the seasonal means and the autoregressive parameters. We also allow a discontinuous trend which can identify changes in level. Second, inside each regime the model structure may be different for each seasonal position (e.g. months) or vary more slowly, changing only according to grouped seasons like quarters, trimesters or semesters.

The number of regimes and change times (or break point) are assumed to be unknown. The problem of their identification can be treated as a statistical model selection problem according to a specified identification criterion [1]. This approach has been used for identification of structural breaks in [5] for piecewise AR process, in [6] for piecewise generalized autoregressive conditionally heteroscedastic (GARCH) processes and in [11] to detect structural breaks in the form of level shifts in climatic time series. In these works a minimum description length (MDL) principle is adopted as identification criterion, and Genetic Algorithms (GAs) are proposed to solve the selection problem.

A classical approach formulates the structural change problems as hypothesis tests where the null is set to 'no structural change' and the alternative contains one or multiple structural breaks. The CUSUM and supF tests are the most popular methods to detect a level change and parameter change in statistical models [1]. In a PAR framework, [12] proposed a modified supF test to detect structural changes for series with autocorrelated and periodic features. In the latter paper the focus is on mean changes and not on autocovariance, trend, or other types of changes. To the best of our knowledge, there are no articles that handle the changing parameters and changing trend problem in PAR models simultaneously. Motivated by this, we proposed a class of GAs to detect the number of regimes of piecewise PAR models and their locations. Our procedure evaluates several regime patterns where the locations that are possibly change times are simultaneously considered. In this way, GAs deal efficiently with the detection of multiple change points. We also allow subset AR models to be selected. Each piecewise subset PAR configuration is evaluated by an AIC identification criterion. So, the GAs

seem able to provide more flexibility and adaptation to the change time detection problem.

A potential drawback of using our model in applications is that the model often requires the use of a substantial number of parameters. This total number of parameters can be large even for moderate values of time series periodicity and no break. Moreover changes in periodic processes from season to season may be slow and more detailed models would be redundant. Many authors have investigated parsimonious versions of PAR models. [17] suggest grouping similar seasons into blocks to reduce parameter totals; alternatively, [4] consider modeling slow seasonal changes in parameters with Fourier series.

In our paper, since the seasonal effect on means, variances and correlations may show different speed and pattern, we propose to join appropriate season parameters into groups for each of these three features. Three different heuristic strategies are implemented to find the optimal season grouping for the mean, the variance and the correlation. The empirical results showed that different season grouping for mean, variance and correlation produces parsimonious piecewise PAR models with better fitting.

2. Model

Suppose that a time series $\{X_t\}$ of N observations is available. The seasonal period of the series is s and is assumed to be known. Assume that there are $m+1$ different regimes, separated by m change times τ_j so that the first regime contains observations from time 1 to $\tau_1 - 1$, the second regime contains data from time τ_1 to $\tau_2 - 1$, the $(j+1)$-th regime contains data from τ_j to $\tau_{j+1} - 1$, and the last regime data from τ_m to N. To ensure reasonable estimates we assume that the minimum regime length is a fixed constant mrl, so that $\tau_j \geq \tau_{j-1} + mrl$ for any j. Any regime assignment is defined by the set $\{\tau_j, j = 1, \ldots, m\}$ subject to

$$mrl < \tau_1 < \tau_2 < \ldots < \tau_m < N - mrl \quad , \quad \tau_j \geq \tau_{j-1} + mrl \ , \ j = 2, \ldots, m.$$

The parameters of the model for regime j will be denoted by a superscript (j).

A linear trend and a different mean for each season is assumed. The residuals are treated as zero mean and described by an autoregressive model with maximum order p, and parameters varying with seasons. Let k_t denote the season of the t-th observation ($1 \leq k_t \leq s$), denote by $a^{(j)} + b^{(j)}t$ the

linear trend in regime j, by $\mu^{(j)}(k)$ the mean of season k in regime j, and by $\phi_k^{(j)}(i)$ the lag-i autoregressive parameter for the model in season k and regime j. Then

$$X_t = a^{(j)} + b^{(j)}t + \mu^{(j)}(k_t) + W_t \ , \ \ \tau_{j-1} \leq t < \tau_j$$

$$W_t = \sum_{i=1}^{p} \phi_{k_t}^{(j)}(i) W_{t-i} + \varepsilon_t \ , \ \ \tau_{j-1} \leq t < \tau_j$$

where $\tau_0 = 1$ and $\tau_{m+1} = N + 1$.

The innovations ε are supposed independent and zero-mean, with variances $\sigma^2(j,k)$ possibly depending on the regime and season.

As far as subset selection is concerned, we introduce also $m+1$ binary vectors $\delta^1, ..., \delta^{m+1}$, which specify presence or absence of autoregressive parameters in each regime as follows: if $\delta^j[p(k_t - 1) + i] = 1$ then $\phi_{k_t}^{(j)}(i)$ is constrained to zero. In summary, a model is identified by the following:

- **external parameters** (fixed and equal for all models)
  N, s, maximum order p, maximum number of regimes M, and minimum number of observations per regime mrl

- **structural parameters** (determining the model structure)

m	number of change points, from 0 to $M-1$
$\tau_1, \tau_2, \ldots, \tau_m$	change times or thresholds
$\delta^1, ..., \delta^{m+1}$	denote which ϕ are zero in each regime and season

- **regression parameters** (to be estimated by Least Squares (LS) or Maximum Likelihood (ML))

$a_1, a_2, \ldots, a_{m+1}$	intercepts
$b_1, b_2, \ldots, b_{m+1}$	slopes
$\mu^{(j)}(k)$	seasonal means, $k = 1, \ldots, s$; $j = 1, \ldots, m+1$
$\phi_k^{(j)}(i)$	AR parameters, $k = 1, \ldots, s$; $j = 1, \ldots, m+1$; $i = 1, \ldots, p$ (some of them may be constrained to zero)
$\sigma^2(j,k)$	innovation variances, regime j, season k

For estimating trend and seasonal means by LS, note that the intercept and the means are linearly dependent, therefore we assume that the seasonal means sum to zero on one cycle:

$$\mu^{(j)}(1) + \mu^{(j)}(2) + \ldots, + \mu^{(j)}(s) = 0 \ \ \forall j. \tag{1}$$

Therefore the following equations are estimated:

$$X_t = b^{(j)}t + c(j,k_t) \quad , \quad \tau_j \leq t < \tau_{j+1} \tag{2}$$

and then the parameter vector is

$$\beta' = \left\{b^{(1)}, b^{(2)}, \ldots, b^{(m+1)}, c(1,1), c(1,2), \ldots, c(1,s), c(2,1), \ldots,\right.$$
$$\left. c(2,s), \ldots, c(m+1,1), \ldots, c(m+1,s)\right\}$$

with dimension $(m+1)\times(s+1)$. Each row of the design matrix relates to a time t from 1 to N and has the value t in the column corresponding to $b^{(j)}$, and 1 in the column $m+(j-1)s+k_t$ that corresponds to parameter $c(j,k_t)$. For determining k_t we compute $\mod(t-1+s,s)$. If the result is 0 we are in the first season, if the result is 1 we are in the second season, and so on. It follows that

$$k_t = \mod(t-1+s,s)+1.$$

From the $\{\hat{c}(j,k)\}$, the intercepts $\hat{a}^{(j)}$ and seasonal means $\hat{\mu}^{(j)}(k)$ are recovered basing on assumption (1). It follows

$$\hat{a}^{(j)} = \frac{1}{s}\sum_{k=1}^{s}\hat{c}(j,k) \quad , \quad \hat{\mu}^{(j)}(k) = \hat{c}(j,k) - \hat{a}^{(j)}.$$

Moreover it is possible to prescribe trend continuity by imposing that, if the number of regimes is larger than one, the trend values of two consecutive regimes coincide on the first observation of the second regime. For imposing a continuity constraint on the trend $a^{(j)} + b^{(j)}\,t$ between consecutive regimes, we consider constrained LS with the constraints $H\beta = 0$, where each row of the matrix H imposes that the trend part coincides at the first time of each successive regime:

$$b^{(j)}\tau_{j+1} - b^{(j+1)}\tau_{j+1} + \frac{1}{s}\sum_{k=1}^{s}c(j,k) - \frac{1}{s}\sum_{k=1}^{s}c(j+1,k) = 0.$$

It follows that H has m rows and the j-th row has value τ_{j+1} in the j-th entry and $-\tau_{j+1}$ in the $(j+1)$-th entry, that the entries from $m+(j-1)\times s+1$ to $m+j\times s$ are all equal to $1/s$ and the entries from $m+j\times s+1$ to $m+(j+1)\times s$ are all equal to $-1/s$.

The trend continuity is controlled by an indicator (if $indT \neq 0$ then the continuity constraints are added). A possible level change at $t=\tau_{j+1}$ is estimated if the trend continuity is not imposed.

2.1. Parameters of the periodic autoregressive model

Conditioning on thresholds, seasonal arrangement and estimated trend and means, the residual series is computed:

$$\hat{W}_t = X_t - \hat{a}^{(j)} - \hat{b}^{(j)}t - \hat{\mu}^{(j)}(k_t).$$

For each regime j and season k a separate autoregressive process is considered:

$$\hat{W}_t = \sum_{i=1}^{p} \phi_{k_t}^{(j)}(i)\hat{W}_{t-i} + \varepsilon_t.$$

We denote by $I(j,k)$ the set of times belonging to regime j and season k ($\tau_j \leq t < \tau_{j+1}$ such that $k_t = k$). For each pair (j,k) the observation indexes belonging to the subseries $I(j,k)$ are selected and the LS estimates of the parameters $\{\phi_k^{(j)}(i), i = 1,\ldots,p\}$ are obtained, together with the related residual variance estimate $\hat{\sigma}^2(j,k)$. For each pair regime j, season k ($j = 1,\ldots,M; k = 1,\ldots,s$), the data are the (j,k) subseries $\{\hat{W}_t, t \in I(j,k)\}$ that we put in a vector z with n_{jk} rows and one column. The design matrix $Z : n_{jk} \times p$ has rows containing the regressors $\hat{W}_{t-1}, \hat{W}_{t-2}, \ldots, \hat{W}_{t-p}$ and the parameter vector to be estimated is $\phi = [\phi_k^{(j)}(1), \phi_k^{(j)}(2), \ldots, \phi_k^{(j)}(p)]$.

In each regime and season, all possible subset models may be estimated. If $q_{jk} = \delta^j[p(k-1)+1] + \delta^j[p(k-1)+2] + \ldots + \delta^j[p(k-1)+p]$ denotes the number of AR parameters constrained to zero in regime j and season k, the constrained lags matrix H is a $q_{jk} \times p$ matrix that designates the AR parameters to be constrained to zero, and depends on the binary sequence δ^j in this way: for each value r from 1 to p, if $\delta^j[p(k-1)+r] = 1$, add a row to H equal to the r-th row of I_p identity matrix of order p.

Then the AR parameters for regime j and season k are estimated by least squares constrained optimization with linear constraint given by the $q_{jk} \times p$ linear system $H\phi = 0$:

$$\hat{\phi} = \phi_{LS} - (Z'Z)^{-1}H'[H(Z'Z)^{-1}H']^{-1}H\phi_{LS},$$

where $\phi_{LS} = (Z'Z)^{-1}Z'z$ are the unconstrained least squares estimates (this ensures that $\hat{\phi}_r = 0$ if $\delta^j[p(k-1)+r] = 1$) and the residuals

$$e = z - Z\,\hat{\phi}$$

give the estimate of the innovations for regime j and season k $\{\varepsilon_t, t \in I(j,k)\}$.

The number of change times m, the time thresholds $\tau_1, \tau_2, \ldots, \tau_m$ and the indicators δ of the $\phi_k^j(i)$ coefficients constrained to zero in each segment are called the structural parameters. They take discrete values and their combinations amount to a very large number. GAs are naturally suitable for the choice of optimal structural parameters.

3. Genetic algorithms

Many optimization problems do not satisfy the necessary conditions to guarantee the convergence of traditional numerical methods. For instance, in order to apply standard gradient methods to maximum likelihood estimation we need a globally convex likelihood function, however there are a number of relevant cases with non-convex likelihood functions or functions with several local optima. Another class of "hard" problems is when the solution space is discrete and large. These problems are known as combinatorial problems. A simple approach for solving an instance of a combinatorial problem is to list all the feasible solutions, evaluate their objective function, and pick the best. However, for a combinatorial problem of a reasonable size, the complete enumeration of its elements is not computationally feasible, and most available searching algorithms are likely to yield some local optimum as a result [13].

GAs are often used to solve such problem instances. These methods do not rely on a set of strong assumptions about the optimization problem, on the contrary they are robust to changes in the characteristics of the problem. But, on the other side, they do not produce a deterministic solution but a high quality stochastic approximation to the global optimum.

GAs have been initially developed by [10] and they are classified as *population based* methods and *evolutionary algorithms*. They work on a whole set of solutions that is adapted simultaneously by imitating the evolutionary process of species that reproduce sexually. We give a brief sketch of this method.

GAs imitate the evolution process of biological systems, to optimize a given function. They use a set of candidate solutions, called *population*, instead of one single current solution. In GA terminology, any candidate solution is encoded via a numerical vector called *chromosome*. The GA proceeds iteratively by updating the population of active chromosomes (the sets of current candidate solutions) in rounds, called generations. In each generation, some of the active chromosomes are selected (parents-chromosomes) to

form the chromosomes of the next generation (children-chromosomes). The selection process is based on an evaluation measure called *fitness function*, linked to the objective function, that assigns to each chromosome a positive number. This fitness is the determining factor for the probability to select a chromosome as a parent. A higher fitness value leads to higher probability that the corresponding chromosome will be one of the parents used to form the children-chromosomes. Children are formed by recombining (*crossover*) the genetic material of their two parents-chromosomes and perhaps after a random alteration of some of the genes (single digits of the chromosome), which is called *mutation* [see 10, 8, for a detailed description].

4. Estimation of structural parameters

This section further describes the algorithms implementation we used for building the model. A successful implementation of GAs is certainly crucial to obtain satisfactory results. Before a GA can be applied to a problem some important decisions have to be made. The GA methods require a suitable encoding for the problem and an appropriate definition of objective function. In addition operators of selection, crossover and mutation have to be chosen. The following sections describe the choices made.

4.1. Solution Encoding

An appropriate encoding scheme is a key issue for GAs. The chromosome genotype is a sequence of binary digits in one-to-one correspondence with the phenotype. In order to set up a GA, each model is represented by a chromosome phenotype that may be associated to the following integer array:

$$m,\ \tau_1,\ \tau_2, \ldots, \tau_{M-1}, \delta^1,\ \delta^2,\ \ldots, \delta^M, \tag{3}$$

where m is between 0 and $M-1$, the tau's satisfy $mrl+1 \leq \tau_1$, $\tau_1+mrl \leq \tau_2$, $\tau_2 + mrl \leq \tau_3$ and finally $\tau_m \leq N - mrl - 1$. The m change times τ_j are coded using m numbers $th_1, th_2, \ldots, th_m$, each with value in (0,1), as in [3].

The deltas are binary sequences specifying what lags are constrained (associated with ϕ's equal zero) in regime j and the consecutive seasons: each delta is a vector of $p \times s$ binary digits and if digit is 0 the corresponding AR parameter is in the model for that season and regime, otherwise it is missing. More precisely, if $\delta^j[p(k-1)+i] = 1$ then $\phi_i^j(k)$ is constrained to zero. Genes are encoded in the following way:

One gene encodes m. If max regimes is $M = 4$ use two bits $g_1(1), g_1(2)$ and $m = 2g_1(1)+g_1(2)$; if max regimes is larger we use three bits $g_1(1), g_1(2), g_1(3)$. Thus encoding m requires 2 or 3 bits.
$M-1$ genes encode $\tau_1, \tau_2, ..., \tau_{M-1}$ by means of $M-1$ real numbers $th_1, th_2, ..., th_{M-1}$, each with value in $(0, 1)$. These latter are constructed to determine the percentage of remaining values to place a changepoint. In fact, when placing a new changepoint there are some illegal positions, due to the constraints on minimum regime length mrl. These imply that mrl observations must be left out from both the beginning and the end of considered segment. This strategy depends on candidate number of regimes, and allows to always provide legal changepoints. For numbers $th_j (j = 1, ..., M-1)$, in order to ensure that any time could be a changepoint a priori, each gene length should be at least $\log_2 N$ bits. The number of bits for each coding may be left as an external parameter $lbit$.
M genes encode $\delta^1, \delta^2, ..., \delta^M$. The deltas are sequences of $p \times s$ binary digits, thus they are directly coded in $p \times s$ bits.
In case of GA including full described phenotype/genotype, the total length of the binary sequence is 2 (or 3) $+lbit \times (M-1) + M \times p \times s$. If all subset models are estimated and the best retained, the deltas are not considered in the chromosome.

From gene 1 one obtains the number of thresholds m by $m = 2g_1(1)+g_1(2)$ or $m = 4g_1(1)+2g_1(2)+g_1(3)$. From genes 2 to M to the change times decode with the routine of [3]:

- If $m = 0$ (only 1 regime) let $\tau_1 = N + 1$.
- If $m = 1$ (two regimes) let $\tau_1 = mrl + 1 + (N - 2mrl) \times th_1$.
- If $m = 2$ (three regimes) let $\tau_1 = mrl + 1 + (N - 3mrl) \times th_1$ and $\tau_2 = \tau_1 + mrl + (N - 2mrl - \tau_1 + 1) \times th_2$
- If $m = 3$ (four regimes) let $\tau_1 = mrl + 1 + (N - 4mrl) \times th_1$ and $\tau_2 = \tau_1 + mrl + (N - 3mrl - \tau_1 + 1) \times th_2$ and $\tau_3 = \tau_2 + mrl + (N - 2mrl - \tau_2 + 1) \times th_3$.

and similar way for more than four regimes. Genes M+1 to 2M give directly the vectors δ^j.

This method guarantees an efficient coding since it produces no illegal chromosome and no redundancy.

4.2. Objective function

The most natural objective in building statistical models is to minimize an identification criterion such as AIC, BIC, ICOMP, MDL. They all are based on the estimated residual variance $\hat{\sigma}^2(j,k)$ and the total number of estimated parameters: there are $m+1$ parameters for trend, $(m+1)\times s$ seasonal means, and in regime j there are $(p+1)\times s - |\delta^j|$ autoregressive parameters (where $|x|^2 = \sum_i x_i^2$). So, the total number of estimated parameters, P, is:

$$P = (m+1)(s+1) + (m+1)ps - |\delta^1| - |\delta^2| - \ldots - |\delta^M|.$$

If continuity constraints on trend are added, the number of parameters decreases by m.

The most obvious generalization of AIC is the NAIC criterion introduced by [18, p. 379] for threshold models:

$$NAIC = [\sum_{j=1}^{m+1}\sum_{k=1}^{s} AIC_{j,k}]/N = \left[\sum_{j=1}^{m+1}\sum_{k=1}^{s} n_{j,k}\log\hat{\sigma}^2(j,k) + \pi\times P\right]/N,$$

where $AIC_{j,k}$ is identification criterion for series of regime j and season k, $\sigma^2(j,k)$ is the related residual variance, P is the total number of parameters, π is the penalization term (equal to 2 in the original Akaike's proposal).

A possible alternative is the weighted Schwarz criterion where each parameter is penalized with the logarithm of the number of observations on which it is estimated. Slopes and means are in number of $(m+1)(s+1)$ and are estimated on N observations, leading to a penalization term of $(m+1)(s+1)\log(N)$ (minus $m\log(N)$ in case of continuity constraints). The AR parameters and residual variance for regime j and season k are estimated on n_{jk} observations and are in number of $p-|\delta_{jk}|+1$, where $|\delta_{jk}|$ is the number of parameters constrained to zero. Thus a generalized Schwarz criterion may be written

$$\sum_j\sum_k n_{jk}\log(\hat{\sigma}^2(j,k)) + \sum_j\sum_k (p-|\delta_{jk}|+1)\log(n_{jk}) \\ +(m+1)(s+1)\log(N) - I_{[cont]}\{m\log(N)\}$$

where $I_{[cont]}$ is the indicator of trend continuity constraints.

A simplified approximate form of Schwarz criterion is obtained considering the $NAIC$ criterion with a penalizing constant $\pi = \log(N/s)$. A further alternative is the MDL criterion (Minimum Description Length criterion) of [15], based on a penalization given by the minimum length in bits necessary for describing the data. For a model $\mathcal{M}$, MDL is the sum of the length to code the model $C(\mathcal{M})$ and the length to code the errors of the model $C(\varepsilon|\mathcal{M})$. Rissanen shows that $C(\varepsilon|\mathcal{M})$ equals $-\frac{1}{2}\log_2(\hat{L})$, where $\hat{L}$ is the maximum likelihood, therefore in our case

$$C(\varepsilon|\mathcal{M}) = \frac{1}{2}\sum_j \sum_k n_{jk} log_2(\hat{\sigma}^2_{jk}),$$

while $C(\mathcal{M})$ is the sum of the binary code length necessary to code the parameters of the models, according to the following rules:

- for an integer parameter p, code length is $\log_2(p)$;
- for an integer with an upper bound U, code length is $\log_2(U)$,
- for a real parameter estimated by ML on n observations, code length is $\frac{1}{2}\log_2(n)$.

The parameters of our model are of 6 kinds :

- Number of seasons s, integer number, code length $\log_2(s)$
- Autoregressive order p, integer number, code length $\log_2(p)$
- Number of change points m, integer, code length $\log_2\{max(1, m)\} = \log_2^+(m)$
- Change times: m integer numbers $\leq N$, code length $m\log_2(N)$
- Trend slope parameters and means. There are $(m + 1)(s + 1)$ real parameters, estimated by least square on N observations, leading to a code length of $\frac{1}{2}(m + 1)(s + 1)\log_2(N)$. If continuity constraints are required, m linear constraints determine the slopes after the first regime, therefore the total code length is decreased by $\frac{1}{2}m\log_2(N)$.
- Autoregressive parameters (subset models). The constrained parameters are determined by the arrays δ_{jk}, binary with length p bits. For

each j, k the number of estimated AR parameters is $p - |\delta_{jk}| + 1$, they are real parameters estimated on n_{jk} observations. Thus the overall code length is $\frac{1}{2}\sum_j \sum_k (p - |\delta_{jk}| + 1) \log_2(n_{jk})$.

Finally we may write:

$$C(\mathcal{M}) = \log_2^+(m) + m \log_2(N) + \log_2(s) + \log_2(p) + \frac{1}{2}(m+1)(s+1)\log_2(N)) + \frac{1}{2}\sum_{j=1}^{m+1}\sum_{k=1}^{s}(p - |\delta_{jk}| + 1)\log_2(n_{jk}) - I_{[cont]}\left\{\frac{1}{2}m\log_2(N)\right\}.$$

Since the identification criteria are to be minimized, the fitness function is a monotonically decreasing function of the identification criterion. We propose two kinds of fitness functions:

1. Exponential: $f = \exp(-NAIC/\alpha)$, where α is a scaling constant
2. Linear: $f = [\log(\sum_{t=1}^{N} X_t^2/N) + \pi/N] - NAIC$, where first term is a possible worst case scenario corresponding to i. i. d. observations, and similar for the other two alternative identification criteria, putting $\log(N)$ in place of π.

4.3. Operators and other implementation issues in the GAs

Since the binary chromosome has independent bits and any sequence of bits is legal, the usual genetic operators may be employed. Any selection method is suitable but we propose three types of selection operators: roulette wheel selection, rank selection and tournament selection. In "roulette wheel" rule the probability of a chromosome being selected as a parent is proportional to its fitness. In ranking selection the individuals are sorted in order of fitness and then reproductive fitness values are assigned according to rank. Then every chromosome is selected with respect to its rank. The basic idea of tournament selection is to select the individual with the highest fitness value from a certain number of individuals (tournament) in the population into the next generation. The number of the individuals taking part in the tournament is called tournament size. In the tournament selection, there is only comparison between individuals by fitness value.

Each selected couple of parents will produce two "children" by methods of crossover and mutation.

Two types of crossover are implemented: random one cut point and uniform crossover. In order to make sense as an exchange of parent phenotypes, crossover should be between genes, rather than bits. In other terms, the cut points should fall between genes. Alternative is uniform crossover - each child receives each (entire) gene from one parent or the other randomly with probability 1/2. Other crossover methods may be employed.

Finally, a probability is chosen for randomly changing the value of each bit of each gene of the child-chromosome (mutation).

The entire population of chromosomes is replaced by the offsprings created by the crossover and mutation processes at each generation except for the best chromosome, which survives to the next generation. This *elitist* strategy ensures that the fitness will never decrease through generations [16].

Another important decision is the initial population. A simple way to start would be to select the initial genotype chromosomes to be random binary sequences of required bits. As to the population size, [14] suggests (for binary random population) that to ensure that the probability of every single bit appearing in the population is at least Q, population size should be at least $1 + \log(-G/\log(Q))/\log 2$ where G is number of bits. For reasonable probability $Q = 0.999$ this rule gives a size around 20.

5. Seasons grouping

The seasonal effect on means, variances and correlations may show different speed and pattern, thus it seems advisable, for each of these features, to use a different splitting of the year, determined by a different length of the season inside which that feature remains constant. For example, if the seasonal period is s, we may have exactly s different models, one for each seasonal position; or rather only s/c different structures, when c consecutive observations are supposed to belong to the same season. E. g. if $s = 12$ (monthly data) and $c = 3$, the same model works for each quarter, therefore there are only $s/c = 4$ different seasons. This may be useful when the seasonal variation is slow and more detailed models would be redundant.

Regime changes may occur at any time, and the data length N is not assumed a multiple of s. It follows that in each regime, the number of observations belonging to different seasons may be different. For full seasonal models, i.e. $c = 1$, the season to which the first observation belongs is labeled '1'. When grouped seasons are considered, i.e. $c > 1$, and the seasons are

only s/c rather than s, the first completely observed season is labeled as '1': it could start at observations $1, 2, \ldots, c$, thus the parameter fo ($\leq c$) specifies what is the index of the first observation belonging to season labeled 1.

We allow a different season grouping for means, correlation and variance. We denote by c_M the number of consecutive observations for which the mean remains constant (c_M divides s); if $c_M > 1$, let fo_M denote the time of the first observation of the first season (the season labeled 1 as far as the means are concerned). Each season for the mean is formed by c_M consecutive observations, and the different seasons are $ss = s/c_M$. In an analogue fashion, we denote by c_{AR} the number of consecutive observations for which the AR parameters remain constant. If $c_{AR} > 1$, let fo_{AR} be the time of the first observation in the first season (i. e. the season labeled 1 concerning the AR model). Each season for the AR model is c_{AR} observation long, and the number of different seasons is $sv = s/c_{AR}$. E. g. for $s = 12$, if $c_{AR} = 1$, each month has a different set of AR parameters, then the variances of the 12 month may a priori be different. If on the contrary $c_{AR} > 1$, the variances of c_{AR} contiguous observations all are proportional, through the same coefficient, to the residual variances, thus a variance instability is equivalent to a residual variance instability. Therefore, we allow the possibility that, inside each single season for the AR model (containing c_{AR} consecutive observations) the residual variances may change. Thus we must consider sub-seasons composed by c_V observations, where c_V divides c_{AR}, and allow the residual variance to change every c_V observations, in a total number of seasons (concerning the residual variance) equal to $svar = s/c_V$. If $c_V > 1$, the first observation of the first season remains at time fo_{AR}.

A complete model with season grouping includes therefore the structural parameters of model without season grouping and, in addition, the parameters $c_M, c_{AR}, c_V, fo_M, fo_{AR}$ subject to the following constraints:

c_M divides s ; c_{AR} divides s ; c_V divides c_{AR} ; $1 \leq fo_M \leq c_M$; $1 \leq fo_{AR} \leq c_{AR}$.

The external parameters are the same as for the model without season grouping but the number and estimation of the regression parameters is considerably different. The regression parameters are

$a_1, a_2, \ldots, a_{m+1}$	intercepts
$b_1, b_2, \ldots, b_{m+1}$	slopes
$\mu^{(j)}(k)$	seasonal means, $k = 1, \ldots, ss$; $j = 1, \ldots, m+1$
$\phi_k^{(j)}(i)$	AR parameters, $k = 1, \ldots, sv$; $j = 1, \ldots, m+1$; $i = 1, \ldots, p$ (some of them may be constrained to zero)
$\sigma^2(j,k)$	innovation variance, $k = 1, \ldots, svar$; $j = 1, \ldots, m+1$

The seasonal means are in number of ss, the AR parameters are sv and the innovation variances are $svar$. Also in this case for the estimation of trend and seasonal means we assume that the seasonal means sum to zero on one cycle because the intercept and the means are linearly dependent:

$$\mu^{(j)}(1) + \mu^{(j)}(2) + \ldots, +\mu^{(j)}(ss) = 0 \quad \forall j.$$

Therefore the following equations are estimated:

$$X_t = b^{(j)}t + c(j, k_t) \quad , \quad \tau_j \leq t < \tau_{j+1} \tag{4}$$

and then the parameter vector is

$$\beta' = \left\{b^{(1)}, b^{(2)}, \ldots, b^{(m)}, c(1,1), c(1,2), \ldots, c(1,ss), c(2,1), \ldots,\right.$$

$$\left. c(2,ss), \ldots, c(m,1), \ldots, c(m,ss)\right\}$$

with dimension $m \times (ss+1)$. Each row of the design matrix relates to a time t from 1 to N and has the value t in the column corresponding to $b^{(j)}$, and 1 in the column $m + (j-1)ss + k_t$ that corresponds to parameter $c(j, k_t)$. For determining k_t, since the first observation of the first season is at time fo_M, we compute $\mod(t - fo_M + s, s)$. If the result is from 0 to $ss - 1$ we are in the first season, if the result is from ss to $2ss - 1$ we are in the second season, and so on. It follows that

$$k_t = \lfloor \frac{1}{c_M} \mod (t - fo_M + s, s) \rfloor + 1,$$

where $\lfloor n \rfloor$ is the largest integer contained in n. The intercepts $\hat{a}^{(j)}$ and seasonal means $\hat{\mu}^{(j)}(k)$ are determined from the $\{\hat{c}(j,k)\}$, assuming that the means sum to zero on a whole seasonal cycle in each regime, as follows

$$\hat{a}^{(j)} = \frac{1}{ss} \sum_{k=1}^{ss} \hat{c}(j,k) \quad , \quad \hat{\mu}^{(j)}(k) = \hat{c}(j,k) - \hat{a}^{(j)}.$$

Moreover, it is possible to prescribe trend continuity, by imposing that (if the number of regimes is larger than one) the trend values of two consecutive regimes coincide on the first observation of the second regime. For imposing a continuity constraint on the trend $a^{(j)} + b^{(j)}\, t$ between consecutive regimes, we consider constrained least squares with the constraints $H\beta = 0$, where each row of the matrix H imposes that the trend part coincides at the first time of each successive regime:

$$b^{(j)}\tau_{j+1} - b^{(j+1)}\tau_{j+1} + \frac{1}{ss}\sum_{k=1}^{ss} c(j,k) - \frac{1}{ss}\sum_{k=1}^{ss} c(j+1,k) = 0.$$

It follows that H has m rows and the j-th row has value τ_{j+1} in the j-th entry and $-\tau_{j+1}$ in the $(j+1)$-th entry, that the entries from $m + (j-1) \times ss + 1$ to $m + j \times ss$ are all equal to $1/ss$ and the entries from $m + j \times ss + 1$ to $m + (j+1) \times ss$ are all equal to $-1/ss$.

Based on thresholds, seasonal arrangement and estimated trend and means, the residual series is computed:

$$\hat{W}_t = X_t - \hat{a}^{(j)} - \hat{b}^{(j)}t - \hat{\mu}^{(j)}(k_t).$$

For each regime j and season k the autoregressive parameters of process:

$$\hat{W}_t = \sum_{i=1}^{p} \phi_{k_t}^{(j)}(i)\hat{W}_{t-i} + \varepsilon_t$$

are estimated in a similar way of the model without season grouping. The main difference is in the selection, for each season $k = 1, 2, \ldots, sv$ and regime j, of observations belonging to that subseries (t such that $k_t = k$). For each pair regime j, season k ($j = 1, \ldots, M; k = 1, \ldots, sv$), the data are the (j,k) subseries $\{\hat{W}_t, t \in I^*(j,k)\}$ where the entries of the time index vector $I^*(j,k)$ are the times from $\ell \times s + (k-1) \times c_{AR} + fo_{AR}$ to $\ell \times s + k \times c_{AR} + fo_{AR} - 1$ for any natural ℓ such that the resulting time is between τ_j and $\tau_{j+1} - 1$.

The estimation of the innovations is obtained using the residuals

$$e = z - Z\,\hat{\phi}$$

that give the estimates of the innovations $\{\varepsilon_t, t \in I^*(j,k)\}$ for regime j and season k.

The optimum value of structural parameters may be determined following different strategies. The first strategy is a GA algorithm where the phenotype

contains also the genes for coding c_M, c_{AR}, c_V, fo_M and fo_{AR}. In this way it optimizes simultaneously all the parameters of a complete model. Many generations may be needed to reach satisfying results.

The second strategy is a two step GA. We try first to determine the best splitting in regimes, considering in the first stage a GA for chromosomes containing only the number of regimes and thresholds $m, \tau_1, ..., \tau_{M-1}$. To ensure that the fitness depends essentially only on the fitting of trend and regimes, we try to account for seasonality and correlation instability as much as possible, considering monthly seasonality ($c_M = c_{AR} = 1$) and only full autoregressive models (with no parameter constrained to zero). Given the best choice of $m, \tau_1, ..., \tau_{M-1}$, a second stage runs a GA for optimizing the remaining features: selection of the best subset model, and season grouping assignment; here the chromosome contains only the genes $\delta^1, ..., \delta^M$ and those for c_M, c_{AR}, c_V, fo_M and fo_{AR}.

The third possible strategy is a hybrid algorithm. As an alternative to the preceding one, if the order p and the seasonal period s are not too large, we can substitute the second stage with a complete enumeration of the possible solutions for the δ's and the parameters c_M, c_{AR}, c_V, fo_M and fo_{AR}.

For all strategies, when computing the fitness, the first term is based on the logarithm of the residual variances of the *svar* sub-seasons,while the second term (penalization) accounts for the parameters of the *sv* seasons related to the AR models, the *ss* means and the trend parameters. Finally, we consider the residual variances as unknown parameters themselves and their values inserted into the fitness as ML estimates, therefore we add in the penalization term their number, equal to *svar* times the number of regimes.

6. Applications

The proposed methods will be illustrated on datasets from the fields of hydrology, economics and climate. Four kinds of periodic autoregressive models will be fitted to the time series: a *complete* model, with a different AR model for each seasonal position (month or quarter) and no constraint on the autoregressive parameters; a *subset* model similar to the previous one, but with some AR parameters constrained to zero in order to maximize fitness; a *grouped* subset model where the seasons are grouped as explained in Section 5; and finally a *constant* seasonality model, subset as well, where the autoregressive parameters remain equal in each regime.

For all those models, the structure was chosen by maximization of the fitness function. We considered the original NAIC criterion ($\pi = 2$), the weighted Schwarz criterion (labeled BIC) and the approximate criterion obtained choosing $\pi = \log(N/s)$. The fitness form was exponential with scale $\alpha = 0.1$. In our framework, a model is better than another model simply if its fitness is larger; however, all the estimated models were subject to a diagnostic checking step, and in particular the usual portmanteau tests were computed on the residuals.

For comparison, more conventional seasonal autoregressive integrated moving average ($SARIMA$) models were also fitted to the series.

The parameters of each model were estimated not on all available data, but leaving some most recent observations for out-of-sample forecasts.

In the present application we adopted the hybrid algorithm strategy: the regime changes are identified on complete models, then all possible season grouping and subset choice are evaluated and the best one retained. The GA population size was 50 and the number of generations 200. We used roulette wheel selection with elitist strategy, uniform cross-over and mutation with probability 0.2.

6.1. Garonne riverflow

The first considered time series contains the natural logarithms of the riverflows of the river Garonne. The data are obtained from daily discharge measurements in cubic meter per second (m^3/s) recorded at the Tonneins gauging station (DIREN-Banque Hydro, French water monitoring). The observations are monthly from January 1959 to December 2012, and the twelve months of 2013 will be used for out-of-sample forecasts. The Garonne time series is reported in Figure 1. Here $N = 648$ and $s = 12$. The maximum allowed order for the autoregressive models is $p = 3$ and the minimum regime length $mrl = 120$ months. The results appear in Table 1. Each column relates to a different fitness type, and each row to a different model and reports the fitness value and the overall residual variance. The first row explains the suggested regime splitting (remember that it is identified on optimizing the complete model): number of regimes, and starting time of each successive regime. For models that group seasons, the grouping is indicated by the triple $(ss, sv, svar)$. In the last row the results, in terms of residual variance and fitness, of an airline model $SARIMA(0,1,1) \times (0,1,1)_{12}$ are shown.

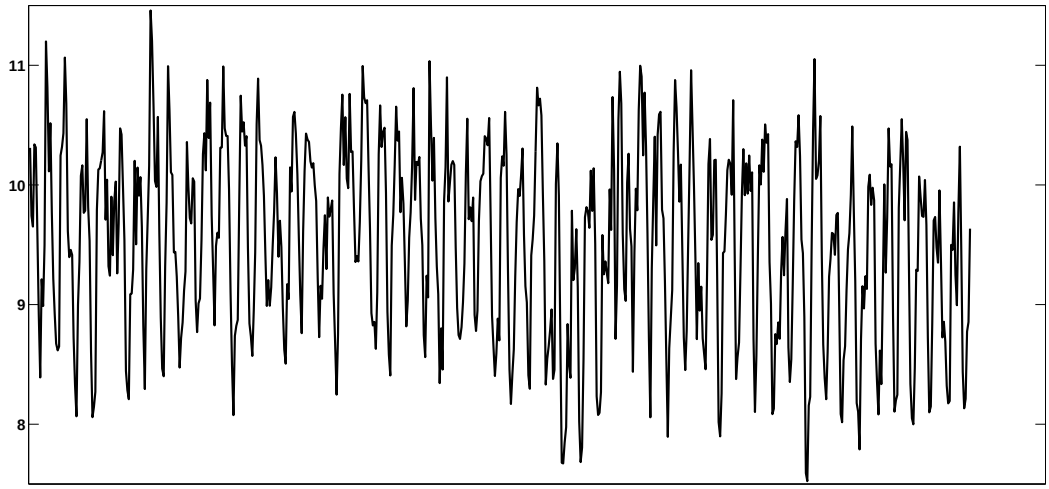

Figure 1: Garonne river log flows, 1959-2012

The optimal regime identification is different for the different fitness functions. The fitness $AIC\{2\}$ favors a less parsimonious model, and points to a regime change at August 1988, while the other two fitness types are more similar and detect no regime change. The resulting two alternative trend estimates are reported in Figure 2. With the fitness $AIC\{2\}$ the best model is one with constant seasons (the seasonal behavior does not change in the two regimes) and grouping $(12, 3, 12)$: means different for each month, but only three different AR models, one for each four months period, and residual variances different in each month. We note that the grouped (non constant) model does not suggest any season grouping neither in mean, nor in AR parameters, nor in residual variances, therefore it is equal to the subset model.

Using the other two fitness forms the results are similar. Here, since only one regime is proposed, the constant model is equal to the grouped model, which is the most fit according to both fitness forms. There is only a slight difference in the number of residual variances detected (quarterly for $AIC\{log(N/s)\}$ and half-yearly for BIC), but both suggest different means for each month, and only two AR models alternating across semesters. An overall observation is that for this series the possible regime change does not appear to have a sensible influence on seasonality.

The $SARIMA(0, 1, 1) \times (0, 1, 1)_{12}$ model shows worse fit when the fitness $AIC\{2\}$ is considered, while for the other two fitness forms it is inferior only to the best periodic model (the grouped season one).

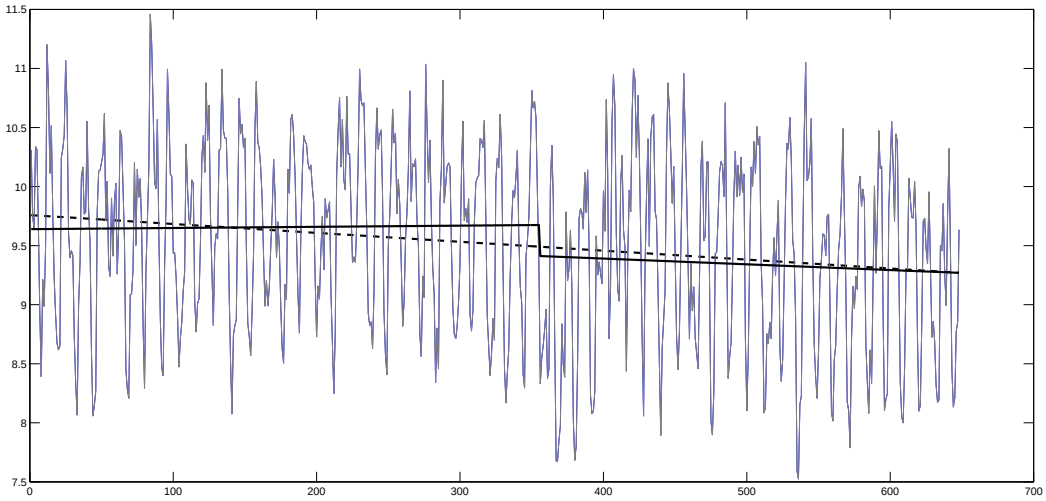

Figure 2: Estimated regimes and trends for Garonne river data. Continuous line: $AIC\{2\}$; dashed line: BIC

The results of a forecasting exercise are illustrated in Table 4. One-step-ahead forecasts for the 12 months of 2013 were computed for each model, and the average square forecast errors appear in the table. It may be seen that, whatever the fitness form, the optimal models achieve also the smallest forecast error. The $SARIMA$ model gives a larger forecast error than all other models.

6.2. Italian Industrial Production Index

We consider the time series of the Italian industrial production index (Ateco branch C), monthly from January 1990 to September 2016 (published by the National Institute of Statistics); the observations from October 2016 to September 2017 will be left for out-of-sample forecasts. The time series is reported in Figure 3. Here $N = 321$, $s = 12$. The maximum allowed order for the autoregressive models is $p = 3$ and the minimum regime length $mrl = 84$ months. The results appear in Table 2. Here also the optimal regime identification is different for the different fitness functions. The fitness $AIC\{2\}$ suggests three regimes, the second starting at July 2002, and the third at September 2009; the other two fitness types are more similar and detect only one regime change at November 2008 (see Figure 4).

With the fitness $AIC\{2\}$ the best model is the subset model, which is similar to the grouped one, since no grouping is suggested. The constant

Table 1: Garonne River log flows, monthly 1959-2012

Model		$AIC\{2\}$	$AIC\{log(N/s)\}$	BIC
Regimes		2: Aug 1988	1	1
Complete	σ^2	0.129	0.150	0.15
	fitness	5.44	7.5	28.9
Subset	σ^2	0.135	0.154	0.154
	fitness	12.5	23.7	91.1
Grouped	group	(12,12,12)	(12,2,4)	(12,2,2)
seasons	σ^2	0.135	0.156	0.158
	fitness	12.5	46.8	240.8
Constant	group	(12,3,12)	-	-
seasons	σ^2	0.154		
	fitness	102.0	-	-
Airline	σ^2	0.195	0.195	0.195
	fitness	1.18	11.1	103

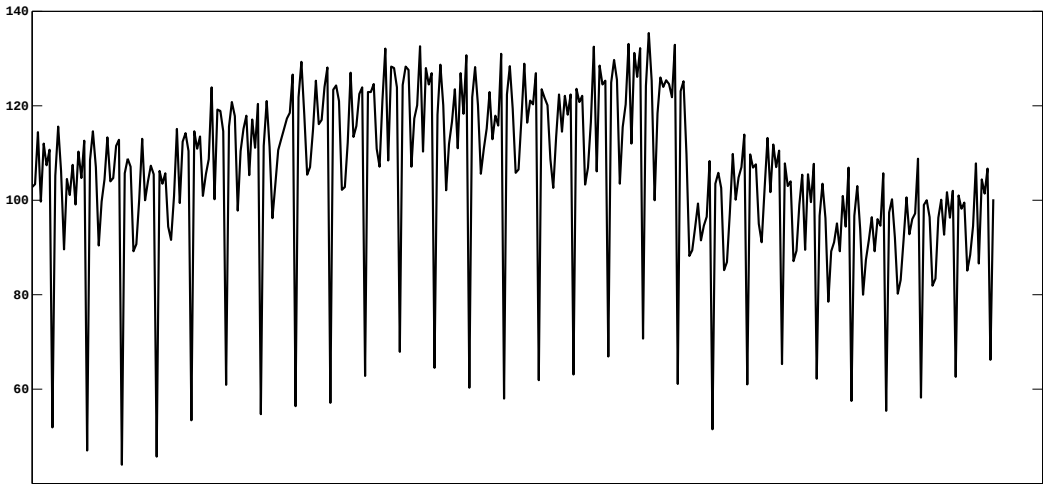

Figure 3: Italian industrial production index, 1990-2016

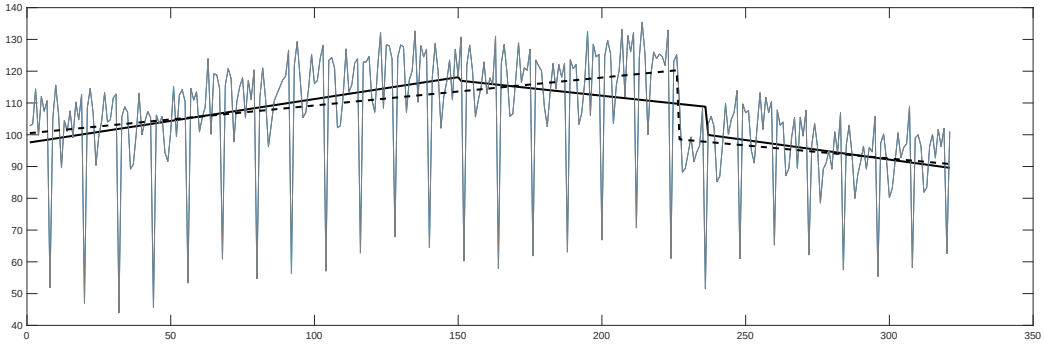

Figure 4: Estimated regimes and trends for IPI data. Continuous line: $AIC\{2\}$; dashed line: BIC

seasonality model shows a fitness considerably smaller than all other models (and a much larger residual variance).

Using the other two fitness forms the results are somewhat similar. The best is the grouped model, and the suggested grouping is in both cases to maintain 12 different means for the months, but to reduce the number of different AR models to 6 or to 2, the fitness BIC being more parsimonious than $AIC\{log(N/s)\}$.

The results of a forecasting exercise are also reported in Table 4. One-step-ahead forecasts from October 2016 to September 2017 were computed for each model, and the average square forecast errors appear in the table. It is apparent that the forecasts obtained with the constant seasonality models are worse than all others, the forecast errors for the other models are not much different, but the complete model seems to perform slightly better. Moreover, the identification of one single regime change seems to reduce the forecast error with respect to the two regime change model.

The $SARIMA(0,1,1) \times (0,1,1)_{12}$ model shows a much smaller fitness than the other models in any case. However, as far as the average forecast errors are concerned, it is comparable to the others, though always worse than the optimal models.

The results as a whole suggest that the industrial production series presents at least one serious regime change, and it influences also the seasonal behavior. Figure 5 shows the estimated monthly means for the optimal model for $AIC\{log(N/s)\}$ and the two regimes, and suggests that the seasonal variations became smaller after the regime change, though maintaining a similar pattern.

Table 2: Italian Industrial Production Index, monthly Jan 1990 : Sept 2016

Model		$AIC\{2\}$	$AIC\{log(N/s)\}$	BIC
Regimes		3: Jul 2002, Sep 2009	2: Nov 2008	2: Nov 2008
Complete	σ^2	5.75	6.46	6.46
	fitness	380	2.76	0.35
Subset	σ^2	6.29	7.17	7.06
	fitness	2710	47.3	2.1
Grouped	group	(12,12,12)	(12,6,6)	(12,2,2)
seasons	σ^2	6.29	8.42	9.99
	fitness	2710	55.1	4.3
Constant	group	(12,12,12)	(12,12,12)	(12,12,12)
seasons	σ^2	16.18	16.35	16.35
	fitness	7.5	2.9	0.36
Airline	σ^2	35.4	35.4	35.4
	fitness	0.003	0.002	0.002

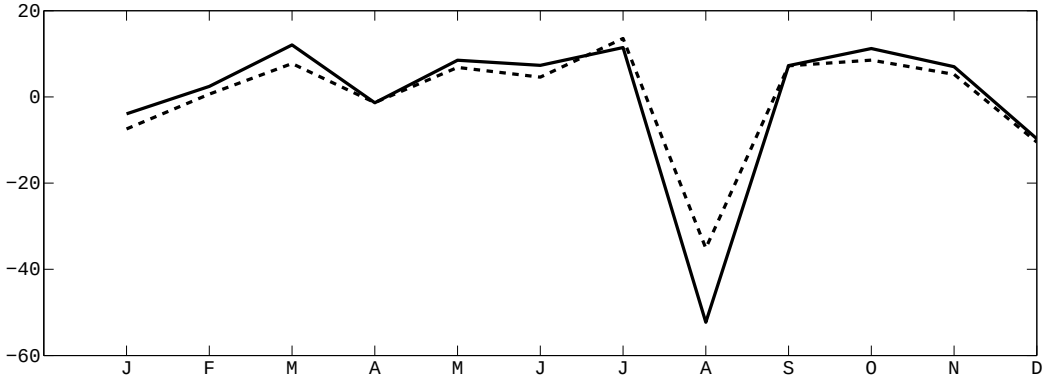

Figure 5: Italian industrial production index: estimated monthly means. Continuous line: first regime, dotted line: second regime

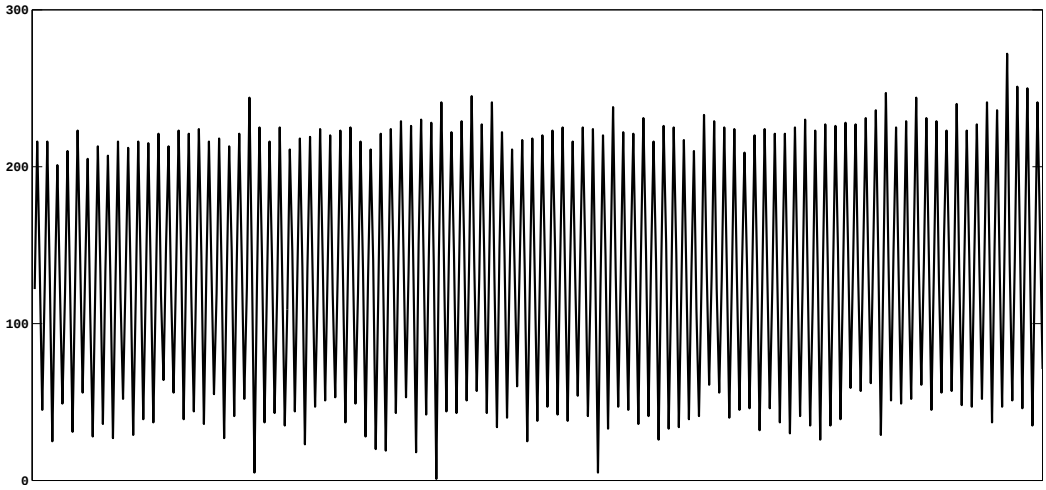

Figure 6: Turin quarterly average temperature, 1907–2006

6.3. Turin temperature

We consider a long temperature recording series starting in 1760, recorded in Turin (Italy). The observations are quarterly average temperatures for years 1760 to 2008, and come from the repository of the Histalp Project [2]. Here we consider hundred years from 1907 to 2006, while the years 2007 and 2008 will be left for out-of-sample forecasts (thus $N = 400$ and $s = 4$). The maximum considered AR order was 2 and the minimum regime length $mrl = 80$ quarters. The time series appears in Figure 6, and the results in Table 3. The optimal regime identification again is different for the different fitness functions. The fitness $AIC\{2\}$ suggests three regimes, the second starting in the second quarter 1942, and the third in the fourth quarter 1970; the other two fitness types are more similar and detect no regime change. These two alternatives are displayed in Figure 7.

With the fitness $AIC\{2\}$ the best model is the subset without any season grouping. The constant model (imposing equal seasonal behavior for each of the 3 regimes) suggests a grouping of season with only one constant AR model, but achieves a considerably smaller fitness value.

Using the other two fitness forms the results are similar. Here, since only one regime is proposed, the constant model is equal to the grouped model, which is the most fit according to both fitness forms. The optimal grouping is equal to the previous one: four quarterly means and four quarterly residual variances, but only one constant AR model.

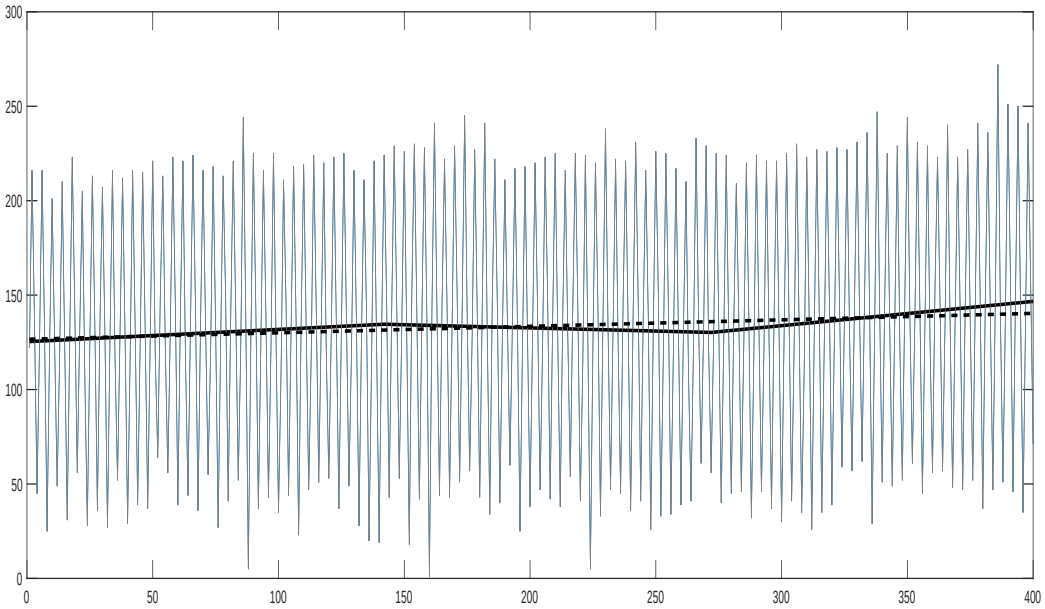

Figure 7: Estimated regimes and trends for Turin temperature data. Continuous line: $AIC\{2\}$; dashed line: BIC

An overall observation is that for this series the possible regime change does not appears to have a sensible influence on seasonality. We note further that in several cases (especially for the more parsimonious fitness types) the autoregressive parameters are all constrained to zero, leading to models with uncorrelated observations and only means depending on the season.

The results of a forecasting exercise are also reported in Table 4. One-step-ahead forecasts for the 8 quarters of years 2007 and 2008 were computed for each model, and the average square forecast errors appear in the table. For the fitness $AIC\{log(N/s)\}$ and BIC the optimal models (the grouped season one) achieve also the smallest forecast error, and with the alternative fitness $AIC\{2\}$ the smallest forecast error is also achieved by the most fit (the subset model). Note that here the models with regime changes ensure a forecast error definitely smaller than those without regime splitting.

The $SARIMA(0,1,1)\times(0,1,1)_4$ model has fitness of type $AIC\{2\}$ much smaller, but a slightly larger fitness than the other models if the other fitness types are considered. However the $SARIMA$ model shows forecast errors uniformly larger than the other models.

We may conclude that for this temperature series the detection of a regime

Table 3: Average temperature in Turin, quarterly 1907 : 2006, piecewise continuous

Model		$AIC\{2\}$	$AIC\{log(N/s)\}$	BIC
Regimes		3: II, 1942; IV, 1970	1	1
Complete	σ^2	78.8	94.2	94.2
	fitness	21.9	4.3	3.2
Subset	σ^2	80.7	94.7	94.7
	fitness	38.2	7.3	5.4
Grouped	group	(4,4,4)	(4,1,4)	(4,1,4)
seasons	σ^2	80.7	94.7	94.7
	fitness	38.2	9.1	6.4
Constant	group	(4,1,4)	-	-
seasons	σ^2	90.2		
	fitness	26.0	-	-
Airline	σ^2	97.0	97.0	97.0
	fitness	11.6	9.5	8.6

change is important also for forecasting purposes, but on the other side this regime change seems to influence more the levels than the seasonal behavior.

7. Conclusions

In this paper we have proposed models that are able to explain, on one side, regime changes and structural breaks, and on the other side a seasonal behavior that evolves in time.

The complex problem of identifying and estimating such models is solved by computational strategies based on GAs. The best model is selected according to a fitness function that is a monotonically decreasing transformation of widely used identification criteria. Experience on real and simulated data suggests that the choice of the fitness function is crucial because a too parsimonious criterion may lead to models that overlook important structure changes.

The results obtained analyzing some hydrological, climatological and economic time series seem to support the usefulness of the proposed methods

Table 4: Average square one-step-ahead forecast errors

Series	Model	$AIC\{2\}$	$AIC\{log(N/s)\}$	BIC
Garonne River	Complete	.269	.274	.274
(1 year)	Subset	.275	.292	.292
	Grouped seas.	.275	.251	.231
	Constant seas.	.243	.251	.231
	Airline	.372	.372	.372
Ind. Prod. Index	Complete	24.4	17.41	17.41
(1 year)	Subset	29.9	15.3	20.2
	Grouped seas.	29.9	26.5	22.7
	Constant seas.	42.8	53.6	53.6
	Airline	26.7	26.7	26.7
Turin temper.	Complete	66.6	72.4	72.4
(2 years)	Subset	61.7	72.0	72.0
	Grouped seas.	61.7	70.9	70.9
	Constant seas.	68.9	70.9	70.9
	Airline	88.9	88.9	88.9

in detecting relevant changes in the structure of the trend and also possible evolution in the seasonal behavior concerning levels, variance and correlation.

The generalized periodic autoregressive models allow a closer analysis of the seasonal behavior, suggesting also the most convenient grouping of seasons in terms of fitness.

In our applications, the proposed models achieved a larger fitness, and a smaller forecast error, than the more common seasonal autoregressive integrated moving average models.

References

[1] A. Aue and L. Horváth. Structural breaks in time series. *Journal of Time Series Analysis*, 34(1):1–16, 2013.

[2] I. Auer, R. Böhm, A. Jurkovic, W. Lipa, A. Orlik, R. Potzmann, W. Schöner, M. Ungersböck, C. Matulla, K. Briffa, P.D. Jones, D. Efthymiadis, M. Brunetti, T. Nanni, M. Maugeri, L. Mercalli, O. Mestre, J-M. Moisselin, M. Begert, G. Müller-Westermeier, V. Kveton, O. Bochnicek, P. Stastny, M. Lapin, S. Szalai, T. Szentimrey, T. Cegnar, M. Dolinar, M. Gajic-Capka, K. Zaninovic, Z. Majstorovic, and E. Nieplova. Histalp— Historical Instrumental Climatological Surface Time Series of the Greater Alpine Region. *International Journal of Climatology*, 27:17–46, 2007.

[3] F. Battaglia and M.K. Protopapas. Multi-regime models for nonlinear nonstationary time series. *Computational Statistics*, 27:319–341, 2012.

[4] P. Bloomfield, H.L. Hurd, and R.B. Lund. Periodic correlation in stratospheric ozone data. *Journal of Time Series Analysis*, 15(2):127–150, 1994.

[5] R.A. Davis, T.C.M. Lee, and G.A. Rodriguez-Yam. Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, 101(473):223–239, 2006.

[6] R.A. Davis, T.C.M. Lee, and G.A. Rodriguez-Yam. Break detection for a class of nonlinear time series models. *Journal of Time Series Analysis*, 29(5):834–867, 2008.

[7] P. H. Franses and R. Paap. *Periodic Time Series Models.* Oxford University Press, New York, 2004.

[8] D.E. Goldberg. *Genetic algorithms in search optimization and machine learning.* Addison-Wesley, Reading, MA, 1989.

[9] K. W. Hipel and A. I. McLeod. *Time Series Modelling of Water Resources and Environmental Systems.* Elsevier, Amsterdam, 1994.

[10] J.M. Holland. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control and AI.* The University of Michigan, Ann Arbor, MI, 1975.

[11] Q. Lu, R.B. Lund, and T.C.M. Lee. An MDL approach to the climate segmentation problem. *The Annals of Applied Statistics*, 4(1):299–319, 2010.

[12] R.B. Lund, X.L. Wang, Q. Lu, J. Reeves, C. Gallagher, and Y. Feng. Changepoint detection in periodic and autocorrelated time series. *Journal of Climate*, 20(20):5178–5190, 2007.

[13] V.J. Rayward-Smith, I.H. Osman, C.R. Reeves, and G.D. Smith. *Modern Heuristic Search Methods.* Wiley, Chichester, New York, 1996.

[14] C.R. Reeves and J.E. Rowe. *Genetic algorithms - Principles and Perspective: A Guide to GA Theory.* Kluwer Academic Publishers, London, 2003.

[15] J. Rissanen. *Information and complexity in statistical modeling.* Springer Science & Business Media, 2007.

[16] G. Rudolph. Convergence analysis of canonical genetic algorithms. *IEEE Transactions on Neural Networks*, 5:96–101, 1994.

[17] R.M. Thompstone, K.W. Hipel, and I. Mcleod. Grouping of periodic autoregressive models. In O.D. Anderson, J.K. Ord, and E.A. Robinson, editors, *Time series analysis: theory and practice 6. Proc. Conference Toronto 1983*, pages 35–49. North-Holland, Amsterdam, 1985.

[18] H. Tong. *Non Linear Time Series: A Dynamical System Approach.* Oxford University Press, Oxford, 1990.

[19] E. Ursu and K. F. Turkman. Periodic autoregressive model identification using genetic algorithms. *Journal of Time Series Analysis*, 33:398–405, 2012.