Fourier-PARMA Models and Their Application to River Flows

Paul L. Anderson*, Yonas Gebeyehu Tesfaye†, and Mark M. Meerschaert‡

20 September 2006

Abstract

For analysis and design of water resource systems, it is sometimes useful to generate high-resolution (e.g., weekly) synthetic river flows. Periodic autoregressive moving average (PARMA) time series models provide a powerful tool for generating synthetic flows. Periodically stationary models are indicated when the basic statistics (mean, variance, and autocorrelation) of the time series exhibit significant seasonal variations. Parameter estimation for high-resolution PARMA models involves numerous parameters, which can lead to over-fitting. Thus, this paper develops a parsimonious method of parameter fitting for high-resolution PARMA models, using discrete Fourier transforms to represent the set of periodic autoregressive and moving average model coefficients. Model parameters are computed via the innovations algorithm, and the asymptotic distributions of the discrete Fourier transform coefficients are obtained. Those asymptotic results are useful to determine the statistically significant Fourier coefficients to include in the model. Effectiveness of the technique is shown using simulated data from different PARMA models. Discharge measurements from the Fraser River in British Columbia are then modelled, firstly as a monthly series and secondly as a weekly series. Diagnostic checks are used to ensure adequacy of the models. Finally, a careful statistical analysis of the PARMA model residuals, including a novel truncated Pareto model for the extreme tails, is combined with the Fourier-PARMA time series model to generate realistic synthetic flows. A key finding is that the Fourier-PARMA method produces superior results as compared to a conventional PARMA model, despite using far fewer parameters.

Key words and phrases: *Discrete Fourier transform, parsimonious PARMA model, parameter estimation, innovations algorithm, asymptotic distribution, Fourier analysis, model identification, simulation study, synthetic river flows.*

1 Introduction

Modeling and simulation of river flow time series is an important step in the planning and operational analysis of water resources systems. Generation of synthetic river flow series may be useful for determining the dimensions of hydraulic works, for risk assessment in urban water supply and irrigation, for optimal operation of reservoir systems, for determining the risk of failure of dependable capacities of hydroelectric systems, for planning capacity expansion of water supply systems, and others (see Salas, 1993). A natural river flow process has significant periodic behavior in the mean, standard deviation, skewness and serial dependence structure (see, for example, Moss and Bryson, 1974). In the area of stochastic hydrology, standardizing or filtering is used to transform

*Department of Mathematics and Computer Science, Albion College, Albion MI 48209.

†Graduate Program in Hydrologic Sciences, University of Nevada, Reno NV 89557. Partially supported by National Science Foundation grant DMS-0139927.

‡Department of Statistics & Probability, Michigan State University, A416 Wells Hall, East Lansing, MI 48824-1027. Partially supported by National Science Foundation grants DMS-0139927 and DMS-0417869.

periodic time series to stationary series before fitting stationary stochastic models (Salas, Delleur, Yevjevich, and Lane 1980; Thompstone, Hipel, and McLeod 1985; Vecchia, 1985a,1985b; Salas, 1993; Chen and Rao, 2002). However, standardizing or filtering of most river flow series will not yield stationary residuals due to periodic autocorrelations. In these cases, the resulting model is misspecified (Tiao and Grupe, 1980). To model such periodicity in autocorrelations, periodic ARMA (PARMA) models can be employed. Several researchers have dealt with periodic time series models (see: Jones and Brelsford, 1967; Pagano, 1978; Troutman, 1979; Tjøstheim and Paulsen, 1982; Salas, Obeysekera, and Smith 1981; Salas, Boes, and Smith, 1982; Salas, Tabios III, and Bartolini, 1985; Vecchia, 1985a,1985b; Vecchia and Ballerini, 1991; Anderson and Vecchia, 1993; Ula, 1990,1993; Ula and Smadi, 1997,2003; Adams and Goodwin, 1995; Anderson and Meerschaert, 1997,1998; Lund and Basawa, 1999,2000; Shao and Lund, 2004; and Tesfaye, Meerschaert, and Anderson, 2006). In most cases, PARMA models have been applied to time series at a time scale of a month or more. However, when the number of periods is large (e.g., weekly data), PARMA models require estimation of an exorbitant number of parameters, heretofore making PARMA modeling virtually impractical. The methods proposed in this paper adhere to the principle of statistical parsimony. Model parsimony is achieved by expressing the periodic model parameters in terms of their discrete Fourier transforms. The practicality of our methodolgy is illustrated in several examples given in Sections 4 and 5.

In most water resources systems design and operation studies, the periodic phenomena have been represented by Fourier functions. Quimpo (1967) has applied Fourier analysis to daily river flow sequences in order to detect significant harmonic components embedded within the sequence considered. Since then Fourier analysis has become a standard tool in any hydrologic study concerning periodicity. Salas, Delleur, Yevjevich, and Lane (1980) proposed a Fourier series approach for reducing the number of parameters in PAR or PARMA models. Vecchia (1985a) also adopted the same approach but used Akaike's information criterion (AIC) for the selection of significant harmonics. Experience in using Fourier analysis for estimating periodic parameters of hydrologic time series shows that for small time interval series, such as daily and weekly series, only the first few harmonics are necessary for a good Fourier series fit in the periodic estimate of model parameters (Salas, Delleur, Yevjevich, and Lane, 1980). This practical criteria should be supplemented by more precise analysis and tests. For instance, Anderson and Vecchia (1993) use asymptotic properties of the discrete Fourier transform of the estimated periodic autocovariance and autocorrelation function for selecting the harmonics in the PARMA model parameters.

This paper presents a comprehensive time series analysis for high resolution periodic processes. In section 2 of the paper, initial parameter estimates are obtained with the innovations algorithm and their statistical properties are discussed. In section 3, we develop the asymptotic distributions of the discrete Fourier transform coefficients of the parameter estimates with the intent of obtaining a parsimonious model for the series. A specific PARMA model, useful in describing river flow series, is discussed in detail. In section 4, we use simulated data to discuss model identification and the efficacy of our estimation techniques. Section 5 gives the practitioner a general approach for modeling river flows, including model validation, and then implements this step-by-step approach by analyzing a river flow series for the Fraser River near Hope, British Columbia. We analyze this data firstly as a monthly series, and secondly as a weekly series. Model validation includes diagnostic checking of the estimated noise series, to verify that no significant serial dependence remains.

2 Parameter Estimation for PARMA Model

A stochastic process $\tilde{X}_t$ is (weakly) periodically stationary if $\mu_t = E\tilde{X}_t$ and $\gamma_t(h) = \text{Cov}(X_t, X_{t+h})$ for $h = 0, \pm 1, \pm 2, ...$ are periodic functions of time t with the same period ν (that is, for some integer ν, for $i = 0, 1, ..., \nu - 1$, and for all integers k and h, $\mu_i = \mu_{i+k\nu}$ and $\gamma_i(h) = \gamma_{i+k\nu}(h)$). The periodic ARMA process $\{\tilde{X}_t\}$ with period ν (denoted by PARMA$_\nu$ (p,q)) has representation

$$X_t - \sum_{j=1}^{p} \phi_t(j) X_{t-j} = \varepsilon_t - \sum_{j=1}^{q} \theta_t(j) \varepsilon_{t-j} \tag{1}$$

where $X_t = \tilde{X}_t - \mu_t$ and $\{\varepsilon_t\}$ is a sequence of random variables with mean zero and scale σ_t such that $\{\delta_t = \sigma_t^{-1}\varepsilon_t\}$ is independent and identically distributed (iid). The notation in (1) is consistent with that of Box and Jenkins (1976). The autoregressive parameters $\phi_t(j)$, the moving average parameters $\theta_t(j)$, and the residual standard deviations σ_t are all periodic functions of t with the same period $\nu \geq 1$. The residual standard deviations parameters σ_t are assumed strictly positive. We also assume the model admits a causal representation

$$X_t = \sum_{j=0}^{\infty} \psi_t(j) \varepsilon_{t-j} \tag{2}$$

where $\psi_t(0) = 1$, $\sum_{j=0}^{\infty} |\psi_t(j)| < \infty$ for all t, and $\psi_t(j) = \psi_{t+k\nu}(j)$ for all j. An additional invertibility condition $\varepsilon_t = \sum_{j=0}^{\infty} \pi_t(j) X_{t-j}$ is also assumed for technical reasons.

The innovations algorithm, developed for PARMA models by Anderson, Meerschaert and Vecchia (1999), is a useful method for parameter estimation. Compute:

$$\begin{aligned} v_{0,i} &= \gamma_i(0) \\ \theta_{k,k-\ell}^{(i)} &= (v_{\ell,i})^{-1} \left[\gamma_{i+\ell}(k-\ell) - \textstyle\sum_{j=0}^{\ell-1} \theta_{\ell,\ell-j}^{(i)} \theta_{k,k-j}^{(i)} v_{j,i} \right] \\ v_{k,i} &= \gamma_{i+k}(0) - \textstyle\sum_{j=0}^{k-1} \left(\theta_{k,k-j}^{(i)} \right)^2 v_{j,i} \end{aligned} \tag{3}$$

where (3) is solved recursively in the order $v_{0,i}$, $\theta_{1,1}^{(i)}$, $v_{1,i}$, $\theta_{2,2}^{(i)}$, $\theta_{2,1}^{(i)}$, $v_{2,i}$, $\theta_{3,3}^{(i)}$, $\theta_{3,2}^{(i)}$, $\theta_{3,1}^{(i)}$, $v_{3,i}, \ldots$ and so forth. These parameter estimates converge to the model parameters in equation (2) as follows:

$$\theta_{k,j}^{(\langle i-k \rangle)} \to \psi_i(j) \quad \text{and} \quad v_{k,\langle i-k \rangle} \to \sigma_i^2 \quad \text{as } k \to \infty \tag{4}$$

for all i, j where $\langle t \rangle$ is the season corresponding to index t, so that $\langle j\nu + i \rangle = i$. If we replace the autocovariances in (3) with the corresponding sample autocovariances, we obtain the innovations estimates $\hat{\theta}_{k,l}^{(i)}$ and $\hat{v}_{k,i}$ based on the time series data. Theoretical results in Anderson, Meerschaert and Vecchia (1999) show that these quantities converge (in probability) to give a consistent estimate of the moving average model parameters from data. A simulation study in Tesfaye (2005) indicates that $k = 10$ to 15 iterations of the algorithm is generally adequate to obtain convergence. Furthermore, Anderson and Meerschaert (2005) show that

$$N_y^{1/2} (\hat{\theta}_{k,u}^{(\langle i-k \rangle)} - \psi_i(u)) \Rightarrow \mathcal{N}\Big(0, \sum_{n=0}^{u-1} \frac{\sigma_{i-n}^2}{\sigma_{i-u}^2} \psi_i^2(n)\Big). \tag{5}$$

Formula (5) can be used to determine which model parameters in the moving average (2) are statistically significantly different from zero, see Tesfaye, Meerschaert, and Anderson (2006) for details. This is useful for model selection (how many non-zero coefficients are needed in (2) for an adequate model). Estimates of the moving average parameters in (2) can also be used to fit the PARMA$_\nu(p,q)$ model parameters in (1). Substitute (2) into (1) and equate the coefficients of ε_t on both sides to get

$$\begin{aligned} \psi_t(0) &= 1 \\ \psi_t(1) - \phi_t(1)\psi_{t-1}(0) &= -\theta_t(1) \\ \psi_t(2) - \phi_t(1)\psi_{t-1}(1) - \phi_t(2)\psi_{t-2}(0) &= -\theta_t(2) \\ \psi_t(3) - \phi_t(1)\psi_{t-1}(2) - \phi_t(2)\psi_{t-2}(1) - \phi_t(3)\psi_{t-3}(0) &= -\theta_t(3) \\ &\vdots \end{aligned} \tag{6}$$

where we take $\phi_t(\ell) = 0$ for $\ell > p$ and $\theta_t(\ell) = 0$ for $\ell > q$. For example, in the special case of a PARMA$_\nu(1,1)$ model

$$X_t = \phi_t X_{t-1} + \varepsilon_t - \theta_t \varepsilon_{t-1} \tag{7}$$

the system of equations (6) reduces to

$$\psi_t(1) - \phi_t = -\theta_t \quad \text{and } \psi_t(2) - \phi_t \psi_{t-1}(1) = 0 \tag{8}$$

which can be solved to obtain estimates for θ_t and ϕ_t. Then a continuous mapping argument shows that the resulting estimates are consistent and asymptotically normal, in particular

$$N_y^{1/2}\left(\hat{\phi}_i - \phi_i\right) \Rightarrow \mathcal{N}\left(0, w_{\phi i}^2\right) \quad \text{and} \quad N_y^{1/2}\left(\hat{\theta}_i - \theta_i\right) \Rightarrow \mathcal{N}(0, w_{\theta i}^2) \tag{9}$$

where

$$\begin{aligned} w_{\phi_i}^2 &= \psi_{i-1}^{-4}(1)\left\{\psi_i^2(2)\sigma_{i-2}^{-2}\sigma_{i-1}^2\left(1 - \frac{2\psi_i(1)\psi_{i-1}(1)}{\psi_i(2)}\right) + \psi_{i-1}^2(1)\sigma_{i-2}^{-2}\sum_{n=0}^{1}\sigma_{i-n}^2\psi_i^2(n)\right\} \\ w_{\theta_i}^2 &= \psi_{i-1}^{-4}(1)\left\{\psi_i^2(2)\sigma_{i-2}^{-2}\sigma_{i-1}^2\left(1 - \frac{2\psi_i(1)\psi_{i-1}(1)}{\psi_i(2)}\right) + \sum_{j=1}^{2}\psi_{i-1}^{4/j}(1)\sigma_{i-j}^{-2}\sum_{n=0}^{j-1}\sigma_{i-n}^2\psi_i^2(n)\right\}. \end{aligned}$$

Complete mathematical details will be included in a forthcoming paper (Tesfaye, Meerschaert and Anderson, 2006a). The α-level confidence intervals for ϕ_i and θ_i are

$$(\hat{\phi}_i - z_{\alpha/2} N_y^{-1/2} w_{\phi i}, \hat{\phi}_i + z_{\alpha/2} N_y^{-1/2} w_{\phi i})$$

$$(\hat{\theta}_i - z_{\alpha/2} N_y^{-1/2} w_{\theta i}, \hat{\theta}_i + z_{\alpha/2} N_y^{-1/2} w_{\theta i})$$

where $z_{\alpha/2}$ is the tail quantile of a standard normal distribution ($z_{\alpha/2} = 1.960$ for $\alpha = 0.05$). Parameters whose confidence interval contains zero can be excluded, to obtain a more parsimonious time series model, see Tesfaye, Meerschaert, and Anderson (2006) for an illustration. For low-resolution (e.g., monthly or quarterly) time series, this method can be effective for model selection and parameter fitting. For high-resolution models (e.g., weekly) the number of nonzero parameters becomes unwieldy, and alternative methods are desirable. In the next section, we discuss discrete Fourier transform methods that are appropriate in such cases.

3 Discrete Fourier Transform Methods

The PARMA$_\nu(p,q)$ model (1) has $(p+q+1)\nu$ total parameters. For example, for a weekly series ($\nu = 52$) where $p = q = 1$, there are 156 parameters. It is not advisable to fit this many parameters, since they are typically highly correlated, and therefore individually meaningless. An alternative is to replace the periodic sequences of parameters with their discrete Fourier transforms. When the period ν is large, the model parameters should vary smoothly with respect to time and can therefore be explained by only a few Fourier coefficients. Experience in using Fourier analysis for estimating periodic parameters of hydrologic time series shows that for high-resolution series, such daily and weekly series, only the first few harmonics are necessary for a good Fourier fit of the model parameters $\theta_t(j)$, $\theta_t(j)$ and σ_t (Salas, Delleur, Yevjevich, and Lane, 1980). In some cases, a larger period ν may actually result in smoother parameter functions, and hence (ironically) fewer Fourier coefficients may be required. In this section, we outline practical statistical tests to determine which Fourier coefficients are needed in the model. These results are useful for model selection. Each test is predicated on the following theorem.

Theorem 3.1 *Let $X_t = \tilde{X}_t - \mu_t$, where X_t is the periodic moving average (2) and μ_t is a periodic mean function with period ν. Then, under mild regularity conditions (see Theorem 2 in Anderson and Meerschaert (2005)) and letting $\hat{\theta}^{(\langle i-k \rangle)}_{k,\ell} = \hat{\psi}_i(\ell)$, we have for any nonnegative integers j and h with $j \neq h$,*

$$N_y^{1/2} \begin{pmatrix} \hat{\psi}(j) - \psi(j) \\ \hat{\psi}(h) - \psi(h) \end{pmatrix} \Rightarrow \mathcal{N}\left(0, \begin{pmatrix} V_{jj} & V_{jh} \\ V_{hj} & V_{hh} \end{pmatrix}\right) \tag{10}$$

and for $x = Min(h,j)$

$$V_{jh} = \sum_{n=0}^{x-1} \left\{ F_{j-1-n} \Pi^{(\nu-1)(j-1-n)} B_{n+1} \left(F_{h-1-n} \Pi^{(\nu-1)(h-1-n)} \right)^T \right\} \tag{11}$$

where

$$\begin{aligned} F_n &= \mathrm{diag}\{\psi_0(n), \psi_1(n), \ldots, \psi_{\nu-1}(n)\} \\ B_n &= \mathrm{diag}\{\sigma_0^2 \sigma_{0-n}^{-2}, \sigma_1^2 \sigma_{1-n}^{-2}, \ldots, \sigma_{\nu-1}^2 \sigma_{\nu-1-n}^{-2}\} \end{aligned} \tag{12}$$

and Π an orthogonal $\nu \times \nu$ cyclic permutation matrix,

$$\Pi = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}. \tag{13}$$

PROOF. This result follows from a rearrangement of rows and columns of the matrix V in Equation (19) from Theorem 2 in Anderson and Meerschaert (2005). This covers the case where the noise sequence has finite fourth moments. A more comprehensive treatment in Anderson, Kavalieris and Meerschaert (2006) shows that the same result holds assuming only finite second moments.

We begin with the moving average model (2), and we define the discrete Fourier transforms

$$\psi_t(j) = c_0(j) + \sum_{r=1}^{k} \left\{ c_r(j) \cos\left(\frac{2\pi r t}{\nu}\right) + s_r(j) \sin\left(\frac{2\pi r t}{\nu}\right) \right\} \tag{14}$$

where $c_r(j)$ and $s_r(j)$ are the Fourier coefficients, r is the harmonic and k is the total number of harmonics, which is equal to $\nu/2$ or $(\nu-1)/2$ depending on whether ν is even or odd, respectively. For instance, for monthly series where $\nu = 12$, we have $k = 6$; for weekly series with $\nu = 52$, $k = 26$ and for daily series with $\nu = 365$, $k = 182$. In practice, a small number of harmonics $k^* < k$ can be used to approximate the periodic function. The Fourier coefficients can be computed by way of the inverse transform

$$\begin{aligned} c_r(j) &= \nu^{-1} \sum_{m=0}^{\nu-1} \cos\left(\frac{2\pi r m}{\nu}\right) \psi_m(j) && (r = 0 \text{ or } \nu/2) \\ c_r(j) &= 2\nu^{-1} \sum_{m=0}^{\nu-1} \cos\left(\frac{2\pi r m}{\nu}\right) \psi_m(j) && (r = 1, 2, \ldots, [(\nu-1)/2]) \\ s_r(j) &= 2\nu^{-1} \sum_{m=0}^{\nu-1} \sin\left(\frac{2\pi r m}{\nu}\right) \psi_m(j) && (r = 1, 2, \ldots, [(\nu-1)/2]) \end{aligned} \tag{15}$$

and similarly, we can write the Fourier coefficients of the estimated model parameters $\hat{\psi}_m(j)$ in terms of their discrete Fourier coefficients $\hat{c}_r(j)$ and $\hat{s}_r(j)$. Using Theorem 3.1 and continuous mapping arguments, it follows that these estimates are consistent and asymptotically normal. Complete mathematical details can be found in a forthcoming paper (Tesfaye, Meerschaert and Anderson, 2006a). These results can also be used to develop a simple test to show which Fourier coefficients need to be included in the time series model. Under the null hypothesis that the mean-standardized process (2) is stationary with $\psi_t(j) = \psi(j)$ and $\sigma_t = \sigma$, the estimated Fourier coefficients $\hat{c}_r(j)$ and $\hat{s}_r(j)$ are asymptotically independent with

$$\begin{array}{lll} N_y^{1/2}\{\hat{c}_m(h) - \mu_m(h)\} & \Rightarrow \mathcal{N}(0, \nu^{-1}\eta_V(h)) & (m = 0 \text{ or } \nu/2) \\ N_y^{1/2}\{\hat{c}_m(h) - \mu_m(h)\} & \Rightarrow \mathcal{N}(0, 2\nu^{-1}\eta_V(h)) & (m = 1, 2, \ldots, [(\nu-1)/2]) \\ N_y^{1/2}\{\hat{s}_m(h) - \mu_m(h)\} & \Rightarrow \mathcal{N}(0, 2\nu^{-1}\eta_V(h)) & (m = 1, 2, \ldots, [(\nu-1)/2]) \end{array} \quad (16)$$

for all $h \geq 1$, where

$$\mu_m(h) = \begin{cases} \psi(h) & (m = 0) \\ 0 & (m > 0) \end{cases} \qquad \text{and} \quad \eta_V(h) = \sum_{n=0}^{h-1} \psi^2(n). \quad (17)$$

For example, if ν is odd then $\{\hat{c}_1(j), \hat{s}_1(h), \ldots, \hat{c}_{(\nu-1)/2}(h), \hat{s}_{(\nu-1)/2}(h)\}$ form $\nu - 1$ asymptotically independent normal random variables with mean zero and standard error $(2\nu^{-1}\hat{\eta}_V(h)/N_y)^{1/2}$. The Bonferroni α-level test statistic rejects the null hypothesis that $c_m(h)$ and $s_m(h)$ are zero for all $m \geq 1$ if $|Z_c| > z_{\alpha'/2}$ and $|Z_s| > z_{\alpha'/2}$, respectively, and

$$Z_c = \frac{\hat{c}_m(h)}{(\lambda\hat{\eta}_V(h)/N_y)^{1/2}} \ , \ Z_s = \frac{\hat{s}_m(h)}{(\lambda\hat{\eta}_V(h)/N_y)^{1/2}} \quad (18)$$

where

$$\lambda = \begin{cases} \nu^{-1} & (m = \nu/2) \\ 2\nu^{-1} & (m = 1, 2, \ldots, [(\nu-1)/2]) \end{cases} \qquad \text{and} \quad \hat{\eta}_V(h) = \sum_{n=0}^{h-1} \hat{\psi}^2(n) \quad (19)$$

and $\alpha' = \alpha/(\nu - 1)$. When $\alpha = 5\%$ and $\nu = 12$, $\alpha' = 0.05/11 = 0.0045$, $z_{\alpha'/2} = z_{0.0023} = 2.84$, and the null hypothesis is rejected when any $|Z_{c,s}| > 2.84$. Hence we should include the corresponding Fourier coefficients in our monthly model if the corresponding test statistic (18) exceeds 2.84 in absolute value. The remaining Fourier coefficients are statistically insignificant, and may be set to zero in our model. As another example, a 99% Bonferroni test for weekly data uses $\alpha = 1\%$ and $\nu = 52$, so that $\alpha' = 0.01/51 = 0.000196$ and $z_{\alpha'/2} = z_{0.000098} = 3.72$, so we would include the corresponding Fourier coefficients in our weekly model (with 99% certainty) if the absolute value of the corresponding test statistic (18) exceeds 3.72. This procedure can be useful to determine how many nonzero coefficients are needed in a moving average model.

For more complicated PARMA$_\nu(p, q)$ models, a similar procedure can be developed. The Fourier series representation of the parameters $\phi_t(\ell), \theta_t(\ell)$ and σ_t in (1) are

$$\begin{aligned} \theta_t(\ell) &= c_{a0}(\ell) + \sum_{r=1}^{k} \left\{ c_{ar}(\ell)\cos\left(\frac{2\pi rt}{\nu}\right) + s_{ar}(\ell)\sin\left(\frac{2\pi rt}{\nu}\right) \right\} \\ \phi_t(\ell) &= c_{b0}(\ell) + \sum_{r=1}^{k} \left\{ c_{br}(\ell)\cos\left(\frac{2\pi rt}{\nu}\right) + s_{br}(\ell)\sin\left(\frac{2\pi rt}{\nu}\right) \right\} \\ \sigma_t &= c_{d0} + \sum_{r=1}^{k} \left\{ c_{dr}\cos\left(\frac{2\pi rt}{\nu}\right) + s_{dr}\sin\left(\frac{2\pi rt}{\nu}\right) \right\} \end{aligned} \quad (20)$$

and k is the total number of harmonics as in (14). The estimated model parameters $\hat{\theta}_t(\ell)$, $\hat{\phi}_t(\ell)$, and $\hat{\sigma}_t$ can be written in terms of their Fourier coefficients, $\hat{c}_{ar}(\ell)$ and $\hat{s}_{ar}(\ell)$ and so forth, in a similar manner. The asymptotic behavior of these parameter estimates can be obtained by writing the Fourier coefficients in terms of the $\psi_t(j)$ weights using (6) and the inverse Fourier transform, and applying continuous mapping arguments, see Tesfaye, Meerschaert and Anderson (2006a). In general, these computations are difficult because the system of equations (6) leads to complex nonlinear relations. In this paper we consider the special case of a PARMA$_\nu(1,1)$ model, which has been proven useful in previous studies of natural river flows (see, e.g., Anderson and Meerschaert (1998)).

Under the null hypothesis that the mean-standardized PARMA$_\nu(1,1)$ model (7) is stationary with $\phi_t = \phi$, $\theta_t = \theta$ and $\sigma_t = \sigma$, the Fourier coefficients are asymptotically independent with

$$\begin{aligned} N_y^{1/2}\{\hat{c}_{am} - \mu_{am}\} &\Rightarrow \mathcal{N}(0, \nu^{-1}\eta_S) \qquad (m = 0 \text{ or } \nu/2) \\ N_y^{1/2}\{\hat{c}_{am} - \mu_{am}\} &\Rightarrow \mathcal{N}(0, 2\nu^{-1}\eta_S) \qquad (m = 1, 2, \ldots, [(\nu-1)/2]) \\ N_y^{1/2}\{\hat{s}_{am} - \mu_{am}\} &\Rightarrow \mathcal{N}(0, 2\nu^{-1}\eta_S) \qquad (m = 1, 2, \ldots, [(\nu-1)/2]) \end{aligned} \tag{21}$$

where

$$\mu_{am} = \begin{cases} \theta & (m = 0) \\ 0 & (m > 0) \end{cases} \tag{22}$$

$$\eta_S = \psi^{-4}(1)\left\{\psi^2(2)\left(1 - \frac{2\psi^2(1)}{\psi(2)}\right) + \sum_{j=1}^{2}\psi^{4/j}(1)\sum_{n=0}^{j-1}\psi^2(n)\right\} \tag{23}$$

while

$$\begin{aligned} N_y^{1/2}\{\hat{c}_{bm} - \mu_{bm}\} &\Rightarrow \mathcal{N}(0, \nu^{-1}\eta_Q) \qquad (m = 0 \text{ or } \nu/2) \\ N_y^{1/2}\{\hat{c}_{bm} - \mu_{bm}\} &\Rightarrow \mathcal{N}(0, 2\nu^{-1}\eta_Q) \qquad (m = 1, 2, \ldots, [(\nu-1)/2]) \\ N_y^{1/2}\{\hat{s}_{bm} - \mu_{bm}\} &\Rightarrow \mathcal{N}(0, 2\nu^{-1}\eta_Q) \qquad (m = 1, 2, \ldots, [(\nu-1)/2]) \end{aligned} \tag{24}$$

where

$$\mu_{bm} = \begin{cases} \phi & (m = 0) \\ 0 & (m > 0) \end{cases} \tag{25}$$

$$\eta_Q = \psi^{-4}(1)\left\{\psi^2(2)\left(1 - \frac{2\psi^2(1)}{\psi(2)}\right) + \psi^2(1)\sum_{n=0}^{1}\psi^2(n)\right\} \tag{26}$$

and $\psi(1) = \phi - \theta$, and $\psi(2) = \phi\psi(1)$ throughout, see Tesfaye, Meerschaert and Anderson (2006a) for complete mathematical details. These asymptotic formulae can be used to construct tests to determine which Fourier coefficients should be included in the time series model. The Bonferroni α-level test statistic rejects the null hypothesis that c_{am} and s_{am} are zero for all $m \geq 1$ if $|Z_c| > z_{\alpha'/2}$ and $|Z_s| > z_{\alpha'/2}$, respectively, where $\alpha' = \alpha/(\nu - 1)$,

$$Z_c = \frac{\hat{c}_{am}(h)}{(\lambda\hat{\eta}_S/N_y)^{1/2}}\ ,\ Z_s = \frac{\hat{s}_{am}(h)}{(\lambda\hat{\eta}_S/N_y)^{1/2}} \tag{27}$$

and λ is from (19). Similarly, the test for c_{bm} and s_{bm} follows the same procedure using the test statistics

$$Z_c = \frac{\hat{c}_{bm}(h)}{(\lambda\hat{\eta}_Q/N_y)^{1/2}}\ ,\ Z_s = \frac{\hat{s}_{bm}(h)}{(\lambda\hat{\eta}_Q/N_y)^{1/2}}. \tag{28}$$

4 Simulation Study

A detailed simulation study was conducted to demonstrate the methods of the previous section using simulated data from different PARMA$_\nu(p,q)$ models. For each model, individual realizations of $N_y = 50$, 100, 300, and 500 years of data (i.e., sample size of $N = N_y\nu$) were simulated and the innovations algorithm was used to obtain parameter estimates for each realization. In each case, estimates were obtained for $k = 15$ iterations. A sensitivity analysis on k indicated that increasing the number of iterations did not significantly improve the results. Then discrete Fourier transformed innovation estimates and model parameters were obtained, and test statistics were computed using (18), (28) and (27) to identify those that were statistically significant. A FORTRAN program was used to simulate the PARMA samples as well as to make all the necessary calculations.

As an example, we summarize here the results of two particular cases of a $\text{PARMA}_{12}(p, q)$ model. We first simulated a $\text{PARMA}_{12}(0, 1)$ monthly moving average model

$$X_{k\nu+i} = \varepsilon_{k\nu+i} + \theta_i \varepsilon_{k\nu+i-1} \tag{29}$$

where we assumed the periodic model parameters

$$\begin{aligned} \theta_i &= c_{a0} + \sum_{r=1}^{2} \left\{ c_{ar} \cos\left(\frac{2\pi k i}{12}\right) + s_{ar} \sin\left(\frac{2\pi r i}{12}\right) \right\} \\ \sigma_i &= 2.00 + 0.15 \cos \tfrac{2\pi i}{12} + 0.90 \sin \tfrac{2\pi i}{12} \end{aligned} \tag{30}$$

with $c_{a0} = 0.45$, $c_{a1} = 0.25$, $s_{a1} = 0.75$, $c_{a2} = 0.80$ and $s_{a2} = 0.50$, and the innovations were simulated by setting $\varepsilon_{k\nu+i} = \sigma_i \delta_{k\nu+i}$ where $\{\delta_t\}$ was generated as an independent and identically distributed sequence of normal random variables with mean zero and standard deviation one. A single realization with $N_y = 500$ years of data (sample size of $N = 6000$) was generated.

Table 1 shows the results after $k = 15$ iterations of the innovations algorithm. Fourier coefficients with test statistics $z > 3.32$ (Bonferroni test for $\nu = 12$ and $\alpha = 1\%$) are considered to be significant, indicated by a * in Table 1. The first thing to note is that most Fourier coefficients are insignificant after lag 1, which reflects the order $q = 1$ of the moving average model. Next, note that the only a few of Fourier coefficients at lag 1 are significant, in particular those with harmonic $m = 1, 2$. This correctly identifies these Fourier coefficients as the ones to include in our time series model. Finally, note that the estimated standard deviation ($\hat{\sigma}_i \approx 1.92 + 0.15\cos\frac{2\pi i}{12} + 0.88\sin\frac{2\pi i}{12}$) is close to the true values. In a real data analysis, the next step would be to analyze the residuals to determine the adequacy of the model. For this moving average model, the standardized residuals can be computed from the theoretical formula

$$\delta_t = \frac{X_t + \sum_{j=1}^{\infty} (-1)^j \theta_t \theta_{t-1} ... \theta_{t-j+1} X_{t-j}}{\sigma_t} \tag{31}$$

Residuals obtained by substituting the estimated parameter values into this formula appear independent and normally distributed with mean zero and variance one, providing further evidence that the model fit is adequate. In this case, considering the discrete Fourier transform reduces the model of the moving average component from 12 to 5 parameters without any real loss in accuracy.

Next we consider a mixed autoregressive moving average $\text{PARMA}_{12}(1, 1)$ model (7) with periodic parameters given by (20) with coefficients $c_{a0} = 0.35$, $c_{a1} = 0.15$, $s_{a1} = 0.40$, $c_{a2} = 0.25$, $s_{a2} = 0.35$, and $c_{b0} = 0.35$, $c_{b1} = 0.25$, $s_{b1} = 0.35$, $c_{b2} = 0.45$, $s_{b2} = -0.15$ (two harmonics), and a constant innovation variance $\sigma_i^2 = 1$ for all i. From the above model, a single realization with $N_y = 500$ years of monthly data (sample size of $N = 6000$) was generated. After $k = 15$ iterations of the innovations algorithm, the discrete Fourier transform of $\hat{\psi}_i$ weights (not shown) do not generally cut-off to statistically zero at a certain lag. In this case, it is not advisable to adopt a low-order moving average model, and model selection must consider various autoregressive or mixed autoregressive moving average models. Fitting an autoregressive model of order one lead to residuals that appeared correlated (not shown), so we next fit a mixed $\text{PARMA}_{12}(1, 1)$ model.

The Fourier coefficients of the model parameter estimates are summarized in Table 2. It it clear that the estimated model parameters ($\hat{\theta}_t$, $\hat{\phi}_t$) of the $\text{PARMA}_{12}(1, 1)$ model are dominated by the

Table 1: Discrete Fourier transform of moving average parameter estimates $\hat{\psi}_i(\ell)$ at season i and lag $\ell = 1, \ldots, 4$, and standard errors (SE), after $k = 15$ iterations of the innovations algorithm applied to $N_y = 500$ years of simulated $\text{PARMA}_{12}(0,1)$ monthly moving average data. Values in $(\cdots)$ are the corresponding test statistics from (18).

		harmonic m						
lag	Coefficient	0	1	2	3	4	5	6
1	$\hat{c}_m(1)$	0.463	0.282*	0.806*	-0.013	-0.008	-0.000	0.029
			(15.434)	(44.143)	(-0.700)	(-0.436)	(-0.016)	(2.230)
	$\hat{s}_m(1)$		0.725*	0.519*	0.005	-0.028	-0.013	
			(39.711)	(28.441)	(0.288)	(-1.544)	(-0.719)	
	SE		0.018	0.018	0.018	0.018	0.018	0.013
2	$\hat{c}_m(2)$	0.015	0.053	0.027	0.049	0.001	-0.011	-0.015
			(2.678)	(1.354)	(2.466)	(0.054)	(-0.536)	(-1.073)
	$\hat{s}_m(2)$		0.001	-0.004	-0.037	-0.008	0.006	
			(0.058)	(-0.210)	(-1.887)	(-0.456)	(0.299)	
	SE		0.020	0.020	0.020	0.020	0.020	0.014
3	$\hat{c}_m(3)$	0.010	0.035	0.091*	0.049	0.038	0.006	0.017
			(1.750)	(4.604)	(2.495)	(1.939)	(0.319)	(1.219)
	$\hat{s}_m(3)$		-0.024	-0.004	0.031	-0.015	-0.007	
			(-1.211)	(-0.220)	(1.579)	(-0.741)	(-0.370)	
	SE		0.020	0.020	0.020	0.020	0.020	0.014
4	$\hat{c}_m(4)$	0.007	0.027	0.045	-0.007	-0.008	-0.018	-0.021
			(1.349)	(2.270)	(-0.352)	(-0.408)	(-0.917)	(-1.490)
	$\hat{s}_m(4)$		0.016	0.067*	0.022	0.036	0.010	
			(0.798)	(3.421)	(1.091)	(1.815)	(0.523)	
	SE		0.020	0.020	0.020	0.020	0.020	0.014
:	:	:	:	:	:	:	:	:

*Fourier coefficients with test statistic ≥ 3.32

first two harmonics, and so we can set the remaining Fourier coefficients equal to zero in our model. Residual analysis based on the formula

$$\hat{\sigma}_t \hat{\delta}_t = X_t - \left(\hat{\phi}_t + \hat{\theta}_t\right) X_{t-1} + \sum_{j=2}^{\infty} (-1)^j \left(\hat{\phi}_{t-j+1} + \hat{\theta}_{t-j+1}\right) \hat{\theta}_t \hat{\theta}_{t-1} \ldots \hat{\theta}_{t-j+2} X_{t-j} \tag{32}$$

indicated that the model residuals are approximately independent and identically distributed, further evidence of an adequate fit. In this case, using discrete Fourier transforms reduces the monthly moving average and autoregressive components of the model from 24 to 10 parameters. For higher resolution models (weekly or daily) we expect that the improvement in efficiency would be even greater. In the next section, we explore this idea using some actual river flow data.

5 Stochastic Modeling of River Flows

In this section, we outline a step-by-step procedure for finding a parsimonious model for a river flow time series. The authors have found from experience that it is prudent to initially fit a $\text{PARMA}_\nu(1,1)$ model to the data. Note that this includes the $\text{PAR}_\nu(1)$ and the $\text{PMA}_\nu(1)$ model. Thus, we start by fitting a $\text{PARMA}_\nu(1,1)$ model to the data using (27) and (28) to determine which Fourier coefficients should be included in describing the estimated autoregressive parameters, $\hat{\phi}_t$, and the estimated moving average parameters, $\hat{\theta}_t$. We next use equation (32) to compute residuals. The validation of a time series model is tantamount to the application of diagnostic checks to the model residuals to see if they resemble white noise. We use the Ljung-Box test to test the white-noise null hypothesis (see Brockwell and Davis, 1991). If the null hypothesis of white-noise

Table 2: Discrete Fourier transform of model parameters estimates and standard errors (SE) for simulated $\text{PARMA}_{12}(1,1)$ data. Values in $(\cdots)$ are the test statistics (28) and (27).

		harmonic m						
Parameter	Statistic	0	1	2	3	4	5	6
$\hat{\theta}_t$	$\hat{c}_{am}$	0.361	0.181*	0.228*	-0.023	-0.048	0.009	-0.020
			(5.711)	(7.204)	(-0.715)	(-1.529)	(0.289)	(-0.895)
	$\hat{s}_{am}$		0.418*	0.314*	0.021	-0.021	-0.047	
			(13.234)	(9.916)	(0.668)	(-0.652)	(-1.473)	
	SE		0.032	0.032	0.032	0.032	0.032	0.022
$\hat{\phi}_t$	$\hat{c}_{bm}$	0.336	0.229*	0.455*	-0.015	0.028	0.009	0.019
			(6.280)	(12.471)	(-0.406)	(0.760)	(0.244)	(0.733)
	$\hat{s}_{bm}$		0.343*	-0.133*	-0.006	-0.004	0.034	
			(9.404)	(-3.652)	(-0.162)	(-0.104)	(0.943)	
	SE		0.022	0.022	0.022	0.022	0.022	0.015

*Fourier coefficients with test statistic ≥ 3.32

residuals is not rejected, and if the autocorrelation and partial autocorrelation functions of the residuals show no evidence of serial correlation, then we judge the model to be adequate. Fitting a suitable distribution to the residuals allows for a faithful simulation based on this model. To obtain additional parsimony, it is also permissible to consider simpler models where some statistically significant model parameters (e.g., high-frequency Fourier coefficients) are set to zero. If the resulting model residuals pass the same diagnostic tests, then the simplified model is also deemed adequate. Finally, if we reject the null hypothesis that the $\text{PARMA}_\nu(1,1)$ model residuals resemble iid noise, we would abandon that model and fit a $\text{PMA}_\nu(q)$ model to the data, $q \geq 2$. Using Theorem 3.1 and the test statistics given in (18), we would identify the order, q, of the pure moving average and then parsimoniously estimate the moving average parameters that we deem to be nonzero, using (18) to determine the statistically significant model parameters. However, in our experience, the $\text{PARMA}_\nu(1,1)$ model is generally adequate to model most river flow time series.

Now we illustrate the general approach with a typical data set. The Fraser River is the longest river in BC, travelling almost 1400 km and sustained by a drainage area covering 220,000 km^2. It rises in the Rocky Mts., at Yellowhead Pass, near the British Columbia-Alta. line and flows northwest through the Rocky Mt. Trench to Prince George, thence south and west to the Strait of Georgia at Vancouver. Its main tributaries are the Nechako, Quesnel, Chilcotin, and Thompson rivers. To begin, we model a monthly river flow time series from the Fraser River at Hope, British Columbia. The data are obtained from discharge measurements in cubic meter per second, averaged over each of the respective months. The resulting series contains 72 years of data from October 1912 to September 1984. In the following analysis, $\nu = 0$ corresponds to October and $\nu = 11$ corresponds to September, corresponding to the usual "water year" designed to minimize serial correlation between years.

A previous study in Tesfaye, Meerschaert, and Anderson (2006) found that a $\text{PARMA}_{12}(1,1)$ model gave a reasonable fit for this monthly series. A statistical analysis of the seasonal mean and covariance functions indicated that a stationary model was not appropriate, due to statistically significant differences in the seasonal mean, and the seasonal lag 1 and lag 2 autocorrelations. Figure 1 shows the autoregressive and moving average parameter functions along with the 95% confidence intervals computed from (9). The monthly variations are fairly smooth, indicating that the Fourier transforms may be dominated by a few terms. Following the same procedure as in the previous section, we calculated the Fourier transformed values of the model parameters, the standard errors and the test statistics. Fourier coefficients with test statistics $z > 3.32$ ($c_{a0} = 0.304$, $s_{a1} = -0.426$, $c_{a3} = 0.665$, $c_{b0} = 0.337$, $s_{b1} = 0.466$, $c_{b2} = 0.408$, $s_{b2} = 0.355$ and $c_{b3} = -0.649$) were considered to be significant, and the remaining coefficients were set to zero to achieve model parsimony. The resulting reduced Fourier approximations are plotted against the original parameter curves in Figure 1. We see from the figure that a parsimonious fit is reasonably achieved using 3 harmonics to govern the periodic coefficients. To validate the reduced model, the standardized residuals were computed using equation (32). The autocorrelation function and partial autocorrelation function (ACF and PACF, not shown) of the residuals show no serial dependence. The p-value from the Ljung-Box test was 0.27 indicating that we do not reject the null hypothesis that the residuals resemble iid white noise. It is interesting to note that in the Fourier-transformed model, the residuals are 'whiter' than in the untransformed case considered in Tesfaye, Meerschaert, and Anderson (2006), where the p-value from the Ljung-Box test was 0.08. Thus, this is a case where the parsimonious Fourier transform model produces a better fit in all facets of the analysis. The application of discrete Fourier transforms in this case reduced the number of autoregressive and moving average parameters in the model from 24 to 8, so that superior results are obtained using fewer parameters, as a result of the discrete Fourier transform.

Next we modeled the same river flow (Fraser River at Hope, British Columbia) at a higher resolution. The weekly river flow series were obtained by averaging daily measurements over 7 days (the last week containing 8 or 9 days depending on leap year) starting from 1 October of each year. The resulting series contained 72 years of data from 1 October 1912 to 30 September 1984. In this analysis, $\nu = 0$ corresponds to the first 7 days of October and $\nu = 51$ corresponds to the last 8 or 9 days of September. The weekly statistics such as the mean, standard deviation, and lag 1 and lag 2 serial correlations are displayed in Figure 2. All of these statistics exhibit statistically significant seasonal variations, so that a stationary time series model is not adequate.

The innovations algorithm ($k = 15$ iterations) was use to estimate parameters for the moving average model (2). A statistical analysis based on the asymptotic formula (5) showed that the $\psi_t(j)$-weights do not die off quickly to (statistically) zero as j increases, so that an autoregressive or mixed autoregressive moving average model is indicated. Based on our experience with the monthly data, we fit a mixed $\text{PARMA}_{52}(1,1)$ model to the weekly data. The parameter estimates for this model, obtained using equations (7) and the estimates of the $\psi_t(j)$-weights, are shown in Figure 3. The 95% confidence bands for these parameter functions were computed from (9). When interpreting these graphs, it is important to take the confidence intervals into account, as fluctuations within these bands are irrelevant to the time series model. Next we applied the inverse Fourier transform to compute the coefficients in (20). Then the statistical test of (27) and (28) was performed to determine which of these Fourier coefficients were most important to the time series model. For the autoregressive parameters, only the zero harmonic was statistically significant, indicating that a model with a constant non-seasonal autoregressive parameter should be adequate. The results of this procedure for the moving average parameters $\hat{\theta}_t$ are summarized in Table 3. The first two cosine harmonics are highly significant, as are the fifteenth and twentieth cosine harmonics and the third sine harmonic. To preserve parsimony, we chose to include only the first two cosine harmonics in the model, leading to the parameter equations

$$\begin{aligned} \hat{\theta}_i &= 0.255 - 0.197\cos\left(\tfrac{\pi i}{26}\right) + 0.121\cos\left(\tfrac{2\pi i}{26}\right) \\ \hat{\phi}_i &= 0.732 \end{aligned} \tag{33}$$

for our weekly time series model (7). We note here that from the Ljung-Box test there was virtually no difference in the parsimonious model fitted above and the model that included the other nonzero harmonics. The p-value of the test actually decreases with the larger set of parameters. Overfitting, we think, tends to model the noise and not the signal. Hence, for the reduced model, the resulting reduced Fourier approximations are plotted against the original parameter curves in Figure 3. Note that all but a few of the approximations are within the 95% confidence bands, indicating a reasonable fit. See Table 4 for remaining model parameters, $\hat{\sigma}_t$ and $\hat{\mu}_t$. To validate the reduced model, the standardized residuals were computed using equation (32). The ACF and PACF (not shown) of the residuals indicate no serial dependence, further evidence that the fit is adequate. The application of discrete Fourier transforms in this weekly model reduced the number of autoregressive and moving average parameters in the model from 104 to 4. It is interesting to note that the (approximate) weekly parameter functions are actually smoother than their monthly counterparts. This is sensible when one considers that a higher-resolution times series model should naturally give rise to more smoothly varying coordinates. For example, the average weekly flow should vary more smoothly than the average monthly flow. Further steps towards parsimony could no doubt be achieved by Fourier transforming the mean and standard deviation parameters.

Table 3: Discrete Fourier transform of moving average parameter estimate, $\hat{\theta}_t$ (with standard error, SE = 0.017 for $m = 26$ and SE = 0.024 for $m \neq 26$) for PARMA$_{52}(1,1)$ model of average weekly flow series for Fraser River at Hope, BC. Note that the value in (.) is the test statistic (27).

	harmonics m								
	0	1	2	3	4	5	6	7	8
$\hat{c}_{am}$	0.255	-0.197*	0.121*	-0.040	0.081	-0.022	-0.031	-0.013	-0.078
		(-8.25)	(5.062)	(-1.670)	(3.388)	(-0.941)	(-1.300)	(-0.531)	(-3.283)
$\hat{s}_{am}$		-0.0198	0.081	-0.117*	0.015	-0.044	-0.041	-0.020	-0.016
		(-0.828)	(3.392)	(-4.89)	(0.633)	(-1.823)	(-1.731)	(-0.832)	(-0.680)
	9	10	11	12	13	14	15	16	17
$\hat{c}_{am}$	0.044	0.052	0.023	-0.071	0.009	0.021	0.139*	-0.053	0.017
	(1.833)	(2.169)	(0.957)	(-2.996)	(0.392)	(0.880)	(5.836)	(-2.240)	(0.696)
$\hat{s}_{am}$	0.018	0.027	-0.059	-0.034	-0.051	-0.071	0.041	0.073	-0.009
	(0.753)	(1.118)	(-2.471)	(-1.436)	(-2.154)	(-2.993)	(1.714)	(3.078)	(-0.396)
	18	19	20	21	22	23	24	25	26
$\hat{c}_{am}$	0.031	-0.007	0.113*	0.019	0.075	-0.016	0.018	-0.065	0.019
	(1.335)	(-0.280)	(4.71)	(0.795)	(3.151)	(-0.660)	(0.769)	(-2.741)	(1.110)
$\hat{s}_{am}$	-0.014	-0.040	-0.076	-0.018	-0.091	-0.041	-0.016	0.008	
	(-0.569)	(-1.696)	(-3.177)	(-0.748)	(-3.829)	(-1.716)	(-0.669)	(0.351)	

*Fourier coefficients with test statistic ≥ 3.72

Finally we applied the weekly time series model (7) with parameters (33) to see how well we could simulate synthetic river flows. The success of this endeavor depends heavily on a suitably accurate distributional model for the standardized residuals. In this case, we found through exploratory data analysis that a mixture of a lognormal (scale = 0.094, location = 2.561 and threshold = -13.01) with truncated Pareto tails gave an adequate fit. For details of this procedure, see Tesfaye, Meerschaert, and Anderson (2006) where a similar distributional model was fit to the residuals from a monthly model. The truncated Pareto distribution function is

$$F_X(x) = P(X \leq x) = \frac{1 - (\gamma/x)^\alpha}{1 - (\gamma/\beta)^\alpha} \tag{34}$$

with $0 < \gamma \leq x \leq \beta < \infty$ and $\gamma < \beta$. In this case, the parameter values for the upper tail distribution ($\hat{\beta} = 8.374$, $\hat{\gamma} = 0.547$, $\hat{\alpha} = 2.068$) were found by maximum likelihood estimation based on the 215 largest positive residuals, following the procedure in Aban, Meerschaert, and Panorska (2005). Similarly, the values for the lower tail distribution ($\hat{\beta} = 8.386$, $\hat{\gamma} = 0.667$, $\hat{\alpha} = 3.459$) were based on the 207 largest negative residuals, after a change of sign. Note that a different time scale yields a different residual distribution, compare the monthly flow model in Tesfaye, Meerschaert and Anderson (2006). Substituting the simulated innovations into the model (7) generates N_y years of simulated river flow. It is advantageous to simulate several extra years of river flows and throw out the initial years (we threw out the first 100 years in this case), to ensure that the simulated series is periodically stationary. Figure 2 shows the main statistical characteristics (mean, standard deviation and autocorrelations) for one typical realization of the synthetic river flow time series obtained by this method, against the same statistical measures for the observed time

Table 4: Other parameter estimates for $\text{PARMA}_{52}(1,1)$ model of average weekly flow series for the Fraser River at Hope, BC.

	Season ν							
Parameter	0	1	2	3	4	5	6	7
$\hat{\sigma}_t$	11341.4	9365.7	10674.6	11173.0	11442.3	10321.3	7949.6	9161.4
$\hat{\mu}_t$	73139.4	69715.0	69133.0	68332.7	66348.2	60664.8	54528.4	51692.9
	8	9	10	11	12	13	14	15
$\hat{\sigma}_t$	8064.8	6196.3	8288.3	7374.0	4567.7	5751.0	3381.8	4368.1
$\hat{\mu}_t$	49291.5	44493.0	40590.5	38387.8	36543.0	35021.9	33342.0	32970.1
	16	17	18	19	20	21	22	23
$\hat{\sigma}_t$	4699.0	3014.5	4614.0	3065.2	2229.2	3958.7	2543.7	2073.1
$\hat{\mu}_t$	32180.6	31958.4	32220.3	30982.1	29701.5	29361.2	28435.7	28440.0
	24	25	26	27	28	29	30	31
$\hat{\sigma}_t$	3805.1	4267.0	5566.1	7582.3	11230.3	17216.4	19467.5	23820.6
$\hat{\mu}_t$	29564.3	32309.3	36677.3	45680.3	61496.0	80419.7	107893.3	136771.3
	32	33	34	35	36	37	38	39
$\hat{\sigma}_t$	28120.6	29511.4	35410.9	29142.7	26127.0	29320.7	18830.2	20127.7
$\hat{\mu}_t$	172356.4	205165.0	227797.0	245244.2	255747.8	256049.1	242545.2	228583.8
	40	41	42	43	44	45	46	47
$\hat{\sigma}_t$	15124.9	14232.0	11492.2	9000.8	10763.7	10741.4	8428.7	11202.4
$\hat{\mu}_t$	211048.4	198195.4	178312.8	155861.5	139551.5	127168.7	117237.5	108533.9
	48	49	50	51				
$\hat{\sigma}_t$	8588.8	10147.6	9350.6	9845.6				
$\hat{\mu}_t$	98497.6	90001.4	81269.0	76323.0				

series. It is apparent that the Fourier-$\text{PARMA}_{52}(1,1)$ model closely reproduces the main statistical characteristics, indicating that the model is trustworthy to produce faithful reproductions of the actual river flows. Finally, Figure 4 shows a side-by-side comparison of the synthetic and actual river flow time series, and it is evident that the two series are statistically similar. This kind of synthetic river flow can be useful for drought and flood modeling, and operational analysis of water projects such as reservoirs, and treatment plants for water supply.

6 Conclusion

For analysis and design of water resource systems, it is sometimes useful to model and simulate river flows with high resolution (e.g., weekly values). Periodic autoregressive moving average (PARMA) time series models can be useful for modeling and simulating these river flows. PARMA models are indicated when the basic descriptive statistics (mean, variance, and autocorrelations) of the river flow exhibit significant seasonal variations. The innovations algorithm can be used to estimate the model parameters. Conventional PARMA models can be used at any time scale, but for high-resolution time series, the model depends on numerous parameters, since the number of parameters grows proportional to the number of seasons (days or weeks) in a year. This proliferation of parameters can lead to over-fitting, or "modeling the noise." To remedy this situation, this paper uses discrete Fourier transforms to efficiently represent the periodic model parameters in terms of just a few Fourier coefficients. Asymptotic statistical tests and confidence intervals are developed to identify the essential Fourier coefficients to include in the model. Since the parameters in a high-resolution model can be expected to vary smoothly over the course of the year, only a few Fourier coefficients are generally required. This results in a parsimonious time series model that captures the essential features of the flow, including seasonal variations in the correlation structure. A simulation study demonstrates the effective application of this procedure. A further practical application shows how the method can be fruitfully applied to generate high-resolution synthetic flows, that faithfully reproduce a weekly series. By combining the Fourier-PARMA modeling with an accurate representation of the model residuals, we generate synthetic flows that reproduce all the statistical features of the data, for an example series representing average weekly flows on the Fraser River at Hope, British Columbia, Canada. Fourier-PARMA models are applicable to any river flow time series, since they accommodate a flexible dependence structure and residual distribution. The results of this research should be useful to practitioners wishing to generate high-resolution synthetic river flows for water resource systems planning and operations analysis, as well as flood and drought modeling.

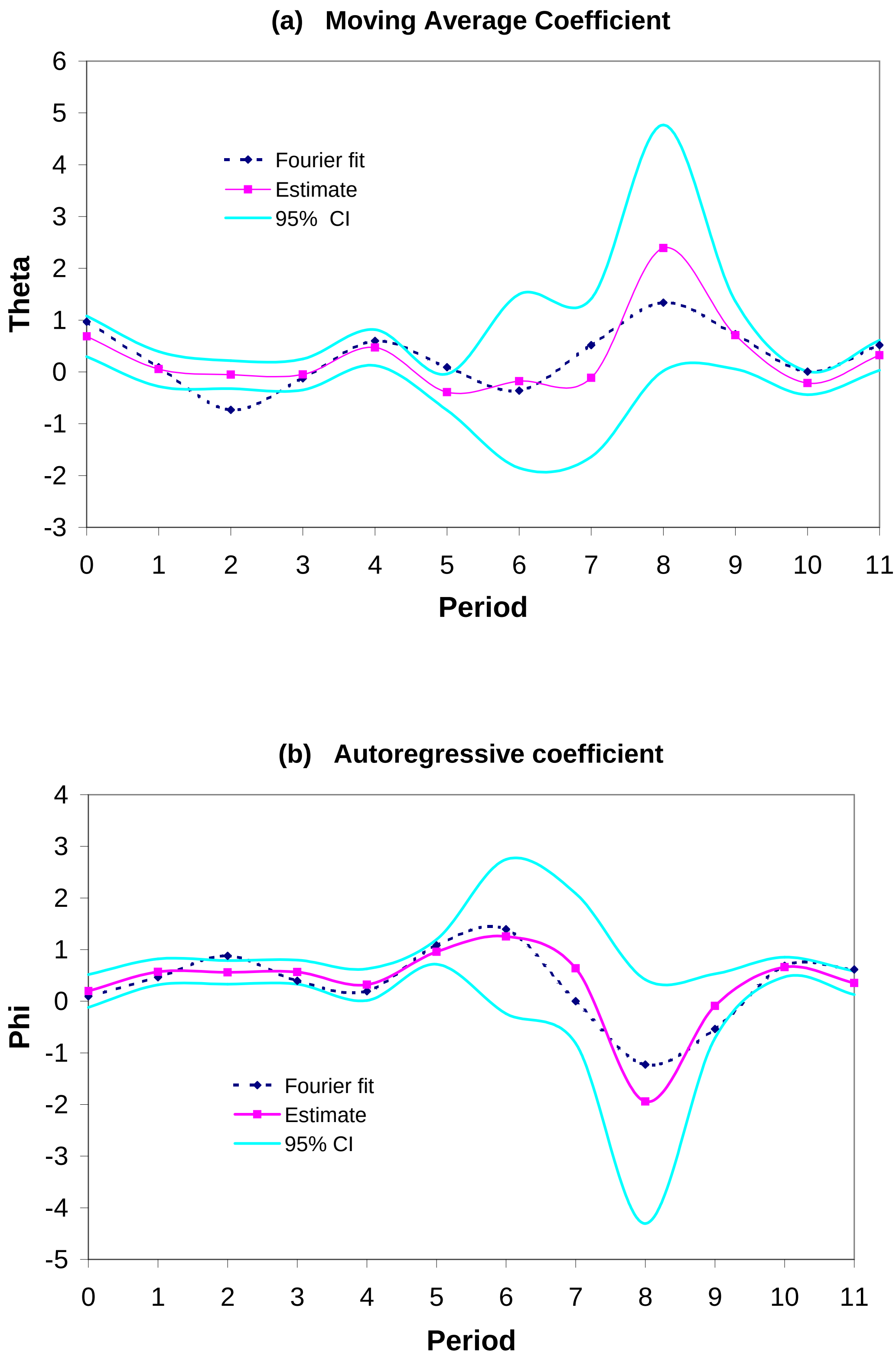


Figure 1: Plot of $PARMA_{12}(1,1)$ model parameters (with their Fourier fit) of average monthly flow series for the Fraser River at Hope, BC.

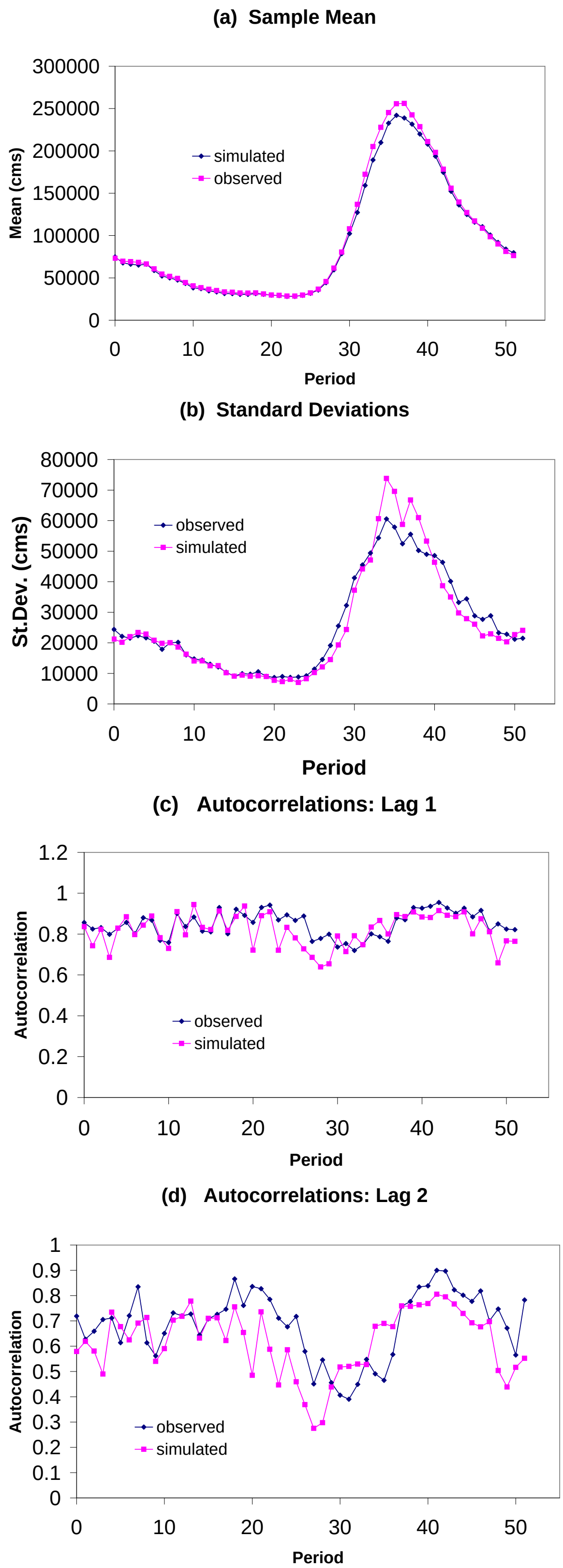


Figure 2: Comparison of mean, standard deviation, and autocorrelations for simulated vs. observed weekly river flow data for the Fraser River at Hope, BC.

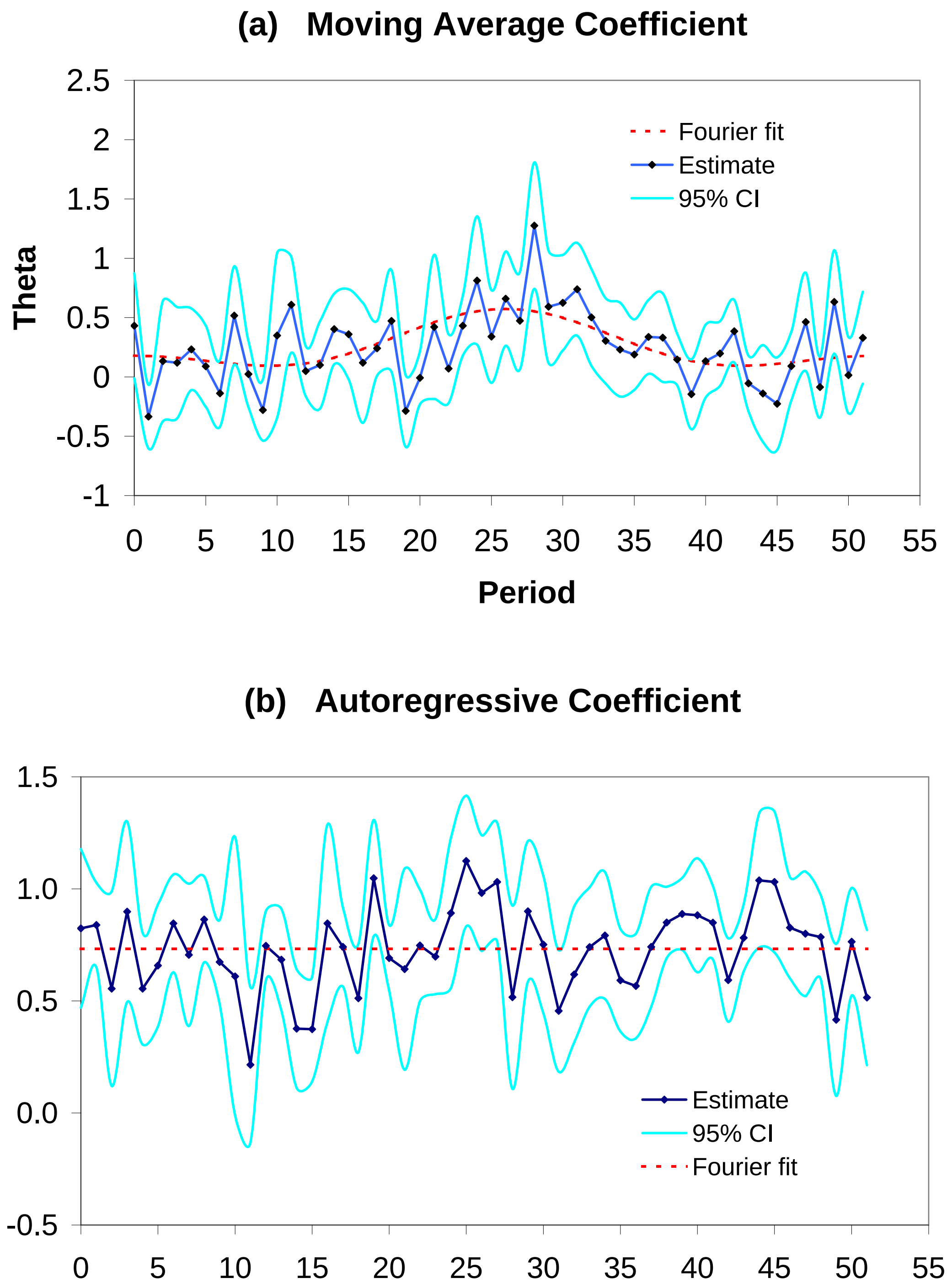
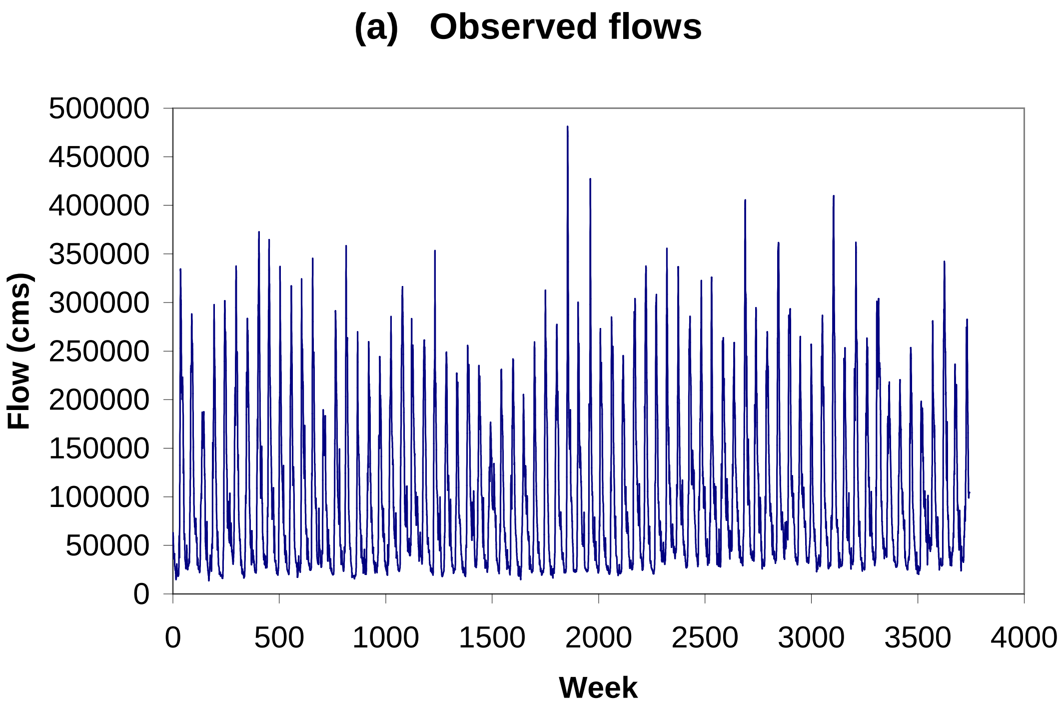


Figure 3: Plot of $\text{PARMA}_{52}(1,1)$ model parameters (with their Fourier fit) of average weekly flow series for the Fraser River at Hope, BC.



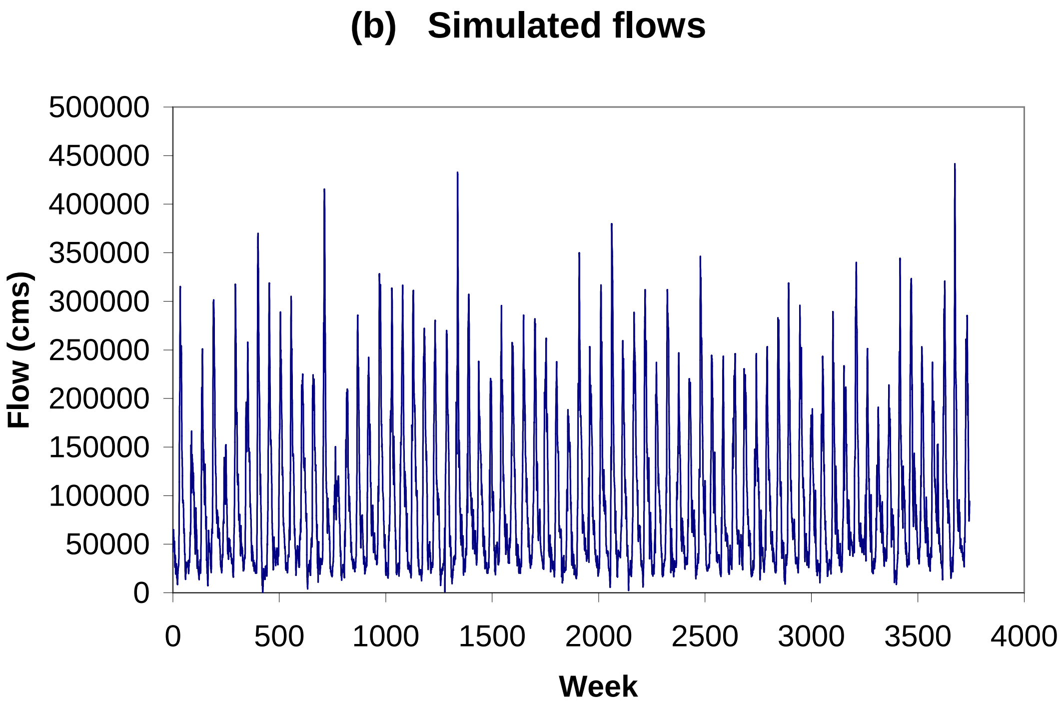


Figure 4: Plot of (a) observed and (b) simulated weekly river flows for the Fraser River at Hope, BC, indicating similarity.

References

[1] Aban, I.B, M.M. Meerschaert and A.K. Panorska (2006) Parameter estimation for the truncated Pareto distribution. *Journal of the American Statistical Association: Theory and Methods, 101*(473), pp.270–277.

[2] Adams, G.J and G.C. Goodwin (1995) Parameter estimation for periodically ARMA models. *Journal of Time Series Analysis, 16*, 127–145.

[3] Anderson, P.L. and A.V. Vecchia (1993) Asymptotic results for periodic autoregressive moving-average processes. *Journal of Time Series Analysis, 14*, 1–18.

[4] Anderson, P.L. and M.M. Meerschaert (1997) Periodic moving averages of random variables with regularly varying tails. *Annals of Statistics, 25*, 771–785.

[5] Anderson, P.L. and M.M. Meerschaert (1998) Modeling river flows with heavy tails. *Water Resources Research*, *34*(9), 2271–2280.

[6] Anderson, P.L., M.M. Meerschaert and A.V. Veccia (1999) Innovations algorithm for periodically stationary time series. *Stochastic Processes and their Applications, 83*, 149–169.

[7] Anderson, P.L. and M.M. Meerschaert (2005) Parameter estimates for periodically stationary time series. *Journal of Time Series Analysis, 26*, 489–518.

[8] Anderson, P.L., L. Kavalieris and M.M. Meerschaert (2006) Innovations Algorithm Asymptotics for Periodically Stationary Time Series with Heavy Tails. Submitted. Preprint available at `www.stt.msu.edu/%7Emcubed/innolimHT.pdf`

[9] Box, G.E., and G.M. Jenkins (1976) *Time Series Analysis: Forcasting and Control.* 2nd edition, San Francisco: Holden-Day.

[10] Brockwell, P.J. and R.A. Davis (1991), *Time Series: Theory and Methods*, 2nd ed. New York: Springer-Verlag.

[11] Chen, H.-L. and A.R. Rao (2002) Testing hydrologic time series for stationarity. *Journal of Hydrologic Engineering,7*(2), 129–136.

[12] Jones, R. H. and W. M. Brelsford (1967) Times series with periodic structure. *Biometrika, 54*, 403–408.

[13] Lund, R. B. and I. V. Basawa (1999) Modeling and inference for periodically correlated time series. *Asymptotics, Nonparameterics, and Time series.* (S. Ghosh, Ed.) 37–62. Marcel Dekker, New York.

[14] Lund, R. B. and I. V. Basawa (2000) Recursive prediction and likelihood evaluation for periodic ARMA models. *Journal of Time Series Analysis*, *20*(1), 75–93.

[15] Moss, M.E. and M.C. Bryson (1974) Autocorrelation structure of monthly streaflows. *Water Resources Research, 10*(4),737–744.

[16] Pagano, M. (1978) On periodic and multiple autoregressions. *Annals of Statistics*, *6*(6), 1310–1317.

[17] Quimpo, R.G. (1967) Stochastic modeling of daily river flow sequences. *Hydrology Paper 18*, Colorado State University, Fort Collins, Colorado.

[18] Salas, J.D, J.W. Delleur, V. Yevjevich and W.L. Lane (1980) Applied modeling of hydrologic time series. *Water Resource Publications*. Littleton, Colorado.

[19] Salas, J.D, J. T.B Obeysekera and R.A. Smith (1981) Identification of streamflow stochastic models. *ASCE Journal of the Hydraulics Division*, *107*(7), 853–866.

[20] Salas, J.D, D.C. Boes and R.A. Smith (1982) Estimation of ARMA models with seasonal parameters. *Water Resources Research*, *18*, 1006–1010.

[21] Salas, J.D, G.Q. Tabios III and P. Bartolini (1985) Approaches to multivariate modeling of water resources time series. *Water Resources Bulletin*, *21*, 683–708.

[22] Salas, J.D. (1993), Analysis and modeling of hydrologic time series. *The McGraw-Hill Handbook of Hydrology*. (D.R. Maidment, Ed.), 19.5–19.9.

[23] Shao, Q. and R.B. Lund (2004) Computation and characterization of autocorrelations and partial autocorrelations in periodic ARMA models. *Journal of Time Series Analysis*, *25*(3), 359–372.

[24] Tesfaye, Y.G. (2005) *Seasonal Time Series Models and Their Application to the Modeling of River Flows*. Doctoral Dissertation, Hydrological Sciences, University of Nevada, Reno. Available at www.stt.msu.edu/%7Emcubed/Yonas_thesis.pdf

[25] Tesfaye, Y.G., M.M. Meerschaert and P.L. Anderson (2006) Identification of PARMA models and their application to the modeling of river flows. *Water Resources Research, 42*(1), W01419 (11pp.), doi:10.1029/2004WR003772.

[26] Tesfaye, Y.G., M.M. Meerschaert and P.L. Anderson (2006a) Asymptotic results for Fourier-PARMA coefficients. In preparation.

[27] Tiao, G.C. and M.R. Grupe (1980) Hidden periodic autoregressive moving average models in time series data. *Biometrika 67*, 365-373.

[28] Tjøstheim, D. and J. Paulsen (1982) Empirical identification of multiple time series. *Journal of Time Series Analysis, 3*, 265–282.

[29] Thompstone, R.M., K.W. Hipel and A.I. McLeod (1985) Forecasting quarter-monthly riverflow. *Water Resources Bulletin*, *25*(5), 731–741.

[30] Troutman, B.M. (1979) Some results in periodic autoregression. *Biometrika, 6*, 219–228.

[31] Ula, T.A. (1990) Periodic covariance stationarity of multivariate periodic autoregressive moving average processess. *Water Resources Research*, *26*(5), 855–861.

[32] Ula, T.A. (1993) Forcasting of multivariate periodic autoregressive moving average processes. *Journal of Time Series Analysis, 14*, 645–657.

[33] Ula, T.A. and A.A Smadi (1997) Periodic stationarity conditions for periodic autoregressive moving average processess as eigenvalue problems. *Water Resources Research*, *33*(8), 1929–1934.

[34] Ula, T.A. and A.A Smadi (2003) Identification of periodic moving average models. *Communications in Statistics: Theory and Methods, 32*(12), 2465–2475.

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

[36] Vecchia, A.V. (1985b) Maximum likelihood estimation for periodic moving average models. *Technometrics*, *27*(4), 375–384.

[37] Vecchia , A.V. and R. Ballerini (1991) Testing for periodic autocorrelations in seasonal time series data. *Biometrika*, *78*(1), 53–63.