Computational Statistics manuscript No.
(will be inserted by the editor)

Detection and estimation of additive outliers in seasonal time series

Francesco Battaglia · Domenico Cucina · Manuel Rizzo

Received: date / Accepted: date

Abstract The detection of outliers in a time series is an important issue because their presence may have serious negative effects on the analysis in many different ways. Moreover the presence of a complex seasonal pattern in the series could affect the properties of the usual outlier detection procedures. Therefore modelling the appropriate form of seasonality is a very important step when outliers are present in a seasonal time series. In this paper we present some procedures for detection and estimation of additive outliers when parametric seasonal models, in particular Periodic AutoRegressive, are specified to fit the data. A simulation study is presented to evaluate the benefits and the drawbacks of the proposed procedure on a selection of seasonal time series. An application to three real time series is also examined.

Keywords Periodic autoregressive process · Periodic autocorrelation · False detection

1 Introduction

The presence of outliers in time series data is generally known as a source of deviation from the assumptions of the statistical model employed for analysing data. An undetected outlier can lead to model misspecification, biased parameter estimates and unreliable forecasting results.

F. Battaglia
Department of Statistical Sciences, University La Sapienza, Rome, Italy
E-mail: francesco.battaglia@uniroma1.it

D. Cucina
Department of Economics and Statistics, University of Salerno, Italy
E-mail: dcucina@unisa.it

M. Rizzo
Department of Statistical Sciences, University La Sapienza, Rome, Italy
E-mail: manuel.rizzo@uniroma1.it

When outliers are present in a seasonal time series, modelling the appropriate form of seasonality may strongly influence the success of the outliers detection procedure. This is also true because the influence of the outlier in the series may be masked by the seasonal effect, which may have a complex structure possibly subject to changes in time. In this paper we present some estimation procedures for additive outliers detection when seasonal models are specified to fit the data.

Among several approaches proposed to deal with outliers estimation we shall focus on methods based on parametric models. In such a context we refer to the classic works by, among others, Tsay (1986); Chang et al. (1988); Chen and Liu (1993); Gómez and Maravall (2001). Under the assumption of data generated by an ARIMA process and a contamination model, for each observation a contamination measure is estimated by least squares. At the times at which the largest measures occur, if above a pre-determined threshold, the corresponding observations can be considered as aberrant. Several different contamination models, corresponding to different outlier types, are considered: additive outliers that influence only one observation, innovation outliers that affect the innovation process of the ARMA model, level shifts that modify the process mean from a certain time on, and transient level changes.

In this paper we extend this kind of outlier detection methodology to Periodic Autoregressive (PAR) processes, that were proved very effective in modelling complicated seasonality patterns. A PAR is a non-stationary process where the observations at each seasonal position (month, quarter etc) are generated by a different autoregressive structure, and may account for seasonality not only in the means, but also in the autocorrelation and the variance. These models were proposed for describing time series arising in different areas such as economics, hydrology, climatology and signal processing (Hipel and McLeod 1994; Franses and Paap 2004; Ursu and Turkman 2012). Robust estimation procedures for the parameters of univariate and multivariate PAR models have been proposed in Shao (2008); Sarnaglia et al. (2010); Ursu and Pereau (2014). For a review of contributions to PAR models see Franses and Paap (2004) and references therein; these models were also considered in a Bayesian framework (e. g. Vosseler and Weber 2018).

The present authors have already considered the case of level shift (including the seasonal level shift proposed by Kaiser and Maravall 2001), in a more general setting of regime-switching PAR models (Battaglia et al. 2018; Cucina et al. 2019). In the present work we shall focus on additive outliers, which correspond to the situation in which an exceptional event, or 'a gross error of observation or recording error, affects a single observation' (Fox 1972).

For illustrating the relevance of seasonality in outlier analysis and the risk that the seasonal effect decreases the evidence of a contaminated observation, we consider a simple example. Starting from the same string of Gaussian white noise innovations, we have simulated a stationary first-order autoregression with parameter -0.5 and a strongly seasonal monthly periodic autoregressive model (referred to as Model 2 in Section 4, with different monthly means and autocorrelations), and added to both series at the same time q an outlier of

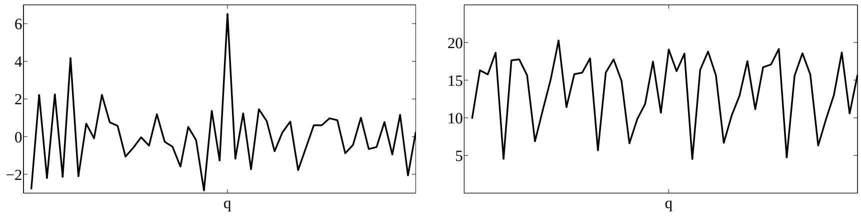


Fig. 1 Detail of simulated series with an additive outlier of size 5 at time q. Left panel: $AR(1)$ with parameter -0.5. Right panel: $PAR(1)$ according to Model 2 specification

equal size 5. The two models have approximately the same overall variance, but the outlier for the AR data is relatively evident at first sight, while it is not noticeable at all in the data generated by the seasonal model, as it may be seen in Figure 1. We computed also the differences $y_t = x_t - x_{t-12}$ of the seasonal series, but the aberrant observation is unnoticeable even in the differenced series.

The simultaneous influence of outlying observations and seasonal effect on the data has already been analysed in Haldrup et al. (2011) where the authors proposed a non parametric approach, based on seasonally differenced data, to detect additive outliers in series with unit roots. The test (from now on HMS test) may be employed also when the variance is subject to seasonal change and for stationary series. It is compared with our methods in the next Section and also through some simulations in Section 4.

The paper is organized as follows: Section 2 outlines the proposed methodology; Section 3 addresses computational issues; Section 4 reports some simulation studies. In Section 5 we illustrate real data applications, and Section 6 concludes the paper.

2 Methodology

In this Section we extend the outlier estimation and detection method of Tsay (1986, 1988) to periodic autoregressive processes, considering a seasonal series with N observations recorded s times a year, generated by the PAR model:

$$\Pi_t(B)[X_t - m(t)] = \varepsilon_t, \tag{1}$$

where $\Pi_t(B) = \sum_{j=0}^{p} \pi_t(j)B^j$ ($\pi_t(0) = 1$), the single parameters $\pi_t(j)$ are periodic with respect to t of period s, the seasonal means $m(t)$ are also periodic of period s and the variables $\{\varepsilon_t\}$ are uncorrelated with means $E(\varepsilon_t) = 0$ and variances $E(\varepsilon_t^2) = \sigma_\varepsilon^2(t)$, also periodic with period s. To allow for stochastic seasonality, $\Pi_t(B)$ may have one or more seasonal unit roots. In this case we assume that the polynomial may be written

$$\Pi_t(B) = \Phi_t(B)\Delta_t(B)$$

where $\Phi_t(B) = \phi_t(0) - \phi_t(1)B - \phi_t(2)B^2 - \ldots - \phi_t(h)B^h$ with $\phi_t(0) \equiv 1$ is the autoregressive part with all roots outside the unit circle, and $\Delta_t(B)$ is periodic of period s and has only k unit roots, possibly different for each t, with $h+k=p$. The problem of seasonal unit roots has been extensively addressed mainly in the econometrics literature (e.g. Ghysels and Osborn 2001) and several tests are available in the literature for testing the existence of seasonal unit roots, the most popular seems to be the HEGY test of Hylleberg et al. (1990). These tests may help to evaluate the opportunity of including a unit root factor $\Delta_t(B)$ in the autoregressive polynomial.

The observed perturbed series is $Y_t = X_t + \delta(t,q)\omega$ where $\delta(\alpha,\beta)=1$ if $\alpha=\beta$ and zero otherwise; $\{Y_t\}$ has an additive outlier of size ω at time q. We suppose for simplicity that the series is observed for M full years, so that $N=Ms$. The model for the observed data Y_t is

$$\Pi_t(B)[Y_t - m(t) - \delta(t,q)\omega] = \varepsilon_t.$$

In order to estimate ω, we assume normality and minimize the sum of squares of residuals in the log-likelihood:

$$\sum_t \varepsilon_t^2 = \sum_t \left\{\Pi_t(B)[Y_t - m(t) - \delta(t,q)\omega]\right\}^2.$$

Denoting the observed residuals by e_t:

$$e_t = \Pi_t(B)[Y_t - m(t)] = \sum_{j=0}^{p} \pi_t(j)[Y_{t-j} - m(t-j)] = \varepsilon_t + \sum_{j=0}^{p} \pi_t(j)\delta(t-j,q)\omega$$

we have

$$\sum_t \varepsilon_t^2 = \sum_t \left\{ e_t - \sum_{j=0}^{p} \pi_t(j)\delta(t-j,q)\omega \right\}^2.$$

Note that

$$\sum_{j=0}^{p} \pi_t(j)\delta(t-j,q) = \begin{cases} 0 & , t<q \\ 1 & , t=q \\ \pi_t(t-q) & , t=q+1,\ldots,q+p \\ 0 & , t>q+p \end{cases}$$

and on equating to zero the derivative with respect to ω we obtain the maximum likelihood estimate $\hat{\omega}$:

$$\hat{\omega}_q = \left\{ \sum_{j=0}^{p} \pi_{q+j}(j) e_{q+j} \right\} \Big/ \left\{ \sum_{j=0}^{p} \pi_{q+j}^2(j) \right\}.$$

Since $E(e_q)=\omega, E(e_{q+j})=\pi_{q+j}(j)\omega$, the above estimate is unbiased. Its variance is (because the e_t's are independent):

$$\mathrm{Var}\{\hat{\omega}_q\} = \left\{ \sum_{j=0}^{p} \pi_{q+j}^2(j)\sigma_\varepsilon^2(q+j) \right\} \Big/ \left\{ \sum_{j=0}^{p} \pi_{q+j}^2(j) \right\}^2$$

and therefore the standardized statistic is

$$\hat{\Omega}_q = \left\{ \sum_{j=0}^{p} \pi_{q+j}(j) e_{q+j} \right\} / \sqrt{\sum_{j=0}^{p} \pi_{q+j}^2(j) \sigma_\varepsilon^2(q+j)} \qquad (2)$$

and an outlier at $t = q$ is detected if $|\hat{\Omega}_q| > \Delta$, where Δ is a predefined threshold.

Under the Gaussian assumption, the statistic $\hat{\Omega}_q$ is normal with mean $E(\hat{\Omega}_q)$ and variance 1. Therefore if there is an outlier of size ω at time q the probability of detecting it is

$$\begin{aligned}\Pr\{|\hat{\Omega}_q| > \Delta\} = 1 - \Pr\{-\Delta < \hat{\Omega}_q < \Delta\} &= 1 - \Pr\{-\Delta - E(\hat{\Omega}_q) < Z < \Delta - E(\hat{\Omega}_q)\} \\ &= 1 - F\{\Delta - E(\hat{\Omega}_q)\} + 1 - F\{\Delta + E(\hat{\Omega}_q)\}\end{aligned}$$

where Z is $N(0,1)$ and F is its cumulative probability function, where

$$E(\hat{\Omega}_q) = \omega \left\{ \sum_{j=0}^{p} \pi_{q+j}^2(j) \right\} / \sqrt{\sum_{j=0}^{p} \pi_{q+j}^2(j) \sigma_\varepsilon^2(q+j)}.$$

If $\omega >> 0$ then $F\{\Delta + E(\hat{\Omega}_q)\} \simeq 1$, while if $\omega << 0$ then $F\{\Delta - E(\hat{\Omega}_q)\} \simeq 1$. It follows that the probability of detecting the outlier at $t = q$ may be written $1 - F\{\Delta - |E(\hat{\Omega}_q)|\}$, or

$$P_q = 1 - F \left\{ \Delta - |\omega| \frac{\sum_{j=0}^{p} \pi_{q+j}^2(j)}{\sqrt{\sum_{j=0}^{p} \pi_{q+j}^2(j) \sigma_\varepsilon^2(q+j)}} \right\}. \qquad (3)$$

Formula (3) details the effect of the chosen threshold Δ, the outlier size ω, the innovation variances $\sigma_\varepsilon^2(t)$ and the model structure $\{\pi_t(j)\}$ on the detection ability.

In order to evaluate the probability of false detections, let us consider the behaviour of the statistic on times $t \neq q$. Since $E(e_t) = 0$ for $t < q$ and $t > q + p$, the mean of $\hat{\omega}_t$ is zero except for $t = q-p, q-p+1, \ldots, q+p$ and for $k = 0, 1, \ldots, p$:

$$E(\hat{\omega}_{q-k}) = \omega \sum_{j=k}^{p} \pi_{q+j-k}(j) \pi_{q+j-k}(j-k) / \left\{ \sum_{j=0}^{p} \pi_{q+j-k}^2(j) \right\}$$

$$E(\hat{\omega}_{q+k}) = \omega \sum_{j=0}^{p-k} \pi_{q+j+k}(j) \pi_{q+j+k}(j+k) / \left\{ \sum_{j=0}^{p} \pi_{q+j+k}^2(j) \right\}.$$

It follows that, especially when $\Pi_t(B)$ contains unit roots, the outlier estimates for times around q may be largely biased.

In practice both the autoregressive parameters and the residual variances have to be estimated from the data. The autoregressive parameters may be estimated, for each season, by least squares on the observations related to that season only and in a similar manner the estimated residual variances may be obtained (see Section 3). Under the hypothesis that the model is correctly identified, both the parameters and residual variances estimates may be easily shown to be consistent as $M \to \infty$. A valid alternative, especially when many outliers are suspected, is using robust estimates (see e.g. Chen and Liu 1993).

The HMS test (Haldrup et al. 2011) is based on the seasonally differenced series, and the size of an additive outlier at $t = q$ is estimated by

$$\hat{d}(q) = \frac{1}{2}\{(1 - B^s)Y_q - (1 - B^s)Y_{q+s}\}.$$

The test statistic is obtained by standardizing $\hat{d}(q)$ with the following estimate of its variance:

$$\hat{s}^2(q) = \frac{1}{2}\{\hat{R}(0) - \hat{R}(s)\}$$

where $\hat{R}(j) = N^{-1}\sum_{t=s+j+1}^{N} \hat{v}_j \hat{v}_{t-j}$ and $\hat{v}_t = (1 - B^s)Y_t - \hat{d}(q)[\delta(t, q) - \delta(t + s, q)]$. Alternatively, in order to allow for seasonally varying variances, the covariances $\hat{R}(0)$ and $\hat{R}(s)$ are estimated using only the observations belonging to the same seasonal position as time q (HMS-PH test for periodic heteroscedasticity). We note that only in the particular case that the entertained PAR model is (1) with $\Pi_t(B) = (1 - B^s)$, $\hat{d}(q)$ would be exactly equal to $\hat{\omega}_q$, while the denominator of the HMS statistic differs from that of $\hat{\Omega}_q$ because it takes into account the possible correlation between e_q and e_{q+s}. Thus, only if the seasonally differenced series is actually generated by a white noise, the HMS statistic and ours give essentially equal values. In all other cases our test is based on the maximum likelihood estimate of the outlier size (given the model), therefore it is likely to produce more precise results, as will be exemplified with some simulations in Section 4.

3 Computational issues

We shall denote by $X_{(n-1)s+k}$ the series value during the $k-$th season, with $k = 1, \ldots, s$, at year $n = 1, 2, \ldots, M$. The periodic model used in this paper assumes a different mean for each season and (optionally) a linear trend. The residuals are treated as zero mean and described by an autoregressive model with order p, and parameters varying with seasons. Then

$$X_{(n-1)s+k} = a + b[(n-1)s + k] + m(k) + W_{(n-1)s+k}, \tag{4}$$

where $n = 1, \ldots, M$; $k = 1, \ldots, s$ and $W_{(n-1)s+k}$ is a PAR process given by:

$$W_{(n-1)s+k} = \sum_{i=1}^{p} \phi_k(i) W_{(n-1)s+k-i} + \varepsilon_{(n-1)s+k}. \tag{5}$$

The innovations $\varepsilon_{(n-1)s+k}$ in equation (5) are uncorrelated with means zero and variance $\sigma^2(k) > 0$, $k = 1, \ldots, s$.

For estimating trend and seasonal means by Ordinary Least Squares (OLS) we assume that the seasonal means sum to zero on one year:

$$m(1) + m(2) + \ldots, +m(s) = 0.$$

We propose to estimate the model parameters according to the following steps:

1. In the first step the following equation is estimated:

$$X_{(n-1)s+k} = b[(n-1)s + k] + c(1) + c(2) + \ldots + c(s) + e_t. \qquad (6)$$

   The design matrix $V : N \times (p+1)$ has first column equal to time, while the other columns are zero-one vectors with value one only at rows corresponding to the appropriate season. The parameter vector with dimension $(s+1)$ is

$$\beta = [b, c(1), c(2), \ldots, c(s)]'$$

   and estimated with

$$\hat{\beta} = [\hat{b}, \hat{c}(1), \hat{c}(2), \ldots, \hat{c}(s)]' = (V'V)^{-1}Vx$$

   where x is the data vector.
2. From the $\{\hat{c}(k)\}$, the intercept $\hat{a}$ and seasonal means $\hat{m}(k)$ are recovered assuming that the means sum to zero on a whole year. It follows

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

3. Based on estimated trend and seasonal means, the residual series is computed:

$$\hat{W}_{(n-1)s+k} = X_{(n-1)s+k} - \hat{a} - \hat{b}[(n-1)s+k] - \hat{m}(k), \;\; n = 1, ..., M, \; k = 1, ..., s$$

4. We denote by $I(k)$ the set of times belonging to season k. For each season k the observations belonging to the subseries $I(k)$ are selected and the least squares estimates of the parameters $\{\phi_k(i), i = 1, \ldots, p\}$ are obtained. For each season k $(k = 1, \ldots, s)$ the subseries $\{\hat{W}_t, t \in I(k)\}$ is arranged in a vector z_k with $n_k = M$ entries. The design matrix $Z : n_k \times p$ has rows containing the regressors $\hat{W}_{t-1}, \hat{W}_{t-2}, \ldots, \hat{W}_{t-p}$ and the parameter vector to be estimated is $\phi_k = [\phi_k(1), \phi_k(2), \ldots, \phi_k(p)]'$. The final estimate $\hat{\phi}_k = [\hat{\phi}_k(1), \ldots, \hat{\phi}_k(p)]'$ of ϕ_k is obtained by $\hat{\phi}_k = (Z'Z)^{-1}Z'z_k$.

5. Lastly, the estimation of error variances $\hat{\sigma}^2(k)$ is performed for each season on the final residuals:

$$\hat{\sigma}^2(k) = \frac{1}{M}\sum_{n=1}^{M}\hat{\varepsilon}^2_{(n-1)s+k},$$

   where $\hat{\varepsilon}_{(n-1)s+k} = \hat{W}_{(n-1)s+k} - \sum_{i=1}^{p}\hat{\phi}_k(i)\hat{W}_{(n-1)s+k-i}$.

The models discussed in this section extend the class of autoregressive (AR) models by allowing the autoregressive parameters to vary with the seasons. If at step 4 we assume that the autoregressive parameter do not change with seasons: $\phi_k(i) \equiv \phi(i)$ we obtain the constant parameter periodic autoregressive model

$$W_{(n-1)s+k} = \sum_{i=1}^{p} \phi(i) W_{(n-1)s+k-i} + \varepsilon_{(n-1)s+k}, \tag{7}$$

where the ε_t have constant variance $\sigma^2 > 0$. It may also be assumed that the seasonal means or the autoregressive parameters, or both, remain equal on groups of adjacent months, so that the final number of seasons ranges from 1 to s (see Thompstone et al. 1985; Battaglia et al. 2018).

The code that implements the proposed procedures is written in Matlab and has been run on MS-Windows-based system.

In our simulations the normal random numbers are generated by the function `randn` of Matlab that uses the method of Marsaglia and Tsang (2000). The simulated monthly series following the PAR models are generated according to equations (4) and (5) with $n = 1, \ldots, 100$, $k = 1, \ldots, 12$, initialized with $W_t = \varepsilon_t = 0, t \leq 0$, and discarding the first 100 terms.

4 Simulation study

We consider monthly series ($s = 12$) recorded on $M = 100$ years for a total of $N = 1200$ observations, and four data generating processes:

Model 1: a first-order periodic autoregressive model with 12 different monthly means (figures are taken from Lu et al. 2010) and autoregressive parameters $\phi_t(1)$ generally different for each month.

Model 2: a first-order periodic autoregressive model with only three different means, related to the months of January to April, May to August, September to December. The autoregressive parameters are also grouped: the first value operates on months from January to March, the second value is used for months from April to July, the third one relates to August–October, and the last one to November and December.

Model 3: it uses the estimated values of a periodic autoregressive model fitted to the Italian industrial production index (see Battaglia et al. 2019): there are 11 different monthly means (May and June share the same value) and the autoregressive structure is of order 3, and varies across months in the following way: a first set of parameters is used for January; a second set operates from February to July; a third set relates to August, and a final set of values is employed for the remaining months.

Model 4: monthly means equal to those of Model 1 and a constant second order autoregression with parameters $\phi(1) = 0.5, \phi(2) = -0.73$ (equal for each month).

The detailed values of the parameters are reported in Table 1. The innovations were always simulated according to standard normal independent variates, and

Table 1 Parameter values used in the simulations

Month	Model 1		Model 2		Model 3		Model 4	
	$m(t)$	$\phi_t(1)$	$m(t)$	$\phi_t(1)$	$m(t)$	$\phi_t(1), \phi_t(2), \phi_t(3)$	$m(t)$	$\phi_t(1), \phi_t(2)$
1	-.61	0.3	0	0.7	104	0.5,-.73,0.0	-.61	0.5, -.73
2	.99	0.3	0	0.7	111	0.0, 0.3, 0.63	.99	0.5, -.73
3	2.35	0.5	0	0.7	112	0.0, 0.3, 0.63	2.35	0.5, -.73
4	4.91	0.3	0	0.3	107	0.0, 0.3, 0.63	4.91	0.5, -.73
5	8.74	0.35	6	0.3	117	0.0, 0.3, 0.63	8.74	0.5, -.73
6	12.15	0.3	6	0.3	117	0.0, 0.3, 0.63	12.15	0.5, -.73
7	15.55	0.25	6	0.3	121	0.0, 0.3, 0.63	15.55	0.5, -.73
8	15.47	0.1	6	-.2	52	0.0, 0.56, 0.42	15.47	0.5, -.73
9	12.79	0.1	2	-.2	116	0.18, 0.0, 0.69	12.79	0.5, -.73
10	7.82	0.1	2	-.2	121	0.18, 0.0, 0.69	7.82	0.5, -.73
11	2.32	0.2	2	0.0	116	0.18, 0.0, 0.69	2.32	0.5, -.73
12	-.25	0.2	2	0.0	98	0.18, 0.0, 0.69	-.25	0.5, -.73

500 replications for each model were generated. Two additive outliers of size ω were added to each series at times selected at random: 302 (month 2) and 969 (month 9) for models 1 and 3, and times 121 (month 1) and 609 (month 9) for models 2 and 4, and the standardized statistics $\hat{\Omega}_t$ in (2) were computed for each t according to the following methods, that fit different models:

method 1: different monthly means and autoregressive parameters are estimated for each month, i.e. we fit a complete periodic autoregressive model
method 2: twelve monthly means are estimated, and a stationary autoregressive model is fitted to the data after removing the seasonal means, i. e. $\phi_t(j) \equiv \phi(j)$ and $\sigma^2_\varepsilon(t) \equiv \sigma^2_\varepsilon$ for any t.
method 3: a parsimonious periodic autoregressive model is considered, estimating only parameters that are actually different according to the data generating process; e. g. for model 2 we estimate only three seasonal means and four different autoregressive parameters, using observations of the appropriate months. Thus we fit a grouped periodic autoregressive model (Thompstone et al. 1985).

Method 2 assumes that the second order properties (variance and autocorrelations) do not vary with the season: the autoregressive parameters and the innovations variance (hence the variance of the series) are equal for each month.

First of all the outlier size estimates $\hat{\omega}_q$ were always found, for any method and model, with good accuracy, with a small bias and without large differences. Also the bias of autoregressive parameters was found moderate and scarcely influenced by the presence and size of the outliers.

As far as the behaviour of the standardized statistics computed at the perturbed times is considered, Figure 2 reports, for each model, the distributions (smoothed with a gaussian kernel density estimator) of the standardized statistics for the three methods and an outlier of size $\omega = 3$. It may be seen that the methods based on periodic autoregressive models generally give larger values of the statistic for model 1 to 3, while in the case of model 4 (constant au-

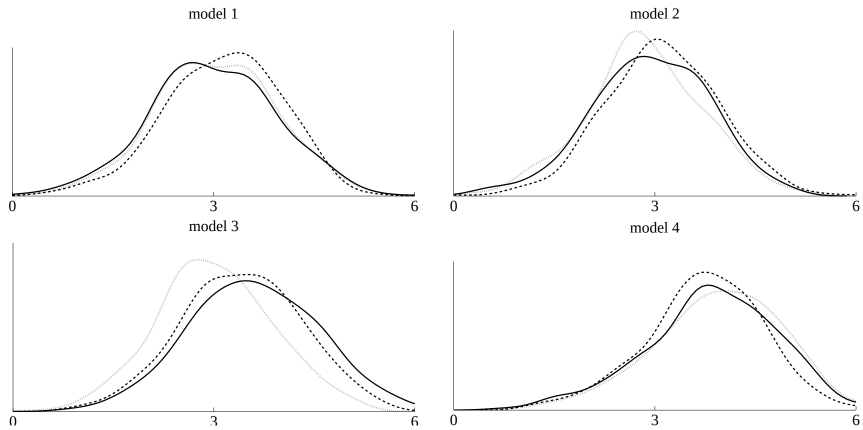


Fig. 2 Distributions of the standardized statistics computed at perturbed times with outlier size $\omega = 3$. Dashed line: method 1, dotted line: method 2; continuous line: method 3

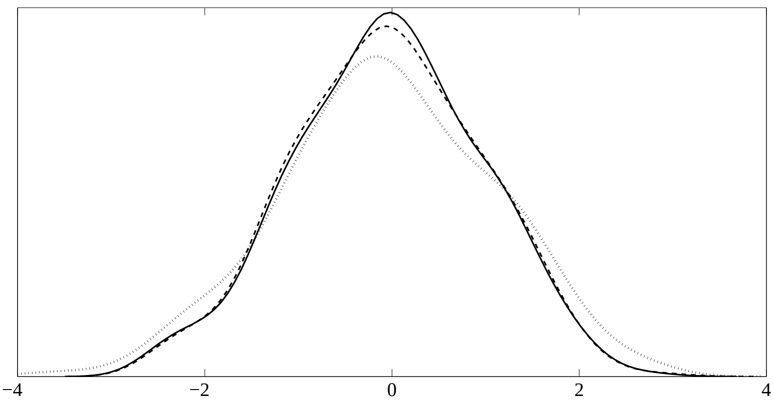


Fig. 3 Distribution of the standardized statistics computed at time 500 (not perturbed) for model 2. Dashed line: method 1, dotted line: method 2; continuous line: method 3

toregressive), method 2 (based actually on a constant autoregressive structure) yields better results, but the method based on grouped periodic autoregression is only slightly less satisfying.

Here, the differences among the three estimation methods are not markedly large. However, the behaviour on clean (not perturbed) observations changes: the distributions of the standardized statistics computed under method 2 (constant autoregressive) tend to show much larger tails (an example for model 2 at $t = 500$ appears in Figure 3), leading to a larger frequency of false detections.

We shall consider now a detection procedure based on the threshold $\Delta = 3.5$, and simulated series not perturbed or with outliers of size 2.0, 2.5, 3.0, 3.5, 4.0 or 4.5. Let us consider first the series without outliers. The number of observed false detections (standardized statistic absolute values larger than

Table 2 Observed frequencies of false detections on the 500 simulated replications

Model	1	2	3	4
method 1	195	225	228	226
method 2	268	439	1597	273
method 3	275	301	273	292

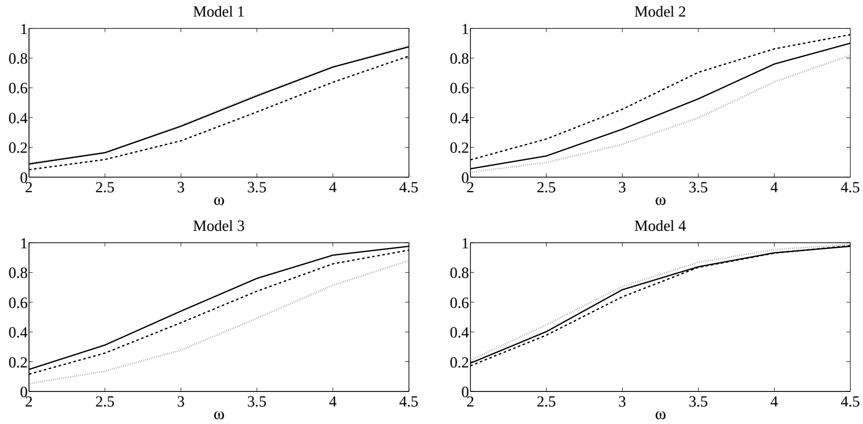


Fig. 4 Relative frequency of detection of an outlier of size ω. Dashed line: method 1, dotted line: method 2; continuous line: method 3

3.5) on the 500 replications (with 1200 observations each) is reported in Table 2. It appears clearly that, especially for the strongly periodic model 3, method 2 (based on constant autoregression) tends to overdetect with respect to the other methods. The frequency of false detections found for series with the outliers were similar to those of Table 2 for all outlier sizes ω and are omitted.

The performance on detecting outlying observations, in dependence of their actual size, is illustrated in Figure 4, where the relative frequency of significant values (larger than 3.5) for the standardized statistics for each model and method is reported (figures computed at $t = 969$ for model 1 and 3, and at $t = 121$ for model 2 and 4). It may be concluded that the differences among the three methods remain similar also when the outlier size ω is increased.

A further important topic is the precision of the detection procedure: if the maximum (in absolute value) of the standardized statistic is at time q, and is larger than the threshold Δ, then q is detected as a perturbed observation, but it could be an actually clean observation. We have computed (for each method, model and outlier size) the percentage of significant replications on which the maximum standardized statistic corresponds to an actual outlier, results are shown in Table 3. Again we may see less satisfying results for method 2, and a good detection ability for the other methods, nearly perfect for outlier size at least 4.

The following conclusions may be tentatively drawn:

Table 3 Percentage of replications in which the maximum significant statistic corresponds to true outlier locations

ω		Model 1	Model 2	Model 3	Model 4
3.0	method 1	71	78	81	89
	method 2	67	50	20	90
	method 3	66	66	85	89
3.5	method 1	85	90	92	96
	method 2	84	67	35	97
	method 3	82	83	94	96
4.0	method 1	94	96	97	99
	method 2	91	85	60	99
	method 3	91	94	98	98
4.5	method 1	98	98	99	99
	method 2	98	94	80	99
	method 3	98	98	99	99

1. Bias in autoregressive parameters and outlier size estimation was found limited. It happens also because we simulated long series with only one or two aberrant observations, and it cannot be excluded that the bias would be more serious for more perturbed and shorter series, and in that case robust variance estimators would be advisable.
2. The detection method based on constant autoregressive parameters and different monthly means (corresponding to the popular idea of stationary seasonal differences) is effective when seasonality is constant, but much worse when the seasonality affects also autocorrelations and variance.
3. The proposed detection methods, based on periodic autoregressive models, are satisfying and precise also for complex seasonality, and yield nearly equivalent results to the previous method when seasonality is constant.
4. The idea of fitting parsimonious periodic autoregressive models, considering different parameters only for groups of similar months, seems not to produce relevant benefits in outlier detection, because the frequencies of correct detection are only slightly larger but coupled with a bit more false detections.

Finally, we compared our proposed methods with the HMS test. First of all we considered the case of series whose seasonal differences are white noise: in this case, as explained in Section 2, we expect equivalent results. A set of simulated series generated by $x_t = x_{t-12} + \varepsilon_t$ with the ε_t's independent unit normal was considered, adding at $t = 121$ an outlier of size 4. The resulting test statistics of the HMS test and our methods applied with $\Pi_t(B) = (1 - B^{12})$ were found nearly equal. The same happened when the variances of the ε_t's were taken different, but periodic with period 12. Then, we considered the more realistic case of autocorrelated series, using the same set of previously simulated series according to models 1-4, but with only one outlier (at time $q = 969$ for models 1 and 3, and at time $q = 121$ for models 2 and 4) and size $\omega_q = 4$. The test statistics of our methods and HMS were computed on each of

Table 4 Comparison of the HMS test with our methods: percentages of detection and number of false detections based on threshold $\varDelta = 3.5$ and 500 replications

		Model 1	Model 2	Model 3	Model 4
% detection	HMS	36	30	17	7
	HMS-PH	46	23	23	9
	method 1	64	86	86	93
	method 2	74	64	71	95
	method 3	74	76	92	93
false detect.	HMS	327	569	321	241
	HMS-PH	637	618	588	612
	method 1	178	258	281	394
	method 2	237	379	1477	565
	method 3	237	294	347	571

the 500 replications, and with the same threshold $\varDelta = 3.5$, we registered the number of replications at which the test statistic in $t = q$ was larger than $\varDelta$ in absolute value (detections) and the number of times the statistics computed for $t \neq q$ were found larger than $\varDelta$ in absolute value (false detections). The results appear in Table 4. It may be concluded that our methods always yield definitely better results than the HMS test, differences being larger for the models with stronger autocorrelations.

5 Case studies

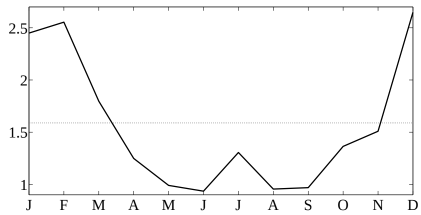
We consider three series widely studied in the literature, that exhibit strong and complex seasonality and have been analysed with periodic autoregressive models: the Central England Temperature series, the Saugeen and the Fraser river flow data.

For identifying the PAR models to be fitted we first applied the HEGY test, that did not provide evidence of seasonal unit roots in any of the three series, thus we consider autoregressive polynomials with roots only outside the unit circle. For selecting the order p we have used (as suggested by several authors, see e. g. Gómez and Maravall 2001) the Bayesian Identification Criterion of Schwarz (1978): $BIC(w) = -2\log(\hat{L}) + w\,\log(N)$ where w is the number of parameters and $\hat{L}$ the maximized likelihood. The results are shown for the three series in Table 5 and suggest for all of them order $p = 1$. A constant autoregressive model (according to method 2) of the same order was also fitted to the series.

The Central England Temperature series is one of the longest existing monthly temperature recording and was recently studied by Proietti and Hillebrand (2017). We take into account data from January 1921 to December 2013, that are also characterized by a slow increase (due to global warming), therefore a linear trend will be included in our models. The estimated residual autocorrelations of the PAR(1) model were small and none of the usual portmanteau tests applied to global residuals suggests to reject this model. The residual variance

Table 5 Identification of the PAR models for the series of Central England Temperature, Saugeen and Fraser riverflows: values of Schwarz's criterion for $p = 1, 2, 3, 4$

order	CET	Saugeen	Fraser
$p = 1$	3822	985	-245
$p = 2$	3883	1028	-188
$p = 3$	3953	1121	-125
$p = 4$	4021	1192	-53



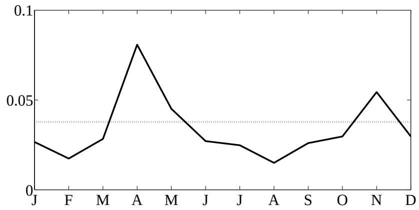


Fig. 5 Residual variances for single months of the PAR model. Dotted line: residual variance of the AR model

of the fitted constant autoregressive model (method 2) is 1.59, and the overall average residual variance of the fitted PAR model is 1.56, but much different across months, ranging from about 1 in May, June, August, September, 1.3 in July, up to about 2.5 for December to February (see Figure 5, left panel). If we consider method 2 based on a constant autoregressive structure several observations have notably large outlier statistics, the largest are at times 1080 (December 2010, $\hat{\Omega} = -4.16$) and 732 (December 1981, $\hat{\Omega} = -3.57$); also, there are standardized values larger than three at 289 (January 1945, $\hat{\Omega} = -3.24$) and 422 (February 1956, $\hat{\Omega} = -3.14$). On the contrary when a periodic autoregressive model is used (method 1) we get mild suspicious values only at times 1080 ($\hat{\Omega} = -3.35$) and 1027 (July 2006, $\hat{\Omega} = 3.22$), while at all other times the values are moderate, due to a better fit and a more accurate residual variance estimate of the PAR model with respect to the constant autoregressive. On the other side, observation 1027 is not indicated by method 2 because of the difference between the uncertainty for July and the global measure (residual variance 1.3 instead of 1.59). The two suggested outliers correspond to July 2006 with an estimated size $\hat{\omega}_{1027} = 3.02$, and December 2010 with estimated size $\hat{\omega}_{1080} = -5.31$. A plot of the series for the years 2004–2012 is reported together with the trend-seasonal means component in Figure 6, where the abnormal size of the average temperature in those two months appears clear. They are actually the largest and the smallest monthly temperature in the whole series.

The Saugeen river data set was considered in Hipel and McLeod (1994); Wong et al. (2007); Ye and Dai (2018). The time series consists of average monthly log river flows in m^3/sec collected at Walkerton, Canada from January 1915 to December 1976. We estimated a PAR(1) model and obtained very

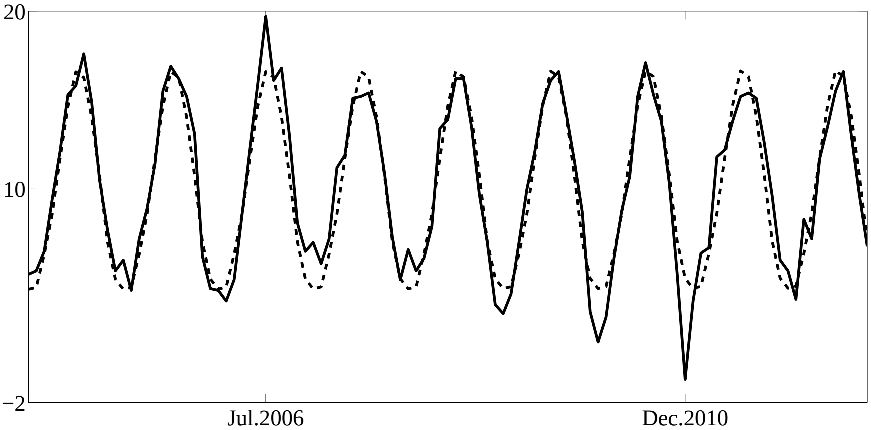


Fig. 6 Observations (continuous line) and trend-seasonal means component (dashed line) of the CET series, years 2004–2012

small residual autocorrelations. The estimated residual variance for a constant autoregressive model is 0.209, while fitting the PAR yields an overall average residual variance equal to 0.179, but varying across months from about 0.1 for June, August and September to 0.25 for December and January, to 0.35 for March. The method based on PAR detects time 478 (October 1954, $\hat{\Omega} = 4.28$) and, less evident, time 512 (August 1957, $\hat{\Omega} = -3.34$). Method 2 based on constant autoregressive structure does highlight time 478 ($\hat{\Omega} = 3.90$) but does not discover time 512, and in addition detects time 303 (March, $\hat{\Omega} = -3.86$) and further (but with standardized values smaller than 3.5) observations 267 and 315, both in March. These disagreements seem to be due to a less precise estimation of the residual variance implied by the non periodic structure.

River Fraser was often analysed in the literature of seasonal time series, and has also been fitted by means of periodic autoregressive models (McLeod 1994; Sarnaglia et al. 2010; Ursu and Turkman 2012). The related time series consists of mean monthly river flows collected at Hope, British Columbia. We study observations from January 1931 to December 1990 (720 monthly observations of the log river flows). The usual portmanteau tests applied to global residuals of the PAR(1) model do not reject this model. The residual variance of the constant autoregressive model is estimated equal to 0.0378, while the overall average residual variance of the fitted PAR model is 0.0338, but there are large differences among the various months: for February and August it is about 0.015, for May and November about 0.05, for April is more than 0.08 and for the other months around 0.025 (see Figure 5, right panel). Method 1, based on a PAR model, detects significant observations at times 374 (February 1962, $\hat{\Omega} = 4.10$) and 108 (December 1939, $\hat{\Omega} = 3.77$), and some other suspicious observations at times 211 (July 1948, $\hat{\Omega} = -3.38$) and 636 (December 1983, $\hat{\Omega} = -3.34$). Method 2 (based on a constant autoregressive structure), does suggest an outlier at 108 and 374 (though with a smaller significance) but detects also time 280 (April 1954, $\hat{\Omega} = -3.9$) and (with standardized values between 3 and 3.5) times 148, 208, 317, 447, 616, 641, 712. These observations generally belong to the month of April, the most difficult to fit (residual variance of April data is 0.08), and may be considered all false detections.

6 Conclusions

Our findings confirm that a complicated and strong seasonal behaviour may produce serious problems for the standard outlier detection procedures based on ARIMA models.

The proposed method, based on periodic autoregressive models, yields better results because, on one hand, it ensures a better general fitting, and on the other hand it allows a more precise estimation of the uncertainty, that may differ from month to month.

In these cases, the use of a stationary autoregressive model, also if modified with seasonal means, or applied to seasonal differences, is subject to several false detections and lack of precise outlier detection.

The present research may be forwarded in several directions, for example including moving average structures (PARMA models), considering periodically integrated autoregressive processes (Franses and Paap 2004), studying outlier patches (e. g. Justel et al. 2001). Moreover, the PAR idea may be extended to other outlier detection procedures, e. g. those based on genetic algorithms (Baragona et al. 2001; Cucina et al. 2014) and to multivariate time series (Ursu and Duchesne 2009).

Acknowledgements The authors wish to thank two referees whose comments greatly helped to improve the paper.

References

Baragona R, Battaglia F, Calzini C (2001) Genetic algorithms for the identification of additive and innovation outliers in time series. Comput Stat Data An 37(1):1–12

Battaglia F, Cucina D, Rizzo M (2018) Periodic autoregressive models with multiple structural changes by genetic algorithms. In: Corazza M, Durbán M, Grané A, Perna C, Sibillo M (eds) Mathematical and Statistical Methods for Actuarial Sciences and Finance: MAF 2018, Springer, Berlin, pp 107–110

Battaglia F, Cucina D, Rizzo M (2019) Parsimonious periodic autoregressive models for time series with evolving trend and seasonality. Stat and Comput. https://doi.org/10.1007/s11222-019-09866-0

Chang I, Tiao GC, Chen C (1988) Estimation of time series parameters in the presence of outliers. Technometrics 30(2):193–204

Chen C, Liu L (1993) Joint estimation of model parameters and outlier effects in time series. J Am Stat Assoc 88(421):284–297

Cucina D, di Salvatore A, Protopapas MK (2014) Outliers detection in multivariate time series using genetic algorithms. Chemometr Intell Lab 132:103–110

Cucina D, Rizzo M, Ursu U (2019) Multiple changepoint detection for periodic autoregressive models with an application to river flow analysis. Stoch Environ Res Risk Assess, forthcoming

Fox AJ (1972) Outliers in time series. J R Stat Soc Series B 34(3):350–363

Franses PH, Paap R (2004) Periodic Time Series Models. Oxford University Press, New York

Gómez V, Maravall A (2001) Automatic modeling methods for univariate series. In: Peña D, Tiao GC, Tsay RS (eds) A Course in Time Series Analysis. Wiley, New York, pp 171–201

Ghysels E, Osborn D (2001) The Econometric Analysis of Seasonal Time Series. Cambridge University Press

Haldrup N, Montañes A, Sansó A (2011) Detection of additive outliers in seasonal time series. J Time Ser Econometrics 3(2)

Hipel KW, McLeod AI (1994) Time Series Modelling of Water Resources and Environmental Systems. Elsevier, Amsterdam

Hylleberg S, Engle R, Granger C, Yoo B (1990) Seasonal integration and cointegration. J Econometrics 44(1):215–238

Justel A, Peña D, Tsay RS (2001) Detection of outlier patches in autoregressive time series. Stat Sinica 11(3):651–673

Kaiser R, Maravall A (2001) Seasonal outliers in time series. Estadistica 53(160):97–142

Lu Q, Lund R, Lee T (2010) An MDL approach to the climate segmentation problem. Ann Appl Stat 4(1):299–319

Marsaglia G, Tsang WW (2010) The ziggurat method for generating random variables. J Statist Softw 5(8):1–7.

McLeod AI (1994) Diagnostic checking of periodic autoregression models with applications. J Time Ser Anal 15(2):221–233

Proietti T, Hillebrand E (2017) Seasonal changes in Central England temperatures. J R Stat Soc Series A 180(3):769–791

Sarnaglia A, Reisen V, Lévy-Leduc C (2010) Robust estimation of periodic autoregressive processes in the presence of additive outliers. J Multivariate Anal 101(9):2168–2183

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6(2):461–464

Shao Q (2008) Robust estimation for periodic autoregressive time series. J Time Ser Anal 29(2):251–263

Thompstone R, Hipel K, McLeod A (1985) Grouping of periodic autoregressive models. In: Anderson O, Ord J, Robinson E (eds) Time Series Analysis: Theory and Practice 6. Proc. Conference Toronto 1983, North-Holland, Amsterdam, pp 35–49

Tsay RS (1986) Time series model specification in the presence of outliers. J Am Stat Assoc 81(393):132–141

Tsay RS (1988) Outliers, level shifts, and variance changes in time series. J Forecasting 7(1):1–20

Ursu E, Duchesne P (2009) On modelling and diagnostic checking of vector periodic autoregressive time series models. J Time Ser Anal 30(1):70–96

Ursu E, Pereau JC (2014) Robust modeling of periodic vector autoregressive time series. J Stat Plan Infer 155:93–106

Ursu E, Turkman KF (2012) Periodic autoregressive model identification using genetic algorithms. J Time Ser Anal 33(3):398–405

Vosseler A, Weber E (2018) Forecasting seasonal time series data: a Bayesian model averaging approach. Comput Stat 33(4):1733–1765

Wong H, Ip WC, Zhang R, Xia J (2007) Non-parametric time series models for hydrological forecasting. J Hydrol 332(3-4):337–347

Ye R, Dai Q (2018) A novel transfer learning framework for time series forecasting. Knowledge-Based Systems 156:74–99