Groupe de Recherche en
Économie Théorique et Appliquée

Application of periodic autoregressive process to the modeling of the Garonne river flows

Eugen URSU

GREThA, CNRS, UMR 5113
Université de Bordeaux
eugen.ursu@u-bordeaux.fr

&

Jean-Christophe PEREAU

GREThA, CNRS, UMR 5113
Université de Bordeaux
jean-christophe.pereau@u-bordeaux.fr

Cahiers du GREThA
n° 2015-14
April

GRETHA UMR CNRS 5113
Université de Bordeaux
Avenue Léon Duguit - 33608 PESSAC - FRANCE
Tel : +33 (0)5.56.84.25.75 - Fax : +33 (0)5.56.84.86.47 - www.gretha.fr

Application of periodic autoregressive process to the modeling of the Garonne river flows

Abstract

Accurate forecasting of river flows is one of the most important applications in hydrology, especially for the management of reservoir systems. To capture the seasonal variations in river flow statistics, this paper develops a robust modeling approach to identify and estimate periodic autoregressive (PAR) model in the presence of additive outliers. Since the least squares estimators are not robust in the presence of outliers, we suggest a robust estimation based on residual autocovariances. A genetic algorithm with Bayes information criterion is used to identify the optimal PAR model. The method is applied to average monthly and quarter-monthly flow data (1959-2010) for the Garonne river in the southwest of France. Results show that forecasts are better off in the robust model rather than the unrobust model. The accuracy of the forecasts is also improved when the model is specified in quarter-monthly flows, especially for the dry seasons.

Keywords: River flows analysis, periodic time series, robust estimation, genetic algorithms, Garonne river

Application des processus périodiques auto-regressifs à la modélisation des débits de la Garonne

Résumé

La prévision des débits d'eau des fleuves est l'une des principales applications en hydrologie et en particulier pour le management des barrages lors des périodes d'étiage. Afin de repérer les variations saisonnières dans les données de débits d'eau, cet article développe une modélisation robuste pour identifier et estimer des modèles périodiques autorégressifs (PAR) en présence de données aberrantes. Dans la mesure où les estimateurs obtenus par la méthode des moindres carrés ne sont pas robustes à la présence de ces données aberrantes, une estimation robuste fondée sur les auto-covariances résiduelles est réalisée. Un algorithme génétique avec le critère d'information de Bayes est utilisé pour identifier le modèle PAR optimal. La méthode est appliquée aux débits moyens par semaine et par mois de la Garonne dans le sud-ouest la France entre 1959 et 2010. Les résultats montrent que les prévisions sont meilleures dans le cas robuste (avec prise en compte des données aberrantes) que dans le cas non robustes. La qualité des prévisions est améliorée quand le modèle est spécifié en semaine par rapport au modèle mensuel, en particulier pour les saisons sèches.

Mots-clés : Analyse de débits d'eau, séries temporelles périodiques, estimation robuste, algorithmes génétiques, Fleuve Garonne

JEL: C22, C53, Q25

Reference to this paper: URSU Eugen, PEREAU Jean-Christophe (2015) Application des processus périodiques auto-regressifs à la modélisation des débits de la Garonne , *Cahiers du GREThA*, n°2015-14.

http://ideas.repec.org/p/grt/wpegrt/2015-14.html.

GRETHA UMR CNRS 5113
Université de Bordeaux
Avenue Léon Duguit - 33608 PESSAC - FRANCE
Tel : +33 (0)5.56.84.25.75 - Fax : +33 (0)5.56.84.86.47 - www.gretha.fr

1. Introduction

In recent years there has been considerable research in the development of time series models with seasonal or periodic properties in hydrology and water resources (Vecchia, 1985a). The main objectives are to detect trends and shifts in river flow records for a better planning and design of water management policies. Issues at stake are related to urban water supply, hydropower, irrigation management, flood and drought control, pollution, protection of endangered fishes or migrating fishes, wetland and habitats conservation (Baker and Vervier, 2004; Caballero et al., 2007; Oeurng et al., 2011; Maire et al., 2013).

In practice and in the case-study of the paper, the Garonne River (located in the southwest of France), river authorities aim at controlling the quantitative management of water to achieve a good water status. The Garonne Water Agency sets at different management points of the river some threshold values for the river-flows to be reached every year during low water periods between the 1st of July to October 31st. To ensure a good functioning of the economic and ecological system downstream the management points, river-flows have to remain above this minimal value. Below this value, water saving measures and water discharges from dams and reservoir systems have to be decided upstream the control points. The desired value of this threshold results from a informal negotiation between conflicting stakeholder economic interests (farmers, hydropower plants, tourism activities ...) together with ecological requirements imposed by the European Water Framework Directive. It consists in a daily value of the river flow measured in cubic meter per second. For a given year, the water management program is said to be efficient if the mean of the lowest stream flow for 10 consecutive days or the mean of the lowest monthly stream flow over the last five years are above 80% of the reference value. The management is sustainable if this condition is fulfilled 8 years out of 10. To deal with such issues, time series models can be used to evaluate and improve the relevance of these water management programs implemented by the Adour-Garonne Water Agency.

Seasonal time series models like the seasonal autoregressive moving average (SARMA) model developed originally by Box and Jenkins (1970, chapter 9) have been extensively studied in the literature and applied to river flows displaying seasonal fluctuations in mean, standard deviations and skewness.

However, as pointed out by McLeod (1993), river flows for a particular season of the year may be statistically similar from year to year, but may depend intrinsically on the season. This feature cannot be captured by SARMA models which represent a class of stationary models with large lag autocorrelations that are invariant with respect to the season. Moreover as the correlation structure of these time series depends on the season, many seasonal time series cannot be filtered to achieve second-order stationarity (Vecchia, 1985b). The majority of river flow time series satisfy the property of periodic stationarity, stating that their mean and covariance functions are periodic with respect to time. Hence the periodic correlation structure of time series justifies the use of periodic autoregressive (PAR) modelling in water resources. Moreover the use of these models appears to be relevant and tractable for analysing river flows at a quarter-monthly periodicity.

The method of moments based on Yule-Walker equations (McLeod, 1994) and the least squares (LS) method in the univariate case (Franses and Paap, 2004) are efficient to estimate PAR models. As mentioned by Hipel and McLeod (1994), when the seasonal data and the model for each season are used rather than the annual data and the associated model, significant gain in parameter efficiency can be achieved. However the main problem in PAR modelling relies in the number of parameters that need to be estimated which increases with the choice of the season for river-flows. Moving from monthly surveys to quarter-monthly river flow data increases both the number of models and the number of parameters to be estimated. To obtain parsimonious models, it is of interest to study situations in which linear constraints on the parameters of a given season are introduced (Ursu and Duchesne, 2009). It also justifies the use of genetic algorithm with Bayes information criterion (BIC) to identify the optimal order of the PAR model. A second problem in the parameter estimation of time series models occurs with the presence of additive outliers that may imply serious problems. In particular the sensitivity of the LS estimation method to outliers requires the use of robust approach (Denby and Martin (1979) for autoregressive models of first order; Ben et al. (1999) for vector autoregressive moving average (VARMA) models; Shao (2007) and Sarnaglia et al. (2010) for univariate PAR models). In periodic vector autoregressive (PVAR) models, Ursu and Pereau (2014) implement a robust estimation method based on residual autocovariances (RA) to deal with additive outliers.

This article is organized as follows. In Section 2, the PAR model is introduced and least squares estimators are computed. In Section 3, a robust estimation in the presence of outliers is developed. Section 4 illustrates the results for the case study of the Garonne River. Section 5 offers some concluding remarks.

2. Periodic models

The class of periodic autoregressive (PAR) models extends the class of autoregressive (AR) models by allowing the autoregressive parameters to vary with the seasons. It is worth pointing out that a PAR model is formed by defining a different AR model for each season of the year. A PAR model with 12 periods can be associated with 12 AR models. It should be noted that, when the number of periods is 1, PAR model becomes AR model.

Let $Y = \{Y_t, t \in \mathbb{Z}\}$ be a periodic autoregressive (PAR) stochastic process given by

$$Y_{ns+\nu} = \sum_{k=1}^{p(\nu)} \phi_k(\nu) Y_{ns+\nu-k} + \epsilon_{ns+\nu}. \tag{1}$$

For fixed ν and predetermined value s, the random variable $Y_{ns+\nu}$ represents the realization during the νth season, with $\nu \in \{1, \dots, s\}$, at year n, $n \in \mathbb{Z}$. With monthly data the value $\nu = 12$ is naturally selected, whereas that for quarter-monthly data $\nu = 48$. The autoregressive model order at season ν is given by $p(\nu)$, whereas $\phi_k(\nu)$, $k = 1, \dots, p(\nu)$, represent the autoregressive model coefficients during season ν, $\nu = 1, \dots, s$. The error process $\epsilon = \{\epsilon_t, t \in \mathbb{Z}\}$ in equation (1) corresponds to a periodic white noise, with $E(\epsilon_t) = 0$ and $\text{var}(\epsilon_{ns+\nu}) = \sigma^2(\nu) > 0$, $\nu = 1, \dots, s$. The random process Y_t in (1) is supposed to have zero mean.

Unless otherwise stated we assume that PAR models are stationary in the periodic sense. Periodic stationarity is discussed in Gladyshev (1961). Typically, the periodic models used in water resources and environmental systems are stationary, in the sense that they do not need to be differenced to achieve stationarity (or to say it differently, data do not have unit roots). In applications, seasonal means are first removed from the time series.

2.1. Identification and estimation for PAR models

This section summarizes without proofs the relevant material on identification, and parameter estimation for PAR models. References that provide detailed proofs are included in the text.

Several estimation techniques are available for PAR models, namely the least-square method (Franses and Paap, 2004; Lütkepohl, 2005), the method of moments based on Yule-Walker estimation (Pagano, 1978; Hipel and McLeod, 1994) and the maximum likelihood estimation (Vecchia, 1985a,b).

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

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

be $N \times 1$, $N \times 1$ and $N \times p(\nu)$ random matrices. By defining the $p(\nu) \times 1$ vector $\boldsymbol{\beta}(\nu)$ of the parameters as:

$$\boldsymbol{\beta}(\nu) = \left(\phi_1(\nu), \ldots, \phi_{p(\nu)}(\nu)\right)^{\top}, \tag{2}$$

the PAR model can be written in the following form:

$$\mathbf{z}(\nu) = \mathbf{X}(\nu)\boldsymbol{\beta}(\nu) + \mathbf{e}(\nu), \; \nu = 1, \ldots, s. \tag{3}$$

From equation (3), the least squares estimators (unconstrained and constrained) of $\boldsymbol{\beta}(\nu)$ can be easily found. For more details we refer the reader to Ursu and Turkman (2012, Section 2).

Various selection criteria using AIC or BIC can be used for PAR model identification. One possible way is to use the BIC selection criterion separately for each of the seasonal components:

$$BIC(\nu) = \log \hat{\sigma}^2(\nu) + \frac{\log(N)}{N} p(\nu), \tag{4}$$

where $\hat{\sigma}(\nu)$ stands for the least squares estimators of $\sigma(\nu)$, and $p(\nu)$ represents the number of autoregressive parameters in season ν (McLeod, 1994).

Even if this method reduces the number of models to be investigated, the number of possible models remains very high. The large number of possible solutions of the PAR selection model suggests that genetic algorithms

(GA) can be useful for an efficient examination of the space of solutions and the selection of the combination of parameters that corresponds to the best model. The GA combined with BIC criterion is a reliable and easy way of identifying PAR models (Ursu and Turkman, 2012).

We briefly summarize our GA procedure for subset PAR modeling.

- *String representation.* Each subset AR model is encoded as a string, each locus in the string is filled with 1 if the parameter is free, and with 0 if the parameter is constrained to zero. Since a maximum search order has to be selected, every string has the same length L. For example, if we take $s = 12$, $\nu = 1$ and $p(1) = 15$, and the model

$$Y_{12n+1} = \phi_6(1)Y_{12n-5} + \phi_7(1)Y_{12n-6} + \epsilon_{12n+1}$$

  then, the string representing our model is

$$000001100000000.$$

  Note that in this case, the number of all possible models is $12 \times 2^{15} = 393216$.

- *Initial population.* An arbitrarily population of chromosomes of size N_p is generated. Each chromosome is encoded as a binary string of length L as described above. The population size N_p and the length of the chromosome L are chosen by the investigator.

- *A fitness function.* Each chromosome is evaluated by means of a positive real-valued function called fitness function. Since the $BIC(\nu)$ may be negative, a natural candidate for the fitness function is an exponential transformation

$$f_j(\nu) = \exp\{BIC_j(\nu)/d\},$$

  where $BIC_j(\nu)$ stands for the $BIC(\nu)$ value for the jth chromosome in the current population and d is a scaling constant. For yet another appropriate fitness function we refer the reader to Gaetan (2000).

- *Generating a new population.* A new population of potential chromosomes is created, using evolutionary operators as : selection, crossover and mutation. This cycle continues until the maximum number of generations N_g is attained, or until a stop condition is reached.

For many variations of the basic GA and detailed explanations, see Goldberg (1989); Mitchell (1996); Sivanandam and Deepa (2008).

2.2. Forecasting with PAR model

Forecasting with PAR models proceeds in the same way than standard AR models. The objective is to obtain a forecast with the lowest possible error, leading to the minimum mean squared error forecast (MMSE). The MMSE forecast is given by its conditional expectation (Hipel and McLeod, 1994).

Assuming that observations and innovations are known up to the n-th year and ν-th season, one takes the conditional expectation of eq. (1) to obtain the MMSE forecast $\hat{Y}_{ns+\nu}(l)$, where $\hat{Y}_{ns+\nu}(l)$ is interpreted as the l-step ahead forecast at the forecast origin $t = ns+\nu$. For example, the 1-step ahead forecast made at the origin $t = ns + \nu$ is

$$\begin{aligned}\hat{Y}_{ns+\nu}(1) &= E[Y_{ns+\nu+1}|Y_{ns+\nu}, Y_{ns+\nu-1}, \ldots] \\ &= \phi_1(\nu)Y_{ns+\nu} + \phi_2(\nu)Y_{ns+\nu-1} + \ldots + \phi_p(\nu)Y_{ns+\nu-p+1}\end{aligned}$$

The causal representation of PAR models (Ursu and Duchesne, 2009, eqn.5) can be use to compute confidence intervals for forecasts but this is beyond the scope of this paper. The best general reference for confidence intervals in periodic models are Hipel and McLeod (1994, chap.15) and Anderson et al. (2013). The forecasting performance of several time series models used in river flow analysis is presented in Noakes et al. (1985). Results suggest that the PAR models provide the most accurate forecasts.

3. Robust modeling of PAR models

As it is well-known, estimation methods may be seriously affected in the presence of additive outliers (Bustos and Yohai, 1986; Shao, 2007). Additive outliers refer to a PAR process with probability $1 - \omega$ and a PAR process plus an error with probability ω. The occurrence of outliers is generally small ($\omega \leq 0.05$).

Robust estimators based on robust autocovariances for ARMA models were proposed by Bustos and Yohai (1986). Their methodology was extended for multivariate PAR models by Ursu and Pereau (2014). The system of equations obtained in Ursu and Pereau (2014, eq.(6) and eq.(9)) can be easily adapted for PAR processes.

Therefore, in order to reduce the influence of the residuals suspected to be outliers, we replace the residuals $\hat{\epsilon}_{ns+\nu}$ defined in Section 2 by their modified residuals $\tilde{\epsilon}_{ns+\nu}$ defined as:

$$\tilde{\epsilon}_{ns+\nu} = \psi\left(\frac{\hat{\epsilon}_{ns+\nu}}{\hat{\sigma}(\nu)}\right), \tag{5}$$

where ψ stands for an odd and bounded function and $\hat{\sigma}(\nu)$ is an robust estimator for $\sigma(\nu)$. An usual choice for the ψ function is the Huber function:

$$\psi_{H,k}(x) = \operatorname{sgn}(x) \min\{|x|, k\},$$

where k is a constant and $\operatorname{sgn}(x)$ is the signum function. Generally, an iterative algorithm is proposed for ARMA models in Bustos and Yohai (1986) and for PAR models in Ursu and Pereau (2014).

The convergence of this algorithm cannot be guaranteed, but the convergence always occurred in all the simulations run in Ursu and Pereau (2014). Other recent works on robustness in periodic time series include the estimator of PAR models proposed by Sarnaglia et al. (2010). Also, a robust estimation for PAR models was discussed in Shao (2007).

4. Case study: Garonne river

The PAR model is applied to the average monthly river flow and average quarter-monthly river flow of the Garonne river in the southwest of France. This river is the third largest river in France in terms of flow, with a catchment area of $56,000km^2$. The Garonne river flows over 647km from its source in Spain to the Atlantic Ocean. It is the main contributor to the Gironde Estuary which is the major European fluvial-estuarine system.

As mentioned in the introduction, the Adour-Garonne Water Agency aims at controlling the river flow at different management points along the river. We select measures recorded at Tonneins which is the outlet of the watershed. Tonneins is the latest measure point located downstream the river before the Gironde Estuary. Data are obtained from daily discharge measurements in cubic meter per second (m^3/s) from January 1959 to December 2010 (DIREN-Banque Hydro, French water monitoring). Daily data flows are then transformed in monthly data, respectively quarter-monthly data, consisting in flows averaged for one month, respectively from the 1st to the 7th, from the 8th to the 15th, from the 16th to the 22nd, and from the 23rd to the end of the month as in Hipel and McLeod (1994). At Tonneins, the threshold value of the daily flow is equal to 110 m^3/s. This objective is difficult to satisfy during the dry season in summer between the 1st of July to

October 31st which corresponds to periods 25 to 40. To reach this objective of minimal flow during the summer the Adour-Garonne Water agency uses the water stored in reservoirs and dams located in the Pyrenees mountains. This quantitative management of water for all the watershed implies high financial costs. During wet years, about 40 million of m^3 are needed for a cost of 2.5 millions of € and during dry years this amounts to 58 millions of m^3 for a cost of 5 millions of €. About 90% of the 58 millions of m^3 come from reservoirs used by hydropower plants. As the water storage capacity is limited, water saving restrictions have to implemented especially in irrigated agriculture.

Figure 1 plots the annual flows of the Garonne river between 1959 and 2010. It shows that several episodes of severe drought occurred in 1989-1990 and in 2005. This figure also shows that annual flows remained below the mean of 600 m^3/s for several years during the last decade.

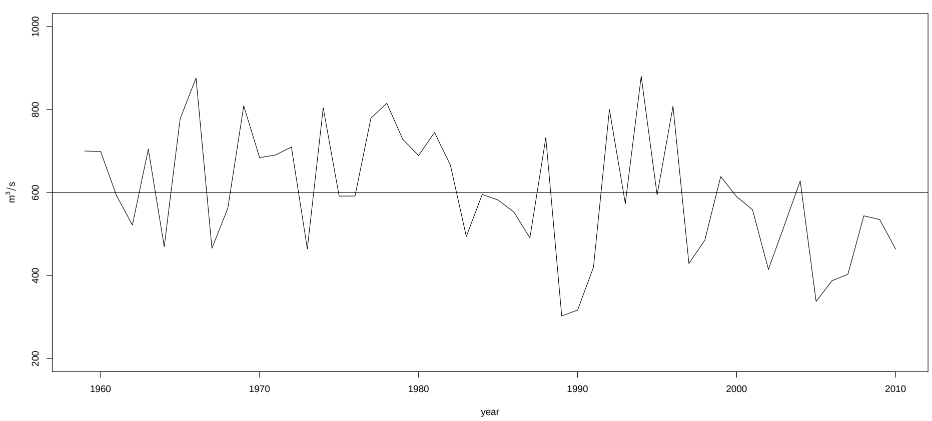


Figure 1: Plot of average annual flows of Garonne river at Tonneins (1959-2010)

To capture the periodic pattern of river flows, average monthly and quarter-monthly flow series are analysed. Table 1 and Table 2 show respectively the sample mean, median and standard deviation for each flow series. A partial plot of monthly and quarter-monthly surveys between 1980 and 2000 is given in Figure 2.

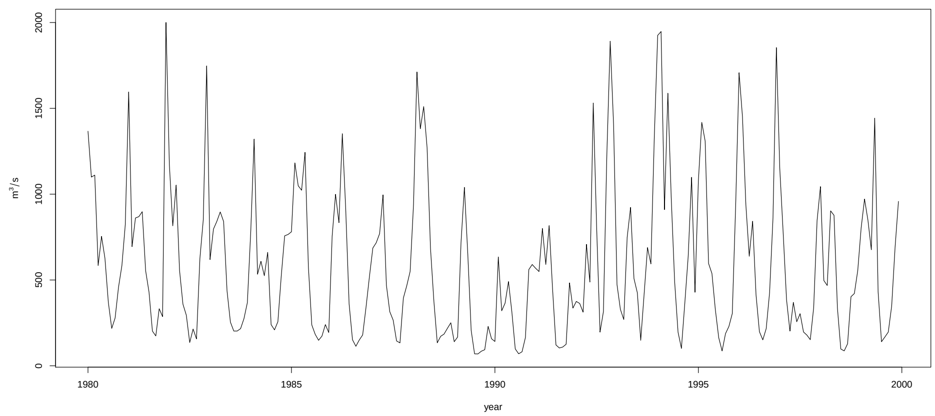
Figure 2 shows that a periodic fluctuation in the means and variances is clearly displayed. For monthly data, river flows are higher in February and much lower in August. For quarter-monthly data, most peaks occurred

period	mean	median	sd	period	mean	median	sd
1	927.89	854.70	493.66	7	258.06	235.60	159.46
2	984.70	876.25	477.15	8	161.63	146.85	81.52
3	824.19	810.80	321.06	9	212.13	175.10	118.54
4	840.39	808.65	331.90	10	344.47	274.75	226.29
5	798.71	756.15	320.76	11	524.22	438.10	322.96
6	527.83	482.65	249.42	12	823.28	721.80	600.96

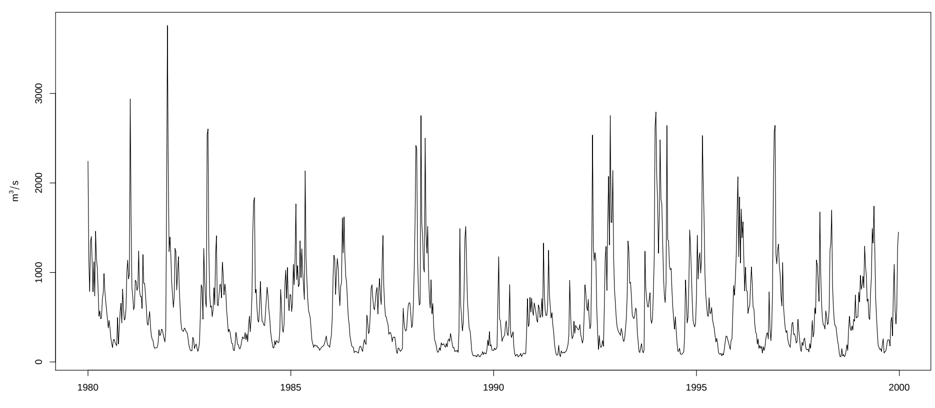
Table 1: Sample mean, median and standard deviation of the average monthly flow series. Period 1 corresponds to January.

period	mean	median	sd	period	mean	median	sd
1	898.87	738.21	749.83	25	334.77	283.57	198.61
2	864.82	682.19	574.70	26	306.38	236.38	273.73
3	907.50	822.93	536.55	27	227.25	207.21	125.45
4	1022.39	911.33	639.42	28	179.41	161.72	87.35
5	1001.22	795.86	675.40	29	166.54	145.43	106.39
6	998.22	778.56	608.97	30	164.21	145.62	84.26
7	995.64	833.93	559.70	31	150.34	130.29	72.73
8	938.86	795.71	507.46	32	164.31	151.00	92.44
9	843.37	754.29	388.77	33	171.39	157.36	91.53
10	778.47	707.50	361.49	34	179.45	156.81	86.64
11	809.78	695.36	494.55	35	226.81	173.57	199.56
12	861.10	722.28	463.38	36	267.63	177.00	221.72
13	810.89	793.29	387.19	37	288.85	212.07	233.54
14	821.18	686.50	433.57	38	316.05	276.94	222.85
15	829.67	764.93	371.26	39	352.91	285.21	210.20
16	894.76	782.81	469.93	40	406.41	299.56	327.58
17	867.98	707.86	458.46	41	441.59	308.57	382.77
18	835.76	777.50	407.20	42	515.26	367.38	352.86
19	768.96	678.14	379.08	43	550.07	470.36	399.15
20	735.05	683.44	349.19	44	582.86	462.31	403.36
21	630.39	596.71	258.53	45	779.49	533.71	678.62
22	612.22	506.50	387.94	46	835.27	563.38	756.48
23	484.57	399.00	256.52	47	829.46	579.00	680.99
24	391.56	351.75	196.47	48	841.88	610.22	735.66

Table 2: Sample mean, median and standard deviation of the average quarter-monthly flow series.



(a) Monthly periods



(b) Quarter-monthly periods

Figure 2: Partial plot of average monthly and quarter-monthly flows of Garonne river at Tonneins between 1980-2000

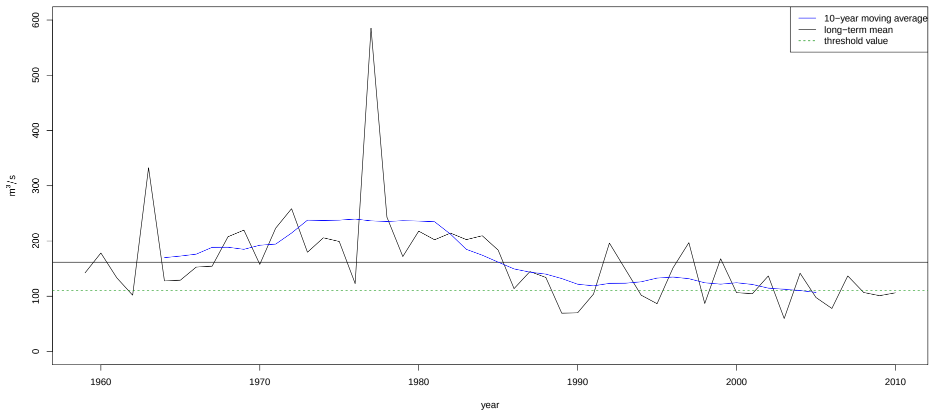
during the 4th, 5th, 6th and 7th period and most troughs occurred in the 30th, 31th, 32th and 33th period. Maximum mean quarter-monthly discharge was observed at 48th period in 1959 as 4059 m^3/s, while a minimum mean quarter-monthly discharge of 55.86 m^3/s was recorded at 27th period in 2003. The driest periods for monthly and quarter-monthly data are in August (8th month) and in period 31 respectively while February (2nd month) and the 4th period are the wettest period for monthly and quarter monthly data.

Figures 3 and 4 display the pattern of the series over the 1959-2010 period. For the driest periods, we observe that the 10-year moving average has increased from about 1959 to about 1975, then it was stabilized between 1975 to 1980 followed by a decrease from 1980 to 1990. By 1990, the 10-year moving average was around 110 m^3/s and seems to remain stable. A similar behavior is observed for the wettest periods. Changes in the average hydrological conditions in the Garonne river flows can be explained by several factors related to natural changes after the severe droughts of 1989-1990 or human activities with increased irrigated agriculture and urban growth. Since the mid 1990s, the stabilization of the 10-year moving average around the value 110 m^3/s is related to the implementation of water management policies.

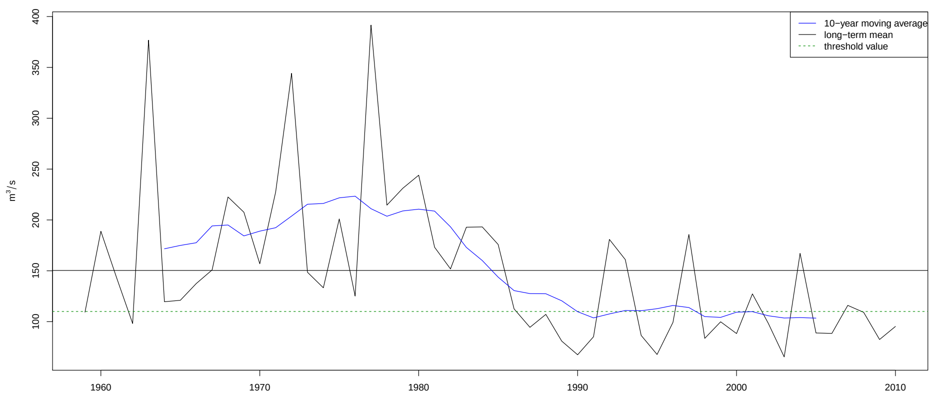
Figure 5 shows how many periods the quarter-monthly flows remain below the threshold value of 110 m^3/s. Between 1959 and 2010, it occurs more frequently during periods 28 to 33 with a maximum of 20 times for period 31.

In the unrobust case modeling, data have been centered by subtracting seasonal means (see Tables 1 and 2). The last year (12 months or 48 observations) has been omitted from the data set. The PAR model is fitted to the truncated series. The number of AR models used in PAR model is equal to the number of season associated to the choice of data set. We obtain 12 AR models with monthly data and 48 models for quarter-monthly data. Note that with the monthly model, the number of possible models to be estimated is approximately $4{*}10^5$ and it increases to $16{*}10^5$ with quarter-monthly data. This large number of possible solutions suggests the use of GA techniques to reduce the space of solutions and to select the combination of parameters which gives the best model. For each season, the parameters for the identified AR model are estimated using the least squares with linear constraints.

Using GA methods only 26 and 135 parameters have been estimated for the 12 and 48 different AR models, respectively. The most complicated model for quarter-monthly data is obtained for the 43th season where an AR

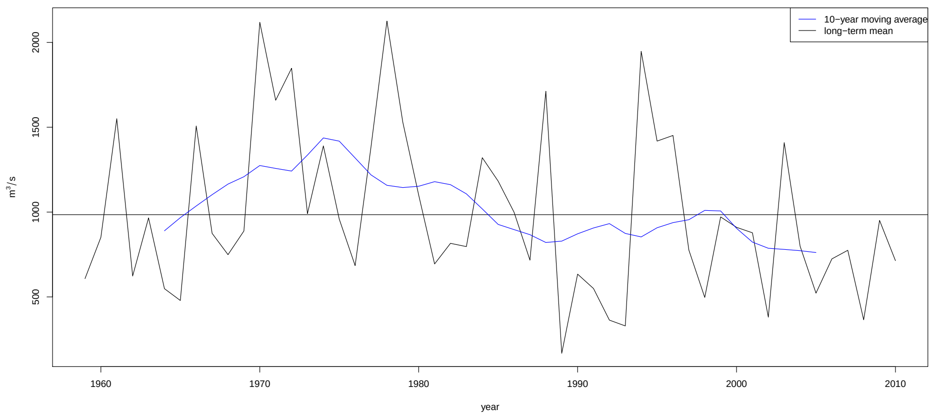


(a) Monthly periods

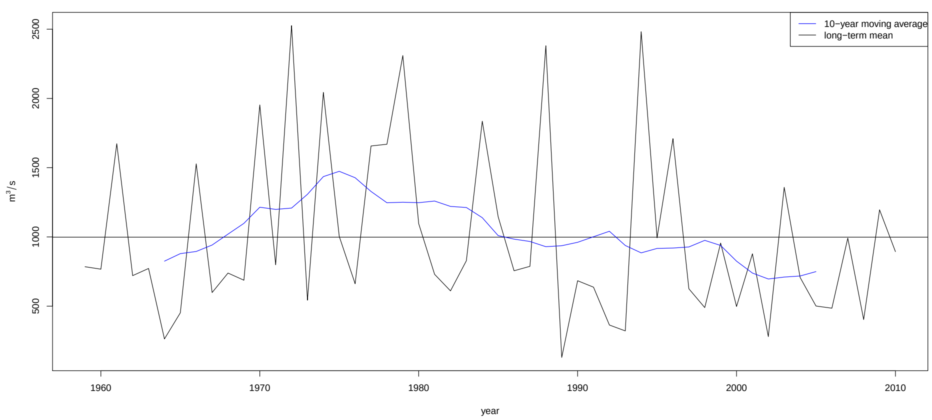


(b) Quarter-monthly periods

Figure 3: The driest monthly (August) and quarter-monthly flows (31th period) of Garonne river at Tonneins (1959-2010). The long term mean for August is 161.27 m$^3/s$ and 150.34 m$^3/s$ for period 31.



(a) Monthly periods



(b) Quarter-monthly periods

Figure 4: The wettest monthly (February) and quarter-monthly flows (4th period) of Garonne river at Tonneins (1959-2010). The long term mean for February is 984.7 m^3/s and 1022.39 m^3/s for period 4.

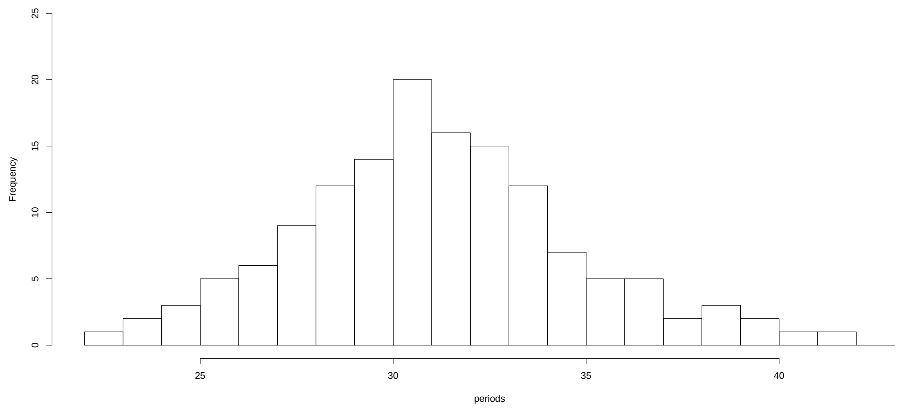


Figure 5: Histogram of driest quarter-monthly flows of Garonne river at Tonneins (1959-2010)

model with 8 parameters was identified. With monthly data, the number of estimated parameters was 6 for the month of October[1]. A parsimonious PAR model may be obtained by using a single model for all seasons in a given group and therefore the number of parameters in a PAR models is decreasing (Hipel and McLeod, 1994, chapter 14).

The proposed models for Garonne are then used to generate one-step-ahead forecasts for both flow series. Figure 7 shows forecast and the observed data for 2010. Figure 1 shows that the river flow in 2010 is about 400 m^3/s which is lower than the mean of the whole sample. It could be considered as a dry year.

4.1. Robust modeling of river flows

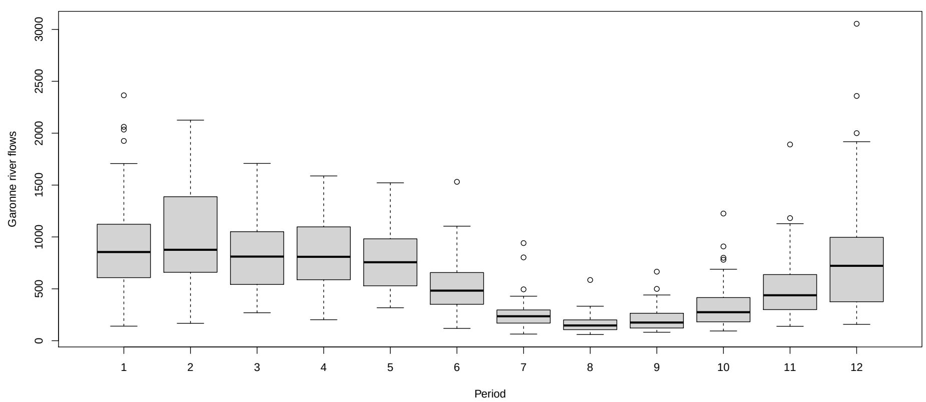
There are several observations that can possible be identified as outliers in the seasonal boxplots in Figure 6 (according to the one and half inter-quartile range rule). Outliers appear to be more numerous in quarter-monthly flows

[1] The identified model associated to this month can be written as

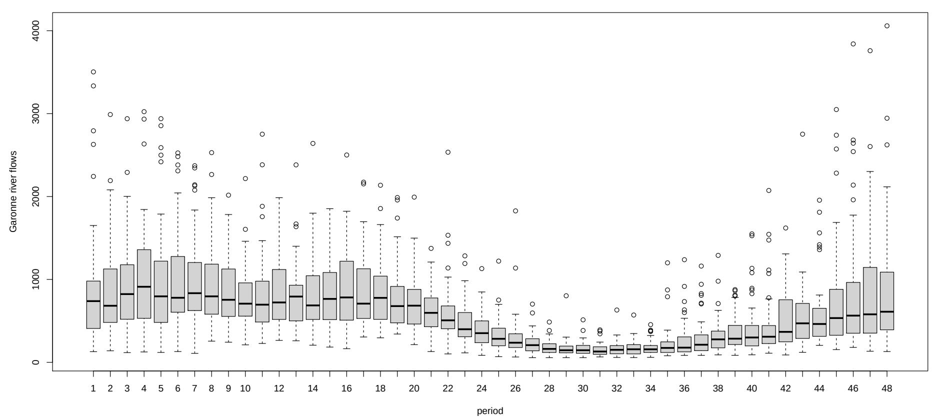
$$Y_{12n+10} = \hat{\phi}_1 Y_{12n+9} + \hat{\phi}_2 Y_{12n+8} + \hat{\phi}_3 Y_{12n+7} + \hat{\phi}_4 Y_{12n+6} + \hat{\phi}_{10} Y_{12n} + \hat{\phi}_{11} Y_{12n-1} + \hat{\epsilon}_{12n+10},$$

where $\hat{\phi}_1 = 1.23$, $\hat{\phi}_2 = -2.07$, $\hat{\phi}_3 = 1.08$, $\hat{\phi}_6 = -0.18$, $\hat{\phi}_{10} = 0.15$ and $\hat{\phi}_{11} = -0.27$.

than in monthly flows in the wettest period (from periods 4 to 10) but also during the driest periods (from periods 36 to 40).



(a) Monthly flows



(b) Quarter-monthly flows

Figure 6: Box-plots of the monthly and quarter-monthly flows of Garonne river at Tonneins

Contrary to the unrobust case, data have been centered by subtracting the seasonal medians instead of the seasonal means. As indicated by Shao (2007), this choice is better since outliers have smaller impact on the medians. The robust procedure described in Section 3 is then applied. As mentioned in Section 2, the recommended approach for identifying the AR parameters required in each season for the PAR model is to use genetic algorithms. It is important to emphasize that the number of estimated parameters for monthly and quarter-monthly data were 23 and 113 respectively. The most complicated AR model (for one period) has 5 parameters (in November) and 6 parameters (27th period) for the monthly and quarter-monthly data, respectively.

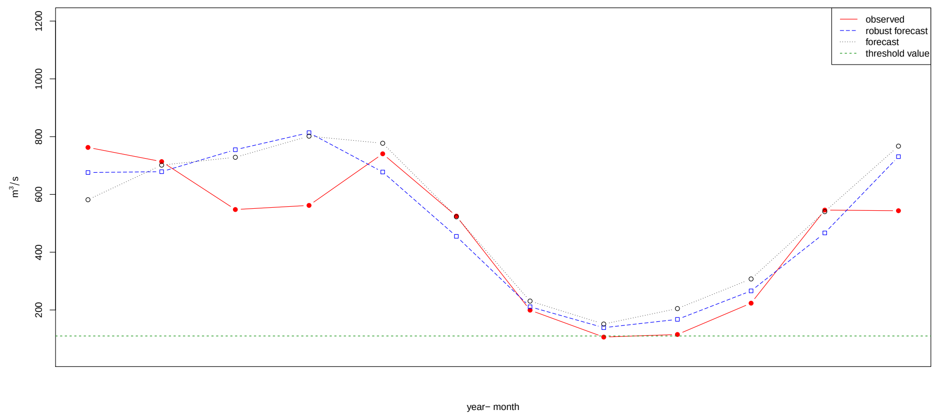
The robust model for Garonne data are used to generate one-step-ahead forecasts for the average flow series. Figure 7 shows robust forecast for 2010 and the observed data.

To evaluate the forecast accuracy for the unrobust and robust models we compute the following measures: root mean square error (RMSE), mean absolute error (MAE), mean percentage error (MPE) and mean absolute percentage error (MAPE). These measures are explicitly defined in Hyndman and Koehler (2006). These criteria have to be interpreted only as an indication to which model perform better, but no statement can be drawn from this comparison. To test the null hypothesis of no difference in the accuracy of the proposed models a Wilcoxon signed rank test for paired data may be used (Noakes et al., 1985). For another measures of forecast accuracy we refer to Hyndman and Koehler (2006).

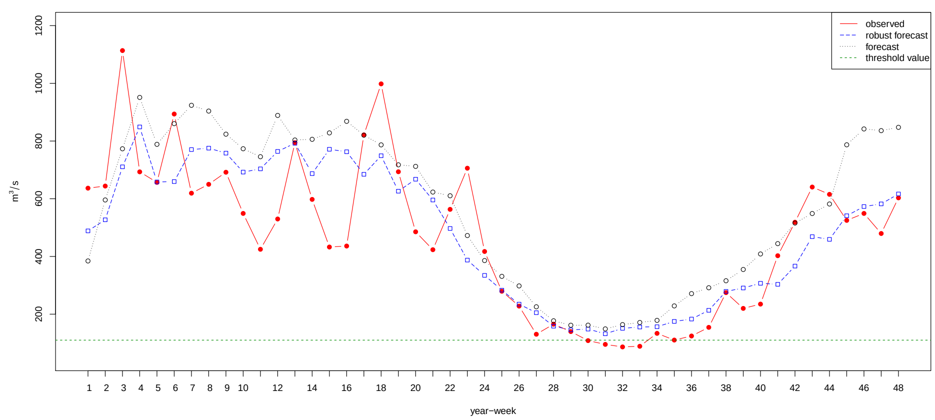
	Monthly		Quarter-monthly		Agg. quarter-monthly	
criterion	unrobust	robust	unrobust	robust	unrobust	robust
MAPE	26.847	22.543	41.755	28.082	33.666	20.947
MPE	-22.366	-13.768	-34.937	-13.813	-31.004	-10.293
MAE	94.305	93.240	152.473	114.230	118.569	80.453
RMSE	126.663	119.332	193.609	151.430	153.045	100.823

Table 3: Accuracy of one-step forecasts of Garonne river flows.

Another monthly predictions can be derived from the aggregation of the quarter-monthly predictions and obtained by taking the average over the four periods of each month. Table 3 gives the measures RMSE, MAE, MAPE and MPE for monthly, quarter-monthly and aggregated quarter-monthly predictions. Results show that the robust model is better with respect to all four criteria whatever the data frequency. It is important to emphasize that results are also better for the aggregated quarter-monthly predictions than the monthly data predictions, meaning that using a quarter-monthly model may improve monthly predictions.



(a) Twelve-months forecast



(b) Forty-eight period forecast

Figure 7: Robust forecast (dashed line) based on 51 years of Garonne river data. The observed data (solid line) and the unrobust forecast (dotted line) are also shown. The observed data were not use in the forecast.

5. Conclusions

Accurate forecasting of river flows is one of the most important applications in modern hydrology, especially for the management of reservoir systems.

Based on a robust modeling approach for identification and estimation of periodic autoregressive time series model, this paper provides an application to the Garonne River over the 1959-2010 period. To deal with the problem of large number of parameters that need to be estimated especially with quarter-monthly models, an automatic method using genetic routines has been developed. Results show that detection of outliers has been higher with quarter-monthly flow data than monthly data, implying better robust estimators than the least square (unrobust) estimators. Then, the forecasting accuracy of unrobust and robust models were investigated for one year. Results show that forecasts are better off in monthly and quarter-monthly robust models. The aggregated quarter-monthly forecasts also show better performance than the monthly forecasts in the robust analysis. The accuracy of the forecast analysis during the driest season is important for the river management authorities to achieve minimal flow objectives during these periods. Future research dealing with neural networks and periodic threshold models is expected to improve river flow fitting and forecasting.

Acknowledgements

The authors thank A. Coynel from university of Bordeaux and EPOC unit for her comments. Garonne flows data are available upon request. This study has been carried out with financial support from the French National Research Agency (ANR) as part of the project ADAPTEAU (ANR-11-CEPL-008) and in the frame of the Cluster of Excellence COTE (ANR-10-LABX-45).

References

Anderson, P.A., Meerschaert, M.M., Zhang, K., 2013. Forecasting with prediction intervals for periodic autoregressive moving average models. Journal of Time Series Analysis 34, 187–193.

Baker, M.A., Vervier, P., 2004. Hydrological variability, organic matter supply and denitrification in the garonne river ecosystem. Freshwater Biology 49(2), 181–190.

Ben, M.G., Martinez, E.J., Yohai, V.J., 1999. Robust estimation in vector autoregressive moving-average time series models. Journal of Time Series Analysis 20, 381–399.

Box, G.E.P., Jenkins, G.M., 1970. Time series analysis, forecasting and control. Holden-Day, San Francisco, CA.

Bustos, O.H., Yohai, V.J., 1986. Robust estimates for ARMA models. Journal of the American Statistical Association 81, 155–168.

Caballero, Y., Voirin-Morel, S., Habets, F., 2007. Hydrological sensitivity of the adour-garonne river basin to climate change. Water Resources Research 43(7).

Denby, L., Martin, R.D., 1979. Robust estimation of the first-order autoregressive parameter. Journal of the American Statistical Association 74, 140–146.

Franses, P.H., Paap, R., 2004. Periodic time series models. Oxford University Press.

Gaetan, C., 2000. Subset ARMA model identification using genetic algorithms. Journal of Time Series Analysis 21, 559–570.

Gladyshev, E.G., 1961. Periodically correlated random sequences. Soviet Mathematics 2, 385–388.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization & Machine Learning. Addison-Wesley.

Hipel, K.W., McLeod, A.I., 1994. Time series modelling of water resources and environmental systems. Elsevier, Amsterdam.

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22, 679–688.

Lütkepohl, H., 2005. New Introduction to Multiple Time Series Analysis. Springer: Berlin.

Maire, A., Buisson, L., Biau, S., Canal, J., Lafaille, P., 2013. A multi-faceted framework of diversity for prioritizing the conservation of fish assemblages. Ecological Indicators 34, 450–459.

McLeod, A.I., 1993. Parsimony, model adequacy, and periodic autocorrelation in time series forecasting. International Statistical Review 61, 387–393.

McLeod, A.I., 1994. Diagnostic checking periodic autoregression models with applications. Journal of Time Series Analysis 15, 221–233.

Mitchell, M., 1996. An introduction to genetic algorithms. MIT Press.

Noakes, D.J., McLeod, A.I., Hipel, K.W., 1985. Forecasting monthly riverflow time series. International Journal of Forecasting 1, 179–190.

Oeurng, C., Sauvage, S., Coynel, A., Maneux, E., Etcheber, H., Sanchez-Perez, J.M., 2011. Fluvial transport of suspended sediment and organic carbon during flood events in a large agricultural catchment in southwest france. Hydrological Processes 25, 2365–2378.

Pagano, M., 1978. On periodic and multiple autoregressions. The Annals of Statistics 6, 1310–1317.

Sarnaglia, A.J.Q., Reisen, V.A., Lévy-Leduc, C., 2010. Robust estimation of periodic autoregressive processes in the presence of additive outliers. Journal of Multivariate Analysis 101, 2168–2183.

Shao, Q., 2007. Robust estimation for periodic autoregressive time series. Journal of Time Series Analysis 29, 251–263.

Sivanandam, S.N., Deepa, S.N., 2008. Introduction to genetic algorithms. Springer: Berlin.

Ursu, E., Duchesne, P., 2009. On modelling and diagnostic checking of vector periodic autoregressive time series models. Journal of Time Series Analysis 30, 70–96.

Ursu, E., Pereau, J.C., 2014. Robust modeling of periodic vector autoregressive time series. Journal of Statistical Planning and Inference 155, 93–106.

Ursu, E., Turkman, K.F., 2012. Periodic autoregressive model identification using genetic algorithm. Journal of Time Series Analysis 33, 398–405.

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

Vecchia, A.V., 1985b. Maximum likelihood estimation for periodic autoregressive moving average models. Technometrics 27, 375–384.

Cahiers du GREThA
Working papers of GREThA

GREThA UMR CNRS 5113

Université de Bordeaux

Avenue Léon Duguit
33608 PESSAC - FRANCE
Tel : +33 (0)5.56.84.25.75
Fax : +33 (0)5.56.84.86.47

http://gretha.u-bordeaux.fr/

Cahiers du GREThA (derniers numéros – last issues)

2015-02 : DOYEN Luc, CISSE Abdoul, SANZ Nicolas, BLANCHARD Fabien, PEREAU Jean-Christophe *The tragedy of ecosystems in open-access,*

2015-03 : BLANCHETON Bertrand, *La loi de 1973 sur l'indépendance de la Banque de France. Le mythe de la fin des avances*

2015-04 : BALLET Jérôme, BARILLOT Sébastien, *Cruauté, Sadisme et Masochisme : Des Dispositions Morales Efficientes pour les Clubs*

2015-05 : RAITERI Emilio, *Un temps pour nourrir ? Evaluer l'effet des marchés publics innovants sur la généralité technologique à partir des données de brevet*

2015-06 : FERRARI Sylvie, *Fondements et enjeux bioéconomiques de la durabilité: L'apport de Nicholas Georgescu-Roegen*

2015-07 : FERRARI Sylvie, *De la Bioéconomie à la Décroissance : proximités et distances entre Nicholas Georgescu-Roegen et Serge Latouche*

2015-08 : CLEMENT Matthieu, BONNEFOND Céline, *Does social class affect nutrition knowledge and food preferences among Chinese urban adults?*

2015-09 : LEVASSEUR Pierre, Causal effects of socioeconomic status on central adiposity: Evidence using panel data from urban Mexico

2015-10 : BERGE Laurent, Network proximity in the geography of research collaboration

2015-11 : MAVEYRAUD Samuel, The international contagion of short-run interest rates during the Great Depression

2015-12 : ROUGIER Eric, "The parts and the whole": Unbundling and re-bundling institutional systems and their effect on economic development

2015-13 : BEDU Nicolas, VANDERSTOCKEN Alexis, L'impact des subventions régionales à la R&D: le cas des PME aquitaines

2015-14 : URSU Eugen, PEREAU Jean-Christophe, Application des processus périodiques auto-regressifs à la modélisation des débits de la Garonne

La coordination scientifique des Cahiers du GREThA est assurée par Emmanuel PETIT. La mise en page est assurée par Julie VISSAGUET