Identification of periodic autoregressive models using genetic algorithms

Eugen Ursu

Outline

Periodic Series

A periodic series is a random sequence in which the first two moments are periodic with period s. Formally, a time series $\{X_t\}$ with bounded second moments satisfying:

$$\mathbb{E}(X_{n+s}) = \mathbb{E}(X_n)$$

$$Cov(X_{n+s}, X_{m+s}) = Cov(X_n, X_m)$$

for all m and n, is called periodic (or periodically stationary) with period s.

Notation

- $N =$ number of years
- $N \times s =$ the number of observations
- The periodic notation $\{X_{ns+k}\}$ denotes the series during the $k-$th season of the $n-$th cycle, where $k = 1, 2, \ldots, s$ and $n = 0, 1, 2, \ldots, N-1$

Notation

We define

$$\mu_k = \mathbb{E}[X_{ns+k}]$$

and the autocovariance function (ACVF) of $\{X_t\}$ at season k and lag $h \geq 0$, as

$$\gamma_h(k) = Cov(X_{ns+k}, X_{ns+k-h}),$$

and the corresponding autocorrelation function (ACF) as

$$\rho_h(k) = \frac{\gamma_h(k)}{\sqrt{\gamma_0(k)\gamma_0(k-h)}}$$

The autocovariance $Cov(X_{ns+k}, X_{ns+k-h})$ of a periodic series does not depend on the cycle n, but does depend on both lag h and season k.

ACVF

- The autocovariance function (ACVF) $\gamma_h(k)$ is not symmetric in h;
- Generally, periodicity in the ACVF of a series cannot always be visually detected from the sample ACVF.
- There are periodicity tests in both time and frequency domains (see Vecchia and Ballerini (1991), Anderson and Vecchia (1993))

Periodically stationary

- Periodically stationary time series are not stationary unless s $= 1$.
- Many periodic time series cannot be transformed to stationary series; therefore, some well established statistical inference techniques for stationary time series are not appropriate in the periodic setting.
- Before doing any statistical analysis on a given series, it may be necessary to confirm that the series is periodic in its first two moments.
- If periodicity is identified in the second moment of a time series, the next step is to find an adequate statistical model that incorporates all relevant information in the observations. Periodic Autoregressive (PAR) models can be used to model a large class of periodic series.

- PAR series should not be confused with seasonal ARMA (SARMA) series.
- SARMA series are stationary (not periodic) with strong correlations(large in absolute value) at lags which are multiples of the period s.
- Specifically, a SARMA series is represented by:

$$\Phi_p(B)\Upsilon_P(B^s)X_t = \Theta_q(B)\Psi_Q(B^s)\epsilon_t, \quad \forall t$$

  where ϵ_t is white noise (stationary).
- McLeod (1993) illustrates the inadequacies of SARMA models when forecasting periodic streamflow series.

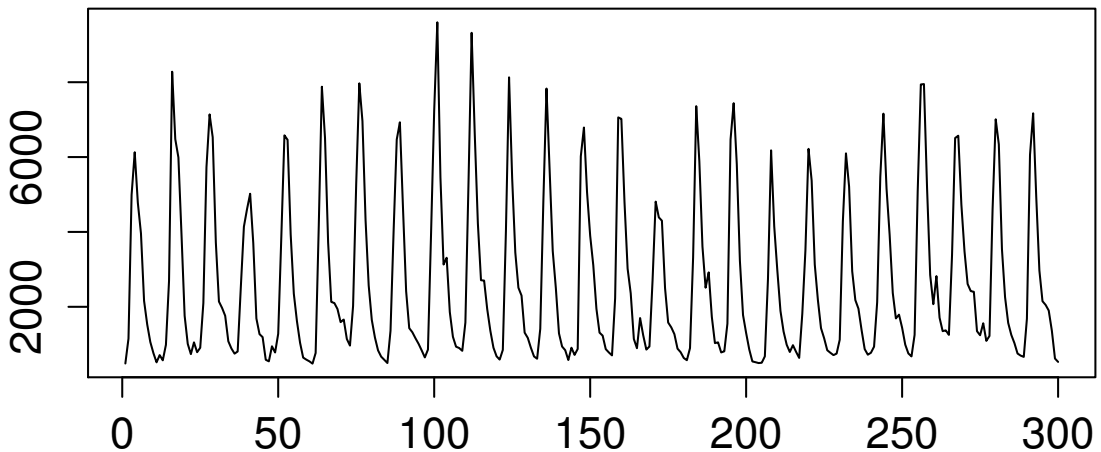
PAR

- Many time series in climatology, hydrology, and economics exhibit periodicity in their autocovariance structure.
- Such series include quarterly river flows, daily temperatures, quarterly profit margins, monthly airline ticket sales etc.
- For example, Lund et al. (1995) have studied monthly ozone concentrations from Arosa, Switzerland, and conclude that the data are well modeled by a time series whose autocovariances vary from month to month within a year, but repeat at the same months in different years.

Seasonal time series

Time series which regularly displays same variations every s time instants

Real examples: hydrology time series



Models:

- *seasonal ARIMA*
- *PAR*

PAR (Franses & Paap, 2004; Hipel & McLeod, 1994)

- Let Y_{ns+k} be a PAR stochastic process given by:

$$Y_{ns+k} = \mu_k + \sum_{i=1}^{p(k)} \phi_i(k) Y_{ns+k-i} + \epsilon_{ns+k}, \tag{1}$$

$$n = 0, ..., N-1, \ k = 1, ..., s,$$

$$E(\epsilon_{ns+k}) = 0, Var(\epsilon_{ns+k}) = \sigma^2(k)$$

- Seasonality affects mean, autocorrelation and residual variance
- A separate model for each season $k = 1, ..., s$ is specified
- Many parameters $\rightarrow$ *subset model*

Alternative Periodic modeling

- Periodic Moving Average (PMA; Cipra, 1985)
- Periodic AutoRegressive Moving Average (PARMA; Vecchia, 1985)

PAR process written as a VAR process

First, we note that equation (1) offers a vector autoregressive (VAR) representation:

$$\mathbf{\Phi}_0^* \boldsymbol{Y}_n^* = \sum_{k=1}^{p^*} \mathbf{\Phi}_k^* \boldsymbol{Y}_{n-k}^* + \boldsymbol{\epsilon}_n^*, \tag{2}$$

where $\boldsymbol{Y}_n^* = (Y_{ns+s}, Y_{ns+s-1}, \ldots, Y_{ns+1})^\top$ and $\epsilon_n^* = (\epsilon_{ns+s}, \epsilon_{ns+s-1}, \ldots, \epsilon_{ns+1})^\top$ are $s \times 1$ random vectors.

Example

Consider the PAR$(2, 1, 3, 1)$ process with $s = 4$ periods:

$$\begin{aligned}
Y_{4n+1} &= \phi_1(1) Y_{4n} + \phi_2(1) Y_{4n-1} + \epsilon_{4n+1}, \\
Y_{4n+2} &= \phi_1(2) Y_{4n+1} + \epsilon_{4n+2} \\
Y_{4n+3} &= \phi_1(3) Y_{4n+2} + \phi_2(3) Y_{4n+1} + \phi_3(3) Y_{4n} + \epsilon_{4n+3} \\
Y_{4n+4} &= \phi_1(4) Y_{4n+3} + \epsilon_{4n+4}
\end{aligned}$$

Example

The VAR representation is:

$$\boldsymbol{\Phi}_0^* \boldsymbol{Y}_n^* = \boldsymbol{\Phi}_1^* \boldsymbol{Y}_{n-1}^* + \boldsymbol{\epsilon}_n^*,$$

with

$$\boldsymbol{\Phi}_0^* = \begin{bmatrix} 1 & -\phi_1(4) & 0 & 0 \\ 0 & 1 & -\phi_1(3) & -\phi_2(3) \\ 0 & 0 & 1 & -\phi_1(2) \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\boldsymbol{\Phi}_1^* = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \phi_3(3) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \phi_1(1) & \phi_2(1) & 0 & 0 \end{bmatrix},$$

$$\begin{aligned} \mathbf{Y}_n^* &= (Y_{4n+4}, Y_{4n+3}, Y_{4n+2}, Y_{4n+1})^\top \quad \text{and} \\ \epsilon_n^* &= (\epsilon_{4n+4}, \epsilon_{4ns+3}, \epsilon_{4ns+2}, \epsilon_{4n+1})^\top. \end{aligned}$$

PAR process written as a VAR process

- In principle, the asymptotic properties of the PAR parameter estimators could be deduced from results for multivariate time series, using the general multivariate representation (2).
- Furthermore, from (2), the multivariate VAR process is not written in a standard VAR form; the process must be pre-multiplied on each side by the matrix $\mathbf{\Phi}_0^{*-1}$.
- This rescaling operation complicates the interpretation of the estimated parameters and the derivation of their statistical properties in the original scale, since the covariance matrix of the error term of the standard VAR model now depends on the autoregressive parameters.
- These considerations also occur in Basawa and Lund (2001). Therefore, it is more informative to work directly with the individual PAR components.

PAR estimation

Consider the time series data $Y_{ns+\nu}$, $n = 0, 1, \ldots, N-1$, $\nu = 1, \ldots, s$. Let

$$\begin{aligned}
\mathbf{z}(\nu) &= \left(Y_\nu, Y_{s+\nu}, \ldots, Y_{(N-1)s+\nu}\right)^\top, \\
\mathbf{e}(\nu) &= \left(\epsilon_\nu, \epsilon_{s+\nu}, \ldots, \epsilon_{(N-1)s+\nu}\right)^\top, \\
\mathbf{X}(\nu) &= \begin{bmatrix}
Y_{\nu-1} & Y_{\nu-2} & \cdots & Y_{\nu-p(\nu)} \\
Y_{s+\nu-1} & Y_{s+\nu-2} & \cdots & Y_{s+\nu-p(\nu)} \\
\vdots & & \ddots & \vdots \\
Y_{(N-1)s+\nu-1} & Y_{(N-1)s+\nu-2} & \cdots & Y_{(N-1)s+\nu-p(\nu)}
\end{bmatrix},
\end{aligned}$$

be $N \times 1$, $N \times 1$ and $N \times \{p(\nu)\}$ random matrices.
The PAR model can be reformulated as:

$$\mathbf{z}(\nu) = \mathbf{X}(\nu)\boldsymbol{\beta}(\nu) + \mathbf{e}(\nu), \ \nu = 1, \ldots, s,$$

where the model parameters are collected in the $p(\nu) \times 1$ vector $\boldsymbol{\beta}(\nu)$ which is defined as:

$$\boldsymbol{\beta}(\nu) = \left(\phi_1(\nu), \ldots, \phi_{p(\nu)}(\nu)\right)^\top.$$

Least squares estimation with linear constraints on the parameters.

We assume that, for a known matrix $\boldsymbol{R}(\nu)$ of rank $K(\nu)$, and a known $p(\nu) \times 1$ vector $\boldsymbol{b}(\nu)$, the following relation is satisfied:

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

where $\boldsymbol{\xi}(\nu)$ represents a $K(\nu) \times 1$ vector of unknown parameters. In general, the matrices $\boldsymbol{R}(\nu)$ and the vectors $\boldsymbol{b}(\nu)$ allow for linear constraints on the parameters of the same season ν, $\nu = 1, \ldots, s$. Proceeding as in the previous section, it is possible to find the least squares estimator $\hat{\boldsymbol{\xi}}(\nu)$ of $\boldsymbol{\xi}(\nu)$ and to determine its asymptotic distribution.

Selection criteria for PAR models

The parameters for one unspecified season can be estimated entirely independently of the parameters of any other season.

1. The BIC criterion was factored to obtained a separate criterion for each period.

$$BIC = \sum_{\nu=1}^{s} BIC(\nu),$$

with

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

where $\hat{\epsilon}_{ns+\nu}$, $n = 0, \ldots, N-1$ denote the residuals of the adjustment, $\hat{\sigma}(\nu)$ corresponds to the least squares estimators of $\sigma(\nu)$, and $p(\nu)$ represents the number of autoregressive parameters in the season ν.

2. the sample periodic partial autocorrelation;

Example

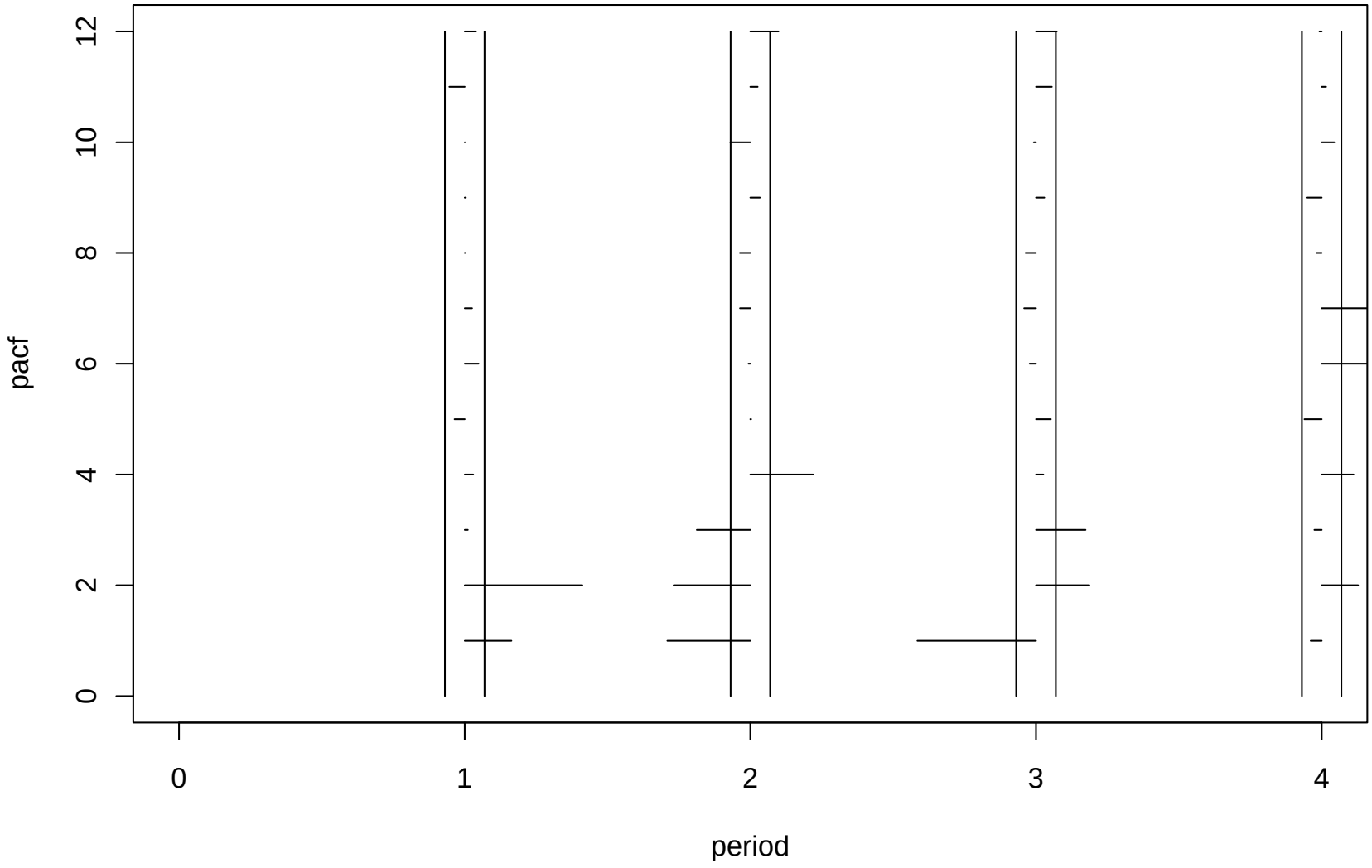
We have simulated the following model:

$$
\begin{aligned}
Y_{4n+1} &= \phi_1(1)Y_{4n} + \phi_2(1)Y_{4n-1} + \epsilon_{4n+1} \\
Y_{4n+2} &= \phi_2(2)Y_{4n} + \phi_4(2)Y_{4n-2} + \epsilon_{4n+2} \\
Y_{4n+3} &= \phi_1(3)Y_{4n+2} + \phi_2(3)Y_{4n+1} + \phi_3(3)Y_{4n} + \epsilon_{4n+3} \\
Y_{4n+4} &= \phi_6(4)Y_{4n-2} + \phi_7(4)Y_{4n-3} + \epsilon_{4n+4},
\end{aligned}
$$

This model represent a PAR$\Big((1,2),(2,4),(1,2,3),(6,7)\Big)$ with the parameters given in:

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7
$\nu = 1$	0.30	0.50	0	0	0	0	0
$\nu = 2$	0	-0.65	0	0.50	0	0	0
$\nu = 3$	-0.80	0.30	0.35	0	0	0	0
$\nu = 4$	0	0	0	0	0	0.81	0.70

Example



Example

The periodic partial autocorrelations suggest a PAR$(2, 4, 3, 7)$.

	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$\hat{\phi}_5$	$\hat{\phi}_6$	$\hat{\phi}_7$
$\nu = 1$	0.314	0.478	0.000	0.000	0.000	0.000	0.000
	(0.031)	(0.023)					
$\nu = 2$	-0.054	-0.656	-0.014	0.521	0.000	0.000	0.000
	(0.087)	(0.048)	(0.068)	(0.075)			
$\nu = 3$	-0.762	0.431	0.297	0.000	0.000	0.000	0.000
	(0.070)	(0.069)	(0.057)				
$\nu = 4$	0.121	0.159	0.085	-0.059	-0.006	0.906	0.786
	(0.078)	(0.112)	(0.118)	(0.090)	(0.093)	(0.111)	(0.082)

In order to propose a more parsimonious model, we could set to zero each autoregressive parameter whose absolute value of the t-statistics (calculated as the value of the estimator divided by its standard error) is smaller than one.

When are Evolutionary Algorithms Useful

1. Intractable objective function;
2. It may be too computationally-intensive to find an exact solution but sometimes a near-optimal solution is sufficient.
3. combinatorial problems:

$$\max_{x} f(\mathbf{x}) : S \rightarrow \mathbb{R} \tag{3}$$

where S=solution space and f=objective function;

When are Evolutionary Algorithms Useful

1. Intractable objective function;
2. It may be too computationally-intensive to find an exact solution but sometimes a near-optimal solution is sufficient.
3. combinatorial problems:

$$\max_{x} f(\mathbf{x}) : S \rightarrow \mathbb{R} \tag{3}$$

where S=solution space and f=objective function;

Generalities

Genetic algorithms (GA) are stochastic optimization algorithms based on mechanisms of natural selection and genetics. The genetic algorithm is very simple, yet it performs well on many types of problems.

- We start with a population of potential solutions (chromosomes) initial arbitrarily chosen.
- Their relative performance (fitness) is evaluated.
- Solutions from one population are taken and used to form a new population. This is motivated by the hope that the new population will be better than the old one. Based on their performance (fitness) we create a new population of potential solutions using simple evolutionary operators: selection, crossover and mutation.
- This cycle continue until we find a satisfactory solution.

Advantages of GA

Fundamental differences between the GA and other optimization methods:

1. GA seek a solution from a population of points and not from a single point;
2. GA do not impose any regularity on the function studied (continuity, differentiability, etc.). GA uses fitness function for evaluation rather than derivatives. This is one of big advantages of genetic algorithms.
3. GA use probabilistic transition rules while conventional methods apply deterministic rules.

Remark

Since the GA is stochastic, it is important to run it multiple times before accepting a solution.

String representation

As the identification will be made period by period, for a given period, we reserve one gene to each possible lag, filling it with 1 if the parameters is free, and with 0 if the parameter is constrained to zero. The obtained string represents the chromosome. For the previous model, for the first period with the equation:

$$Y_{4n+1} = \phi_1(1) Y_{4n} + \phi_2(1) Y_{4n-1} + \epsilon_{4n+1}$$

the chromosome is $chrs = (1100000)$, for the second period with the equation:

$$Y_{4n+2} = \phi_2(2) Y_{4n} + \phi_4(2) Y_{4n-2} + \epsilon_{4n+2}$$

the chromosome is $chrs = (0101000)$ and so on (we consider the maximum order equal to 7 for each period). The problem is reduced to the use of an independent GA for each period. The structure of the genetic algorithms depends on the choice of a maximum possible order for each period.

Genetic algorithms step by step

String No.	Initial population	BIC	fitness value	$prob_i$	Expected count	Actual count
1	110001000101101	286.054	38.182	0.236	0.945	1
2	010110000001010	312.307	11.928	0.074	0.296	0
3	010011110101100	213.715	110.521	0.684	2.735	3
4	101000001001000	323.236	1.000	0.006	0.024	0
Sum		1135.312	161.631	100%	4	4

Table: Selection

Obtain the decoded values for the initial population generated. For string 1, the following parameters are different from zero:

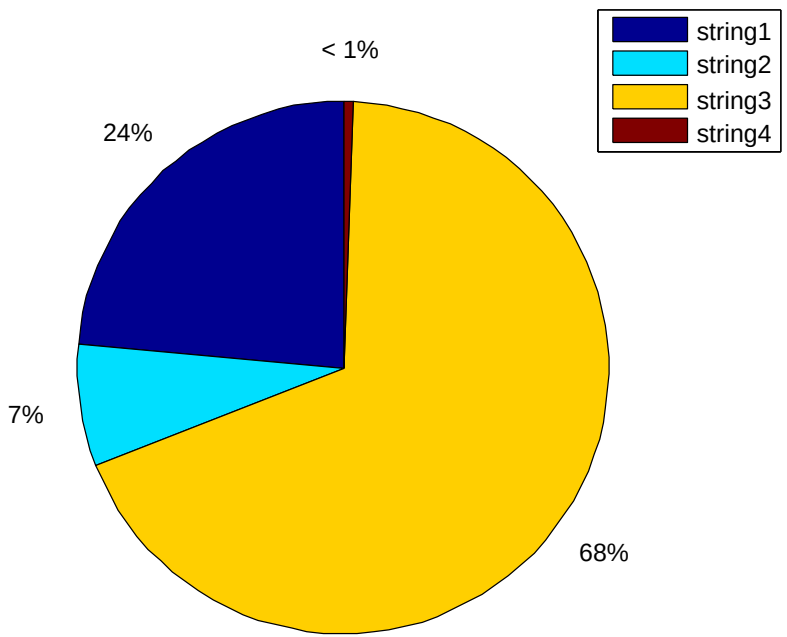
$$\phi_1(1) \quad \phi_2(1) \quad \phi_6(1) \quad \phi_{10}(1) \quad \phi_{12}(1) \quad \phi_{13}(1) \quad \phi_{15}(1)$$

Genetic algorithms step by step

Compute the probability of selection,

$$prob_i = \frac{f(x_i)}{\sum_{i=1}^{N_p} f(x_i)},$$

where N_p represents the size of the population, $f(x_i)$ is the fitness value corresponding to a particular individual in the population. We will select the individual which would participate in the crossover cycle using Roulette Wheel selection.



Genetic algorithms step by step

The crossover point is specified and based on this crossover point, a single point crossover is performed and new offspring is produced

String No.	Mating pool	Cross point	Offspring after crossover	BIC	Fitness value
1	110001000101101	8	110001000101100	281.174	1.000
3	010011110101100	8	010011110101101	218.528	63.646
3	010011110101100	11	010011110101100	213.715	68.459
3	010011110101100	11	010011110101100	213.715	68.459
Sum				927.132	201.565

Table: Crossover

Genetic algorithms step by step

Mutation operation is performed to produce new offsprings after crossover.

String No.	Offspring after crossover	Mutation chromosome for flipping	Offspring after mutation	BIC	Fitness value
1	110001000101100	000000000001000	110001000100100	277.261	1.000
2	010011110101101	000000000000000	010011110101101	218.528	59.733
3	010011110101100	000000000000000	010011110101100	213.715	64.546
4	010011110101100	100000000000000	110011110101100	216.451	61.810
Sum				925.954	187.088

Table: Mutation

Remark

The "elitist" strategy applies: The best chromosome that may be found in each iteration is always maintained in the current population unless an even better string appears due to the genetic operators. If is not the case, then the best chromosome replaces the chromosome with the worst fitness value. Such individuals can be lost if they are not selected to reproduce or if crossover or mutation destroys them.

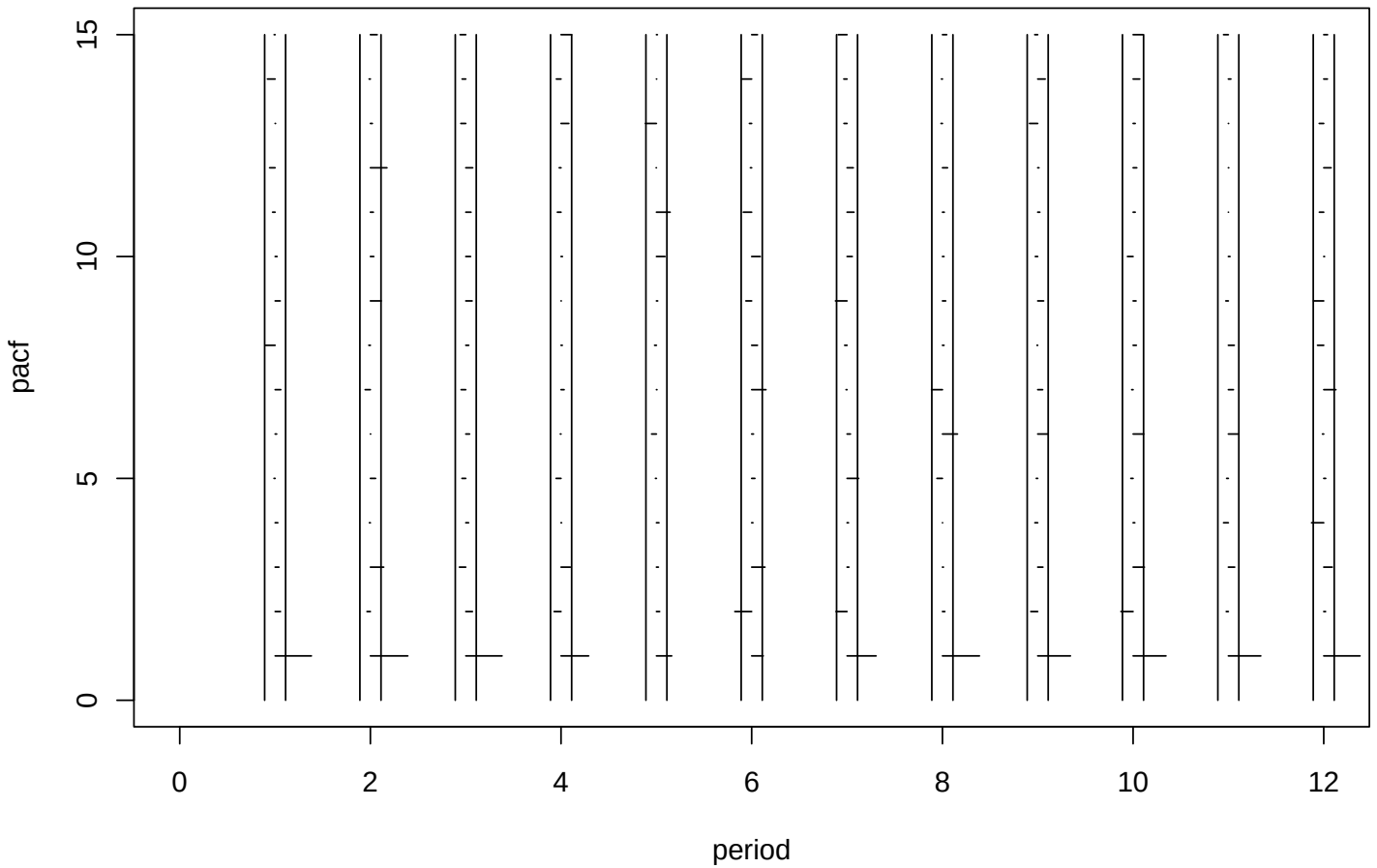
Simulations experiments

The selection algorithm was applied to 100 independent simulations with $N = 200$ observations by period of the model given in previous example. Jump to model
We let the seasonal orders to vary from 0 to 15, leading to a chromosome of length $L = 15$. The size of the GA population is chosen in the set $\{20, 40\}$, the crossover probability $P_c = 0.8$ and the mutation probability $P_m = 0.01$, the maximum allowed number of generations is equal to 25, 50 or 100 and the number of elite individuals is 1.

			q_1	q_2	q_3
$\mathcal{SIMULATED}$ $\mathcal{MODEL}$	$N_p = 20$	$N_g = 25$	74	23	3
		$N_g = 50$	98	2	0
		$N_g = 100$	100	0	0
	$N_p = 40$	$N_g = 25$	99	1	0
		$N_g = 50$	100	0	0
		$N_g = 100$	100	0	0

Fraser river time series

The time series of mean monthly flows of the Fraser river at Hope, British Columbia from March 1912 to December 1990 include 946 observation. Using the plot of the sample periodic partial autocorrelation a PAR$(1, 1, 1, 1, 1, 3, 2, 1, 1, 3, 1, 1)$ was selected.



Fraser river time series using GA

We ran the GA algorithm for 50 iterations with parameters $L = 15$, $N_p = 20$, $P_c = 0.8$, $P_m = 0.01$ and with one elite individual. Note that in this case the number of possible models is of order $4 * 10^5$. The model found is a PAR $\Big((1), (1, 12), (1), (1), (1, 11), (1, 2, 7), (1), (1), (1), (1, 2, 3), (1), (1)\Big)$. The numbers which appear in parentheses correspond to the unconstrained autoregressive parameters in corresponding period. For one additional parameter, the BIC is decreased from -3149.37 to -3175.53.

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}	ϕ_{11}	ϕ_{12}
$\nu = 1$	0.664	0	0	0	0	0	0	0	0	0	0	0
	(0.063)											
$\nu = 2$	0.733	0	0	0	0	0	0	0	0	0	0	0.229
	(0.063)											(0.065)
$\nu = 3$	0.813	0	0	0	0	0	0	0	0	0	0	0
	(0.074)											
$\nu = 4$	0.765	0	0	0	0	0	0	0	0	0	0	0
	(0.123)											
$\nu = 5$	0.177	0	0	0	0	0	0	0	0	0	0.360	0
	(0.065)										(0.137)	
$\nu = 6$	0.278	-0.211	0	0	0	0	0.181	0	0	0	0	0
	(0.077)	(0.049)					(0.056)					
$\nu = 7$	0.715	0	0	0	0	0	0	0	0	0	0	0
	(0.105)											
$\nu = 8$	0.751	0	0	0	0	0	0	0	0	0	0	0
	(0.067)											
$\nu = 9$	0.751	0	0	0	0	0	0	0	0	0	0	0
	(0.089)											
$\nu = 10$	1.159	-0.682	0.364	0	0	0	0	0	0	0	0	0
	(0.137)	(0.210)	(0.167)									
$\nu = 11$	0.754	0	0	0	0	0	0	0	0	0	0	0
	(0.091)											
$\nu = 12$	0.745	0	0	0	0	0	0	0	0	0	0	0
	(0.072)											

Multiregime PAR model with linear trend

$$X_{ns+k} = a^j + b^j(ns+k) + Y_{ns+k},\ j = 1, 2, \ldots, m+1,$$

$$Y_{ns+k} = \mu_k^j + \sum_{i=1}^{p^j(k)} \phi_i^j(k) Y_{ns+k-i} + \epsilon_{ns+k},\ Var(\epsilon_{ns+k}) = \sigma_j^2(k)$$

- Structural change can affect mean, autocorrelation and residual variance
- m structural changes, $m+1$ regimes
- $\tau_1, ..., \tau_m$ changepoint times (at the end of the year)
- $\tau_j \geq \tau_{j-1} + \omega,\ \forall j$
- Linear trend in each regime

Model parameters identification and estimation

Fixing a maximum number of regimes M $(= 8)$, maximum autoregressive order p $(= 3)$ and a minimum regime length ω:

Model parameters identification and estimation

Fixing a maximum number of regimes $M\ (=8)$, maximum autoregressive order $p\ (=3)$ and a minimum regime length ω:

a) *Regression parameters:*

a^j, b^j	trend parameters; regime j
μ_k^j	seasonal means; regime j, season k
$\phi_i^j(k)$	AR parameters; regime j, season k, lag i
$\sigma_j^2(k)$	residual variance; regime j, season k

- Analytically estimated

Model parameters identification and estimation

Fixing a maximum number of regimes $M\ (=8)$, maximum autoregressive order $p\ (=3)$ and a minimum regime length ω:

a) *Regression parameters:*

a^j, b^j	trend parameters; regime j
μ_k^j	seasonal means; regime j, season k
$\phi_i^j(k)$	AR parameters; regime j, season k, lag i
$\sigma_j^2(k)$	residual variance; regime j, season k

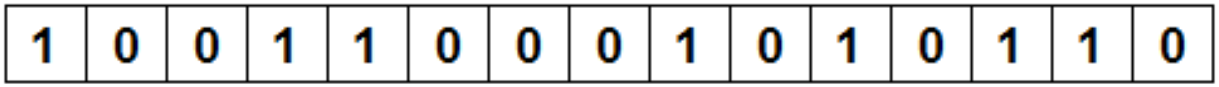
- Analytically estimated

b) *Structural parameters:*

m	number of changepoints
$\tau_1, \tau_2, \ldots, \tau_m$	changepoint times
$\underline{\delta}^1, \ldots, \underline{\delta}^{m+1}$	Subset PAR indicators

- Identification is a complex combinatorial optimization problem $\Rightarrow$ **GAs**

GA solutions encoding (binary)



$$m,\ \tau_1, ..., \tau_m,\ \underline{\delta}^1, ..., \underline{\delta}^{m+1}$$

- m: 3 bits for encoding up to 7 changepoints
- $\underline{\delta}^1, ..., \underline{\delta}^{m+1}$ directly denotes presence (1) or absence (0) of $\phi_i^j(k)$
- $\tau_1, ..., \tau_m$ must ensure:

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

GA solutions encoding

For obtaining $\tau_1, ..., \tau_m$ the chromosome encode m numbers $th_i \in (0, 1)$ which represent "the percentage of remaining values to be attributed to the corresponding i-th segment."

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

Fitness

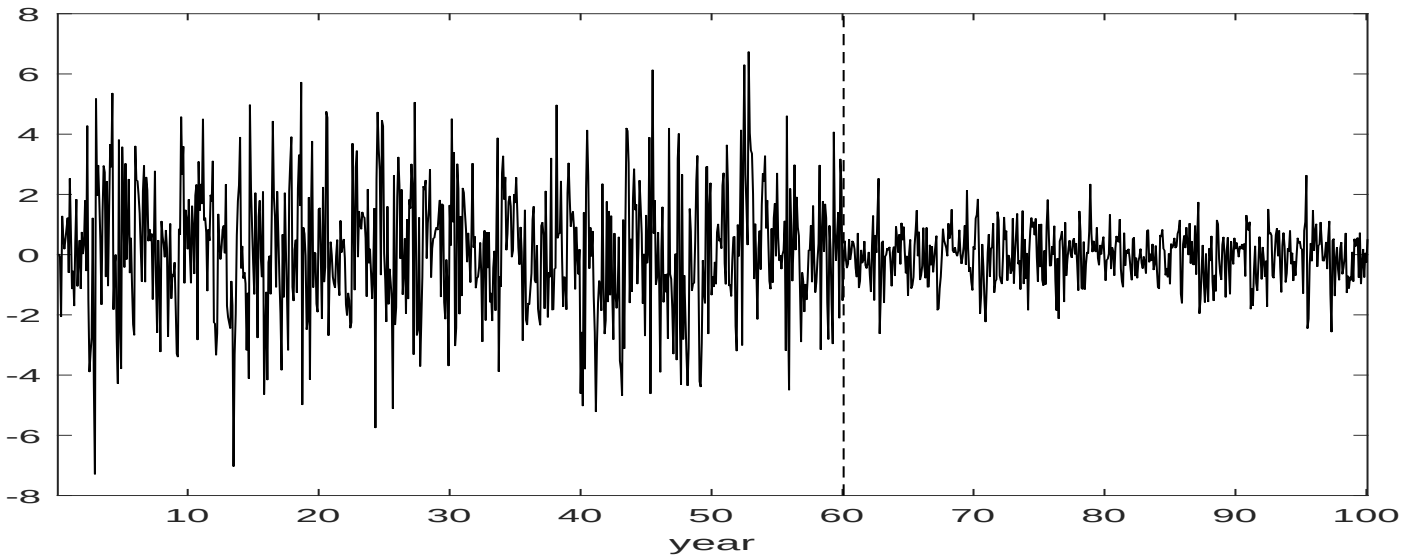
- Analytic estimation of regression parameters:
  - Trend parameters $\underline{\hat{a}}$, $\underline{\hat{b}}$ (LS, separately for each regime)
  - Seasonal means $\underline{\hat{\mu}}$ (separately for each regime and season)
  - Autoregressive parameters $\underline{\hat{\phi}}$ (constrained LS)
  - Residual variance $\underline{\hat{\sigma}}^2$
- Goodness of fit evaluation, inspired by NAIC criterion (Tong, 1991):

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

- Final fitness function:

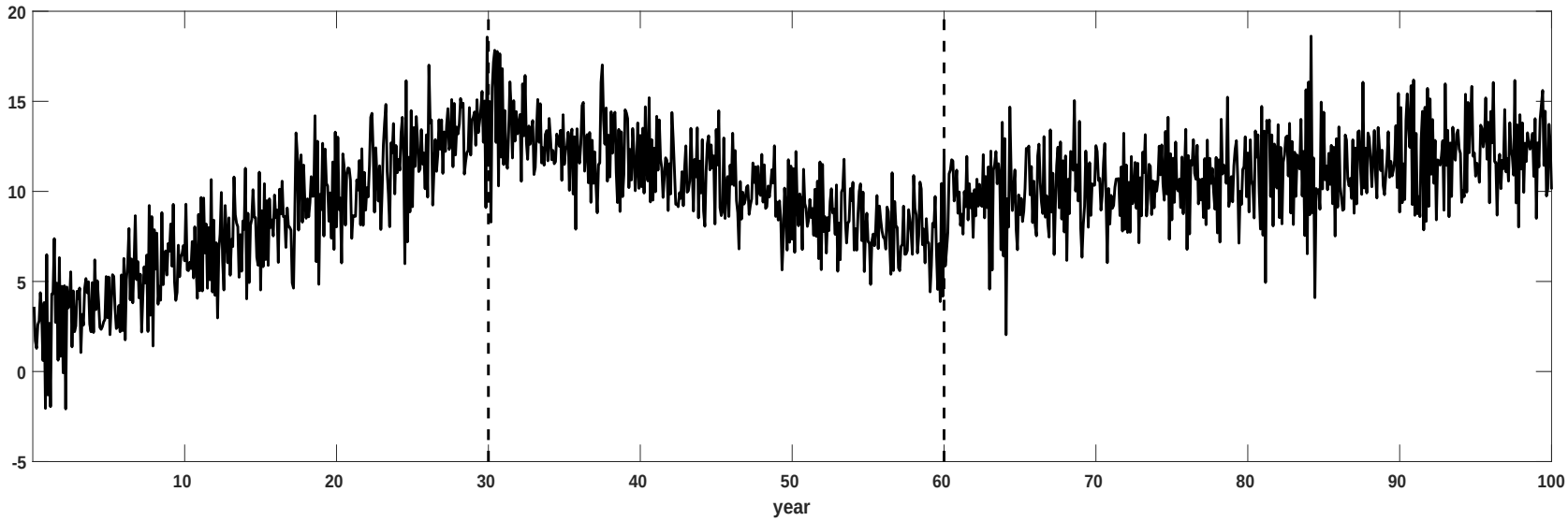
$$f = \exp\{-g/\beta\}$$

Simulations: 1 changepoint with no trend



Number of changepoints	IC=ln N			IC=3			IC=2		
	%	mean	se	%	mean	se	%	mean	se
0									
1	100	60.97	0.68	73.0	60.96	0.59	0.1	24	0
2				19.5	50.40 70.34	15.29 9.05	4.3	48.76 69.86	16.43 9.62
3				6.2	50.90 67.11 81.73	14.50 9.88 8.18	11.3	48.27 65.18 79.92	17.45 12.12 9.73
≥ 4				1.3			84.4		

Simulations: 2 changepoints with different trends



Number of changepoints	IC=ln N			IC=3			IC=2		
	%	mean	se	%	mean	se	%	mean	se
0									
1	0.01	32.00	0.00						
2	90.4	30.94 60.97	1.48 0.45	38.1	30.86 60.99	1.43 0.14	1.4	31.00 61.00	1.17 0
3	8.0	27.51 51.05 71.40	6.33 12.58 10.42	22.7	28.07 50.19 70.01	6.23 11.95 9.96	4.1	25.58 44.51 65.68	7.64 12.07 8.45
≥ 4	1.5			39.2			94.5		

Case study: monthly flows for Colorado river

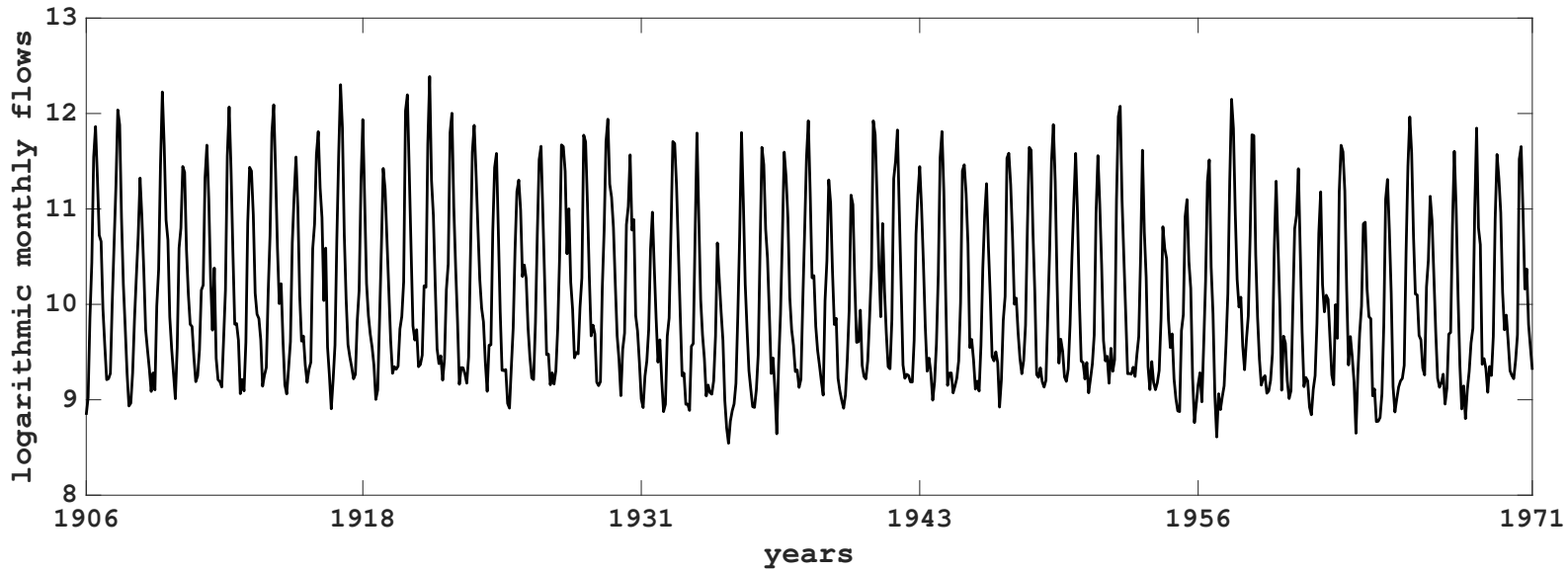


Figure: Logarithmic monthly flows for the Colorado river.

- Monthly flows from January 1906 to December 1970
- $N = 65$ years, $s = 12$, $\omega = 7$

Case study: monthly flows for Colorado river

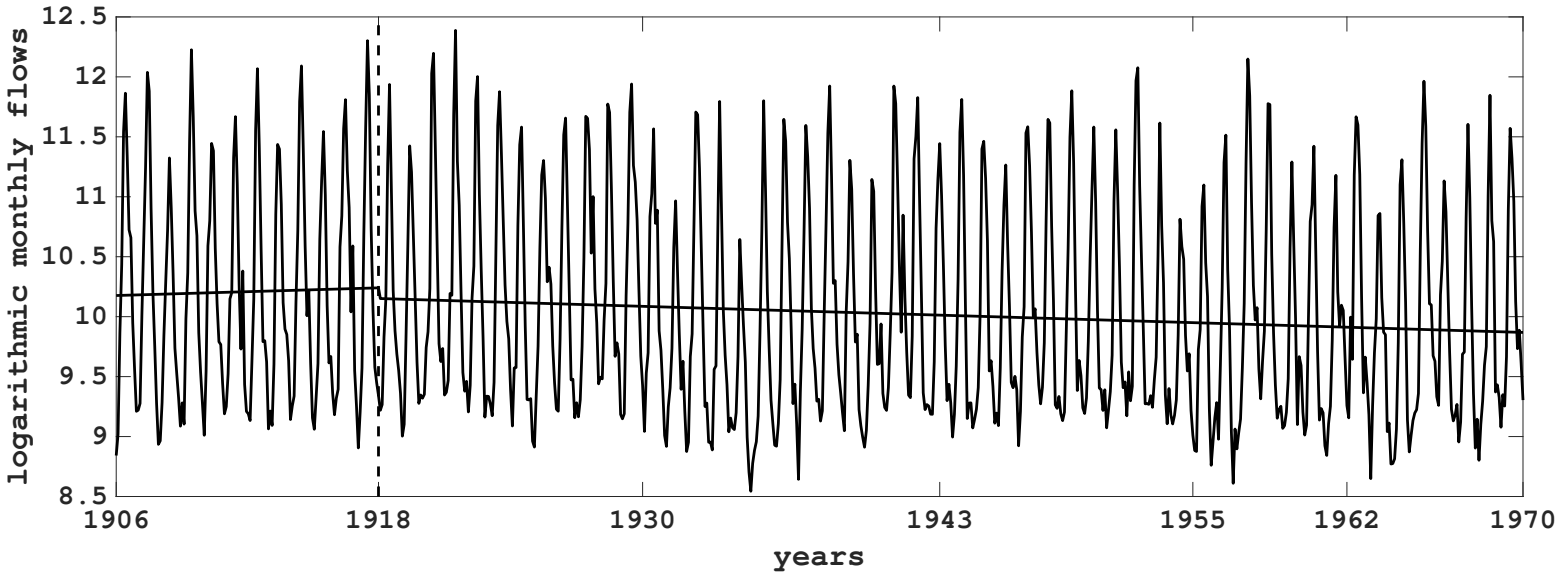


Figure: Changepoint detected on year 1918 for Colorado river

- The first model considered in our analysis is a PAR without changepoints (denoted by *Model 1*)
- 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, whereas with $p = 3$ (*Model 3*) we detect year 1931.

Case study: monthly flows for Colorado river

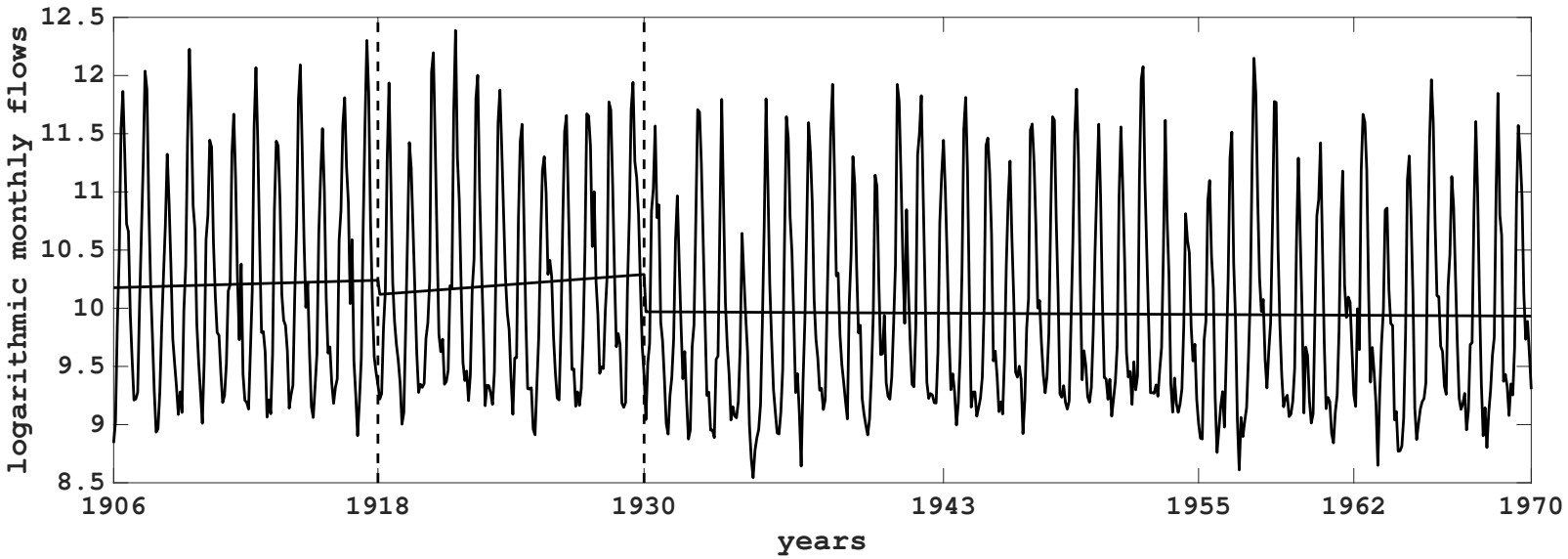


Figure: Changepoints detected on years 1918 and 1930 for Colorado river

- Our method combined with $IC = 3$ (*Model 4*) and $p = 1$ detects 1918 and 1930 as change times.

Case study: monthly flows for Colorado river

In order to evaluate the accuracy of one step-ahead logarithmic forecasts, the RMSE, MAE and MAPE are presented in Table below. *Model 2* performs better in terms of forecasting in comparison with other models.

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

Concluding remarks ad future work

- Allowing changepoints in the middle of the year
- Improving model parsimony by grouping seasons
- Long range dependence and structural changes in PAR modelling for hydrological data? (Song & Bondon, 2013)