PERIODIC ARMA MODELS: APPLICATION TO PARTICULATE MATTER CONCENTRATIONS

A. J. Q. Sarnaglia[*†], *V. A. Reisen*[*†], *P. Bondon*[‡]

[*] Federal University of Espírito Santo, Department of Statistics, Vitória, ES, Brazil
[†] Federal University of Minas Gerais, Graduate Program in Statistics, Belo Horizonte, MG, Brazil
[‡] CNRS - CentraleSupélec - Université Paris-Sud, 91192 Gif-sur-Yvette, France

ABSTRACT

We propose the use of multivariate version of Whittle's methodology to estimate periodic autoregressive moving average models. In the literature, this estimator has been widely used to deal with large data sets, since, in this context, its performance is similar to the Gaussian maximum likelihood estimator and the estimates are obtained much faster. Here, the usefulness of Whittle estimator is illustrated by a Monte Carlo simulation and by fitting the periodic autoregressive moving average model to daily mean concentrations of particulate matter observed in Cariacica, Brazil. The results confirm the potentiality of Whittle estimator when applied to periodic time series.

Index Terms— Cyclostationarity, periodic stationarity, PARMA models, Whittle estimation, particulate matter.

1. INTRODUCTION

Seasonal phenomena are frequently observed in many fields such as hydrology, climatology, air pollution, radio astronomy, econometrics, communications, signal processing, among others. A standard approach in the literature is to fit a stationary seasonal model after removing any trend. As pointed out by [1], this strategy can be suggested by standard time series tools even if the true covariance structure has a periodic (or cyclic) nonstationary behaviour. In this case, adjusting a seasonal model is inappropriate and deteriorates the forecast performance and this model mispecification is not revealed by the usual residual diagnostic checking. Some authors have proposed methods to identity hidden periodic covariances in a signal, see e.g., [2].

Processes with periodically varying covariances are denominated Periodically Correlated (PC) (also known as periodically stationary or cyclostationary) and were introduced in the seminal paper [3]. The occurrence of PC processes is corroborated by real applications in many areas, for example, [4] investigate cyclostationarity in electrical engineering and [5] study stratospheric ozone data. See, e.g., [6] and [7] for recent reviews on PC processes.

The simplest way to build models for PC processes is to allow the parameters of stationary models to vary periodically with time. In this context, the Periodic Autoregressive (PAR) model emerges as an extension of the well-known autoregressive framework. The parameter estimation of PAR model is already well documented in the literature, see e.g. [8] and references therein. However, some data sets require large periodic autoregressive orders to provide an adequate fit. Thus, a more parsimonious model can be built by considering jointly Autoregressive and Moving Average (ARMA) coefficients, which leads naturally to the periodic ARMA (PARMA) model.

Surprisingly, even if the PARMA model is more parsimonious, it has not been widely used yet in real applications. The main reasons of this may be due to the difficulty of implementation of estimation methods and their computational efforts. Since PC processes are nonstationary, in essence, the methodology of [9] seems to be a promising tool for fitting models to PC data. However, we point out that, in general, PC processes do not satisfy the local stationarity assumption which is requested in [9]. On the other hand, generalizations of stationarity-based estimation methods have been developed in the literature for PC processes. For example, the exact Gaussian PARMA likelihood is derived by [10], but the method requires the Choleski decomposition of a matrix whose dimension is the number of data. This can be a serious handicap for large data sets and [11] propose an efficient algorithm to evaluate the Gaussian likelihood which does not require any matrix inversion.

All the estimation methods discussed in the previous paragraph are based on the time domain. To our knowledge, the estimation of PARMA models in the frequency domain proposed here has not been investigated yet. The advantage of the spectral approach is to circumvent the inversion of the covariance matrix by using the well-known Whittle approximation. Since the publication of the seminal paper [12], Whittle type estimation has been studied in many domains, see e.g. [13], [14] and [15]. Here we introduce a Whittle estimator for the parameters of a PARMA model.

The rest of the paper is organized as follows. PC processes and PARMA models are defined in Section 2. Whittle estimator is described in Section 3. Section 4 presents a

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

Monte Carlo simulation study. Section 5 illustrates the usefulness of Whittle estimator through an application to air pollution data set. Concluding remarks are given in Section 6.

2. PC PROCESSES AND PARMA MODELS

Let $\mathbb{Z}$ be the set of integer numbers and $(X_t)_{t\in\mathbb{Z}}$ be a real valued stochastic process satisfying $\mathrm{E}(X_t^2) < \infty$ for all $t \in \mathbb{Z}$. Let $\mu_t = \mathrm{E}(X_t)$ and $\gamma_t(\tau) = \mathrm{Cov}(X_t, X_{t-\tau})$. We say that (X_t) is PC with period T (PC-T) if for every $(t,\tau) \in \mathbb{Z}^2$,

$$\mu_{t+T} = \mu_t \quad \text{and} \quad \gamma_{t+T}(\tau) = \gamma_t(\tau), \tag{1}$$

and there are no smaller values of $T > 0$ for which (1) hold. This definition implies that μ_t and $\gamma_t(\tau)$ are periodic functions in t and need to be known only for $t = 1, \dots, T$. If (X_t) is PC-1 then it is weakly stationary in the usual sense. In the following, we assume without loss of generality that $\mu_t = 0$ for all $t \in \mathbb{Z}$.

The univariate sequence (X_t) is remarkably related by T-blocking to the T-variate sequence $(\mathbf{X}_n)$ defined by $\mathbf{X}_n' = (X_{nT+1}, \dots, X_{nT+T})$, where $\mathbf{X}_n'$ denotes the transpose of $\mathbf{X}_n$. In other words, $\mathbf{X}_n$ is the vector of all observations of cycle n. In particular, (X_t) is PC-T if its T-vector counterpart $(\mathbf{X}_n)$ is stationary in the usual (vector) sense. In addition, the causality and invertibility of PC-T processes can be naturally defined from the causality and invertibility of the vector process $(\mathbf{X}_n)$. For more details, we refer to [3] and [7].

As previously mentioned, the most natural way to build parametric models for PC-T processes is to allow periodically varying coefficients. In this context, the PARMA model emerges as a powerful tool. A time series (X_t) satisfying $\mathrm{E}(X_t) = 0$ and $\mathrm{E}(X_t^2) < \infty$ for all $t \in \mathbb{Z}$ is said to be a PARMA(p,q) process with period $T > 0$ (PARMA$(p,q)_T$) if it is a solution to the periodic linear difference equation

$$X_{nT+\nu} + \sum_{j=1}^{p} \phi_{\nu,j} X_{nT+\nu-j} = \varepsilon_{nT+\nu} + \sum_{j=1}^{q} \theta_{\nu,j} \varepsilon_{nT+\nu-j},$$

where $n \in \mathbb{Z}$, $\nu = 1, \dots, T$, and $(\varepsilon_{nT+\nu})$ is a sequence of zero mean uncorrelated random variables with $\mathrm{E}(\varepsilon_{nT+\nu}^2) = \sigma_\nu^2$. The autoregressive and moving average model orders are p and q, respectively, and $\phi_{\nu,1}, \dots, \phi_{\nu,p}$ and $\theta_{\nu,1}, \dots, \theta_{\nu,q}$ are the autoregressive and moving average parameters, respectively, during season ν. There is no mathematical loss of generality in assuming that p and q are constant in the season ν. The period T is taken to be the smallest positive integer satisfying the above difference equation so that its definition is unambiguous. When $T = 1$, the PARMA(p,q) model corresponds to the usual ARMA(p,q) model.

When (X_t) is a PARMA$(p,q)_T$ process, it is well-known that the sequence $(\mathbf{X}_n)$ satisfies the VARMA difference equation

$$\textstyle\sum_{k=0}^{P} \boldsymbol{\phi}_k \mathbf{X}_{n-k} = \sum_{k=0}^{Q} \boldsymbol{\theta}_k \boldsymbol{\varepsilon}_{n-k}, \tag{2}$$

where $(\boldsymbol{\varepsilon}_n) = (\varepsilon_{nT+1}, \dots, \varepsilon_{nT+T})'$, $[\boldsymbol{\phi}_k]_{l,m} = \phi_{l,kT+l-m}$ and $[\boldsymbol{\theta}_k]_{l,m} = \theta_{l,kT+l-m}$ with the conventions that $\phi_{\nu,j} = 0$ when $j \notin \{0, \dots, p\}$, $\phi_{\nu,0} = 1$, $\theta_{\nu,j} = 0$ when $j \notin \{0, \dots, q\}$, $\theta_{\nu,0} = 1$, for every $\nu = 1, \dots, T$. The orders are $P = \lceil p/T \rceil$ and $Q = \lceil q/T \rceil$, wherein $\lceil x \rceil$ stands for the smallest integer greater than or equal to x. Observe that the covariance matrix $\mathrm{E}(\boldsymbol{\varepsilon}_n \boldsymbol{\varepsilon}_n') = \Sigma$ is diagonal with $[\Sigma]_{l,l} = \sigma_l^2$.

In general, $\boldsymbol{\phi}_0$ and $\boldsymbol{\theta}_0$ are not equal to the identity matrix $\mathbf{I}$. However, since $\boldsymbol{\phi}_0$ and $\boldsymbol{\theta}_0$ are unit lower triangular matrices, they are invertible and $(\mathbf{X}_n)$ satisfies the VARMA difference equation

$$\textstyle\mathbf{X}_n + \sum_{k=1}^{P} \boldsymbol{\phi}_k^* \mathbf{X}_{n-k} = \boldsymbol{\xi}_n + \sum_{k=1}^{Q} \boldsymbol{\theta}_k^* \boldsymbol{\xi}_{n-k}, \tag{3}$$

where $\boldsymbol{\phi}_k^* = \boldsymbol{\phi}_0^{-1} \boldsymbol{\phi}_k$, $\boldsymbol{\theta}_k^* = \boldsymbol{\phi}_0^{-1} \boldsymbol{\theta}_k \boldsymbol{\theta}_0^{-1} \boldsymbol{\phi}_0$ and $\boldsymbol{\xi}_n = \boldsymbol{\phi}_0^{-1} \boldsymbol{\theta}_0 \boldsymbol{\varepsilon}_n$ with $\mathrm{Var}(\boldsymbol{\xi}_n) = \Sigma^* = \boldsymbol{\phi}_0^{-1} \boldsymbol{\theta}_0 \Sigma \boldsymbol{\theta}_0' \boldsymbol{\phi}_0^{-1'}$. Nevertheless, deducing a unique representation (2) from (3) is not always possible.

We now turn to causality (and invertibility) of PARMA processes. Let us define, for $z \in \mathbb{C}$, the polynomials

$$\textstyle\boldsymbol{\phi}(z) = \sum_{k=0}^{P} \boldsymbol{\phi}_k z^k, \quad \boldsymbol{\theta}(z) = \sum_{k=0}^{Q} \boldsymbol{\theta}_k z^k,$$
$$\textstyle\boldsymbol{\phi}^*(z) = \mathbf{I} + \sum_{k=1}^{P} \boldsymbol{\phi}_k^* z^k, \quad \boldsymbol{\theta}^*(z) = \mathbf{I} + \sum_{k=1}^{Q} \boldsymbol{\theta}_k^* z^k.$$

As previously mentioned, causality of $(X_{nT+\nu})$ and $(\mathbf{X}_n)$ are equivalent and, by [16, Theorem 11.3.1], it is ensured whenever $\det \boldsymbol{\phi}^*(z) \neq 0$ for $|z| \leq 1$, or equivalently, $\det \boldsymbol{\phi}(z) \neq 0$ for $|z| \leq 1$. This theorem also shows that the weights of the causal representation of $(\mathbf{X}_n)$ can be obtained from the power series expansions of $\boldsymbol{\phi}(z)^{-1} \boldsymbol{\theta}(z)$ or $\boldsymbol{\phi}^*(z)^{-1} \boldsymbol{\theta}^*(z)$, depending on which white noise $(\boldsymbol{\varepsilon}_n)$ or $(\boldsymbol{\xi}_n)$ is used. The same arguments jointly with [16, Theorem 11.3.2] show that (X_t) is invertible when $\det \boldsymbol{\theta}(z) \neq 0$ for $|z| \leq 1$. In addition, the spectral density of $(\mathbf{X}_n)$ is

$$\begin{aligned}\mathbf{f}(\omega) &= \frac{1}{2\pi} \boldsymbol{\phi}^*(e^{-\mathrm{i}\omega})^{-1} \boldsymbol{\theta}^*(e^{-\mathrm{i}\omega}) \Sigma^* \boldsymbol{\theta}^*(e^{-\mathrm{i}\omega})^{\dagger} \boldsymbol{\phi}^*(e^{-\mathrm{i}\omega})^{-1\dagger} \\ &= \frac{1}{2\pi} \boldsymbol{\phi}(e^{-\mathrm{i}\omega})^{-1} \boldsymbol{\theta}(e^{-\mathrm{i}\omega}) \Sigma \boldsymbol{\theta}(e^{-\mathrm{i}\omega})^{\dagger} \boldsymbol{\phi}(e^{-\mathrm{i}\omega})^{-1\dagger}\end{aligned}$$

According to [16, page 431], causality and invertibility do not ensure that Σ, $\boldsymbol{\phi}(z)$ and $\boldsymbol{\theta}(z)$ are uniquely determined by $\mathbf{f}$. This identifiability problem results in a likelihood surface with more than one maximum. Further restrictions have to be imposed in order to obtain identifiable models, see [13] and [14]. The identifiability of the parameters in the VARMA model (2) is not easily handled, since (2) is not in the usual form (3). We shall not pursue this topic here and will tacitly assume that model (2) is identifiable.

3. WHITTLE ESTIMATION

Let $\mathbb{P} \subset \mathbb{R}^{(p+q+1)T}$ be the parameter space. Set the vectors of AR and MA parameters of season ν as $\phi_\nu = (\phi_{\nu,1}, \dots, \phi_{\nu,p})'$ and $\theta_\nu = (\theta_{\nu,1}, \dots, \theta_{\nu,q})'$, respectively. Typical points of $\mathbb{P}$

are $\varphi = (\varphi'_\phi, \varphi'_\theta, \varphi'_\sigma)'$, where $\varphi'_\phi = (\phi'_1, \ldots, \phi'_T)$, $\varphi'_\theta = (\theta'_1, \ldots, \theta'_T)$ and $\varphi'_\sigma = (\sigma^2_1, \ldots, \sigma^2_T)$. We shall denote the true parameter vector by $\varphi_0 \in \mathbb{P}$. For simplicity, assume that a time series $\mathbf{X} = (X_1, \ldots, X_{NT})'$ of length NT is available.

Let $\mathbf{\Gamma}_{N,\varphi}$ be the $(NT \times NT)$ matrix with (l,m)th block entry given by

$$(\mathbf{\Gamma}_{N,\varphi})_{l,m} = \int_0^{2\pi} \mathbf{f}_\varphi(\omega) e^{\mathbf{i}(m-l)\omega} d\omega, \;\; l, m = 1, \ldots, N,$$

where we use the subscript φ to emphasize the dependency of the spectral density $\mathbf{f}$ on the parameter vector φ. Note that $\mathbf{\Gamma}_{N,\varphi_0} = \text{Cov}(\mathbf{X}, \mathbf{X})$. Consider

$$\hat{\mathcal{L}}(\varphi, \mathbf{X}) = \frac{1}{N} \log\det \mathbf{\Gamma}_{N,\varphi} + \frac{1}{N} \mathbf{X}' \mathbf{\Gamma}^{-1}_{N,\varphi} \mathbf{X}, \; \varphi \in \mathbb{P},$$

and observe that $\hat{\mathcal{L}}(\varphi, \mathbf{X})$ is the Gaussian loglikelihood multiplied by $-2/N$. The well-known Gaussian Maximum Likelihood Estimator (MLE) over the parameter space $\mathbb{P}$ is defined as

$$\hat{\varphi} = \underset{\varphi \in \mathbb{P}}{\text{argmin}}\, \hat{\mathcal{L}}(\varphi, \mathbf{X}).$$

However, asymptotics for $\hat{\varphi}$ do not necessarily requires that $\mathbf{X}$ be Gaussian, see for example, [13].

In most cases this minimization is performed through optimization algorithms, which can demand high computational effort, since *a priori* it is necessary to invert $\mathbf{\Gamma}_{N,\varphi}$. One alternative is to resort to exact efficient algorithms such as in [11]. However, even these tools can be troublesome for large sample sizes. Therefore, to circumvent this difficulty, we propose to use the multivariate version of Whittle's methodology to approximate $\hat{\mathcal{L}}(\varphi, \mathbf{X})$.

The discrete Fourier transform and the periodogram of $\mathbf{X}_n$, $n = 1, \ldots, N$, at the elementary frequencies $\omega_j = \frac{2\pi j}{N}$, $j = 0, 1, \ldots, N-1$, are defined, respectively, by

$$W(\omega_j) = \sum_{n=1}^{N} \frac{\mathbf{X}_n e^{-\mathbf{i}n\omega_j}}{\sqrt{2\pi N}} \;\text{ and }\; \mathbf{P}(\omega_j) = W(\omega_j) W(\omega_j)^\dagger.$$

We shall consider the Whittle likelihood

$$\begin{aligned}\tilde{\mathcal{L}}(\varphi, \mathbf{X}) &= \log\det \Sigma^*_\varphi + \frac{1}{N} \sum_{j=0}^{N-1} \text{tr}\{\mathbf{f}_\varphi(\omega_j)^{-1} \mathbf{P}(\omega_j)\} \\ &= \sum_{\nu=1}^{T} \left[\log \sigma^2_\nu + \frac{2\pi}{\sigma^2_\nu N} \sum_{j=0}^{N-1} \left\| \left({\Theta^{(j)}_{\varphi_\theta}}^{-1} \Phi^{(j)}_{\varphi_\phi} \mathbf{W}_j \right)_\nu \right\|^2 \right], \quad (4)\end{aligned}$$

where $\Theta^{(j)}_{\varphi_\theta} = \boldsymbol{\theta}_{\varphi_\theta}(e^{-\mathbf{i}\omega_j})$, $\Phi^{(j)}_{\varphi_\phi} = \boldsymbol{\phi}_{\varphi_\phi}(e^{-\mathbf{i}\omega_j})$ and $\mathbf{W}_j = W(\omega_j)$. Among several Whittle-type likelihoods, $\tilde{\mathcal{L}}$ is particularly interesting due to its computational advantages. The Whittle Likelihood Estimator (WLE) over $\mathbb{P}$ is now defined as $\tilde{\varphi} = \text{argmin}_{\varphi \in \mathbb{P}} \tilde{\mathcal{L}}(\varphi, \mathbf{X})$. Differentiating (4) with respect to σ^2_ν and noticing that the sum in the second term of (4) is independent of σ^2_ν, the values of σ^2_ν minimizing $\tilde{\mathcal{L}}(\varphi, \mathbf{X})$ are

		Cycle			
Model	Coef	1	2	3	4
	$\phi_{\nu,1}$	-1.4	-0.5	-	-
	$\theta_{\nu,1}$	-0.5	-1.0	-	-
1	σ^2_ν	1	1	-	-
	$\phi_{\nu,1}$	-1.5	-1.2	-0.8	-0.5
	$\theta_{\nu,1}$	-0.6	-0.8	-0.9	-1.1
2	σ^2_ν	1	1	1	1

Table 1. Parameters of Models 1 and 2.

$$\tilde{\sigma}^2_\nu(\varphi_\phi, \varphi_\theta) = \frac{2\pi}{N} \sum_{j=0}^{N-1} \left\| \left({\Theta^{(j)}_{\varphi_\theta}}^{-1} \Phi^{(j)}_{\varphi_\phi} \mathbf{W}_j \right)_\nu \right\|^2, \quad (5)$$

for every $\nu = 1, \ldots, T$. Replacing (5) in (4), we see that the values of $(\varphi_\phi, \varphi_\theta)$ minimizing $\tilde{\mathcal{L}}(\varphi, \mathbf{X})$ are the values which minimize the following "reduced" Whittle Likelihood

$$\tilde{\mathcal{L}}_{\mathfrak{R}}(\varphi_\phi, \varphi_\theta) = \sum_{\nu=1}^{T} \log \tilde{\sigma}^2_\nu(\varphi_\phi, \varphi_\theta).$$

Therefore, WLE is given by $\tilde{\varphi} = (\tilde{\varphi}'_\phi, \tilde{\varphi}'_\theta, \tilde{\varphi}'_\sigma)'$, where $(\tilde{\varphi}_\phi, \tilde{\varphi}_\theta) = \text{argmin}\, \tilde{\mathcal{L}}_{\mathfrak{R}}(\varphi_\phi, \varphi_\theta)$ and $\tilde{\varphi}'_\sigma = (\tilde{\sigma}^2_1, \ldots, \tilde{\sigma}^2_T)$, wherein $\tilde{\sigma}^2_\nu = \tilde{\sigma}^2_\nu(\tilde{\varphi}_\phi, \tilde{\varphi}_\theta)$, $\nu = 1, \ldots, T$.

Observe that $(\varphi'_\phi, \varphi'_\theta)$ involves $(p+q)T$ parameters whereas the dimension of φ is $(p+q+1)T$. Then minimizing $\tilde{\mathcal{L}}_{\mathfrak{R}}$ is simpler than minimizing $\hat{\mathcal{L}}$. This may explain why $\tilde{\varphi}$ is obtained faster than $\hat{\varphi}$, even when the efficient algorithm proposed by [11] is used (see Section 4).

4. MONTE CARLO STUDY

This section presents a Monte Carlo simulation study to investigate the finite sample performance of WLE. For comparison purposes, we also consider the exact MLE obtained with the algorithm in [11]. In each of the 1000 replications, a time series with $N = 100$ samples is generated. We consider two different $\text{PARMA}(1,1)_T$ models, see Table 1.

The parameters in Table 1 are chosen in order to make Models 1 and 2 comparable in the sense that the respective roots of $\det \Phi(z) = 0$ and $\det \Theta(z) = 0$ are almost the same for both models. Empirical root mean squared errors (RMSE) of both estimators are displayed in Table 2. In addition, the mean computation time in seconds for the MLE and the WLE are, respectively 0.995 and 0.572 for Model 1, and 5.077 and 0.635 for Model 2.

We observe from Table 2 that, for both models, MLE and WLE present very similar performances in terms of empirical RMSE. Nevertheless, WLE runs almost twice faster than MLE for Model 1. This difference increases dramatically for

	Model 1		Model 2			
Coef	$\nu = 1$	$\nu = 2$	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\nu = 4$
$\phi_{\nu,1}$	1.2 (1.3)	0.8 (0.9)	1.1 (1.1)	0.7 (0.7)	0.6 (0.6)	0.8 (0.8)
$\theta_{\nu,1}$	1.2 (1.4)	1.3 (1.6)	1.1 (1.1)	0.8 (1.1)	0.8 (1.0)	1.2 (1.3)
σ^2_ν	1.4 (1.8)	1.4 (1.8)	1.4 (2.0)	1.4 (1.6)	1.5 (1.5)	1.4 (1.5)

Table 2. RMSE ($\times 10$) of MLE and WLE, in parentheses, of Models 1 and 2.

higher dimensional parameter spaces such as in Model 2. In the next section, we will see that MLE is also much more time consuming than the WLE when real data are considered.

We point out that the same experiment was carried out for larger sample sizes. For space limitation, these results are not reported here, but they corroborate the fact that both estimators are consistent since their empirical RMSE decrease when N increases. Also the difference in terms of computation time between MLE and WLE increases as N increases.

5. APPLICATION

In this section we analyze the Particulate Matter with an aerodynamic diameter smaller than or equal to 10 μm (PM_{10}). The series is observed between January 1, 2005 and December 31, 2009 at the monitoring station of Environment and Water Resources State Institute located in Cariacica, ES, Brazil. The first 1603 observations are used for fitting the model and the remaining 223 observations are used for the out-of-sample forecast study.

Since the data are collected daily, a PARMA_T model with $T = 7$ seems to be appropriated to fit the series. The sample periodic autocorrelation and partial autocorrelation functions indicate a PARMA models with orders $p_\nu = 1, 0, 3, 1, 1, 1, 1$ and $q_\nu = 0, 1, 0, 1, 0, 0, 0$. We set all the initial AR and MA parameters as zero. Additionally, we observe that the choice of the initial values for the white noise variances $(\sigma^2_1, \ldots, \sigma^2_T)$ has an important influence on the computation time of MLE which is not the case for WLE. Indeed, for MLE, taking as initial values $(1, \ldots, 1)$, the computation time is 294.9 seconds, while taking $(\hat{\sigma}^2_X, \ldots, \hat{\sigma}^2_X)$ as these initial values, where $\hat{\sigma}^2_X$ is the empirical variance of the data, gives a computation time of 88.9 seconds. However, these different initial values do not seem to play any substantial role in MLE estimates, even for σ^2_ν. On the other hand, since the numerical optimization is not performed to obtain the WLE of σ^2_ν, there is no need of these initial values for calculating the WLE and the computation time is only 2.5 seconds. Therefore, WLE is at least 35 times faster than MLE. As a consequence, model selection through information criteria like Akaike or Schwarz are unfeasible using MLE in this application. Finally, observe that the estimates obtained by both methods are almost the same, see Table 4.

Residual diagnostic checking (not reported here) show that the correlation structure is well accommodated by the model regardless of the estimation method. However, the residuals present asymmetric behaviour which is an expected result due to the feature of the PM_{10} series.

	In-sample		Out-of-sample	
Estim.	RMSE	SMAPE	RMSE	SMAPE
MLE	10.82	9.61	11.79	10.07
WLE	10.82	9.61	11.78	10.07

Table 3. One-step-ahead forecasting performance.

We now turn to the forecasting performance. The RMSE and symmetric mean absolute percentage error (SMAPE) statistics are defined by

$$\text{RMSE} = \sqrt{\sum_{t=1}^{NT} \frac{e_t^2}{NT}}, \quad \text{SMAPE} = \frac{100}{NT} \sum_{t=1}^{NT} \frac{|e_t|}{X_t + \hat{X}_t},$$

where $\hat{X}_t$ is the forecast of X_t and $e_t = X_t - \hat{X}_t$. As we see in Table 3, RMSE and SMAPE are almost the same when $\hat{X}_t$ is calculated from the model fitted by MLE and WLE, respectively. Hence, both models have almost the same predictive performance. Figure 1 plots the remaining 233 data and their one-step-ahead forecasts obtained from the model fitted by WLE. Similar results are obtained with the MLE. Visual inspection of this figure shows that the forecasts follow satisfactorily the actual data.

6. CONCLUSION

This paper deals with the Whittle estimator for PARMA models. The method is introduced and a Monte Carlo simulation study is performed to evaluate its finite sample properties and compare it with Gaussian MLE. The results show that WLE is very competitive in terms of RMSE compared with MLE, while the former is much more faster to calculate than the latter, mainly in large dimensional parameter spaces. We also present an application which strongly support the usefulness of this method to fit PARMA models to real data sets.

REFERENCES

[1] G. C. Tiao and M. R. Grupe, "Hidden periodic autoregressive-moving average models in time series data," *Biometrika*, vol. 67, pp. 365–73, 1980.

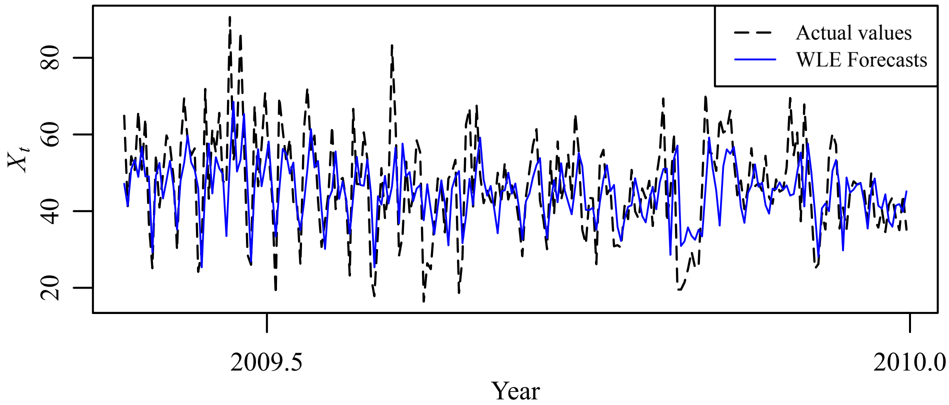

Fig. 1. PM_{10} concentrations not used in estimation and their one-step-ahead forecasts.

ν	Estimates				
	$\phi_{\nu,1}$	$\phi_{\nu,2}$	$\phi_{\nu,3}$	$\theta_{\nu,1}$	σ_ν^2
1	-0.40 (-0.39)	—	—	—	102.05 (102.07)
2	—	—	—	0.41 (0.40)	104.59 (104.08)
3	-0.13 (-0.13)	-0.05 (-0.05)	-0.35 (-0.35)	—	109.22 (108.30)
4	-0.96 (-0.94)	—	—	-0.57 (-0.55)	135.74 (135.49)
5	-0.58 (-0.58)	—	—	—	112.37 (111.89)
6	-0.53 (-0.53)	—	—	—	128.02 (127.59)
7	-0.45 (-0.45)	—	—	—	130.75 (130.37)

Table 4. MLE and WLE, in parentheses, for a PARMA model fitted to the PM_{10} series.

[2] A. V. Vecchia and R. Ballerini, "Testing for periodic autocorrelations in seasonal time series data," *Biometrika*, vol. 78, pp. 53–63, 1991.

[3] E. G. Gladyshev, "Periodically correlated random sequences," *Sov. Math.*, vol. 2, pp. 385–388, 1961.

[4] W. A. Gardner and L. E. Franks, "Characterization of cyclostationary random signal processes," *IEEE Transactions on Signal Processing*, vol. 21, pp. 4–14, 1975.

[5] P. Bloomfield, H. L. Hurd, and R. B. Lund, "Periodic correlation in stratospheric ozone data," *Jour. of Tim. Ser. Anal.*, vol. 15, pp. 127–150, 1994.

[6] W. A. Gardner, A. Napolitano, and L. Paura, "Cyclostationarity: Half a century of research," *Signal Processing*, vol. 86, pp. 639–697, 2006.

[7] H. L. Hurd and A. Miamee, *Periodically correlated random sequences: Spectral theory and practice*, Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken, NJ, 2007.

[8] A. J. Q. Sarnaglia, V. A. Reisen, and C. Lévy-Leduc, "Robust estimation of periodic autoregressive processes in the presence of additive outliers," *Journal of Multivariate Analysis*, vol. 101, pp. 2168–2183, 2010.

[9] R. Dahlhaus, "Fitting time series models to nonstationary processes," *The Annals of Statistics*, vol. 25, no. 1, pp. 1–37, 1997.

[10] W. K. Li and Y. V. Hui, "An algorithm for the exact likelihood of periodic autoregressive moving average models," in *Communications in Statistics*, vol. 17, pp. 1483–94. Taylor and Francis, 1988.

[11] R. B. Lund and I. V. Basawa, "Recursive prediction and likelihood evaluation for periodic ARMA models," *Jour. of Tim. Ser. Anal.*, vol. 21, pp. 75–93, 2000.

[12] P. Whittle, "The analysis of multiple stationary time series," *Journal of the Royal Statistical Society. Series B*, vol. 15, pp. 125–139, 1953.

[13] W. Dunsmuir and E. J. Hannan, "Vector linear time series models," *Adv. in Applied Probability*, vol. 8, no. 2, pp. 339–364, 1976.

[14] M. Deistler, W. Dunsmuir, and E. J. Hannan, "Vector linear time series models: Corrections and extensions," *Adv. in Applied Probability*, vol. 10, pp. 360–372, 1978.

[15] R. Fox and M. S. Taqqu, "Large-sample properties of parameters estimates for strongly dependent stationary gaussian time series," *The Annals of Statistics*, vol. 14, no. 2, pp. 517–532, 1986.

[16] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer Series in Statistics. Springer Science, New York, NY, 2nd edition, 2006.