Noname manuscript No.
(will be inserted by the editor)

Multiple changepoint detection for periodic autoregressive models with an application to river flow analysis

Domenico Cucina · Manuel Rizzo · Eugen Ursu

Received: date / Accepted: date

Abstract River flow data are usually subject to several sources of discontinuity and inhomogeneity. An example is seasonality, because climatic oscillations occurring on inter-annual timescale have an obvious impact on the river flow. Further sources of alteration can be caused by changes in reservoir management, instrumentation or even unexpected shifts in climatic conditions. When such changes are ignored the results of a statistical analysis can be strongly misleading, so flexible procedures are needed for building the appropriate models, which may be very complex. This paper develops an automatic procedure to estimate the number and locations of changepoints in Periodic AutoRegressive (PAR) models, which have been extensively used to account for seasonality in hydrology. We aim at filling the literature gap on multiple changepoint detection by allowing several time segments to be detected, inside of which a different PAR structure is specified, with the resulting model being employed to successfully capture the discontinuities of river flow data. The model estimation is performed by optimization of an objective function based on an information criterion using genetic algorithms. The proposed methodology is evaluated by means of simulation studies and it is then employed in the analysis of two river flows: the South Saskatchewan, measured at Saskatoon,

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

M. Rizzo (✉)
Department of Statistical Sciences, Sapienza University of Rome, Piazzale Aldo Moro, 5, 00100 Rome, Italy
E-mail: manuel.rizzo@uniroma1.it

E. Ursu
GREThA UMR-CNRS 5113, Université de Bordeaux, Avenue Léon Duguit, 33608 Pessac CEDEX, France

Canada, and the Colorado, measured at Lees Ferry, Arizona. For these river flows we build changepoint models, discussing the possible events that caused discontinuity, and evaluate their forecasting accuracy. Comparisons with the literature on river flow analysis and on existing methods for changepoint detection confirm the efficiency of our proposal.

Keywords Periodic time series · Changepoint detection · Genetic algorithm · River flows

1 Introduction

Discontinuities are often introduced into climatic or hydrologic time series as a result of anthropogenic impacts or changes in instrumentation and location. Further plausible reasons are modifications in the reservoir system management or new water pricing. As defined by Li and Lund (2012), a changepoint is "a time where the structural pattern of a time series first shifts". In many cases changepoints are located at known times (dam construction, measuring instrument change) and it is easy to take their effects into account. When changepoints are located at unknown times and their features are ignored, the time series estimation can be misleading (Lu and Lund, 2007; Lund et al, 2007). In fact, an undetected changepoint can lead to misinterpretation of the model, biased estimates and less accurate forecasting (Hansen, 2001). Identifying a changepoint is also an important task because rapid environmental change acts differently on river flows: the magnitude and timing of river flows change and new patterns of extreme droughts or floods are observed. The pressure generated by economic factors raises the question of whether there will be enough water during drought years for irrigation, recreational uses and hydroelectric power generation. Given all of this, changepoint detection becomes a demanding job, especially if the identification is required soon after occurrence (e.g. flood predictions). Therefore, the successful application of time series models with changepoints could lead to improvements in operating the reservoir system, which can result in cost savings (an overview of optimization methods used in reservoir operations can be found in Fayaed et al (2013)).

Over the past four decades several techniques have been employed for changepoint detection. As far as hydrological applications are concerned, many authors have considered the problem of detecting a single changepoint (Cobb, 1978; Buishand, 1984; Hipel and McLeod, 1994; Rao and Tirtotjondro, 1996), but very few have analyzed more realistic multiple changepoint situations. Bai and Perron (1999) proposed a test to evaluate the hypothesis that changes have occurred. The problem of modeling a class of nonstationary time series using parametric models is considered in Davis et al (2006, 2008) and Yau et al (2015). They combine a model selection approach with an algorithm devoted to changepoint detection. Nonparametric estimation of both the number and locations of changepoints was proposed by Kawahara and Sugiyama

(2012) and Matteson and James (2014), while Shaochuan (2019) considered a Bayesian approach. A comprehensive review of several changepoint estimation methods may be found in Aue and Horváth (2013).

A further and common source of inhomogeneity in hydrological data is the seasonal effect (seasonality), for which the time series of interest tends to repeat a similar behaviour after regular time periods. This is generally caused by climatic oscillations occurring on inter-annual time scale, so it is a crucial issue in statistical river flow analysis. Models accounting for seasonality, like the seasonal autoregressive integrated moving average (SARIMA) developed originally by Box and Jenkins (1970), have been broadly used in the literature of hydrologic models (Mishra and Desai, 2005; Durdu, 2010; Wang et al, 2014). One problem with such methods is that they cannot be filtered to achieve second-order stationarity, and this is because the autocorrelation structure of these time series depends on the season (Vecchia, 1985a,b). Also, Delleur et al (1976) observed that using seasonal differenced models like SARIMA can deteriorate river flow forecasting ability, which is a major problem in water resources planning and operating. Periodic time series models (Vecchia, 1985a,b; McLeod, 1993) have been introduced to be able to also analyze a seasonal dependent autocorrelation structure, and they have met with success in many hydrological applications (Hipel and McLeod, 1994; Maçaira et al, 2017; Pereira and Veiga, 2018). General overviews of periodic models and their applications are presented in Hipel and McLeod (1994), Franses and Paap (2004) and Mondal and Wasimi (2006).

In this research paper we will consider Periodic AutoRegressive (PAR) models, for which each seasonal position is related to a possibly different AR model, and multiple changepoints are allowed. We propose a procedure based on Genetic Algorithms (GAs) to detect the changepoints, which will specify several segments, and estimate the resulting PAR models on each segment. GAs are well suited to complex global optimization, as they have been widely applied to intractable or challenging identification and estimation problems. In our case the GA will optimize an identification criterion, introduced to select the best model to describe the data, a choice which has been already adopted in the literature. For example, a recent paper by Doerr et al (2017) used GAs to build a model with multiple mean shifts in the series, whereby each segment is allowed to have a distinct mean. Other proposals of changepoint detection for time series by means of GAs can be found in Jeong and Kim (2013); Song and Singh (2010) and Ursu and Pereau (2015), among others. In order to handle the complexity of the identification problem, GAs combined with the Minimum Description Length principle (MDL; Rissanen (1978)) have often been employed. Such a criterion has been proven useful in several multiple changepoint detection proposals (Davis et al, 2008; Li and Lund, 2012; Yau et al, 2015).

In a contribution closely related to our research work, Lu et al (2010) proposed to combine the MDL criterion and a GA to determine the number and the positions of changepoints in PAR models. They proposed a model with an unknown number of changepoints and with each segment being allowed

to have a different mean, but the same autocovariance structure. Such a proposal fills a research gap in the literature of changepoint detection methods by allowing changes in the mean of PAR models. Our target is to extend the approach of Lu et al (2010) to a more general framework, in which each segment is allowed to have a possibly different model structure, that includes the trend term, the means, the PAR parameters and the residual variances. In fact, the same research paper states that 'in other applications (...) it may be more realistic to keep mean process levels fixed and allow the time series parameters to change at each changepoint time'. Our proposal will include this scenario as a particular case.

There are other papers in which researchers have proposed to allow a different model structures for each detected segment, although no periodic modelling features are concerned. For example, Koutroumanidis et al (2009) analyzed the mean-monthly discharge of Nestos River, which was divided into five segments characterized by different trends, means and ARMA models; Piyoosh and Ghosh (2017) studied changes in mean and trend for seasonal and annual rainfall. Furthermore, in order to extend periodic models, Hipel and McLeod (1994) defined periodic intervention models. In this context, when the complexity of periodic models is to be increased (e.g. when the noise term is affected by an intervention), they suggested allowing all of the parameters in the periodic model to change as time progresses.

Taking all these factors into account, we believe it may be meaningful to also allow the PAR, the trend parameters and the residual variances, along with the means, to change within each segment. Therefore our procedure will be built to detect this kind of changes. The method will also be able to perform subset selection, as we allow intermediate AR parameters to be constrained to zero. This modification can lead to more parsimonious models and could also contribute to improving the forecasting ability. We will evaluate the performance of the procedure by means of simulation studies.

Our proposal is also motivated by the need to generate different scenarios of hydrological inflows, for a better understanding of the hydrological processes and their reaction to climate change and human activity. In this respect we will employ our method to analyze the average monthly flows of two rivers: the South Saskatchewan (measured at Saskatoon, Canada) and the Colorado (measured at Lees Ferry, Arizona). The forecasting ability of the resulting models will then be evaluated by means of standard performance measures and the model adequacy checked by a portmanteau test.

The rest of the article is organized as follows: Section 2 describes the proposed model, the computational procedure, the forecasting method and the model validation technique; in Section 3 simulation studies are reported to empirically evaluate the performance of the procedure; Section 4 introduces the study area and the river flow data; the results of the applications and the related discussions are outlined in Section 5; a summary of the conclusions closes the paper in Section 6.

2 Methodology

We consider the problem of modeling a seasonal non-stationary time series by specifying several time segments, in each of which a possibly different PAR process is specified. Our way of proceeding is similar to the approach of Davis et al (2008), whereby a model class is specified to describe the whole series and the model parameters are subject to change in each time segment. In our case we refer to the class of PAR models with seasonal means and a deterministic trend term. The approach of Lu et al (2010), which considers changepoints in PAR models, is different because only mean shifts are allowed in each segment, while the rest of the model parameters are kept fixed.

2.1 Model description

The considered time series is observed for N years and the period s is assumed to be known. We will adopt the same notation as in Lu et al (2010), where the observation in season k of year n is denoted by $X_{(n-1)s+k}$, with $n = 1, 2, \ldots, N$ and $k = 1, \ldots, s$.

Before introducing the model we will describe the structure of segments. There are M different segments, each of which includes an integer number of years, and τ_{j-1} denotes the first year of segment j ($j = 1, 2, \ldots, M$). Therefore the first segment includes years from $\tau_0 = 1$ to $\tau_1 - 1$, the second segment contains years from τ_1 to $\tau_2 - 1$, the third segment contains years from τ_2 to $\tau_3 - 1$, and so on. If $m = M - 1$ is defined as the number of changepoint times, the segment structure is defined as follows:

$$1 \equiv \tau_0 < \tau_1 < \ldots < \tau_m < \tau_M \equiv N + 1. \tag{1}$$

In order to ensure reasonable estimates, each segment is required to contain at least a minimum number ω of years, therefore $\tau_j \geq \tau_{j-1} + \omega$ for any segment j. We define $R^j = \{\tau_{j-1}, \tau_{j-1} + 1, \ldots, \tau_j - 1\}$ as the set of years included in segment j: therefore, if year n belongs to R^j then the time $(n-1)s + k$ is in segment j. For the sake of simplicity we will assume that the total number of observations T is a multiple of s ($T = N \times s$).

Conditional on the segment structure, we propose to model the observed time series by a process $X_{(n-1)s+k}$ given by:

$$X_{(n-1)s+k} = a^j + b^j[(n-1)s + k] + W_{(n-1)s+k}, \tag{2}$$

where $n \in R^j$, $j = 1, \ldots, M$, $k = 1, \ldots, s$ and $W_{(n-1)s+k}$ refers to the detrended observations:

$$W_{(n-1)s+k} = Y_{(n-1)s+k} + \mu_k^j. \tag{3}$$

The process $\{Y_{(n-1)s+k}\}$ is a PAR given by:

$$Y_{(n-1)s+k} = \sum_{i=1}^{p^j(k)} \phi_i^j(k) Y_{(n-1)s+k-i} + \epsilon_{(n-1)s+k}. \tag{4}$$

In the model proposed by Lu et al (2010) the only parameter that depends on the segment is the mean μ_k^j, which is allowed to shift its value by the same factor Δ_j for all seasons $k = 1, \ldots, s$. We assume that the trend parameters a^j and b^j depend only on the segment, whereas means μ_k^j are also allowed to change with seasons. The AR order at season k in the j-th segment is given by $p^j(k)$, so that $\phi_i^j(k)$, $i = 1, \ldots, p(k)$, represent the PAR coefficients during season k of the j-th segment. In our identification procedure we will set the same maximum AR order p for all segments and seasons, and will let the coefficients $\phi_i^j(k)$ be constrained to zero, in order to get more parsimonious models. This will be accomplished by introducing a binary vector δ^j of length $s \times p$ (named PAR lags indicator) in each segment j, which specify the presence or absence of $\phi_i^j(k)$ parameters. The first p digits represent the subset PAR model for period 1, subsequent p digits are related to period 2 and so on. For example, in the case of $s = 4$ and $p = 2$, the indicators $\delta^1 = (11010010)$ and $\delta^2 = (10000001)$ imply that parameters $\phi_1^1(1), \phi_2^1(1), \phi_2^1(2), \phi_1^1(4), \phi_1^2(1), \phi_2^2(4)$ are constrained to zero.

The error process $\epsilon = \{\epsilon_t, t \in \mathbb{Z}\}$ in equation (4) is a periodic white noise, with $E(\epsilon_{(n-1)s+k}) = 0$ and $\text{var}(\epsilon_{(n-1)s+k}) = \sigma_j^2(k) > 0$, $n \in R^j$, $j = 1, \ldots, M$, $k = 1, \ldots, s$. It is worth noting that the error variances $\sigma_j^2(k)$ are allowed to change with the segment, and our method is also able to detect this kind of change. Unless otherwise stated we assume that the subseries $Y_{(n-1)s+k}$ belonging to each segment are periodic stationary with period s, in the sense that:

$$Cov(Y_{n+s}, Y_{m+s}) = Cov(Y_n, Y_m), \tag{5}$$

for all integers $n, m, n + s$ and $m + s$ belonging to the same segment. The periodic stationarity can be checked using the same arguments as in Lund and Basawa (1999). More on causality and invertibility conditions for PAR models has been derived in Lund and Basawa (2000); Bentarzi and Hallin (1993).

2.2 Model estimation

The number of changepoints m, the changepoint locations $\tau_1, \tau_2, \ldots, \tau_m$ and the PAR lags indicators $\delta^1, ..., \delta^M$ are named as structural parameters. They can take discrete values and the number of possible combinations is very large. Once such parameters are determined (see subsection 2.2), the trend intercepts a^j, the slopes b^j, the seasonal means μ_k^j, the AR parameters $\phi_i^j(k)$ and the error variances $\sigma_j^2(k)$ (for segment j, season k and lag i) are analytically estimated. Assuming that the structural parameters are known, we propose to estimate the model parameters according to the following steps:

1. The trend parameters $a = (a^1, ..., a^M)$ and $b = (b^1, ..., b^M)$ are estimated by the Ordinary Least Squares (OLS) method:

$$\min_{a,b} \sum_{j=1}^{M} \sum_{k=1}^{s} \sum_{n \in R^j} \left(X_{(n-1)s+k} - a^j - b^j[(n-1)s + k] \right)^2, \tag{6}$$

that leads to the detrended data

$$\hat{W}_{(n-1)s+k} = X_{(n-1)s+k} - \hat{a}^j - \hat{b}^j[(n-1)s+k]. \tag{7}$$

2. The seasonal means $\hat{\mu}_k^j$ are computed on the resulting detrended data as follows:
$$\hat{\mu}_k^j = \frac{1}{\tau_j - \tau_{j-1}} \sum_{n \in R^j} \hat{W}_{(n-1)s+k} \tag{8}$$
and implying: $\hat{Y}_{(n-1)s+k} = \hat{W}_{(n-1)s+k} - \hat{\mu}_k^j$.
3. The AR parameters are estimated separately for each segment and season. Each specific series z_k^j is selected from $\hat{Y}$ and is incorporated in a design matrix Z of dimensions $(\tau_j - \tau_{j-1}) \times p$, that includes lagged observations. The subset selection constraints are specified by a $(p-q) \times p$ matrix H, where q is the number of free parameters. These constraints are designated on the basis of PAR lags indicator δ^j as follows:
   - For each lag i, the element $[p(k-1)+i]$ of δ^j vector is evaluated
   - If the value is equal to 1 then a row equal to the i-th row of the identity matrix I_p is added to H.

   The final estimate $\hat{\phi}^j(k) = (\hat{\phi}_1^j(k), \ldots, \hat{\phi}_p^j(k))$ of $\phi^j(k)$ is obtained by constrained optimization, with linear constraint given by $H\phi^j(k) = 0$. Explicitly (in matrix form):
$$\hat{\phi}^j(k) = \phi^{j,LS}(k) - (Z'Z)^{-1}H'[H(Z'Z)^{-1}H']^{-1}H\phi^{j,LS}(k), \tag{9}$$
   where $\phi^{j,LS}(k) = (Z'Z)^{-1}Z'z_k^j$ is obtained by OLS estimation.

4. Lastly, the estimation of error variances $\hat{\sigma}_j^2(k)$ (named residual variances) is performed for each segment and season on the final residuals:
$$\hat{\sigma}_j^2(k) = \frac{1}{\tau_j - \tau_{j-1}} \sum_{n \in R^j} \hat{\epsilon}_{(n-1)s+k}^2, \tag{10}$$
   where $\hat{\epsilon}_{(n-1)s+k}^2 = \hat{Y}_{(n-1)s+k} - \sum_{i=1}^{p^j(k)} \hat{\phi}_i^j(k) Y_{(n-1)s+k-i}$.

The selection of optimal structural parameters, on the other side, is a complex problem for which no closed form solution is available. In that it involves the evaluation of a very large number of possible combinations, GAs are naturally suitable for this issue.

2.3 Identification of structural parameters

The GA is a nature-inspired optimization method, introduced by Holland (1975), often employed when it is required to find an optimal solution from a prohibitively large discrete set. For a maximization problem it serves to approximate the optimal solution, which maximizes an objective function (named

fitness in the GA terminology), through a simple procedure. In the generic iteration (named generation), a population of binary encoded solutions (called chromosomes) is subdued to the so-called genetic operators: the selection randomly chooses chromosomes for the subsequent steps, usually proportionally to their fitness value; by crossover two solutions are allowed to combine together, with a fixed rate pC, exchanging part of their values and creating two new solutions; lastly, the mutation step allows each binary value to flip its value from 0 to 1 (or vice versa) with a fixed probability pM, providing a further exploration of the search space (bit-flip mutation). The resulting population replaces the previous one, and the flow of generations stops if a certain condition is met, for example a fixed number of generations. It is also possible, adopting the elitist strategy, to maintain the best chromosome found up to the current generation, irrespective of the effect of operators.

On the basis of pilot experiments, we decided to employ hybrid GA strategies for the structural parameter identification problem. According to a first possible strategy, we propose to encode the number of changepoints m and changepoint locations $\tau_1, ..., \tau_m$ in a binary chromosome, while the PAR lags indicators are obtained by enumerating all possible subset models at each fitness evaluation step and returning only the fittest one. This task is computationally feasible when the maximum AR order p is small, in that 2^p models must be evaluated for each segment and season. An alternative would be to make the search of changepoints conditional on complete PAR models, and perform the subset selection only for the best model found. This strategy may prevent possible interactions between changepoint detection and subset selection, but could also make the algorithm converge to a suboptimal solution. According to these ways of proceeding, the optimal values of structural parameters are obtained by combining an exact method (exhaustive enumeration) with an approximation procedure (GA).

In both strategies the binary chromosomes encode a candidate segmentation $[m, \tau_1, ..., \tau_m]$ as follows: the first three bits give the number of changepoints m (limited to a maximum of 7 in our study, so that a number of segments up to 8 is allowed); subsequent bit intervals, whose length is custom fixed, produce changepoint times $\tau_1, ..., \tau_m$. This part of encoding must ensure the following constraints:

$$\omega + 1 \leq \tau_1,\ \omega + \tau_1 \leq \tau_2,\ ...,\ \omega + \tau_{m-2} \leq \tau_{m-1},$$
$$\omega + \tau_{m-1} \leq \tau_m \leq N - \omega - 1,$$

due to the fact that a minimum number ω of years must be contained in each segment. In order to accommodate such constraints, the bit intervals directly encode m real numbers $th_i \in (0, 1)$, $i = 1, ..., m$, constructed to determine the percentage of remaining values to be attributed to the corresponding i-th segment. In fact, when placing a new changepoint there are some illegal positions, due to the constraints specified above: this implies that ω years must be left out from both the beginning and the end of the considered segment. This strategy depends on the candidate number of segments, and the changepoints are uniquely identified in the following ways:

- If $m = 0$ (one segment) then $\tau_1 = N + 1$.
- If $m = 1$ (two segments) then $\tau_1 = \omega + 1 + (N - 2\omega) \times th_1$
- If $m = 2$ (three segments) then:
  - $\tau_1 = \omega + 1 + (N - 3\omega) \times th_1$
  - $\tau_2 = \omega + \tau_1 + (N - 2\omega - \tau_1 + 1) \times th_2$
- If $m = 3$ (four segments) then:
  - $\tau_1 = \omega + 1 + (N - 4\omega) \times th_1$
  - $\tau_2 = \omega + \tau_1 + (N - 3\omega - \tau_1 + 1) \times th_2$
  - $\tau_3 = \omega + \tau_2 + (N - 2\omega - \tau_2 + 1) \times th_3$
- For a general m, we obtain the generic changepoint τ_j as:
  - $\tau_j = \omega + \tau_{j-1} + [N - (m + 2 - j)\omega - \tau_{j-1} + 1] \times th_j$.

Such an encoding procedure, introduced in Battaglia and Protopapas (2012b), always provides legal solutions and helps to save computational time.

As the fitness function measures the goodness of solutions, in our model identification problem it will include a term linked to the goodness of fit and a part related to a penalization on the number of parameters. Many options are available: we will consider a criterion inspired by the Normalized Akaike's Information Criterion (NAIC), introduced by Tong (1990) for threshold models, given by:

$$g(IC) = [\sum_{j=1}^{M} \sum_{k=1}^{s} n_{jk} \log(\hat{\sigma}_j^2(k)) + IC \sum_{j=1}^{M} \sum_{k=1}^{s} P_{jk}]/T, \tag{11}$$

where $\hat{\sigma}_j^2(k)$ is the model residual variance of series in segment j and season k, n_{jk} and P_{jk} are, respectively, the sample size and the number of parameters of segment j and season k, IC is the penalization term. The choice of IC specifies the magnitude of penalization on the number of parameters: for example a value equal to 2 resembles the structure of an Akaike's Information Criterion (AIC), while $IC = ln(N)$ leading to the analogous to Bayesian Information Criterion (BIC).

As an alternative, one can refer to the MDL criterion, which is based on the penalization given by the minimum length in bits necessary for describing the data. For a model $\mathcal{M}$, the MDL is given by:

$$MDL = C(\mathcal{M}) + C(\varepsilon|\mathcal{M}), \tag{12}$$

where $C(\mathcal{M})$ is the sum of the length to code the model and $C(\varepsilon|\mathcal{M})$ is the length to code the model errors. Rissanen (1978) showed that $C(\varepsilon|\mathcal{M}) = -\frac{1}{2}\log_2(\hat{L})$, where $\hat{L}$ is the maximum likelihood, therefore in our case:

$$C(\varepsilon|\mathcal{M}) = \frac{1}{2} \sum_j \sum_k n_{jk} log_2(\hat{\sigma}_j^2(k)), \tag{13}$$

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 maximum likelihood on n observations, code length is $\frac{1}{2}\log_2(n)$.

In our case we will write:

$$C(\mathcal{M}) = \log_2(max\{1,m\}) + m\log_2(N) + \log_2(s) + \log_2(p) + \\ + \frac{1}{2}M(s+1)\log_2(N) + \frac{1}{2}\sum_{j=1}^{M}\sum_{k=1}^{s}(p-|\delta_k^j|)\log_2(n_{jk}), \tag{14}$$

where the terms on the right refer to the encoding of, respectively, the number of changepoints m, the changepoint locations $\tau_1,\ldots,\tau_m$, the number of seasons s, the AR order p, the trend parameters and seasonal means, the subset PAR parameters (where $(p-|\delta_k^j|$ denotes the number of AR parameters in the subset model for segment j and season k).

If we adopt $g(IC)$ or MDL as identification criterion, the fitness will be based on a scaled transformation of the criterion. This is a common and widely discussed procedure in GAs (Goldberg, 1989; Kreinovich et al, 1993; Baragona et al, 2011, p.53), because it always provides positive values of the fitness to control the shape of the function and the pressure of the selection operator without changing the solutions ranking. It is generally recognized that the best scaling choice depends on the nature of the problem. We adopt a scaled exponential transformation, for maximization purposes: $f = \exp(-g(IC)/\beta)$ or $f = \exp(-MDL/\beta)$, where β is a problem dependent constant.

2.4 Choices of algorithm configurations for model identification

The nature of the GA allows us to introduce a large variety of algorithms, depending on the choices of configurations, operators and their related probabilities. As far as the selection operator is concerned, for example, one can adopt the roulette wheel strategy, whereby solutions are randomly selected with repetition proportionally to their fitness; according to the tournament selection, a single solution is compared with a group of solutions, or with another single one: if it wins, i.e., it has a better fitness than competitors, it is selected with a fixed probability, and rejected with complementary probability. Also many types of crossover operations are available: the single point crossover, which uses a common randomly chosen cutting point in the so-called parent chromosomes, and two new solutions built by taking the left part from the first parent and the right part from the other, and vice versa; an alternative is the *uniform*, which allows each gene of parent chromosomes to be individually swapped, with probability 0.5 (also a generic rate could be adopted, leading to the *parametrized uniform crossover*). There is no general dominant choice, so pilot studies are needed in order to understand the nature of the problem at hand (Goldberg, 1989; Eiben and Smith, 2003). On this basis, our GA showed no decisive dependence on the choice of operators if the number of generations

is chosen to be sufficiently large: therefore we will adopt the usual strategies like the ones described above.

Concerning our specific problem, the choice of the external parameters ω (minimum segment length) and M (maximum number of segments) can be crucial. In fact, our model allows us to split the time series into segments, each of which refers to a possibly different set of parameters related to the trend, the means, the autocorrelation structure and the error variances. In each segment these parameters are estimated on the basis of the subseries related to each season k, therefore ω refers to the minimum number of years used to estimate such parameters. As far as the AR parameters estimation is concerned, it is generally known that a very small number of observations could lead to unreliable estimates (Box and Jenkins, 1970). It is also clear that if the time series consists of N years, the constraint $M \cdot \omega \leq N$ must be satisfied. Therefore, depending on the sample size of the considered series, we must set M in such a way that the minimum segment length ω is able to provide reliable estimates.

In order to avoid having too few observations in any segment, Davis et al (2006) and Lu et al (2010) and Yau et al (2015) imposed a *minimum span*. This could be between 10 and 14 for a reasonable p (Davis et al, 2006), 12 (Lu et al, 2010), 40 (Yau et al, 2015) or a minimum span function depending on the order of the model for each segment (Song and Bondon, 2013). In our case we will estimate each AR parameter and seasonal mean on the subseries related to the segment j and season k. As the minimum segment length refers to the number of years, we suggest a minimum span $\omega = 12$ for PAR models with $p = 1$ and $\omega = 15$ for $p = 3$.

2.5 Forecasting and performance assessment

The forecasting method employed is the standard one-step-ahead procedure. We will remove the last year from the dataset (corresponding to 12 observations) in order to estimate the model, and use those data points for evaluating the forecasting performance. The logarithm of data is adopted as a Box-Cox transformation and performed before fitting the model. It is the most widely used transformation in monthly river flow analysis and it ensures that the model residuals are approximately normally distributed and homoscedastic (Eshete and Vandewiele, 1992; McLeod and Gweon, 2013).

The forecasting accuracy of the resulting models is evaluated with respect to the following measures:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{12}(y_i - \hat{y}_i)^2}{12}}, \tag{15}$$

$$MAE = \frac{\sum_{i=1}^{12}|y_i - \hat{y}_i|}{12}, \tag{16}$$

$$MAPE = \frac{\sum_{i=1}^{12} \frac{|y_i - \hat{y}_i|}{y_i}}{12} \times 100, \tag{17}$$

where $y_1, ..., y_{12}$ and $\hat{y}_1, ..., \hat{y}_{12}$ are, respectively, true and predicted values of last 12 observations, on logarithmic scale. Such measures are explicitly defined in Hyndman and Koehler (2006). These criteria must be interpreted only as an indication of model performance comparison, but no statement can be made from this comparison. All the measures are computed using the *hydroGOF* or the *forecast* packages of R software. Other measures of forecast accuracy can be found in Hyndman and Koehler (2006) and Krause et al (2005).

2.6 Model validation

An analysis of the residuals of the estimated model is needed to test its relevance. The stationarity of the residuals allows us to apply the standard 95% confidence limits, that is $1.96\sqrt{Ns}$. In order to test the joint statistical significance of the residual autocorrelations, McLeod (1994) proposed a Ljung-Box portmanteau test for PAR models:

$$Q_L(k) = N \sum_{l=1}^{L} \frac{N}{N - \lfloor (l-k+s)/s \rfloor} r_l(k)^2, \tag{18}$$

where k is the period, $r_l(k)$ is the autocorrelation coefficient at lag l, L is the maximum time lag considered and $\lfloor x \rfloor$ represents the integer part of the real number x. The test statistic $Q_L(k)$ follows approximately a Chi-squared distribution $\chi^2_{L-p(k)}$ with $L - p(k)$ degrees of freedom. We will perform this kind of diagnostic checking for each segment of an estimated model.

3 Simulation study

We will now present some simulation studies in order to illustrate the efficiency of the proposed procedure. Model identification will be performed by using the GA dealing with complete PAR models on a set of simulated datasets. We will simulate 500 time series consisting in a century ($N = 100$) of monthly data ($s = 12$). As far as the fitness function is concerned, we will study the sensitivity of the penalization IC in the criterion (11) by considering the following options: values of IC equal to 2 and $ln(N)$, which resemble the generalization of AIC and BIC criteria (hereinafter referred to simply as AIC and BIC), and also $IC = 3$, successfully adopted in Battaglia and Protopapas (2012a,b) for the identification of nonstationary nonlinear models by GAs. Moreover, we will consider the MDL criterion, therefore the possible fitness functions will be: $f = \exp(-g(IC)/\beta)$ and $f = \exp(-MDL/\beta)$, with $IC = 2, 3, \ln(N)$ and scaling constant $\beta = 10$.

Initially we will discuss the ability of our method to detect changes in the seasonal mean, and compare the performance with the procedure proposed by Lu et al (2010) (LLL in the following), which is designed to detect mean shifts in PAR models. The GA adopted by LLL uses the MDL as a fitness function and has different configurations from our procedure: the encoding is integer, therefore the model structure is directly coded in the chromosome; the crossover and the mutation act on a single pair of solutions, chosen by rank selection, to generate one new solution; a subpopulations-based strategy with periodic migrations is also employed. Our GA uses binary encoding and it is generational.

In the first simulation experiment (Simulation A) we perform a comparison between the LLL method and our procedure, considering a generating model with six level shifts at times: $\tau_1 = 15, \tau_2 = 30, \tau_3 = 45, \tau_4 = 60, \tau_5 = 75, \tau_6 = 87$, each having the same shift magnitude of about 3 (corresponding to $K = 2$ in the LLL set-up). We also consider two other model scenarios: one change in the AR structure (Simulation B) at time $\tau_1 = 61$, and two changes in the error variance (Simulation C), occurring at times $\tau_1 = 31$ and $\tau_2 = 61$. For these two scenarios we also evaluate the forecasting accuracy of the resulting models: therefore we remove the last year in the model identification step, and we use it to evaluate the forecasting (see subsection 2.5). A summary of the parameters used to generate the models is reported in Table 1. In all scenarios we maintained a common vector of seasonal means specified by μ_k^j and used the error variances $\sigma_j^2(k)$ employed in Lu et al (2010), unless otherwise specified. In Simulation B the first segment has a PAR structure specified by the parameters indexed by j_1, while in the second we use the remaining set of parameters (j_2). In Simulation C we used the j_2 PAR structure, and the error variances multiplied by 0.25 in the second segment and by 4 in the third. Examples of time series generated according to these models are reported in Figure 1.

Concerning the choices of GA configurations, we will use a population of 50 solutions and employ the operators of tournament selection, bit-flip mutation (rate 0.1) and parametrized uniform crossover (rate 0.7). The elitist strategy will also be employed. We will allow a maximum AR order of $p = 1$ for Simulation A and $p = 3$ for Simulations B and C, while ω is 12 for Simulation A and 15 for the other two scenarios. Results are summarized in Tables 2, 3 and 4: they report the percentage of detection of the exact real change times, the same indicator allowing for an absolute error of one year, the percentage of detection of the number of changes and an average over the 500 replications of the RMSE and MAE indexes for Simulations B and C. As LLL method allows for mean shifts at any month, we will report only the percentages of detection of the real changes with the error of one year as a matter of comparison.

The results of Simulation A show that using either AIC or $IC = 3$, our method performs extremely well, with 100% and 94.8% correct numbers of estimated changepoints. The BIC criterion detects the true number of changepoints only in 6.2% of replications, while for 176 series 2 changepoints are found. As with our method combined with the MDL criterion, the results

	μ_k^j	$\sigma_j^2(k)$	$\phi_1^{j1}(k)$	$\phi_2^{j1}(k)$	$\phi_3^{j1}(k)$	$\phi_1^{j2}(k)$	$\phi_2^{j2}(k)$	$\phi_3^{j2}(k)$
1	1	2.713	0.3	0.5	0	0.1	0.3	-0.4
2	1	2.748	0.42	0	0	0.1	0.3	-0.4
3	1	1.871	-0.8	0.4	0.35	0.1	0.3	-0.4
4	2	1.717	-0.3	0	0	0.22	-0.1	-0.5
5	2	2.474	0.7	-0.35	0.4	0.22	-0.1	-0.5
6	2	2.403	0.4	-0.5	0	0.22	-0.1	-0.5
7	3	2.569	0.7	0	0	-0.4	0.23	0.25
8	3	1.910	-0.6	0	0	-0.4	0.23	0.25
9	3	2.826	0.4	0.3	0.4	-0.4	0.23	0.25
10	4	2.488	0.9	0	0	-0.5	0.4	0.1
11	4	2.394	-0.6	0.4	0	-0.5	0.4	0.1
12	4	2.256	0.72	0	0	-0.5	0.4	0.1

Table 1: Summary of the parameters used in simulations: seasonal means μ_k^j, error variances $\sigma_j^2(k)$ and PAR parameters

are better than those obtained using BIC, with 37.6% correct numbers of estimated changepoints. Therefore our method combined with BIC or MDL seems to underestimate the number of changepoints. When the number of changepoints is correctly identified, the best results in terms of changepoint time detection are achieved using AIC and $IC = 3$. The LLL procedure detects the real number of changepoints in 42% of replications, and the changepoint locations are correctly identified between 60% and 70% of cases.

	% 15	% 15 ± 1	% 30	% 30 ± 1	% 45	% 45 ± 1
AIC	67.4	91.2	74.4	90.6	70.8	90.2
IC=3	64.4	87.8	68.4	88.2	73.4	92.4
BIC	37.2	41.6	22.0	25.8	31.2	34.4
MDL	47.0	60.8	39.0	48.8	42.4	53.4.0
LLL	/	60.4	/	65.0	/	60.4

	% 60	% 60 ± 1	% 75	% 75 ± 1	% 87	% 87 ± 1
AIC	69.4	89.8	78.6	94.4	75.0	100.0
IC=3	72.0	90.8	82.2	95.6	75.0	100.0
BIC	38.0	42.4	39.4	42.4	52.2	55.8
MDL	53.6	62.8	65.6	71.4	71.0	81.2
LLL	/	62.2	/	70.0	/	62.4

	$m = 0$	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m > 6$
AIC	0	0	0	0	0	0	100	0
IC=3	0	0	0	0	2.6	2.6	94.8	0
BIC	6.0	17.2	35.2	18.4	12.6	4.4	6.2	0
MDL	3.4	6.0	18.4	13.0	18.8	3.6	37.6	0
LLL	2.2	10.0	17.4	13.8	5.6	4	42.4	4.6

Table 2: Simulation A: percentages of exact detection of the real changepoint years $\tau_1 = 15$,$\tau_2 = 30$, $\tau_3 = 45$, $\tau_4 = 60$, $\tau_5 = 75$ and $\tau_6 = 87$ over 500 replications, allowing an absolute error of one year, the percentages of detection of m changepoints

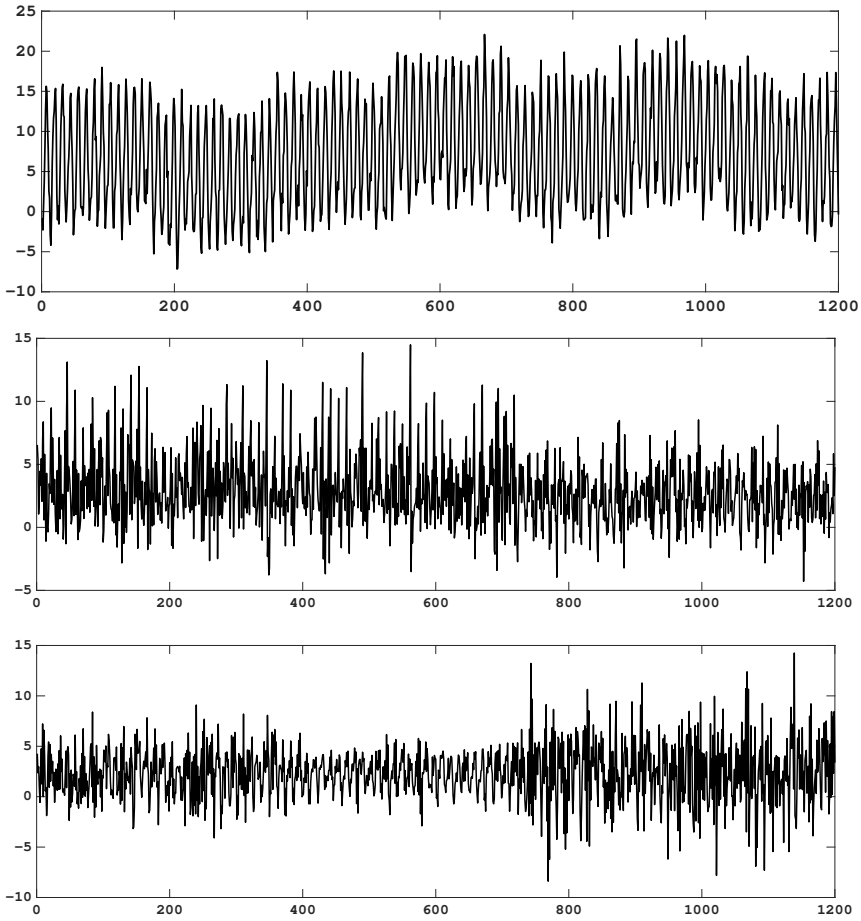

Fig. 1: Examples of time series related to Simulations A, B and C

When we apply our method to the realizations in Simulation B we observe that BIC, $IC = 3$ and MDL always detect the correct number of changepoints. AIC has a correct identification rate of 72.6% and sometimes overestimates the number of changepoints, as in 26.0% of the replications 2 changes are found. The LLL method fails to detect the changepoint, because we choose different autocovariance structures for each segment. We can see that our procedure performs very well in locating the changepoint for all the criteria. As far as the forecasting accuracy is concerned, we also observe that $IC = 3$ and MDL criteria provide slightly better results with respect to AIC and BIC. Such differences are due to the subset model selection, whose results depend on the penalization introduced in the fitness.

	% 61	% 61 ± 1
AIC	95.4	100
IC=3	96.4	100
BIC	96.4	100
MDL	96.4	100

	$m=1$	$m=2$	$m=3$	$m=4$	$m>4$	RMSE	MAE
AIC	72.6	26.0	1.00	0.40	0	1.6603	1.3538
IC=3	100	0	0	0	0	1.6502	1.3485
BIC	100	0	0	0	0	1.6665	1.3610
MDL	100	0	0	0	0	1.6558	1.3530

Table 3: Simulation B: percentages of exact detection of the real changepoint year $\tau_1 = 61$ over 500 replications, allowing an absolute error of one year, the percentages of detection of m changepoints, averages over 500 replications of RMSE and MAE indices

Table 4 lists the results for Simulation C. Our method combined with the $IC = 3$ criterion gives the correct number of changepoints for 99.2% of the 500 realizations. We can see that BIC and MDL seem to underestimate the number of changepoints, while AIC leads to good results (correct rate 78.6%). The changepoint at year $\tau_2 = 61$ is easier to detect, because the error variance of the third segment is 16 times larger than that in the second. This is confirmed by the results in Table 4, which show that such changepoint is correctly identified in at least 96% of the replications. The changepoint $\tau_1 = 31$ has a lower correct detection rate, with the $IC = 3$ criterion providing the best results. Concerning the forecasting accuracy, the best results are observed with AIC and $IC = 3$, possibly because they have a higher rate of detection of the correct number of changepoints.

	% 31	% 31 ± 1	% 61	% 61 ± 1
AIC	46	79	96.4	99.4
IC=3	47	83	98.4	100
BIC	10.8	20	97.6	99.4
MDL	14.2	25.4	97.4	99.6

	$m=1$	$m=2$	$m=3$	$m=4$	$m>4$	RMSE	MAE
AIC	0	78.6	19	2.4	0	2.8468	2.3162
IC=3	0.8	99.2	0	0	0	2.8367	2.3147
BIC	76.8	23.2	0	0	0	2.8645	2.3405
MDL	68.6	31.4	0	0	0	2.8514	2.3296

Table 4: Simulation C: percentages of exact detection of the real changepoint years $\tau_1 = 31$ and $\tau_2 = 61$ over 500 replications, allowing an absolute error of one year, the percentages of detection of m changepoints, averages over 500 replications of RMSE and MAE indices

A generic indication of these simulation studies is that the results are strongly dependent on the penalization in the fitness. This is a subjective

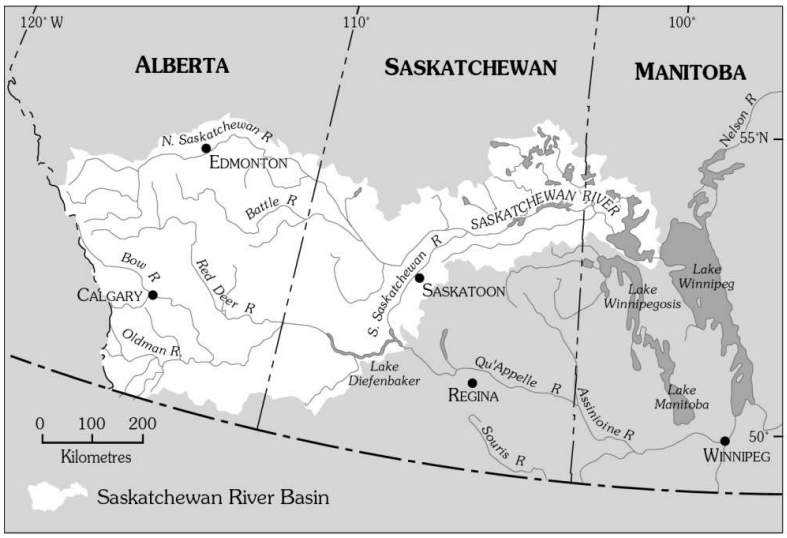

Fig. 2: Saskatchewan river basin.
Source: South East Alberta Watershed Alliance.

element, because in real applications the researcher decides on the compromise between accuracy and parsimony, given by the criterion. Moreover, the success in using one or the other penalization criteria seems to depend in turn on the true (and unknown in real applications) number of changepoints. Therefore we suggest that more than one identification criterion should be employed in real applications.

4 Study area and research data

We will now evaluate the effectiveness of the proposed methodology in river flow analysis. Data related to two rivers displaying a similar behaviour during winter and summer (Fig. 6), of the same length, with different means of annual flows, located in different regions, will be examined. They consist of:

- flows of South Saskatchewan river, measured at Saskatoon, Canada;
- flows of Colorado river, measured at Lees Ferry, Arizona.

The South Saskatchewan river originates in the Rocky Mountains and drains an area of about 139,600 km^2 (Fig. 2). It passes through the Canadian prairies, a major agricultural region with high hydrological variability (Gober and Wheater, 2014). The South Saskatchewan river is one of the largest and most important rivers in Saskatchewan, as almost 50% of the province's population depends on the river for daily needs. The main uses of water from the river are: agricultural irrigation, industry (power production, petroleum related operations) and municipal uses.

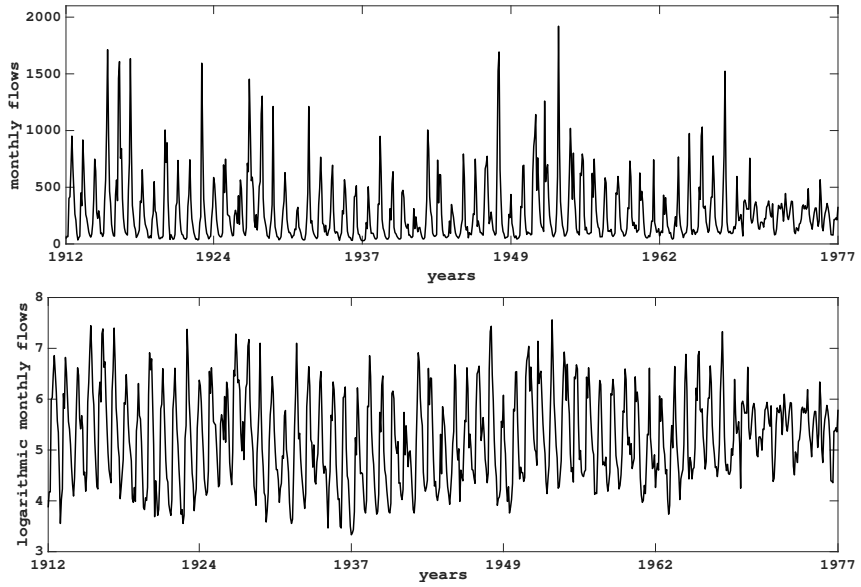

Fig. 3: Monthly flows (up) and logarithmic monthly flows (down) for the South Saskatchewan river.

Lake Diefenbaker is a reservoir lake formed by the construction of Gardiner Dam and Qu'Appelle River Dam across the South Saskatchewan and Qu'Appelle rivers, respectively. Lake Diefenbaker's Gardiner Dam has heavily modified a series of extreme events (floods occurred in summer when rainfall coincided with snow melt). From Lake Diefenbaker the river flows towards the City of Saskatoon and continues north to become Saskatchewan river at the confluence with the North Saskatchewan river. After the confluence, the river passes through Saskatchewan Delta, into Lake Winnipeg.

Although a large fraction of global water resources is available in Canada, the South Saskatchewan river exemplifies the multiple threats to water security: extreme events, rapid population growth and economic development, increasing pollution (Gober and Wheater, 2014). In an intervention analysis study, Hipel and McLeod (1994) used South Saskatchewan river data to determine the alteration of the average monthly flows following the Gardiner Dam operations.

In this paper we analyze the time series of mean monthly flows of the South Saskatchewan river measured at Saskatoon, Canada. The collection period ranges from January 1912 to December 1976 (780 observations, 65 years), and coincides with the available data.[1] The monthly series and the logarithmic monthly flows are reported in Figure 3. Before the creation of Lake Diefen-

[1] Source: http://www.stats.uwo.ca/faculty/mcleod/epubs/mhsets/readme-mhsets.html

baker, the South Saskatchewan river had higher flows from April to August, with declining flows during the fall and low flows in winter (Fig. 6).

The Colorado river (Fig. 4), which flows through seven U.S. states and two Mexican states, is one of the principal rivers of southwestern U.S. and northern Mexico, as it provides water to 40 million people. The Colorado river begins at La Poudre Pass in the Southern Rocky Mountains of Colorado, flows southwest across the Colorado Plateau and through the Grand Canyon, reaches Lake Mead and forms a large estuary before emptying into the Gulf of California, Mexico after a trip of 2330 km. The Upper Colorado river, where upper refers to the course upstream of the Green River (Van Steeter and Pitlick, 1998), passes through the Grand Valley, a major farming and ranching region. After providing water to numerous towns (like Bullhead City, Needles and Lake Havasu), the Lower Colorado river irrigates California's Imperial Valley, the most productive winter agricultural region in the United States. The remaining flow is diverted to irrigate the Mexicali Valley, which is among the most fertile agricultural lands in Mexico. The Colorado river faces many of the same challenges as the Saskatchewan river: persistent drought, climate change, population growth. Water demand impacts the regional economies, challenges food production, degrades the environment, and limits recreational opportunities. Therefore, the Colorado river is stretched to its limit.

In this paper we consider monthly data for a 65-years period (1906-1970) measured at the Lees Ferry station.[2] The data and the log transformed data are displayed in Figure 5. The reason for selecting such observations is due the fact that data starting from 1971 are obtained by *Natural Flow And Salt Calculation* models (Prairie and Callejo, 2005), which are always subject to changes in successive updates.

5 Results and discussion

The South Saskatchewan river flows series is discussed in Noakes et al (1985). The authors used nine different seasonal models to generate thirty-six one-step-ahead forecasts for the logarithmic flows in the South Saskatchewan river, and found that PAR models gave the best results with respect to the criterion of RMSE. The Colorado river is discussed in Srivastav et al (2016), McKee et al (2000) and Van Steeter and Pitlick (1998), among others.

Estimation and forecasting will be performed as discussed in Section 2. We will employ the hybrid GA strategy that enumerates all possible subset models at each fitness evaluation step. As far as the choice of genetic operators is concerned, we propose tournament selection, bit-flip mutation and a modified single-point crossover: instead of allowing all of the bits in the chromosome to be selected as possible cutting points, we will consider only the bits that subdivide the segmentation $[m, \tau_1, ..., \tau_m]$. In such a way the parameter structures can be naturally inherited by the solutions' offspring, avoiding destroying so-

[2] Source: https://www.usbr.gov/lc/region/g4000/NaturalFlow/current.html

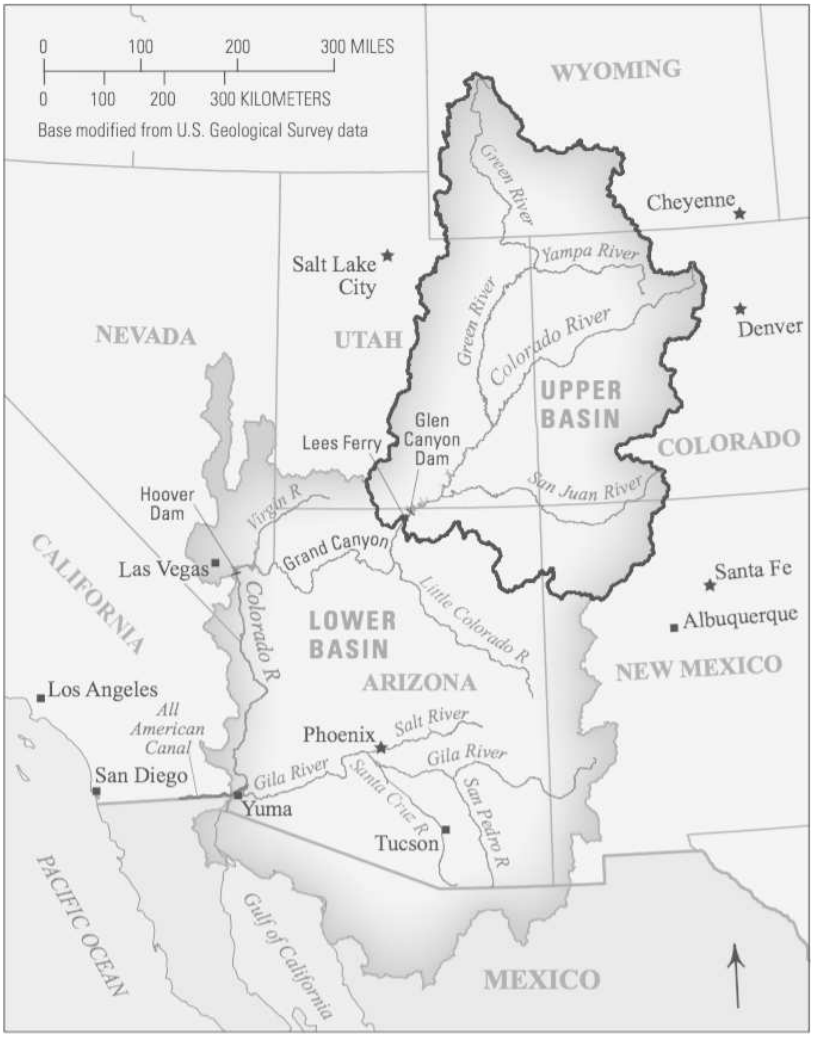

Fig. 4: Colorado river basin.
Source: U.S. Department of the interior.

lutions. The elitist strategy is also adopted. In these analyses we will employ also the changepoint detection method of LLL for comparison purposes.

In the fitness function we will use $IC = 3$, BIC or the MDL criterion. Several experiments will be conducted considering various combinations of parameters p and ω, in order to provide a variety of explanatory models. The forecasting accuracy of these models will be then evaluated, and the portmanteau test-based diagnostic checking discussed in subsection 2.6 will also be conducted. Computations will be performed using Matlab and R softwares.

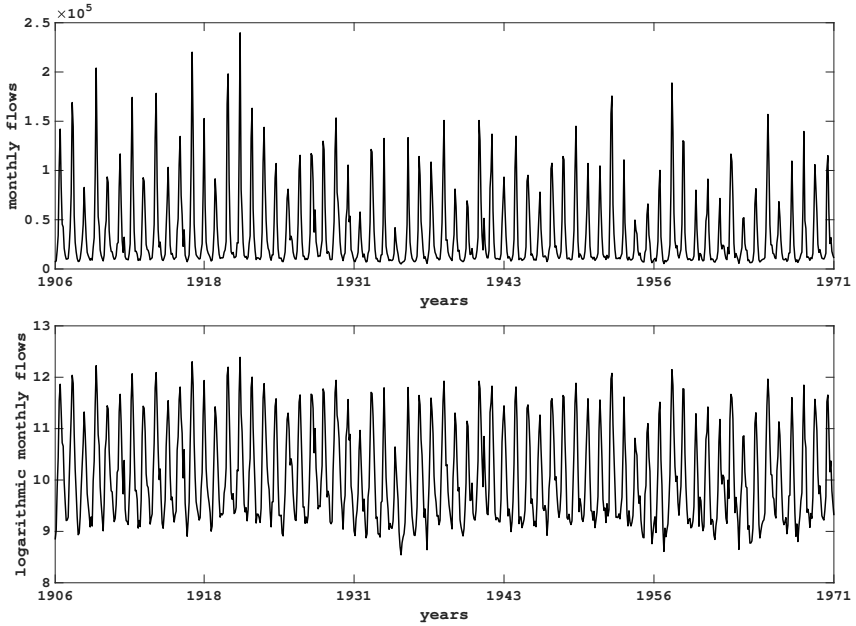

Fig. 5: Monthly flows (up) and logarithmic monthly flows (down) for the Colorado river.

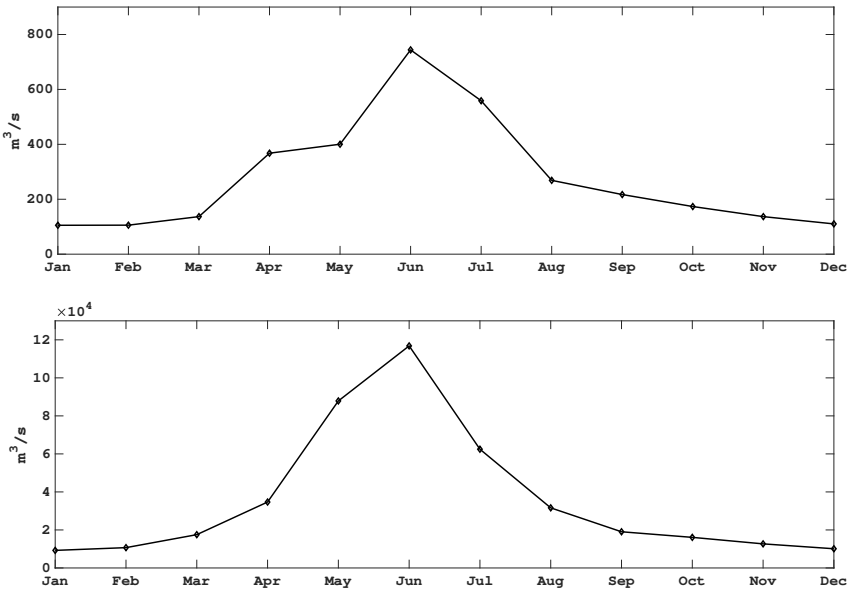

Fig. 6: Monthly means of the South Saskatchewan river (up) and Colorado river (down)

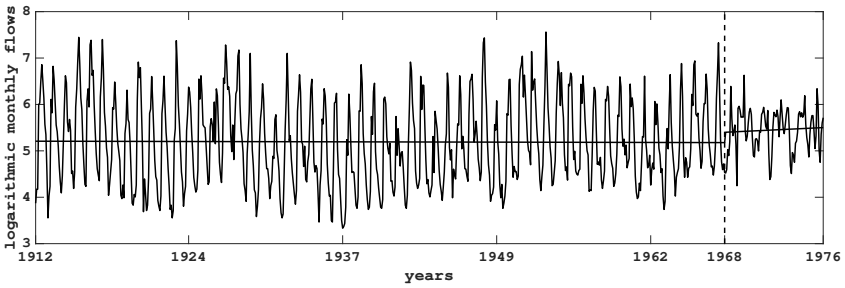

Fig. 7: Changepoint detected on year 1968 for Saskatchewan river

5.1 South Saskatchewan river

Noakes et al (1985) fitted the log transformed data with nine models and they recommended the PAR model. Using a PAR without changepoints (denoted by *Model 1*), we obtain a similar model to that in Noakes et al (1985); the slight differences arise from the fact that our model allows for intermediate constraints and we use more data. In addition, our method was applied to estimate a PAR model with at least one changepoint, using BIC with $p = 1$ and $\omega = 7$ (*Model 2*) and $IC = 3$ with $p = 3$ and $\omega = 7$ (*Model 3*). As far as very short time segments could be identified, we will interpret the results with caution and with the descriptive purpose of understanding the hydrological processes of the river flow. Results are reported in Table 6.

Using our model combined with BIC we detect year 1968 as the only changepoint (Figure 7). We also note that our method combined with MDL and the LLL method locate the same year of change. January 1969 corresponds to a modification in the reservoir system management: the Gardiner Dam came into full operation. It is the third largest embankment dam in Canada and one of the largest in the world. The reservoir provided valuable benefits to the community: power generation, recreational benefits and also a decrease in the magnitude of floods, with minimum flows downstream guaranteed throughout the year. Before the creation of the dam, snowmelt in conjunction with summer rains led to heavy flooding. To model the effect of the operation of the Gardiner Dam on the average monthly flows of the South Saskatchewan river, Hipel et al (1977) developed an intervention model. Using flows measured at Saskatoon from 1942 to 1974, they found that the operation of the Gardiner Dam significantly affected the average monthly flows. They increase from November to March and decrease from April to September.

To ascertain the type of changes in the time series data due to detected changepoints, we calculate the 12 estimated seasonal means for all the years up to changepoint 1968, and from 1968 onwards. We plot the seasonal means for each segment in Figure 8. This kind of graphic representation has been used by Gober and Wheater (2014) to show that the construction of Gardiner Dam has heavily modified conditions downstream. We observe that the seasonal

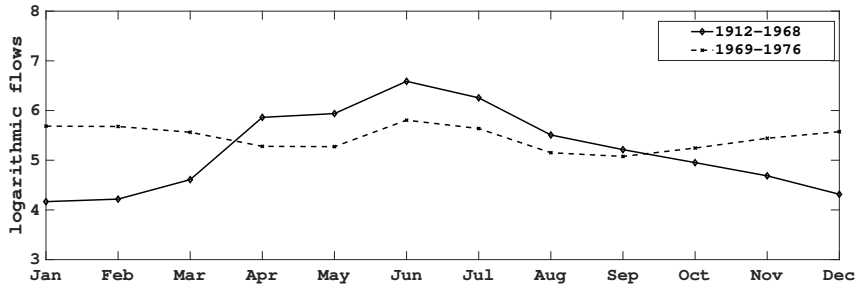

Fig. 8: Monthly means of the logarithmic flows of South Saskatchewan river before and after changepoints detection

Month	percentage change
Jan.	36.46
Feb.	34.65
Mar.	20.68
Apr.	-9.94
May	-11.21
June	-11.87
July	-9.88
Aug.	-6.48
Sep.	-2.62
Oct.	5.92
Nov.	16.14
Dec.	29.18

Table 5: Percentage change in mean before and after 1968 for the logarithmic flows of South Saskatchewan river

means decreased from April to September after the detection of the changepoint, compared with the previous period. Table 5 lists the percentage change in mean monthly flows between different segments (a negative sign indicates decrease in flows). We can explain the detection of the changepoint using the LLL method by observing the magnitude of such changes in mean.

Using our method combined with $IC = 3$ we detect three changepoints, corresponding to 1937, 1961 and 1968 (Figure 9). Looking into the history, the changepoint corresponding to 1937 could be linked to the drought conditions on the Saskatchewan prairies during the Dust Bowl years of the 1930s. The drought came in three waves: 1934, 1936, and 1939-1940, but some regions experienced drought conditions for as many as eight years. The changepoint corresponding to 1961 could arise as an effect of the beginning of dam construction in 1964. We note that several corrections have been made to the monthly flows from January 1964 to December 1968. For this reason 1961 could be viewed as an artificial changepoint introduced as a direct effect of the means correction.

The performance of the models was considered in terms of fitness, goodness of estimation, and RMSE, MAE and MAPE for evaluating the accuracy

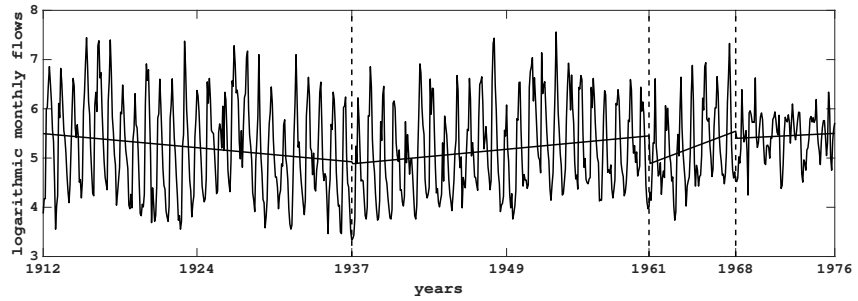

Fig. 9: Changepoint detected on years 1937, 1961 and 1968 for Saskatchewan river

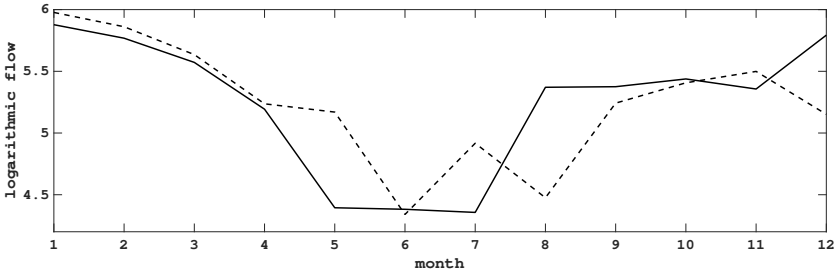

Fig. 10: Twelve-month forecast (dashed line) based on 64 years of Saskatchewan river data. The actual data (solid line) were not used in the forecast.

	Years of changepoint	*RMSE*	*MAE*	*MAPE*	*Fitness*
Model 1	/	0.66428	0.5126	10.2718	1.2253
Model 2	1968	0.4360	0.3167	6.3646	1.2553
Model 3	1937, 1961, 1968	**0.4275**	**0.2933**	**5.8488**	1.2661

Table 6: Results of the evaluation criteria of the logarithmic forecast errors for Saskatchewan

of forecasts. The fitness values are not comparable, as different penalizations have been used. Concerning forecasting, the South Saskatchewan flow series was split into two sets and several PAR models were fitted for the first set of the data. Then the fitted models were used to generate one-step-ahead logarithmic forecasts for the second set of the data (the last year or 12 observations). In practical applications, the one-step-ahead forecasts are very important when accurate forecasts for the inflows are crucial (flood prediction, hydro-electric power generation). The PAR model with 4 segments (*Model 3*) seems to perform better in terms of forecasting than the other models (Table 6). Figure 10 shows forecasts for *Model 3* and the observed data for the last year of the data set (1976).

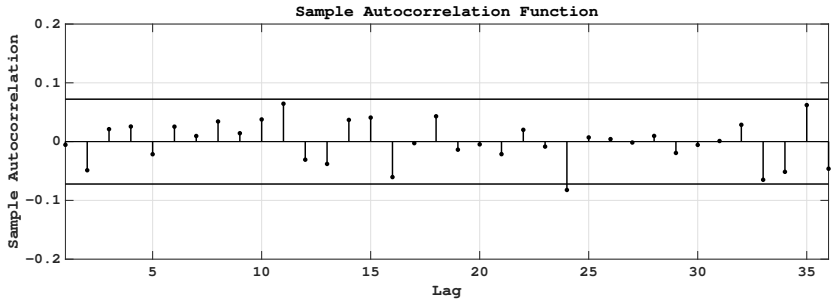

Fig. 11: Autocorrelation function (ACF) of the residuals of the fitted PAR models with year 1968 detected as the changepoint to the South Saskatchewan flow.

	1912-1968	1969-1975
January	0.9285	0.2376
February	0.0381	0.3727
March	0.7502	0.3601
April	0.4695	0.3692
May	0.4121	0.3199
June	0.3694	0.1132
July	0.6612	0.7599
August	0.2049	0.2018
September	0.3204	0.2626
October	0.2399	0.4064
November	0.2295	0.3871
December	0.1816	0.0218

Table 7: P-values of the portmanteau test defined in eq. (18) with $L = 15$.

In order to check for the whiteness of the residuals, their autocorrelations up to lag 36 were computed for *Model 2*. The two full lines indicate lower and upper bounds of the ACF assuming that the residuals are white noise. No significant autocorrelations were found (Figure 11). This graphic checking provides some evidence on the adequacy of the proposed PAR model. Table 7 shows the P-values of the portmanteau test (18): they suggest that the proposed model is not rejected at the 5% significance level, except for February 1912-1968 and December 1969-1975, for which the P-values are larger than 2%, which does not strongly suggest model inadequacy. In this situation, the PAR model without changepoints seems inappropriate at the 5% nominal level.

5.2 Colorado river

We transformed the original data from ft^3 to m^3/s, and considered a standard PAR model without changepoints (denoted by *Model 1* as a basis in Table 9). Considering a BIC penalization, when we impose $p = 1$ as upper bound for the order of PAR models (*Model 2*), we found 1918 as changepoint (Figure 12),

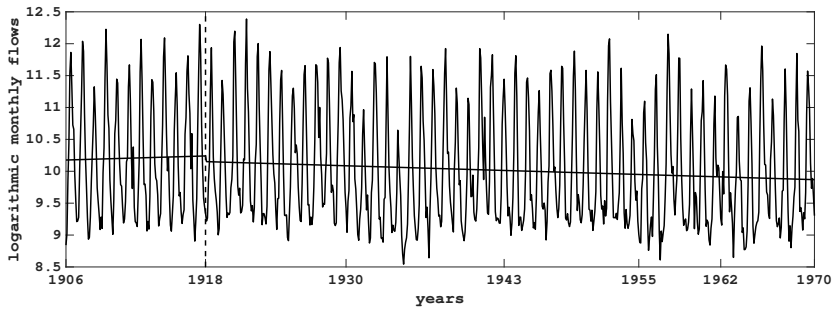

Fig. 12: Changepoint detected on year 1918 for Colorado river

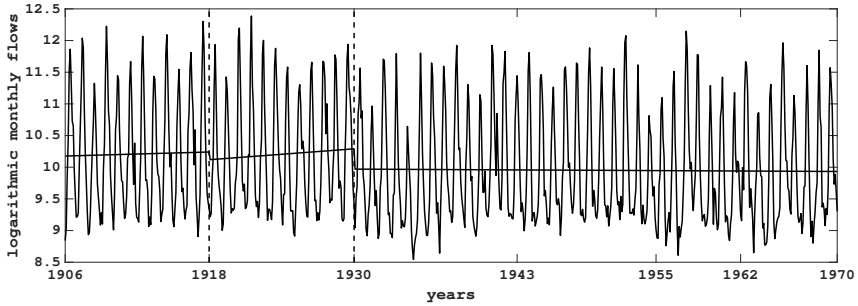

Fig. 13: Changepoints detected on years 1918 and 1930 for Colorado river

whereas with $p = 3$ (*Model 3*) we detected year 1931. On the other hand, our method combined with $IC = 3$ (*Model 4*) and $p = 1$ detected 1918 and 1930 as change times (Figure 13). Using the LLL method no changepoints were detected.

As for the Saskatchewan river, we calculate the 12 seasonal means for all years up until the first changepoint, and for each period until a new changepoint is detected. Table 8 lists the percentage changes in mean and variance of the logarithmic monthly flows between different segments for *Model 2*. We observe that there are no significant changes in mean after the detection of the changepoint, compared with the previous period. This may explain why the method of LLL was unable to detect this changepoint. On the other hand, the changes in variance are noteworthy and this explains why the changepoints were detected with our method. We note that a similar behaviour is observed for the other models.

There are several arguments supporting our findings. Looking into the history, the changepoint corresponding to 1930 could be linked to the most widespread and longest lasting drought (1930-1940) and to the longest wet period (1905-1929) in Colorado history (McKee et al, 2000). Woodhouse et al (2016) examined the influence of precipitation, temperature, and antecedent oil moisture on the flows of the Colorado river at Lees Ferry, from year 1906

Month	percentage change in mean	percentage change in variance
Jan.	-0.18	43.72
Feb.	-0.20	72.98
Mar.	-2.92	-39.44
Apr.	-2.06	225.46
May	-0.95	115.84
June	-2.25	42.12
July	-3.44	-2.01
Aug.	-2.62	25.29
Sep.	-3.38	54.62
Oct.	-4.09	82.35
Nov.	-0.96	482.86
Dec.	0.39	290.71

Table 8: Percentage changes in mean and variance before and after 1918 for the logarithmic flows of the Colorado river

to 2012. They divided what they called "anomalous flow years" into four categories: two categories for which the flow was greater than expected (above and below the median) and two categories where the flow was less than expected (above and below the median), given that year's cool-season precipitation. They found that most of the years falling in the first two categories are between 1918 and 1930. In general, there is a clustering of anomalous years of flow in the early part (1910s and 1920s), followed by an interval with fewer flow anomalies (1930s to 1970s).

Novak et al (2012) found that the runoff efficiency (which is the ratio of flow volume to precipitation volume) is strongly correlated with temperature in a study of the Upper Colorado river basin for the years 1906-2006. They estimated a 14% reduction in the annual Colorado streamflow with each 1° C of warming. Other climatic factors, such as late spring and summer precipitation could influence the runoff efficiency and so the streamflow (Woodhouse and Pederson, 2018). We represent the difference between the average of each year and the average over all the years of the temperature in Coconino County, Arizona (Figure 14), which also includes the gauging station at Lees Ferry. We observe that the temperature increases after 1930, confirming the slight decreasing trend in the Colorado river flow found in *Model 4*. This result is also confirmed by the examination of temperature, precipitation and streamflow time series for the period 1906-2012, which indicates three periods: 1906-1933 (cooler than average temperature and streamflow higher than precipitation), 1934-1987 (near average temperature period) and 1988-2012 (warmer-than-average temperature) (McCabe et al, 2017). Among other factors that could also influence the flow, we report the SPI (standardized precipitation index) and the accumulated precipitation deficit (McKee et al, 2000). In general, a changepoint in river flows associated with precipitation reduction in combination with increased temperatures will likely result in severe droughts. Water managers tasked with providing sufficient water for the regional economy, food production, power generation, and household needs, have worries about the

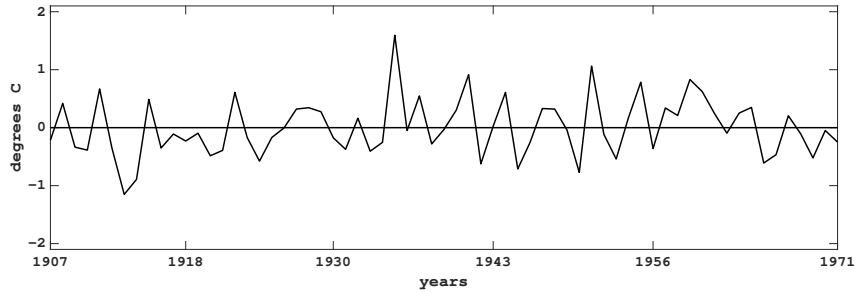

Fig. 14: Temperature evolution

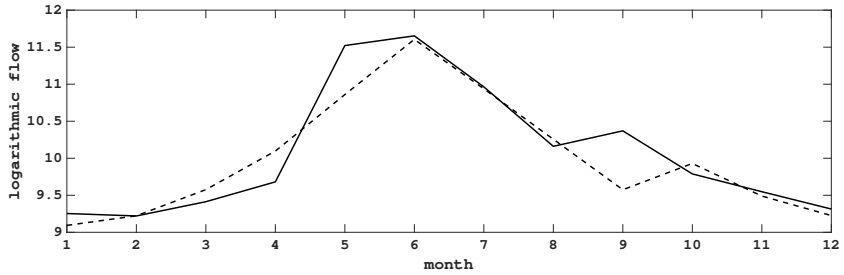

Fig. 15: Twelve-month forecast (dashed line) based on 64 years of Colorado river data. The actual data (solid line) were not used in the forecast.

large uncertainty in climate change projections. Therefore, careful detection of changepoints warranted when placing uncommon events, such as concurrent droughts across the Colorado basin, could be very useful for water resources managers.

In order to evaluate the accuracy of forecasts, the RMSE, MAE and MAPE were computed, as the last year (12 observations) is omitted from the dataset. Several PAR models are fitted to the logarithms of the truncated time series and one-step-ahead logarithmic forecasts are computed. *Model 2* seems to perform better in terms of forecasting than the other models (Table 9). The results suggest that a PAR model with changepoints provides more accurate forecasts than a standard PAR model. Figure 15 shows the forecasts for the logarithmic transformed data, compared with the observed data, for the last year of the data set (1970).

The autocorrelations of residuals up to lag 36 were computed for *Model 4* for the sake of diagnostic checking (see Figure 16). In terms of P-values of the portmanteau test (Table 10), *Model 4* seems to perform very well, as such values suggest that the proposed model is not rejected at a 5% significance level, except for a few cases where the P-values are smaller than 5%.

	Years of changepoint	$RMSE$	MAE	$MAPE$	$Fitness$
Model 1	/	0.3372	0.2276	2.2219	1.3092
Model 2	1918	**0.3341**	**0.2214**	**2.1499**	1.3107
Model 3	1931	0.3716	0.2530	2.4766	1.3159
Model 4	1918,1930	0.3460	0.2266	2.1990	1.3198

Table 9: Results of the evaluation criteria of the logarithmic forecast errors for the Colorado river

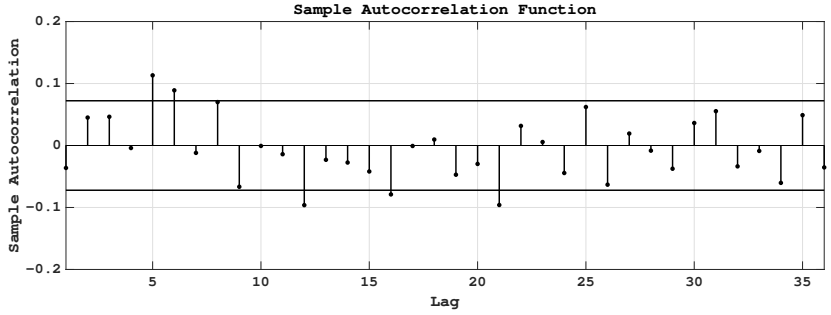

Fig. 16: Autocorrelation function (ACF) of the residuals of the fitted PAR *Model 3*.

	1906-1918	1919-1930	1931-1970
January	0.1721	0.1507	0.1441
February	0.4599	0.8637	0.2720
March	0.0639	0.9160	0.3094
April	0.8100	0.2124	0.0356
May	0.4777	0.1437	0.7697
June	0.5451	0.4226	0.6387
July	0.5714	0.2637	0.5772
August	0.0269	0.3994	0.5772
September	0.8335	0.0158	0.9879
October	0.0828	0.1353	0.5596
November	0.0295	0.1449	0.1789
December	0.6963	0.2722	0.1921

Table 10: P-values of the portmanteau test defined in eq. (18) with $L = 10$.

6 Conclusions

The objective of our research was to develop a computational procedure to estimate the number of changepoints and their locations in time series with a periodic structure. We will now summarize the contributions and the findings of the paper:

1. The proposed model accounts for both seasonality and changes in the PAR model structure, allowing a description of several sources of discontinuity and inhomogeneity.

2. The computational method served to build the model by the use of GAs according to several strategies. Simulation studies showed the efficiency of the method in detecting changes in mean, autocorrelation and error variance.
3. Our procedure estimated changepoints for real time series related to the river flows of South Saskatchewan (Canada) and Colorado (U.S.). This allowed us to discuss the hydrological process of such river flows in relation to both human activities and climatic oscillations. The comparison of our findings with other methods and studies in the literature confirmed the efficiency of our procedure.

In our study we examined monthly data with changepoints allowed only at the end of the year (that is, a multiple of the number of seasons). Modifications to the method proposed in the present paper are under study: techniques for monthly, weekly or daily time series with a periodic structure allowing changepoints at any season are worth pursuing. For example, detecting a changepoint in the middle of a year will prevent the dispersal of its effect over adjacent seasons. Moreover, as PAR models are based on a large number of parameters, one could question whether it is necessary to consider a separate AR model for each season: we allowed subset PAR models to be built in order to conveniently decrease the number of parameters, but a further and considerable gain in parsimony would be achieved by reducing the number of seasons in PAR model (Franses and Paap (2004) and Hipel and McLeod (1994) proposed several statistical hypothesis tests). Lastly, it is known that a stationary AR process has a short memory (Brockwell and Davis (1991); Robinson (2003)). Time series which exhibit long-range dependence are characterized by autocorrelations which decay very slowly, while a stationary AR process has rapidly decaying autocorrelations. Many kinds of time series, including hydrological ones, exhibit structural changes and long-range dependence (Song and Bondon, 2013). Therefore, a long memory process with a periodic structure and changepoints could be appropriate for hydrological data, and could also be extended to other fields (e.g. internet traffic data).

Acknowledgements The authors wish to thank Francesco Battaglia for his valuable and constructive remarks, QiQi Lu for providing us with the Fortran code related to the algorithm in Lu et al (2010) and several anonymous referees for their useful comments. Part of this work has been carried out with the financial support of the CNRS and the French National Research Agency (ANR) in the framework of the Investments for the Future Program, within the Cluster of Excellence COTE (ANR-10-LABX-45).

Conflict of Interest Statement
The authors declare that they have no conflict of interest.

References

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

Bai J, Perron P (1999) Estimating and testing linear models with multiple structural changes. Econometrica 66:47–78
Baragona R, Battaglia F, Poli I (2011) Evolutionary Statistical Procedures. Oxford University Press
Battaglia F, Protopapas MK (2012a) An analysis of global warming in the alpine region based on nonlinear nonstationary time series models. Statistical Methods & Applications 21(3):315–334
Battaglia F, Protopapas MK (2012b) Multi-regime models for nonlinear nonstationary time series. Computational Statistics 27:319–341
Bentarzi M, Hallin M (1993) On the invertibility of periodic moving-average models. Journal of Time Series Analysis 15:263–268
Box GEP, Jenkins GM (1970) Time series analysis, forecasting and control. Holden-Day, San Francisco, CA
Brockwell PJ, Davis RA (1991) Time series: Theory and Methods. Springer Science & Business Media
Buishand TA (1984) Tests for detecting a shift in the mean of hydrological time series. Journal of Hydrology 75:51–69
Cobb GW (1978) The problem of the Nile: conditional solution to a changepoint problem. Biometrika 65:243–252
Davis RA, Lee TCM, Rodriguez-Yam GA (2006) Structural break estimation for nonstationary time series models. Journal of the American Statistical Association 473:223–239
Davis RA, Lee TCM, Rodriguez-Yam GA (2008) Break detection for a class of nonlinear time series models. Journal of Time Series Analysis 29:834–867
Delleur JW, Tao PC, Kavvas ML (1976) An evaluation of the practicality and complexity of some rainfall and runoff time series models. Water Resources Research 12(5):953–970
Doerr B, Fischer P, Hilbert A, Witt C (2017) Detecting structural breaks in time series via genetic algorithms. Soft Computing 21(16):4707–4720
Durdu OF (2010) Application of linear stochastic models for drought forecasting in the Buüyük Menderes river basin, western Turkey. Stochastic Environmental Research and Risk Assessment 24:1145–1162
Eiben AE, Smith JE (2003) Introduction to Evolutionary Computing. Springer
Eshete Z, Vandewiele GL (1992) Comparison of non-gaussian multicomponent and periodic autoregressive models for river flow. Stochastic Hydrology and Hydraulics 6:223–238
Fayaed SS, El-Shafie A, Jaafar O (2013) Reservoir-system simulation and optimization techniques. Stochastic Environmental Research and Risk Assessment 27:1751–1772
Franses PH, Paap R (2004) Periodic time series models. Oxford University Press
Gober P, Wheater HS (2014) Socio-hydrologu and the science-policy interface: a case study of the Saskatchewan River basin. Hydrol Earth Syst Sci 18:1413–1422
Goldberg DE (1989) Genetic Algorithms in Search, Optimization & Machine Learning. Addison-Wesley
Hansen BE (2001) The new econometrics of structural change: dating breaks in U.S. labor productivity. Journal of Economic Perspectives 15:117–128
Hipel KW, McLeod AI (1994) Time series modelling of water resources and environmental systems. Elsevier, Amsterdam
Hipel KW, McLeod AI, McBean EA (1977) Stochastic modelling of the effects of reservoir operation. Journal of Hydrology 32:97–113
Holland JH (1975) Adaptation in natural and artificial systems. University of Michigan Press
Hyndman RJ, Koehler AB (2006) Another look at measures of forecast accuracy. International Journal of Forecasting 22:679–688
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
Kawahara Y, Sugiyama M (2012) Sequential change-point detection based on direct density-ratio estimation. Statistical Analysis and Data Mining 5(2):114–127
Koutroumanidis T, Sylaios G, Zafeiroiou E, Tsihrintzis V (2009) Genetic modeling for the optimal forecasting of hydrologic time-series: Application in Nestos River. Journal of Hydrology 368:156–164

Krause P, Boyle DP, Bäse F (2005) Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences 5:89–97
Kreinovich V, Quintana C, Fuentes O (1993) Genetic algorithms: what fitness scaling is optimal? Cybernetics and Systems 24(1):9–26
Li S, Lund R (2012) Multiple changepoint detection via genetic algorithms. Journal of Climate 25:674–686
Lu Q, Lund R (2007) Simple linear regression with multiple level shifts. Canadian Journal of Statistics 37:447–458
Lu Q, Lund R, Lee TCM (2010) An MDL approach to the climate segmentation problem. The Annals of Applied Statistics 4:299–319
Lund RB, Basawa IV (1999) Modeling and inference for periodically correlated time series. In: Gosh S (ed) Asymptotics, Nonparametrics and Time Series, Marcel Dekker, New York, Statistics : textbooks and monographs, vol 158, pp 37–62
Lund RB, Basawa IV (2000) Recursive prediction and likelihood evaluation for periodic ARMA models. Journal of Time Series Analysis 21:75–93
Lund RB, Wang XL, Lu Q, Reeves J, Gallagher C, Feng Y (2007) Changepoint detection in periodic and autocorrelated time series. Journal of Climate 20:5178–5190
Maçaira PM, Oliveira FLC, Ferreira PGCF, de Almeida FVN, Souza RC (2017) Introducing a causal PAR(p) model to evaluate the influence of climate variables in reservoir inflows: a brazilian case. Pesquisa Operacional 37:107–128
Matteson DS, James NA (2014) A nonparametric approach for multiple change point analysis of multivariate data. Journal of the American Statistical Association 109(505):334–345
McCabe GJ, Wolock DM, Pederson GT, Woodhouse CA, McAfee S (2017) Evidence that recent warming is reducing Upper Colorado river flows. Earth Interaction 21:1–14
McKee TB, Doesken NJ, Kleist J, Shrier CJ, Stanton WP (2000) A history of drought in Colorado: lessons learned and what lies ahead. Colorado Water Resources Research Institute 9:1–20
McLeod AI (1993) Parsimony, model adequacy, and periodic autocorrelation in time series forecasting. International Statistical Review 61:387–393
McLeod AI (1994) Diagnostic checking periodic autoregression models with applications. Journal of Time Series Analysis 15:221–233
McLeod AI, Gweon H (2013) Optimal deseasonalization for monthly and daily geophysical time series. Journal of Environmental Statistics 4:1–11
Mishra AK, Desai VR (2005) Drought forecasting using stochastic models. Stochastic Environmental Research and Risk Assessment 19:326–339
Mondal MS, Wasimi SA (2006) Generating and forecasting monthly flows of the Ganges river with PAR model. Journal of Hydrology 323:41–56
Noakes DJ, McLeod AI, Hipel KW (1985) Forecasting monthly riverflow time series. International Journal of Forecasting 1:179–190
Novak K, Hoerling M, Rajagopalan B, Zagona E (2012) Colorado river basin hydroclimatic variability. Journal of Climate 25:4389–4403
Pereira G, Veiga Á (2018) Par (p)-vine copula based model for stochastic streamflow scenario generation. Stochastic Environmental Research and Risk Assessment 32(3):833–842
Piyoosh AK, Ghosh SK (2017) Effect of autocorrelation on temporal trends in rainfall in a valley region at the foothills of Indian Himalayas. Stochastic Environmental Research and Risk Assessment 31:2075–2096
Prairie J, Callejo R (2005) Natural flow and salt computation methods, calendar years 1971-1995. All US Government Documents (Utah Regional Depository) 35
Rao AR, Tirtotjondro W (1996) Investigation of changes in characteristics of hydrological time series by Bayesian methods. Stochastic Hydrology and Hydraulics 101:295–317
Rissanen J (1978) Modeling by shortest data description. Automatica 14:465–471
Robinson PM (2003) Time Series with Long Memory. Advanced Texts in Econometrics
Shaochuan L (2019) A bayesian multiple changepoint model for marked poisson processes with applications to deep earthquakes. Stochastic Environmental Research and Risk Assessment 33(1):59–72
Song L, Bondon P (2013) Structural changes estimation for strongly-dependent processes. Journal of Statistical Computation and Simulation 83:1783–1806

Song S, Singh VP (2010) Frequency analysis of droughts using the Plackett copula and parameter estimation by genetic algorithm. Stochastic Environmental Research and Risk Assessment 24:783–805

Srivastav R, Srinivasan K, Sudheer KP (2016) Simulation-optimization framework for multi-site multi-season hybrid stochastic streamflow modeling. Journal of hydrology 542:506–531

Tong H (1990) Non-linear time series: a dynamical system approach. Oxford University Press

Ursu E, Pereau JC (2015) Application of periodic autoregressive process to the modeling of the Garonne river flows. Stochastic Environmental Research and Risk Assessment 30(7):1785–1795

Van Steeter MM, Pitlick J (1998) Geomorphology and endangered fish habitats of the upper Colorado River. Historic changes in streamflow, sediment load, and channel morphology. Water Resources Research 34:287–302

Vecchia AV (1985a) Periodic autoregressive-moving average (PARMA) modeling with applications to water resources. Water Resources Bulletin 21:721–730

Vecchia AV (1985b) Maximum likelihood estimation for periodic autoregressive moving average models. Technometrics 27:375–384

Wang Y, Guo S, Chen H, Zhou Y (2014) Comparative study of monthly inflow prediction methods for the three gorges reservoir. Stochastic Environmental Research and Risk Assessment 28:555–570

Woodhouse CA, Pederson GT (2018) Investigating runoff efficiency in upper Colorado River streamflow over past centuries. Water Resources Research 54:286–300

Woodhouse CA, Pederson GT, Morino K, McAfee S, McCabe GJ (2016) Increasing influence of air temperature on upper Colorado River streamflow. Geophysical research letter 43:2174–2181

Yau CY, Tang CM, Lee TCM (2015) Estimation of multiple-regime threshold autoregressive models with structural breaks. Journal of the American Statistical Association 511:1175–1186