



HAL
open science

M-regression spectral estimator for periodic ARMA models. An empirical investigation

Alessandro José Queiroz Sarnaglia, Valdério Anselmo Reisen, Pascal Bondon,
Céline Lévy-Leduc

► **To cite this version:**

Alessandro José Queiroz Sarnaglia, Valdério Anselmo Reisen, Pascal Bondon, Céline Lévy-Leduc. M-regression spectral estimator for periodic ARMA models. An empirical investigation. Stochastic Environmental Research and Risk Assessment, 2021, 35 (3), pp.653-664. 10.1007/s00477-020-01958-y. hal-03185742

HAL Id: hal-03185742

<https://centralesupelec.hal.science/hal-03185742>

Submitted on 30 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

M-regression spectral estimator for periodic ARMA models. An empirical investigation

Alessandro José Queiroz Sarnaglia^{1,2} · Valdério Anselmo Reisen^{1,2,3,4} ·
Pascal Bondon⁴ · Céline Lévy-Leduc⁵

Received: date / Accepted: date

Abstract The *M*-regression estimator has recently been widely used to build spectral estimators in time series models. In this paper, we extend this approach when the data follow a periodic autoregressive moving average (PARMA) process. We introduce an estimator of the parameters based on the classical Whittle estimator. The finite sample size performances of the proposed estimator are analyzed under the scenarios of PARMA processes with and without additive outliers (AO). Under the non-contaminated scenario, our estimator and the maximum Gaussian and Whittle likelihood estimators have similar behaviors. However, in the contaminated case, the two last estimators are severely biased, while the proposed estimator is robust. As a real data application, carbon monoxide (CO) concentrations are analyzed. A PARMA model is fitted and the data are forecasted with the model.

Keywords Periodic stationarity · PARMA models · Robust estimation · Outliers · Whittle estimation · CO concentration

1 Introduction

Stochastic processes exhibiting periodic correlation are frequently named as periodically correlated, periodically stationary (PS) or cyclostationary. Tiao and Grupe (1980) point out that periodic correlation may be neglected and misspecified as stationary seasonality when the standard time series tools are used. Since the introduction of PS processes in the literature by Gladyshev (1961), many authors have identified the periodic correlation phenomenon in time series of different areas, see e.g. Gardner and Franks (1975) and Bloomfield et al. (1994). Recent reviews on PS processes can be found, for instance, in Gardner et al. (2006) and Hurd and Miamer (2007).

Cyclostationary counterparts of standard parametric stationary models, e.g., ARMA processes, are obtained by authorizing the parameters to vary periodically in time. In this context, the PARMA framework is a natural candidate for parsimoniously fitting PS time series. Estimation methods for PARMA models have been investigated in the literature. For example, Lund and Basawa (2000) have considered the Gaussian maximum likelihood estimator (GMLE), Basawa and Lund (2001) have studied the least square method, and Sarnaglia et al. (2015) have proposed a Whittle likelihood estimator (WLE). These authors have shown that GMLE and WLE display similar performances in finite sample sizes, while WLE is faster to calculate than GMLE, mainly in large dimensional parameter spaces.

¹Department of Statistics, Federal University of Espírito Santo, Vitória, ES, Brazil.

²Post Graduate Program in Statistics, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil.

³PPGEA and PPGECON, UFES, Brazil.

⁴Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France.

⁵UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France.

Sarnaglia, A.J.Q.
E-mail: alessandro.sarnaglia@ufes.br

Reisen, V.A. (✉)
E-mail: valderio.reisen@ufes.br

Bondon, P.
E-mail: pascal.bondon@l2s.centralesupelec.fr

Lévy-Leduc, C.
E-mail: celine.levy-leduc@agroparistech.fr

The methodology proposed in this paper is not only motivated by its analytical and empirical features, but also from a real problem in the health and environmental fields. The concern about air pollution problems has increased considerably in the last 50 years. Especially in developing countries, the air quality has been degraded as a result of industrialization, population growth, high rates of urbanization, and inadequate or non-existent policies to control air pollution. The problems caused by air pollution produce local, regional and global impacts. Among different environmental pollution problems, air pollution is reported to cause the greatest damage to health and loss of quality of life from environmental causes.

Air quality control is accomplished by monitoring stations, which constitute the air quality monitoring network, where each station may monitor different pollutants according to the needs of the region where it is installed. Once data are collected from a monitoring system, they must be stored in data management systems and databases. Subsequently, the data must be retrieved and analyzed to see what they reveal about the effectiveness of regulatory standards, the accuracy of modeling, impacts on health endpoints, among others.

The records over time of particulate matter (PM), sulfur dioxide (SO₂) and CO, among other atmospheric pollutants, present some interesting features to investigate by time series tools, such as stationarity and non-stationarity (local trend, unit root), periodicity, asymmetry, weak and strong (long-memory) dependence. For example, He and Lu (2012) have studied the periodic variation of the concentrations of CO, carbon dioxide (CO₂) and PM at a typical traffic intersection in Hong Kong. In addition, the pollution series may present observations with high levels of concentrations producing sample densities with heavy tails. Besides the consequence of the atypical concentrations of the air pollutant on the environment and on the public health, these high levels of concentrations are not sequentially repetitive events, and can be identified as outliers which affect the statistical properties of sample functions, such as the standard mean and covariance. Since the estimation of time series models is connected with these sample functions, the estimated model can be strongly affected by the large peaks of the concentration. Therefore, it is important to develop statistical tools for a time series with these characteristics in order to describe the dynamics of the pollutants as well as to forecast future concentrations.

There are several types of outliers which cause different effects on the estimates. In general, the following three types are considered (Denby and Martin, 1979):

innovation outliers, which affect all subsequent observations; AO and replacement outliers, which have no effect on subsequent observations. AO affect the parameter estimates more than innovation outliers, and they have the same effect as replacement outliers (Ma and Genton, 2000). In the case of PS processes, the effect of AO was discussed by Sarnaglia et al. (2010). These authors have proposed a robust autocovariance function for PS processes which is used in the periodic Yule-Walker equations to provide robust estimates for PAR models. Shao (2008) has also suggested a robust estimation method for PAR models. Solci et al. (2020) have presented a comparative simulation study considering the robust methods proposed by Ma and Genton (2000), Shao (2008) and Sarnaglia et al. (2010) in the scenario of PAR processes with asymmetric errors and atypical observations. Recent studies in the environmental area related to the use of robust methods to estimate and forecast time series data are, for example, Ursu and Perea (2016), Sguera et al. (2016), Cabana et al. (2020), and references therein.

In the frequency domain, the periodogram is a powerful tool in the context of spectral analysis. The periodogram is a function of a time series sample and is strongly affected by outliers. One way to circumvent the outlier's effect is to use robust methods to compute alternative spectral estimators and these functions have been recently introduced as alternatives to the classical periodogram for stationary linear time series. It is well known that the periodogram is related to the least square estimator of the coefficients of a linear regression model with sine and cosine regressors, see e.g., Priestley (1981). Alternatively, several authors have defined M -periodogram by using the non-linear method of M -regression, see e.g. Li (2008) and Li (2010). In Reisen et al. (2017) and Fajardo et al. (2018), the authors have studied the M -periodogram for long-memory processes based on the M -regression approach discussed by Koul (1992). Reisen et al. (2019) have discussed this spectral estimator for the short-memory case such as ARMA processes. Therefore, based on the above references, the asymptotic theory related to M -regression spectral estimation method in the standard stationary univariate linear time series context is now fully established.

Here, we extend the M -regression method to PARMA models by introducing a multivariate M -periodogram spectral estimator in the Whittle likelihood function, denoted as WLE^M. The empirical performance of WLE^M is evaluated through Monte Carlo simulations. The results show that WLE^M behaves similarly to GMLE and WLE for uncontaminated time series. On the other hand, for contaminated time series with AO, GMLE and WLE are seriously affected, while WLE^M gives

close estimates to the ones in the uncontaminated data case. Although the asymptotic properties of WLE^M method are not derived here due to their level of complexities, the finite sample size investigation supports its use in practical problems.

As an application, we consider CO daily maximum concentrations. PARMA models are fitted to the data with WLE and WLE^M, respectively, and the models are used to forecast the data. In addition, we investigate whether possible atypical observations cause effects on the identification of the model and the estimation of its parameters. To the best of our knowledge, the M -regression estimation method for PARMA model and its application have not been discussed in the literature yet.

The rest of the paper is structured as follows: Section 2 describes the PARMA model with AO; Section 3 introduces the M -Whittle estimation method; The finite sample performance of the robust estimator is investigated through a Monte Carlo study in Section 4; An application of the methodology to CO daily mean concentrations is presented in Section 5.

2 PARMA model with AO

Let \mathbb{Z} be the set of integer numbers and (Z_t) , $t \in \mathbb{Z}$, be a real valued stochastic process satisfying $\mathbb{E}(Z_t^2) < \infty$ for all $t \in \mathbb{Z}$. Let $\mu_{Z,t} = \mathbb{E}(Z_t)$ and $\gamma_{Z,t}(\tau) = \text{Cov}(Z_t, Z_{t-\tau})$. We say that (Z_t) is PS with period \mathcal{S} (PS _{\mathcal{S}}) if for every $(t, \tau) \in \mathbb{Z}^2$,

$$\mu_{Z,t+\mathcal{S}} = \mu_{Z,t} \quad \text{and} \quad \gamma_{Z,t+\mathcal{S}}(\tau) = \gamma_{Z,t}(\tau), \quad (1)$$

and there are no smaller values of $\mathcal{S} > 0$ for which (1) holds. This definition implies that $\mu_{Z,t}$ and $\gamma_{Z,t}(\tau)$ are periodic functions in t and need to be known only for $t = 1, \dots, \mathcal{S}$. If (Z_t) is PS₁ then it is weakly stationary in the usual sense. In the following, we assume without loss of generality that $\mu_{Z,t} = 0$ for all $t \in \mathbb{Z}$, and we use the notation $t = (r-1)\mathcal{S} + \nu$ where $r \in \mathbb{Z}$ and the season $\nu = 1, \dots, \mathcal{S}$.

One of the most popular PS _{\mathcal{S}} process is the PARMA model which generalizes the ARMA model, see e.g. Vecchia (1985). (Z_t) is said to be a PARMA model if it satisfies the difference equation

$$\sum_{j=0}^{p_\nu} \phi_{\nu,j} Z_{(r-1)\mathcal{S}+\nu-j} = \sum_{k=0}^{q_\nu} \theta_{\nu,k} \varepsilon_{(r-1)\mathcal{S}+\nu-k}, \quad (2)$$

where, for each season ν , p_ν and q_ν are the AR and MA orders, respectively, $\phi_{\nu,1}, \dots, \phi_{\nu,p_\nu}$ and $\theta_{\nu,1}, \dots, \theta_{\nu,q_\nu}$ are the AR and MA coefficients, respectively, and $\phi_{\nu,0} = \theta_{\nu,0} = 1$. The innovations (ε_t) satisfy $\varepsilon_{(r-1)\mathcal{S}+\nu} = \sigma_\nu \varepsilon_{(r-1)\mathcal{S}+\nu}$ where (ε_t) is a sequence of zero-mean independent and

identically distributed random variables with unit variance. In the following, we set $p = \max_\nu p_\nu$, $q = \max_\nu q_\nu$, $\phi_{\nu,j} = 0$ for $j > p_\nu$, $\theta_{\nu,k} = 0$ for $k > q_\nu$, and we refer to (2) as the PARMA(p, q) _{\mathcal{S}} model.

Let (\mathbf{Z}_r) , $r \in \mathbb{Z}$, be the \mathcal{S} -variate time series defined by $\mathbf{Z}'_r = [Z_{(r-1)\mathcal{S}+1}, \dots, Z_{(r-1)\mathcal{S}+\mathcal{S}}]$, where \mathbf{Z}'_r denotes the transpose of \mathbf{Z}_r . It is well known that (Z_t) is PS _{\mathcal{S}} if and only if (\mathbf{Z}_r) is weakly stationary. The covariance matrix function of (\mathbf{Z}_r) is $\Gamma_{\mathbf{Z}}(\tau) = \text{Cov}(\mathbf{Z}_r, \mathbf{Z}_{r-\tau})$ and is related to $\gamma_{Z,t}(\tau)$ by $[\Gamma_{\mathbf{Z}}(\tau)]_{l,m} = \gamma_{Z,l}(\tau\mathcal{S} + l - m)$ for every $l, m = 1, \dots, \mathcal{S}$. Now, (2) is equivalent to the vector ARMA (VARMA) difference equation

$$\sum_{j=0}^P \Phi_j \mathbf{Z}_{r-j} = \sum_{k=0}^Q \Theta_k \varepsilon_{r-k}, \quad (3)$$

where $P = \lceil p/\mathcal{S} \rceil$, $Q = \lceil q/\mathcal{S} \rceil$ and $\lceil x \rceil$ denotes the smallest integer greater than or equal to x . The entries of matrix Φ_j are $[\Phi_j]_{l,m} = \phi_{l,j\mathcal{S}+l-m}$ with the convention that $[\Phi_0]_{l,m} = 0$ when $l < m$. The definition of Θ_k is similar. The white noise vector process (ε_r) is defined by $\varepsilon_r = [\varepsilon_{(r-1)\mathcal{S}+1}, \dots, \varepsilon_{(r-1)\mathcal{S}+\mathcal{S}}]'$ and has the covariance matrix $\Sigma_\varepsilon = \text{diag}(\sigma_1^2, \dots, \sigma_\mathcal{S}^2)$.

For all $z \in \mathbb{C}$, let

$$\Phi(z) = \sum_{j=0}^P \Phi_j z^j \quad \text{and} \quad \Theta(z) = \sum_{k=0}^Q \Theta_k z^k,$$

and assume that $\det \Phi(z)\Theta(z) \neq 0$ for $|z| \leq 1$. Therefore, (\mathbf{Z}_r) is causal and invertible and the spectral density matrix of (\mathbf{Z}_r) is

$$\mathbf{f}_{\mathbf{Z}}(\omega) = \frac{1}{2\pi} \Phi^{-1}(e^{-i\omega}) \Theta(e^{-i\omega}) \Sigma_\varepsilon \Theta'(e^{i\omega}) \Phi'^{-1}(e^{i\omega}),$$

where $\omega \in (-\pi, \pi]$. The causality and invertibility do not ensure that Σ_ε , $\Phi(z)$ and $\Theta(z)$ are uniquely determined by the covariance matrix function $\Gamma_{\mathbf{Z}}(\tau)$, or equivalently the spectral matrix $\mathbf{f}_{\mathbf{Z}}(\omega)$, see e.g. (Brockwell and Davis, 2006, page 431). Additional restrictions have to be imposed in order to obtain identifiable models, see Dunsmuir and Hannan (1976) and Deistler et al. (1978) for general conditions for VARMA models. Sarnaglia et al. (2020) have developed specific identifiability conditions for PARMA models. In the following, we assume that model (3) is identifiable.

Example 1 When $p_\nu = q_\nu = 1$ for all $\nu = 1, \dots, \mathcal{S}$, we have $p = q = 1$, $P = Q = 1$,

$$\Phi_0 = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ \phi_{2,1} & 1 & \cdots & \cdots & 0 \\ 0 & \phi_{3,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \phi_{\mathcal{S},1} & 1 \end{bmatrix}, \Phi_1 = \begin{bmatrix} 0 & \cdots & \cdots & \phi_{1,1} \\ 0 & \cdots & \cdots & 0 \\ \vdots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}$$

and

$$\det \Phi(z) = \det(\Phi_0 + \Phi_1 z) = 1 + (-1)^{\mathcal{S}+1} \left(\prod_{\nu=1}^{\mathcal{S}} \phi_{\nu,1} \right) z.$$

Similar results can be derived for Θ_0 , Θ_1 and $\det \Theta(z)$. The causality condition $\det \Phi(z) \neq 0$ for $|z| \leq 1$, and the invertibility condition $\det \Theta(z) \neq 0$ for $|z| \leq 1$ are equivalent respectively to

$$\vartheta_\phi = \prod_{\nu=1}^{\mathcal{S}} |\phi_{\nu,1}| < 1 \quad \text{and} \quad \vartheta_\theta = \prod_{\nu=1}^{\mathcal{S}} |\theta_{\nu,1}| < 1. \quad (4)$$

As discussed in Section 1, we shall consider AO since they cause the more harmful effect in the inference of time series. Sarnaglia et al. (2010) have shown theoretically and empirically, that AO can induce a spurious memory loss by increasing the variance of the process. Let (Y_t) be defined by

$$Y_t = Z_t + \eta W_t, \quad (5)$$

where $\eta > 0$ is the magnitude of the outlier and (W_t) is a sequence of independent and identically distributed random variables taking the values $-1, 0, 1$ with probabilities $\mathbb{P}(W_t = 0) = 1 - \zeta$ and $\mathbb{P}(W_t = -1) = \mathbb{P}(W_t = 1) = \zeta/2$, where $\zeta \in (0, 1)$ is the probability of occurrence of an outlier. We assume that (W_t) and (Z_t) are independent processes. The probabilities for positive and negative outliers are the same. If $\zeta = 0$ or $\eta = 0$, then $Y_t = Z_t$ and (Y_t) is uncontaminated. It is worth to point out that in model (5), the position (in time) of the outliers is random, which seems to be appropriate in real applications, since the position and the occurrence of the outliers in the sample is usually unknown. There are other ways to describe atypical observations. For example, one can consider heavy-tailed distributions for the white noise process (ε_t) , see e.g. Katkovnik (1998). However, in this kind of outlier generating mechanism, there is no explicit definition for the magnitude of the outliers and their investigation can not be performed directly.

The autocovariance function $\gamma_{Y,t}(\tau) = \text{Cov}(Y_t, Y_{t-\tau})$ of the contaminated process (Y_t) in (5) is given by

$$\gamma_{Y,t}(\tau) = \begin{cases} \gamma_{Z,t}(0) + \eta^2 \zeta, & \tau = 0, \\ \gamma_{Z,t}(\tau), & \tau \neq 0, \end{cases}$$

while $\mathbb{E}(Y_t) = \mu_{Y,t} = \mu_{Z,t}$. Therefore, (Y_t) is also a $\text{PS}_{\mathcal{S}}$ process but with larger variance than (Z_t) . Let $\mathbf{Y}'_r = [Y_{(r-1)\mathcal{S}+1}, \dots, Y_{(r-1)\mathcal{S}+\mathcal{S}}]$ and $\Gamma_{\mathbf{Y}}(\tau) = \text{Cov}(\mathbf{Y}_r, \mathbf{Y}_{r-\tau})$. Then

$$\Gamma_{\mathbf{Y}}(\tau) = \Gamma_{\mathbf{Z}}(\tau) + \mathbf{D} \mathbf{1}_{\{\tau=0\}}, \quad (6)$$

where \mathbf{D} is the $(\mathcal{S} \times \mathcal{S})$ diagonal matrix with diagonal entries $(\mathbf{D})_{l,l} = \eta^2 \zeta$, $l = 1, \dots, \mathcal{S}$. Therefore, the spectral density matrix of the contaminated vector process (\mathbf{Y}_r) is given by

$$\mathbf{f}_{\mathbf{Y}}(\omega) = \mathbf{f}_{\mathbf{Z}}(\omega) + \frac{1}{2\pi} \mathbf{D}, \quad \omega \in (-\pi, \pi].$$

Note that, letting $\eta \rightarrow \infty$ makes the diagonal matrix \mathbf{D} to dominate $\mathbf{f}_{\mathbf{Y}}$, which becomes close to the spectral density of a vector white noise process. This is the frequency domain counterpart of the memory loss property of processes with AO.

3 The Whittle M -estimator

We now introduce the estimation method proposed in this paper. Firstly, we define a robust alternative to the Fourier transform based on the non-linear M -regression approach. Next, the robust Whittle-type method is presented.

3.1 M -Fourier transform

Let (\mathbf{X}_r) , $r \in \mathbb{Z}$, be any \mathcal{S} -dimensional vector process and $\mathbf{X} = [\mathbf{X}'_1, \dots, \mathbf{X}'_N]'$ be a sample of size N observed from (\mathbf{X}_r) . The discrete Fourier transform of \mathbf{X} , at the frequency $\omega \in (-\pi, \pi]$, is defined by

$$\mathbf{W}_{\mathbf{X}}(\omega) = (2\pi N)^{-1/2} \sum_{r=1}^N \mathbf{X}_r e^{-ir\omega}. \quad (7)$$

Expression (7) can be fastly computed at the Fourier frequencies $\omega_j = 2\pi j/N$, $j = 1, \dots, N'$, where $N' = \lfloor (N-1)/2 \rfloor$ is the greatest integer smaller than or equal to $(N-1)/2$.

For any fixed frequency ω_j , the \mathcal{S} -dimensional vector $\mathbf{W}_{\mathbf{X}}(\omega_j)$ can be viewed as a linear regression vector as follows. Let $X_{r,\nu}$ be the ν th component of vector \mathbf{X}_r , $r = 1, \dots, N$, $\nu = 1, \dots, \mathcal{S}$. Define the vector of covariates $C_{r,j} = [\cos(r\omega_j), \sin(r\omega_j)]'$ and consider the linear model

$$X_{r,\nu} = C'_{r,j} \beta_\nu(\omega_j) + \xi_{r,\nu},$$

where $\xi_{r,\nu}$ is a random error term and the coefficient vector $\beta_\nu(\omega_j) = (\beta_{\nu,1}(\omega_j), \beta_{\nu,2}(\omega_j))'$ can be seen as describing the impact of the j th harmonic on the time series $X_{1,\nu}, \dots, X_{N,\nu}$. The classical least square estimator of the vector $\beta_\nu(\omega_j)$ is given by

$$\hat{\beta}_\nu(\omega_j) = \operatorname{argmin}_{\beta_\nu(\omega_j) \in \mathbb{R}^2} \left[\sum_{r=1}^N (X_{r,\nu} - C'_{r,j} \beta_\nu(\omega_j))^2 \right]. \quad (8)$$

Now, define the vector $\mathbf{d}(\omega_j) = [d_1(\omega_j), \dots, d_S(\omega_j)]'$, where $d_\nu(\omega_j) = \hat{\beta}_{\nu,1}(\omega_j) - i\hat{\beta}_{\nu,2}(\omega_j)$. It can be shown that

$$\mathbf{W}_{\mathbf{X}}(\omega_j) = \sqrt{N/8\pi} \mathbf{d}(\omega_j), \quad j = 1, \dots, N'.$$

It is well known that $\hat{\beta}_\nu(\omega_j)$ is not robust to atypical observations and heavy-tailed distributions. To improve robustness, one idea is to replace $\hat{\beta}_\nu(\omega_j)$ by a non-linear M -regression estimator in $\mathbf{W}_{\mathbf{X}}(\omega_j)$. This will lead to the robust periodogram for PS processes proposed here. The key idea is to replace the quadratic loss function in (8) by an alternative function $\rho(\cdot)$, which gives

$$\operatorname{argmin}_{\beta_\nu(\omega_j) \in \mathbb{R}^2} \left[\sum_{r=1}^N \rho(X_{r,\nu} - C'_{r,j} \beta_\nu(\omega_j)) \right].$$

Equivalently, one can define the M -estimator $\hat{\beta}_{\nu,\psi}(\omega_j)$ of $\beta_\nu(\omega_j)$ as the solution of

$$\sum_{r=1}^N C_{r,j} \psi(X_{r,\nu} - C'_{r,j} \hat{\beta}_{\nu,\psi}(\omega_j)) = 0,$$

where ψ is the first derivative of the Huber function given by

$$\rho(x) = \begin{cases} x^2/2, & |x| \leq \delta, \\ \delta(|x| - \delta/2), & |x| > \delta, \end{cases} \quad (9)$$

see Huber (1964). The choice of the tuning parameter $\delta > 0$ is quite important and provides the compromise between robustness and efficiency of the M -estimator. Larger values of δ would lead to the Huber loss closer to the square loss, which would lead to closer performances between WLE and WLE^M. On the other hand, small value of δ would improve robustness, but at the price of efficiency loss in the uncontaminated Gaussian processes. This conclusion is corroborated by other authors, see e.g. Maronna et al. (2019). Following the usual choice given in the literature, the tuning parameter is set to 1.345, which ensures efficiency of 95% of the regression estimator in the Gaussian case.

Finally, by defining $\mathbf{d}_\psi(\omega_j)$ similarly to $\mathbf{d}(\omega_j)$ with $\hat{\beta}_\nu(\omega_j)$ replaced by $\hat{\beta}_{\nu,\psi}(\omega_j)$, the robust alternative to $\mathbf{W}_{\mathbf{X}}(\omega_j)$ proposed here is given by

$$\mathbf{W}_{\mathbf{X},\psi}(\omega_j) = \sqrt{N/8\pi} \mathbf{d}_\psi(\omega_j), \quad j = 1, \dots, N'. \quad (10)$$

3.2 Whittle M -estimator of PARMA parameters

Assume that Z_1, \dots, Z_n is a sample from a PARMA process with known orders p and q . For simplicity, suppose that $n = NS$, such that every season $\nu = 1, \dots, S$ is observed N times. As previously, define the vector $\mathbf{Z}_r = [Z_{(r-1)S+1}, \dots, Z_{(r-1)S+S}]'$ corresponding to the r th cycle, the full sample being given by $\mathbf{Z} = [\mathbf{Z}'_1, \dots, \mathbf{Z}'_N]'$. The parameter vector of model (2) is $\varphi = (\varphi'_\phi, \varphi'_\theta, \varphi'_\sigma)'$ where φ_ϕ and φ_θ contain all the AR and MA parameters, respectively, and $\varphi_\sigma = (\sigma_1^2, \dots, \sigma_S^2)'$.

We define the parameter space $\mathcal{P} \subset \mathbb{R}^{(p+q)S}$ as the set of points $(\varphi'_\phi, \varphi'_\theta)'$ for which model (2) is identifiable in the sense of Deistler et al. (1978). We denote by $\mathbb{R}_{>0}$ the set of positive real numbers. For any $\varphi \in \mathcal{P} \times \mathbb{R}_{>0}^S$, let $\Gamma_N(\varphi)$ be the $NS \times NS$ matrix with $\Gamma_{\mathbf{Z}}(m-l)$ in the (l, m) th block of $S \times S$ elements, $1 \leq l, m \leq N$. The Gaussian log-likelihood multiplied with $-2N^{-1}$ is given by

$$\hat{\mathcal{L}}_N(\varphi) = N^{-1} \log \det \Gamma_N(\varphi) + N^{-1} \mathbf{Z}' \Gamma_N^{-1}(\varphi) \mathbf{Z}.$$

We denote by φ_0 the true parameter vector φ from which the sample Z_1, \dots, Z_n is generated. We assume that $(\varphi'_{\phi_0}, \varphi'_{\theta_0})' \in \mathcal{P}$, and we have $\Gamma_N(\varphi_0) = \operatorname{Cov}(\mathbf{Z}, \mathbf{Z})$. The GMLE of φ_0 is $\hat{\varphi}_N = \operatorname{argmin}_{\varphi \in \mathcal{P} \times \mathbb{R}_{>0}^S} \hat{\mathcal{L}}_N(\varphi)$.

To obtain $\hat{\varphi}_N$, an optimization algorithm is used and can demand high computational effort due to the fact that $\Gamma_N(\varphi)$ has to be inverted. To circumvent this difficulty, we use the multivariate version of Whittle's methodology to approximate $\hat{\mathcal{L}}_N(\varphi)$, see Dunsmuir and Hannan (1976). For a PARMA process, it was shown by Sarnaglia et al. (2015) that the corresponding Whittle likelihood estimator (WLE) of φ_0 is $\tilde{\varphi}_N = (\tilde{\varphi}'_{\phi_N}, \tilde{\varphi}'_{\theta_N}, \tilde{\varphi}'_{\sigma_N})'$ where

$$(\tilde{\varphi}'_{\phi_N}, \tilde{\varphi}'_{\theta_N})' = \operatorname{argmin}_{(\varphi'_\phi, \varphi'_\theta)' \in \mathcal{P}} \sum_{\nu=1}^S \log \tilde{\sigma}_{N,\nu}^2(\varphi_\phi, \varphi_\theta),$$

$$\tilde{\sigma}_{N,\nu}^2(\varphi_\phi, \varphi_\theta) =$$

$$\frac{1}{2\pi N'} \sum_{j=1}^{N'} |(\Theta^{-1}(e^{-i\omega_j}) \Phi(e^{-i\omega_j}) \mathbf{W}_{\mathbf{Z}}(\omega_j))_\nu|^2,$$

$$\nu = 1, \dots, S \quad (11)$$

and $\mathbf{W}_{\mathbf{Z}}(\omega_j)$ is given by (7) in which \mathbf{X}_r is replaced by \mathbf{Z}_r , and the ν th component of $\tilde{\varphi}_{\sigma_N}$ is $\tilde{\sigma}_{N,\nu}^2(\tilde{\varphi}_{\phi_N}, \tilde{\varphi}_{\theta_N})$ for $\nu = 1, \dots, S$.

The Whittle M -estimator (WLE^M) of φ_0 , $\tilde{\varphi}_{N,\psi}$, is obtained by replacing $\mathbf{W}_{\mathbf{Z}}(\omega_j)$ in (11) by $\mathbf{W}_{\mathbf{Z},\psi}(\omega_j)$ defined in (10) in which \mathbf{X}_r is replaced by \mathbf{Z}_r .

It was pointed out by Sarnaglia et al. (2015) that $(\varphi'_\phi, \varphi'_\theta)$ involves $(p+q)S$ parameters whereas the dimension of φ is $(p+q+1)S$. Then, $\tilde{\varphi}_N$ is easier to

calculate and is obtained faster than $\hat{\varphi}_N$. The same remark applies to the calculation of $\tilde{\varphi}_{N,\psi}$. However, the computation time of $\tilde{\varphi}_{N,\psi}$ may be larger than $\tilde{\varphi}_N$ because a numerical optimization method is needed to obtain $\mathbf{W}_{\mathbf{Z},\psi}(\omega_j)$ since this function does not have a closed form expression.

4 Numerical experiments

In this section, we compare the finite sample behaviors of WLE^M , GMLE and WLE. Samples of the processes (Z_t) and (Y_t) were generated according to (2) and (5), respectively, with $p = 1$, $q \in \{0, 1\}$, $\mathcal{S} = 2$, $\zeta = 0.01$ and $\eta = 10$. The parameter values are given in Table 1. The sample sizes are $n = NS = 300, 800$, and the Huber function (9) is used with $\delta = 1.345$. For each model, 200 sample paths were generated. Other PARMA models were also considered and the results lead to similar conclusions.

We evaluate the finite sample performances of the estimators by computing their sample root mean square errors (RMSE), when the random variables (ϵ_t) are Gaussian and non-Gaussian in (2). The results are displayed in Tables 2, 3, 4 and 5 when (ϵ_t) are Gaussian; in Tables 6, 7, 8 and 9 when (ϵ_t) follow a standardized Student's t distribution with 5 degrees of freedom; in Tables 10, 11, 12 and 13 when (ϵ_t) follow a Laplace distribution with location 0 and scale $\sqrt{2}$. The values with “*” refer to the sample RMSE for the contaminated series.

In the case of uncontaminated data, the three estimators present very similar sample RMSE in all cases. This quantity decreases as the sample size increases, and increases with the orders of the models, which are expected results. Relating to the estimation of the variance of the innovations, GMLE and WLE are more precise than WLE^M when the data are Gaussian (Tables 2 to 5), and this is also an expected result. But, when the data are not Gaussian (Tables 6 to 13), the RMSE of the estimation of the variance of the innovations are similar for the three estimators.

In the case of contaminated data with AO, the sample RMSE of GMLE and WLE increase substantially, which means that these estimators are totally corrupted by the atypical observations, while WLE^M displays similar sample RMSE values as in non-contaminated scenario. WLE^M presents generally accurate estimates, even for a large number of outliers. Its superiority over GMLE and WLE appears clearly in the tables for Gaussian and non-Gaussian innovations. This empirical study indicates that GMLE and WLE should be avoided when the series contains AO, and WLE^M is an alternative estimator to deal with a periodic time series with and without

possible AO. These results are similar to those obtained with the M -spectral estimator in the context of standard stationary time series with short and long-range dependence discussed by Reisen et al. (2019), Reisen et al. (2017) and Fajardo et al. (2018).

5 Analysis of air pollution data

We analyze the daily maximum concentrations of CO collected at the station located in Enseada do Suá area which belongs to the automatic air quality monitoring network of the Greater Vitória Region, ES, Brazil. This monitoring network has eight monitoring stations which are located at urban sites of four cities and monitors PM_{10} concentrations, total suspended particles (TSP), ozone (O_3), nitrogen oxides (NO_x), CO, hydrocarbons (HC) and meteorological variables.

The data consist of $n = NS = 728$ observations measured from January 3, 2010 to December 31, 2011, and 31 observations from January 1, 2012 to January 31, 2012. The first 728 data were used to fit the model, and the last 31 records were used to perform the forecast study. The sample mean and standard deviation of the first 728 data are $2501.35 \mu\text{g}/\text{m}^3$ and $869.41 \mu\text{g}/\text{m}^3$, respectively. The plot of the data over time and the box-plots for each day of the week are displayed in Figures 1 and 2, respectively. These figures show that the series displays a weekly pattern (seasonality) without shifts or “abrupt” changes in the mean and long-term trends. Figure 2 shows that the data are generally symmetric for each day of the week.

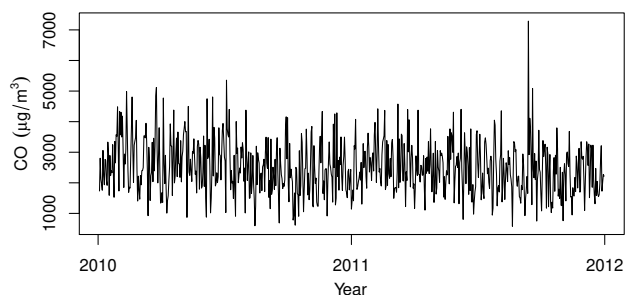


Fig. 1 Daily maximum concentrations of CO concentrations in Vitória, ES, Brazil.

We observe a high level of CO on Tuesday, September 13, 2011 (619th observation), which is approximately 5 standard deviations from the sample mean. Although, high levels of pollutants are an important information that should be considered in the context of air pollution and health impact, these observations may be identi-

Table 1 Parameters.

Model	$\nu = 1$			$\nu = 2$			Eq. (4)	
	$\phi_{1,1_0}$	$\theta_{1,1_0}$	$\sigma_{1_0}^2$	$\phi_{2,1_0}$	$\theta_{2,1_0}$	$\sigma_{2_0}^2$	ϑ_ϕ	ϑ_θ
1	-0.2	0.0	1.0	-0.5	0.0	1.0	0.1	0.0
2	-0.2	-0.5	1.0	-0.5	-0.2	1.0	0.1	0.1
3	-1.0	0.0	1.0	-0.5	0.0	1.0	0.5	0.0
4	-1.0	-0.5	1.0	-0.5	-0.2	1.0	0.5	0.1

Table 2 RMSE of the estimates for Model 1 with Gaussian innovations.

Method	n	$\phi_{1,1}$	σ_1^2	$\phi_{2,1}$	σ_1^2
GMLE	300	0.067; 0.121*	0.117; 1.366*	0.079; 0.252*	0.111; 1.363*
	800	0.048; 0.101*	0.079; 1.122*	0.046; 0.239*	0.074; 1.253*
WLE	300	0.068; 0.121*	0.117; 1.368*	0.079; 0.252*	0.111; 1.364*
	800	0.048; 0.101*	0.079; 1.122*	0.046; 0.239*	0.074; 1.253*
WLE ^M	300	0.067; 0.067*	0.147; 0.179*	0.083; 0.089*	0.147; 0.189*
	800	0.051; 0.054*	0.118; 0.149*	0.051; 0.058*	0.108; 0.152*

Table 3 RMSE of the estimates for Model 2 with Gaussian innovations.

Method	n	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{1,1}$	σ_1^2
GMLE	300	0.364; 1.393*	0.371; 1.398*	0.120; 1.433*	0.638; 2.219*	0.649; 2.249*	0.114; 1.257*
	800	0.171; 0.492*	0.184; 0.500*	0.065; 1.118*	0.167; 1.007*	0.171; 1.031*	0.064; 1.102*
WLE	300	0.369; 0.981*	0.377; 0.985*	0.119; 1.433*	0.545; 1.309*	0.559; 1.340*	0.114; 1.255*
	800	0.171; 0.439*	0.184; 0.448*	0.065; 1.119*	0.167; 0.466*	0.171; 0.504*	0.064; 1.102*
WLE ^M	300	0.357; 0.344*	0.371; 0.363*	0.135; 0.164*	0.747; 0.704*	0.760; 0.714*	0.150; 0.180*
	800	0.178; 0.186*	0.193; 0.201*	0.101; 0.132*	0.189; 0.200*	0.194; 0.206*	0.101; 0.131*

Table 4 RMSE of the estimates for Model 3 with Gaussian innovations.

Method	n	$\phi_{1,1}$	σ_1^2	$\phi_{2,1}$	σ_1^2
GMLE	300	0.064; 0.394*	0.107; 1.770*	0.047; 0.159*	0.107; 1.327*
	800	0.036; 0.373*	0.069; 1.642*	0.030; 0.142*	0.070; 1.279*
WLE	300	0.066; 0.397*	0.116; 1.780*	0.047; 0.159*	0.107; 1.328*
	800	0.037; 0.374*	0.073; 1.646*	0.030; 0.142*	0.070; 1.279*
WLE ^M	300	0.080; 0.107*	0.193; 0.334*	0.058; 0.067*	0.156; 0.227*
	800	0.053; 0.077*	0.168; 0.317*	0.038; 0.047*	0.134; 0.215*

Table 5 RMSE of the estimates for Model 4 with Gaussian innovations.

Method	n	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{1,1}$	σ_1^2
GMLE	300	0.233; 1.539*	0.252; 1.594*	0.118; 1.307*	0.124; 0.319*	0.150; 0.350*	0.107; 1.337*
	800	0.140; 0.278*	0.146; 0.376*	0.068; 1.230*	0.077; 0.134*	0.089; 0.181*	0.071; 1.141*
WLE	300	0.236; 0.792*	0.255; 0.864*	0.121; 1.313*	0.125; 0.238*	0.150; 0.280*	0.107; 1.336*
	800	0.141; 0.276*	0.147; 0.372*	0.068; 1.233*	0.077; 0.133*	0.089; 0.180*	0.071; 1.141*
WLE ^M	300	0.272; 0.296*	0.288; 0.311*	0.153; 0.207*	0.134; 0.140*	0.148; 0.152*	0.141; 0.183*
	800	0.150; 0.149*	0.155; 0.154*	0.112; 0.169*	0.082; 0.085*	0.094; 0.097*	0.107; 0.145*

Table 6 RMSE of Model 1 with Student's t innovations.

Method	n	$\phi_{1,1}$	σ_1^2	$\phi_{2,1}$	σ_2^2
GMLE	300	0.0730; 0.1242*	0.176; 1.296*	0.0816; 0.2505*	0.215; 1.493*
	800	0.0444; 0.0960*	0.124; 1.110*	0.0491; 0.2480*	0.125; 1.197*
WLE	300	0.0727; 0.1235*	0.176; 1.297*	0.0815; 0.2505*	0.215; 1.494*
	800	0.0445; 0.0961*	0.124; 1.111*	0.0495; 0.2481*	0.125; 1.197*
WLE ^M	300	0.0729; 0.0732*	0.206; 0.185*	0.0896; 0.0898*	0.191; 0.183*
	800	0.0450; 0.0454*	0.198; 0.171*	0.0573; 0.0570*	0.145; 0.115*

fied, from a statistical point of view, as being AO if they are not sequentially repetitive events.

To verify if the 619th observation can be identified as an AO or not, a new series was generated from the original data by replacing this observation by the Hu-

Table 7 RMSE of Model 2 with Student's t innovations.

Method	n	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
GMLE	300	0.291; 1.425*	0.305; 1.431*	0.181; 1.256*	0.617; 3.456*	0.625; 3.502*	0.204; 1.437*
	800	0.154; 0.426*	0.164; 0.461*	0.135; 1.202*	0.158; 1.431*	0.163; 1.443*	0.139; 1.118*
WLE	300	0.225; 0.256*	0.240; 0.286*	0.181; 1.263*	0.262; 0.241*	0.280; 0.294*	0.203; 1.453*
	800	0.133; 0.150*	0.141; 0.192*	0.135; 1.205*	0.137; 0.170*	0.144; 0.184*	0.139; 1.121*
WLE ^M	300	0.223; 0.231*	0.237; 0.248*	0.204; 0.181*	0.255; 0.256*	0.270; 0.270*	0.217; 0.187*
	800	0.145; 0.148*	0.154; 0.155*	0.180; 0.150*	0.145; 0.155*	0.149; 0.159*	0.183; 0.154*

Table 8 RMSE of Model 3 with Student's t innovations.

Method	n	$\phi_{1,1}$	σ_1^2	$\phi_{2,1}$	σ_2^2
GMLE	300	0.0640; 0.3882*	0.2213; 1.8123*	0.0502; 0.1705*	0.194; 1.386*
	800	0.0395; 0.3895*	0.1160; 1.6595*	0.0322; 0.1398*	0.127; 1.314*
WLE	300	0.0659; 0.3915*	0.2208; 1.8239*	0.0501; 0.1705*	0.194; 1.382*
	800	0.0406; 0.3911*	0.1188; 1.6650*	0.0321; 0.1398*	0.127; 1.314*
WLE ^M	300	0.0812; 0.1013*	0.1717; 0.2800*	0.0582; 0.0697*	0.170; 0.176*
	800	0.0495; 0.0702*	0.0932; 0.1882*	0.0358; 0.0421*	0.118; 0.110*

Table 9 RMSE of Model 4 with Student's t innovations.

Method	n	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
GMLE	300	0.227; 1.102*	0.230; 1.154*	0.196; 1.247*	0.124; 0.208*	0.151; 0.244*	0.194; 1.407*
	800	0.133; 0.263*	0.139; 0.366*	0.143; 1.197*	0.073; 0.119*	0.088; 0.158*	0.167; 1.151*
WLE	300	0.216; 0.216*	0.218; 0.269*	0.196; 1.251*	0.125; 0.161*	0.154; 0.215*	0.194; 1.411*
	800	0.125; 0.140*	0.130; 0.221*	0.143; 1.198*	0.072; 0.093*	0.088; 0.153*	0.167; 1.152*
WLE ^M	300	0.217; 0.221*	0.227; 0.228*	0.181; 0.165*	0.120; 0.127*	0.145; 0.151*	0.198; 0.174*
	800	0.127; 0.132*	0.133; 0.138*	0.129; 0.101*	0.071; 0.073*	0.087; 0.087*	0.160; 0.129*

Table 10 RMSE of Model 1 with Laplace innovations.

Method	n	$\phi_{1,1}$	σ_1^2	$\phi_{2,1}$	σ_2^2
GMLE	300	0.0725; 0.1149*	0.200; 1.334*	0.0746; 0.2470*	0.182; 1.309*
	800	0.0431; 0.0940*	0.114; 1.151*	0.0535; 0.2493*	0.100; 1.237*
WLE	300	0.0728; 0.1155*	0.200; 1.335*	0.0747; 0.2468*	0.182; 1.309*
	800	0.0432; 0.0940*	0.115; 1.151*	0.0535; 0.2494*	0.100; 1.237*
WLE ^M	300	0.0715; 0.0701*	0.294; 0.266*	0.0894; 0.0888*	0.222; 0.199*
	800	0.0418; 0.0398*	0.272; 0.241*	0.0647; 0.0599*	0.190; 0.155*

Table 11 RMSE of Model 2 with Laplace innovations.

Method	n	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
GMLE	300	0.319; 1.311*	0.327; 1.297*	0.190; 1.350*	2.087; 3.162*	2.088; 3.192*	0.166; 1.377*
	800	0.167; 0.699*	0.177; 0.740*	0.113; 1.173*	0.162; 1.409*	0.161; 1.401*	0.110; 1.125*
WLE	300	0.247; 0.224*	0.252; 0.263*	0.190; 1.353*	0.247; 0.260*	0.271; 0.307*	0.165; 1.397*
	800	0.145; 0.153*	0.154; 0.201*	0.113; 1.175*	0.145; 0.154*	0.146; 0.178*	0.110; 1.128*
WLE ^M	300	0.246; 0.247*	0.259; 0.260*	0.288; 0.264*	0.268; 0.272*	0.283; 0.285*	0.263; 0.233*
	800	0.137; 0.139*	0.146; 0.149*	0.244; 0.211*	0.148; 0.155*	0.147; 0.155*	0.232; 0.200*

Table 12 RMSE of Model 3 with Laplace innovations.

Method	n	$\phi_{1,1}$	σ_1^2	$\phi_{2,1}$	σ_2^2
GMLE	300	0.0615; 0.3880*	0.178; 1.696*	0.0499; 0.1661*	0.172; 1.405*
	800	0.0392; 0.3823*	0.104; 1.626*	0.0304; 0.1423*	0.102; 1.286*
WLE	300	0.0642; 0.3913*	0.187; 1.715*	0.0498; 0.1664*	0.172; 1.403*
	800	0.0391; 0.3838*	0.105; 1.632*	0.0305; 0.1423*	0.103; 1.286*
WLE ^M	300	0.0793; 0.1003*	0.189; 0.297*	0.0518; 0.0622*	0.191; 0.174*
	800	0.0477; 0.0705*	0.104; 0.200*	0.0350; 0.0430*	0.138; 0.106*

ber location M -estimator of the corresponding season ($\nu = 3$, Tuesday). The sample autocovariance functions

Table 13 RMSE of Model 4 with Laplace innovations.

Method	n	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
GMLE	300	0.203; 2.041*	0.215; 2.080*	0.170; 1.468*	0.124; 0.634*	0.147; 0.644*	0.168; 1.273*
	800	0.131; 0.286*	0.133; 0.381*	0.102; 1.280*	0.072; 0.123*	0.085; 0.176*	0.116; 1.164*
WLE	300	0.192; 0.205*	0.200; 0.260*	0.171; 1.475*	0.123; 0.186*	0.147; 0.230*	0.168; 1.281*
	800	0.125; 0.142*	0.126; 0.221*	0.102; 1.283*	0.071; 0.098*	0.085; 0.173*	0.116; 1.165*
WLE ^M	300	0.201; 0.204*	0.215; 0.218*	0.196; 0.191*	0.124; 0.130*	0.147; 0.151*	0.250; 0.217*
	800	0.127; 0.131*	0.134; 0.132*	0.154; 0.121*	0.069; 0.069*	0.082; 0.081*	0.221; 0.187*

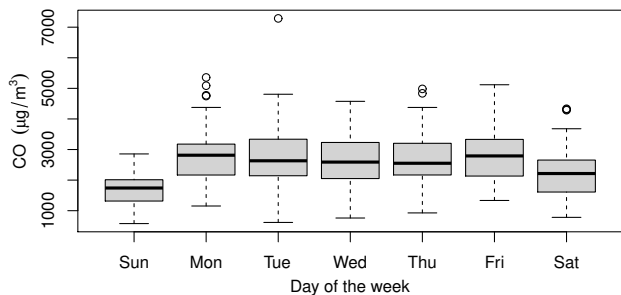


Fig. 2 Daily maximum concentrations of CO by day of the week.

(ACFs) of the original and modified series were computed for each season (day of the week), and are displayed in Table 14. These values reveal that the 619th observation has some impact on the ACFs, especially at lag 1 for the seasonal periods $\nu = 3, 4$. In these cases, the original series has smaller ACF values than the modified series. This means that an AO generates a memory-loss, see Fajardo et al. (2009) and Sarnaglia et al. (2010). The ACF at the other periods are the same. This simple empirical investigation supports the idea that the largest observation may be considered as an AO.

Table 14 ACFs of the original and modified series.

ν	Original			Modified		
	lag 1	lag 2	lag 3	lag 1	lag 2	lag 3
1	0.130	0.141	-0.048	0.130	0.141	-0.048
2	0.045	0.178	0.181	0.045	0.178	0.181
3	0.206	0.138	0.105	0.313	0.096	0.153
4	0.205	0.110	0.052	0.219	0.110	0.052
5	0.238	0.127	0.029	0.238	0.046	0.029
6	0.554	0.267	0.151	0.554	0.267	0.108
7	0.314	0.184	0.160	0.314	0.184	0.160

Now we investigate whether this observation has an impact on the identification and estimation of the model. The WLE was calculated from the centered original data obtained by subtracting the sample mean of each period. The WLE^M was calculated from the centered original data obtained by subtracting the Huber location M -estimator of each period, and taking

$\delta = 1.345$ in (9). The modified WLE (MWLE) was obtained by calculating the WLE with the centered modified data got by subtracting the modified sample mean of each period. As part of the modeling strategy, the values of the Bayesian information criterion (BIC) were computed using each estimator. The GMLE was not considered in this study since GMLE and WLE have similar behaviors, see Section 4.

For WLE, BIC was computed by

$$\text{BIC} = N \sum_{\nu=1}^S \log(\tilde{\sigma}_{N,\nu}^2) + \log(N) \sum_{\nu=1}^S (p_\nu + q_\nu), \quad (12)$$

where, for each $\nu = 1, \dots, S$, $\tilde{\sigma}_{N,\nu}^2$ is the residual variance. For WLE^M, BIC was obtained by replacing $\tilde{\sigma}_{N,\nu}^2$ by $\tilde{\sigma}_{N,\nu,\psi}^2$ in (12) and is denoted by BIC^M. In the case of MWLE, BIC was computed from (12), where $\tilde{\sigma}_{N,\nu}^2$ was replaced by the residual variance obtained with the modified series, and is denoted by MBIC. For the three estimators, the initial values for AR and MA estimates were set as the AR and MA coefficients estimated from a stationary ARMA(p, q) model with $p = \max_\nu p_\nu$ and $q = \max_\nu q_\nu$, respectively, and for white noise variances set as $S_{X,\nu}^2$, which is the data sample variance of the period ν . Identifiable models were fitted using the conditions proposed by Sarnaglia et al. (2020). BIC selected a PAR model with orders $(p_1, \dots, p_S) = (0, 0, 0, 0, 1, 1, 1)$, while BIC^M and MBIC selected a PAR model with orders $(p_1, \dots, p_S) = (0, 0, 1, 1, 1, 1, 1)$. The parameters of these models and the (in sample) RMSE and median absolute deviation (MAD) between the predicted values calculated from the model and the actual values, are displayed in Table 15.

First, we observe that BIC suggested a smaller model order than MBIC. This is in agreement with the results in Table 14, and was also noticed by Sarnaglia et al. (2010) who discussed the reduction of the order of PAR models with AO when using Akaike information criterion and BIC. Since BIC has selected a different model than BIC^M and MBIC, the WLE is not expected to have similar values to WLE^M and MWLE, as seen in Table 15, especially for $\nu = 3$, the season in which the atypical observation is present, and for $\nu = 4$. Notice that even though only one observation is suspected to

Table 15 WLE, WLE^M and MWLE for each season ν .

ν	WLE		WLE ^M		MWLE		$S_{X,\nu}$
	$\phi_{\nu,1}$	σ_ν	$\phi_{\nu,1}$	σ_ν	$\phi_{\nu,1}$	σ_ν	
1	-	475.43	-	503.20	-	475.43	475.43
2	-	835.60	-	830.62	-	835.60	835.61
3	-	901.63	-0.313	767.41	-0.293	742.58	901.63
4	-	781.63	-0.246	770.44	-0.219	762.73	781.63
5	-0.244	775.66	-0.296	794.00	-0.243	775.66	798.65
6	-0.581	696.79	-0.550	769.66	-0.581	696.79	837.09
7	-0.278	701.43	-0.274	695.97	-0.278	701.43	738.88
RMSE	749.03		744.71		720.77		869.41
MAD	482.31		471.95		469.32		583.98

be atypical, it has an impact on the estimation of the parameters. Finally, WLE^M and MWLE gave similar results, which corroborates the robustness of WLE^M.

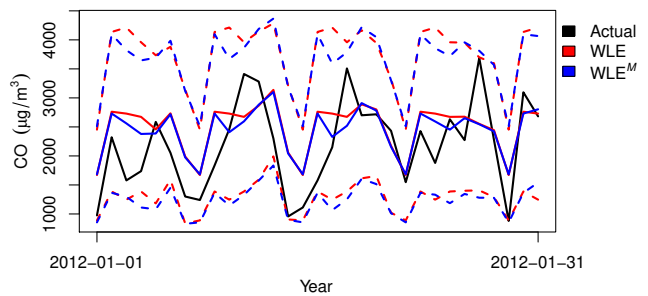
The next step is to investigate whether the aberrant observation has impact on the forecasts. Figure 3 displays the one-step-ahead forecasts and their respective 90% forecast intervals, and Table 16 presents the RMSE and MAD for h -step-ahead forecasts for the models fitted by WLE and WLE^M. The one-step-ahead 90% forecast intervals were calculated as $\hat{Z}_{(r-1)S+\nu-1}(1) \pm 1.64\sigma_\nu$, where $\hat{Z}_{t-1}(1) = E(Z_t|Z_{t-1}, Z_{t-2}, \dots)$ denotes the one-step-ahead point forecast of Z_t . Deviations from normality could be addressed by forecast intervals calculated with the methods addressed in Sarnaglia et al. (2018). Similar conclusions related to the impact of the outlier in the confidence intervals are expected. We observe that the model fitted by WLE^M provides more accurate forecasts across the h -step-ahead, $h = 1, \dots, 7$, since both RMSE and MAD are smaller using this robust approach.

Table 16 RMSE and MAD for h -step-ahead forecasts of the CO concentrations of January 2012.

h	RMSE		MAD	
	WLE	WLE ^M	WLE	WLE ^M
1	660.29	630.70	559.26	442.78
2	661.70	655.85	569.76	547.25
3	662.41	652.47	559.26	539.53
4	638.23	628.78	526.07	492.95
5	624.54	616.44	492.88	457.57
6	636.13	628.26	526.07	498.55
7	631.67	625.05	492.88	457.57

6 Conclusions

This paper proposes to use the M -regression method to estimate PARMA models in the frequency domain. Contrarily to the usual GMLE and WLE, the proposed spectral estimator is robust against AO. Moreover, it

**Fig. 3** Actual observations (solid black line), one-step-ahead point forecasts (solid line), and 90% forecast intervals (dashed lines) of the CO concentrations of January 2012.

has the same finite sample performance as GMLE and WLE in the absence of outliers. Therefore, when the presence of outliers is suspected, the robust spectral estimator should be preferred. As an application to air pollution data, daily maximum CO concentrations are analyzed. The results strongly suggest that the high levels of CO concentration lead to similar features as AO do. Fitting a PARMA model to these data using the robust estimator provides also more accurate multistep predictors.

Acknowledgements The authors would like to thank the agencies CNPq, CAPES and FAVES, Brazil and CNRS and CentraleSupélec, France for their financial support. This research was also partially supported by DATAIA Convergence Institute as part of the “Programme d’Investissement d’Avenir, (ANR17-CONV-0003) operated by CentraleSupélec”, and by the iCODE Institute, research project of the IDEX Paris-Saclay, and by the Hadamard Mathematics LabEx (LMH) through the grant number ANR-11-LABX-0056-LMH in the “Programme d’Investissement d’Avenir”. The authors also thank the referees and the editor for their valuable suggestions which have contributed significantly for the improvement of the paper.

References

Basawa IV, Lund RB (2001) Large sample properties of parameter estimates for periodic ARMA models.

- Journal of Time Series Analysis 22:651–663
- Bloomfield P, Hurd HL, Lund RB (1994) Periodic correlation in stratospheric ozone data. *Journal of Time Series Analysis* 15:127–150
- Brockwell PJ, Davis RA (2006) *Time Series: Theory and Methods*, 2nd edn. Springer Series in Statistics, Springer Science, New York, NY
- Cabana E, Lillo RE, Laniado H (2020) Robust regression based on shrinkage with application to living Environment Deprivation. *Stochastic environmental research and risk assessment* pp 1–18
- Deistler M, Dunsmuir W, Hannan EJ (1978) Vector linear time series models: Corrections and extensions. *Advances in Applied Probability* 10:360–372
- Denby L, Martin RD (1979) Robust estimation of the first order autoregressive parameter. *Journal of the American Statistical Association* 74:140–46
- Dunsmuir W, Hannan EJ (1976) Vector Linear Time Series Models. *Advances in Applied Probability* 8(2):339
- Fajardo F, Reisen VA, Cribari-Neto F (2009) Robust estimation in long-memory processes under additive outliers. *Journal of Statistical Planning and Inference* 139:2511–2525
- Fajardo FA, Reisen VA, Lévy-Leduc C, Taqu MS (2018) M-periodogram for the analysis of long-range-dependent time series. *Statistics* 52(3):665–683
- Gardner WA, Franks LE (1975) Characterization of cyclostationary random signal processes. *IEEE Transactions on Signal Processing* 21:4–14
- Gardner WA, Napolitano A, Paura L (2006) Cyclostationarity: Half a century of research. *Signal Processing* 86:639–697
- Gladyshev EG (1961) Periodically correlated random sequences. *Sov Math* 2:385–388
- He HD, Lu WZ (2012) Spectral analysis of vehicle pollutants at traffic intersection in Hong Kong. *Stochastic environmental research and risk assessment* 26(8):1053–1061
- Huber PJ (1964) Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics* 35(1):73–101
- Hurd HL, Miamee A (2007) *Periodically correlated random sequences: Spectral theory and practice*. Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, NJ
- Katkovnik V (1998) Robust M-Periodogram. *IEEE Transactions on Signal Processing* 46(11):3104–3109
- Koul HL (1992) M-estimators in linear models with long range dependent errors. *Statistics and probability letters* 14(2):153–164
- Li TH (2008) Laplace periodogram for time series analysis. *Journal of the American Statistical Association* 103(482):757–768
- Li TH (2010) A nonlinear method for robust spectral analysis. *Signal Processing, IEEE Transactions on* 58(5):2466–2474
- Lund RB, Basawa IV (2000) Recursive prediction and likelihood evaluation for periodic ARMA models. *Journal of Time Series Analysis* 21:75–93
- Ma Y, Genton M (2000) Highly robust estimation of the autocovariance function. *Journal of Time Series Analysis* 21:663–684
- Maronna RA, Martin RD, Yohai VJ, Salibián-Barrera M (2019) *Robust statistics: theory and methods* (with R). John Wiley & Sons
- Priestley MB (1981) *Spectral Analysis and Time Series*. Academic press
- Reisen VA, Lévy-Leduc C, Taqu MS (2017) An M-estimator for the long-memory parameter. *Journal of Statistical Planning and Inference* 187(2):44–55
- Reisen VA, Lévy-Leduc C, Bondon P, Cotta H, Ispány M, Prezotti P (2019) An overview of robust spectral estimators. In: *Time Series and Cyclostationary process*, Springer series
- Sarnaglia AJQ, Reisen VA, Lévy-Leduc C (2010) Robust estimation of periodic autoregressive processes in the presence of additive outliers. *Journal of Multivariate Analysis* 101:2168–2183
- Sarnaglia AJQ, Reisen VA, Bondon P (2015) Periodic ARMA models: Application to particulate matter concentrations. 23rd European Signal Processing Conference pp 2226–2230
- Sarnaglia AJQ, Monroy NAJ, da Vitória AG (2018) Modeling and forecasting daily maximum hourly ozone concentrations using the regar model with skewed and heavy-tailed innovations. *Environmental and Ecological Statistics* 25(4):443–469
- Sarnaglia AJQ, Reisen VA, Bondon P (2020) Identifiability and Whittle Estimator of periodic ARMA models, submitted to *Brazilian Journal of Probability and Statistics*
- Sguera C, Galeano P, Lillo RE (2016) Functional outlier detection by a local depth with application to no x levels. *Stochastic environmental research and risk assessment* 30(4):1115–1130
- Shao Q (2008) Robust estimation for periodic autoregressive time series. *Journal of Time Series Analysis* 29(2):251–263
- Solci C, Reisen VA, Sarnaglia AJQ, Bondon P (2020) Empirical study of robust estimation methods for PAR models with application to PM₁₀ data. *Communication in Statistics – Theory and Methods* 49(1):152–168
- Tiao GC, Grupe MR (1980) Hidden periodic autoregressive-moving average models in time series

data. *Biometrika* 67:365–73

Ursu E, Perea JC (2016) Application of periodic autoregressive process to the modeling of the garonne river flows. *Stochastic Environmental Research and Risk Assessment* 30(7):1785–1795

Vecchia AV (1985) Periodic autoregressive-moving average (PARMA) modeling with applications to water resources. *Water Resour Bull* 21:721–30