ROBUST BAYESIAN APPROACH FOR AR(p) MODELS APPLIED TO STREAMFLOW FORECASTING

†GUILHERME DE A. BARRETO ‡MARINHO G. DE ANDRADE

†Department of Teleinformatics Engineering
Federal University of Ceará (UFC), Fortaleza, CE, Brazil
guilherme@deti.ufc.br

‡Institute of Computer and Mathematical Sciences (ICMC)
University of São Paulo (USP), São Carlos, SP, Brazil
marinho@icmc.sc.usp.br

SUMMARY

In this work, we propose a Bayesian approach for the parameter estimation problem of autoregressive models of order p, AR(p), applied to streamflow forecasting. Bayesian statistics and Monte Carlo Markov Chain (MCMC) simulation give more flexibility to the estimation task, allowing several hypotheses for the prior density of the parameters to be tested and compared. The point estimates for the parameters and their credible intervals can be easily obtained from samples of the posterior density through MCMC simulation. Procedures for model selection, parameter estimation, forecasting and robustness evaluation of the estimator through MCMC are presented in detail. The proposed approach provides better results when compared with the classical maximum likelihood estimation method on a monthly streamflow time series from Furnas Reservoir in Brazil. Furthermore, MCMC simulation shows that the Bayesian estimators are robust to changes in the location parameter of the priors.

Keywords
Autoregressive models, parameter estimation, Markov Chain Monte Carlo (MCMC), streamflow forecasting, Bayesian robustness.

1 INTRODUCTION

Streamflow time series play a fundamental role in planning and operation of hydroelectric systems. For instance, a generation expansion plan must meet some reliability constraints as risk of deficit or expected energy not supplied, that could be estimated by the simulation of system operation over a large number of streamflow sequences (Maceira and Bezerra, 1997).

The streamflow forecasting problem has been commonly tackled using linear autoregressive models of order p (Awwad et al., 1994; Cheng, 1994). The purpose of these models is to extract as much information as possible from the data, by analyzing the historical record as a sample of a stochastic process and trying to estimate the parameters of this process. When the order of the model and its parameters are identified, the model can then be used to determine the probability distribution of future inflows or to generate samples of synthetic streamflows. In other words, the model provides a mathematical basis for short-, mid- or long-term planning and operation studies of the hydroelectric system. Thus, an "inadequate" streamflow model tends to affect the decision making process.

Recently, there have been an increasing interest in Bayesian statistical methods in application domains as diverse as epidemiology, biostatistics, engineering, computer science and time series (Gelman et al., 1995; Gamerman, 1997; Brooks, 1998). In addition, one of the fastest growing areas of application

and research of Markov Chain Monte Carlo methods is in time series analysis (Tweedie, 1998), specially as faster and more reliable computational resources, such as computer hardware and numerical statistical tools become widely available. In this paper, we describe and apply Bayesian statistics and Markov Chain Monte Carlo (MCMC) simulation to the problem of forecasting monthly mean streamflows for the Furnas Reservoir in Brazil. The proposed Bayesian estimation technique is compared with the classic Maximum Likelihood Estimation (MLE), also known as the Box-Jenkins method (Box et al., 1994).

We also study the problem of robustness of the Bayesian estimators with respect to the dependence of the posterior mean and variance on the discrepancy between the prior location parameter and the observed sample mean. For a normal location parameter problem with known variance, posterior moments of the parameter of interest generally do not have explicit and analytical expressions without a normal prior assumption. The use of double-exponential prior as a basis for constructing a location invariant family of priors that are almost non-informative (i.e. near ignorant) and that allow a robust Bayesian analysis of credible interval has been considered in (Pericchi and Walley, 1991). Pericchi and Smith (1992) derived the exact expressions for the first two posterior moments analytically in the problem where this prior is assumed. Choy and Smith (1997) extended the analysis to other scale mixture of normal distribution such as the exponential power and the symmetric stable distribution.

A Bayesian approach for robustness analysis has also been adopted by McCulloch and Tsay (1993) and Barbieri and Conigliani (2000). McCulloch and Tsay (1993) adopted non-informative prior distributions for the parameters of the models and use MCMC methods (i.e. Gibbs Sampler) to carry out the computations. The identification step was performed on the original series, potentially affected by a change in the mean, whose presence may result in a misidentication of the model. More recently, Barbieri and Conigliani (2000) dealt with the identification of a stationary autoregressive model for time series and the contemporary detection of a change in its mean. They also adopted the Bayesian approach with non-informative prior about the parameters of the models under comparison and an exact form of the likelihood function. To solve computational issues, they resort to *Importance Sampling* (Gelman et al., 1995). In this paper, we will use the approximate likelihood function and informative prior distributions for the parameters of the AR(p) model, namely, the Normal-Gamma and t-Gamma priors. MCMC methods, such as the Gibbs Sampler and the Metropolis-Hastings algorithms, will play an essential role during all the stages of the streamflow forecasting problem, specially to assess the robustness of the Bayesian estimators.

The remaining part of the paper is organized as follows. In Section 2, we discuss the differences between the Box-Jenkins and the Bayesian approaches to the parameter estimation problem. In Section 3, we present concepts related to stochastic simulation via MCMC methods. The model selection and the forecasting problem are addressed in Section 4 and 5, respectively. In Section 6, the simulation results are presented. The paper is concluded in Section 7.

2 PARAMETER ESTIMATION OF AR(p) MODELS

Consider the following stochastic process $\{Z_t, t \in N_+\}$. Thus, a linear, autoregressive model of order p, is that one for which the current value Z_t of the process is expressed as a linear combination of its p past values Z_{t-1}, Z_{t-2}, $\ldots$, Z_{t-p} and of a white noise a_t. The white noise process $\{a_t, t \in N_+\}$ is assumed to be Gaussian, independent and identically distributed (i.i.d), with zero mean and variance σ_a^2. The linear autoregressive model can then be written as follows:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + a_t \tag{2.1}$$

where $\Phi = (\phi_1, \phi_2, \ldots, \phi_p) \in R^p$ and $\sigma_a^2 = 1/\tau$ are the parameters of the model. The problem of fitting an AR(p) model to a given time series can be divided into the following stages:

(i) Estimation of the parameters $\Phi = (\phi_1, \phi_2, \ldots, \phi_p)$, and $\tau = 1/\sigma_a^2$ for several values of p.

(ii) Selection of the most suitable value of p according to some statistical criteria.

(iii) Use of the estimates of the parameters to forecast new values for the time series, using (2.1) recursively.

The standard approach to parameter estimation of an AR(p) model starts by determining the likelihood function $L(\mathbf{Z}|\Phi,\tau)$, where $\mathbf{Z} = (Z_1,\, Z_2,\, \ldots\, ,\, Z_N)'$ is the $N \times 1$ observed time series vector and $\Theta = (\Phi, \tau)$ is the vector of unknown parameters. It can be shown (Box et al., 1994) that the approximate likelihood function, i.e., the likelihood function for the AR(p) model in (2.1), given the first p observations $\mathbf{Z}_p = (Z_1,\, Z_2,\, \ldots\, ,\, Z_p)$, $p \ll N$, is as follows:

$$L(\mathbf{Z}|\Theta) \propto \tau^{\frac{N-p}{2}} exp\left\{-\frac{\tau}{2}\sum_{t=p+1}^{N}\left(Z_t - \phi_1 Z_{t-1} - \ldots - \phi_p Z_{t-p}\right)^2\right\} \tag{2.2}$$

The likelihood function in (2.2) has the form of a probability density, and it is supposedly responsible for the generation of the data $\mathbf{Z}$ given the parameters Θ. The problem consists of finding an estimate $\widehat{\Theta} = (\widehat{\Phi}, \widehat{\tau})$ that maximizes the probability of the data $\mathbf{Z}$ being actually generated by $L(\mathbf{Z}|\Theta)$. This technique is referred to as the *maximum likelihood estimation* and the resulting estimates are calculated as follows (Box et al., 1994; Davis and Vinter, 1985):

$$\widehat{\Phi} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{Z}\right) \tag{2.3}$$

$$\widehat{\sigma}_a^2 = \widehat{\tau}^{-1} = \frac{1}{N-p}\left(\mathbf{Z} - \mathbf{X}\widehat{\Phi}\right)'\left(\mathbf{Z} - \mathbf{X}\widehat{\Phi}\right) \tag{2.4}$$

where

$$\mathbf{Z} = \begin{pmatrix} Z_{p+1} \\ Z_{p+2} \\ \vdots \\ Z_N \end{pmatrix}_{(N-p)\times 1} \qquad \mathbf{X} = \begin{pmatrix} Z_p & \cdots & Z_1 \\ Z_{p+1} & \cdots & Z_2 \\ \vdots & \vdots & \vdots \\ Z_{N-1} & \cdots & Z_{N-p} \end{pmatrix}_{(N-p)\times p} \tag{2.5}$$

Equation (2.2) can be rewritten in matrix form using (2.3) and (2.5), as follows:

$$L(\mathbf{Z}|\Theta) \propto \tau^{\frac{N-p}{2}} \exp\{-\frac{\tau}{2}[(\Phi - \widehat{\Phi})\mathbf{X}'\mathbf{X}(\Phi - \widehat{\Phi}) + (\mathbf{Z} - \widehat{\mathbf{Z}})'(\mathbf{Z} - \widehat{\mathbf{Z}})]\} \tag{2.6}$$

The MLE method considers the parameters of the AR(p) model as constant values to be calculated. On the other side, the Bayesian Inference approach understands the parameters of the model as random variables (Gelman et al., 1995) and, as every random variable, the parameters follow a particular probability distribution. Thus, in addition to the determination of a likelihood function, the process of Bayesian inference requires the specification of a prior probability density, $p\,(\Theta)$, that summarizes the previous knowledge of the researcher about the distribution of the parameters. In other words, the prior density express our uncertainty about Θ before taking the data $\mathbf{Z}$ into account.

The likelihood function $L(\mathbf{Z}|\Theta)$ and the prior density $p\,(\Theta)$ are combined using the Bayes theorem from probability theory to yield the posterior probability density $f\,(\Theta|Z)$ (Gelman et al., 1995; Gamerman, 1997):

$$f\,(\Theta|\mathbf{Z}) = \frac{p\,(\Theta)\,L\,(\mathbf{Z}|\Theta)}{p\,(\mathbf{Z})} \tag{2.7}$$

where $p\,(\mathbf{Z})$ is a normalization factor to ensure that $\int_\Theta f\,(\Theta|\mathbf{Z})\,d\Theta = 1$. This factor is also called marginal likelihood or prior predictive density. In general, we omit the denominator in (2.7), since it is independent of Θ, and write:

$$f\,(\Theta|\mathbf{Z}) \propto p\,(\Theta)\,L\,(\mathbf{Z}|\Theta) \tag{2.8}$$

The posterior density $f\,(\Theta|\mathbf{Z})$ shows how the random variable Θ is distributed after observing the data $\mathbf{Z}$. In other words, the posterior density expresses our uncertainty about after taking the data Z into account. A lot of information can be extracted from the posterior in (2.8) (Tierney, 1994), such as the

point estimates of the parameters (given by $\widehat{\Theta}_b = E[f(\Theta|\mathbf{Z})]$), the variance of the estimator $Var(\widehat{\Theta}_b)$ and the credible interval (Bayesian equivalent of the confidence interval).

It is worth emphasizing two points here. First, the Bayesian approach is more flexible than the maximum likelihood one, since it allows to test several hypotheses for the prior density $p(\Theta)$ and choose the one that results in a posterior density suitable to the problem at hand. Second, and perhaps more important issue, it is possible that the resulting $f(\Theta|\mathbf{Z})$ does not have a standard form. Hence, mean values and variances should be obtained through the evaluation of integrals, which can become a very complex undertaking. A possible solution is to estimate the form of the posterior density $f(\Theta|\mathbf{Z})$ by generating random samples from it by means of MCMC simulation methods.

3 MARKOV CHAIN MONTE CARLO METHODS

The aim of Markov chain Monte Carlo simulation is to generate random samples $\{\Theta^{(j)} = (\Phi^{(j)}, \tau^{(j)}), j = 1, ..., m\}$ from the conditional densities $f(\Phi|\tau, \mathbf{Z})$, and $f(\tau|\Phi, \mathbf{Z})$ obtained from the joint posterior density $f(\Phi, \tau|\mathbf{Z})$ in (2.8). This sample is then used to infer the form of the probability densities governing the behavior of the parameters and to obtain the point estimates, the variance of the point estimates and their credible intervals. For instance, the Bayesian point estimates $\widehat{\Phi}_b = (\widehat{\phi}_{b,1}, \widehat{\phi}_{b,2}, ..., \widehat{\phi}_{b,p})'$ and $\widehat{\tau}_b$ are given by:

$$\widehat{\phi}_{b,i} = \frac{1}{m}\sum_{j=1}^{m} \phi_i^{(j)} \tag{3.1}$$

$$\widehat{\tau}_b = \frac{1}{m}\sum_{j=1}^{m} \tau^{(j)} \tag{3.2}$$

where $i = 1, \ldots, p$. In the next two subsections, we introduce a Bayesian estimation procedure for the parameters of the AR(p) model assuming two different prior densities, namely a conjugate Normal-Gamma prior and a t-Student prior. MCMC algorithms for obtaining the corresponding random samples, $\{\Theta^{(j)}\}$, are also presented.

3.1 Conjugate Prior Normal-Gamma

For a given choice of likelihood $L(\mathbf{Z}|\Theta)$, there is a family of priors which gives rise to a posterior $f(\Theta|\mathbf{Z})$ having a standard functional form. A prior $p(\Theta)$ with this property is said to be a *conjugate prior* (Gelman et al., 1995). In general, the resulting posterior and the conditional densities $f(\Phi|\tau, \mathbf{Z})$ and $f(\tau|\Phi, \mathbf{Z})$ also have standard forms from which a random number can be easily generated. The form of the likelihood in (2.6) suggests the following probability models: a Normal density for Φ and a Gamma density for τ. This prior will be referred to as the Normal-Gamma prior, being given by:

$$p(\Theta) = p_1(\Phi|\tau)p_2(\tau) \tag{3.3}$$

where

$$\begin{aligned} p_1(\Phi|\tau) &\propto \tau^{\frac{p}{2}} \exp\left\{-\frac{\tau}{2}\left[(\Phi - \boldsymbol{\mu})' P^{-1}(\Phi - \boldsymbol{\mu})\right]\right\} \\ (\Phi|\tau) &\sim \mathit{Normal}(\boldsymbol{\mu}, \mathbf{P}^{-1}) \end{aligned} \tag{3.4}$$

$$\begin{aligned} p_2(\tau) &\propto \tau^{\alpha-1} \exp\{-\beta\tau\} \\ (\tau) &\sim \mathit{Gamma}(\alpha, \beta) \end{aligned} \tag{3.5}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ is the location parameter vector and $\mathbf{P}^{-1}$ is the covariance matrix of the Normal density. These parameters, together with α and β of the Gamma density, are generically known as the hyperparameters of the prior $p(\Theta)$ and are specified by the user. For the Normal-Gamma prior, the joint posterior calculated according to (2.8) is given by:

$$f(\Phi, \tau|\mathbf{Z}) \propto \tau^{\frac{(N+2\alpha)}{2}-1} \exp\{-\tau D - \frac{\tau}{2}(\Phi - \widehat{\Phi}_b)'\mathbf{V}(\Phi - \widehat{\Phi})\} \tag{3.6}$$

where $\widehat{\Phi}_b$ is the theoretical Bayesian estimate of Φ:

$$\widehat{\Phi}_b = (\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Z} + \boldsymbol{\mu}\mathbf{P}) \tag{3.7}$$

and

$$\mathbf{V} = (\mathbf{X}'\mathbf{X} + \mathbf{P}) \tag{3.8}$$

$$D = \beta + \frac{1}{2}\left[(\mathbf{Z}'\mathbf{Z} + \boldsymbol{\mu}'\mathbf{P}\boldsymbol{\mu}) - (\mathbf{X}'\mathbf{Z} + \mathbf{P}\boldsymbol{\mu})'(\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Z} + \mathbf{P}\boldsymbol{\mu})\right] \tag{3.9}$$

Equation (3.7) presents the Bayesian estimate of $\widehat{\Phi}_b$ as the weighted mean of the MLE estimate in (2.3) and the location parameter $\boldsymbol{\mu}$:

$$\widehat{\Phi}_b = (\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}[\mathbf{X}'\mathbf{X}\widehat{\Phi} + \boldsymbol{\mu}\mathbf{P}] \tag{3.10}$$

Rearranging the terms in (3.10), as suggested by Pericchi and Smith (1992), we have:

$$\widehat{\Phi}_b = \widehat{\Phi} + (\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}\mathbf{P}(\boldsymbol{\mu} - \widehat{\Phi}) \tag{3.11}$$

As can be noted in (3.11), the difference $\widehat{\Phi}_b - \widehat{\Phi}$ is linearly proportional to $\boldsymbol{\mu} - \widehat{\Phi}$. This relationship will be confirmed later in Section 6 through MCMC simulation, and will be used to evaluate the robustness of the Bayesian estimator, i.e. the sensitivity of the estimator to changes in the location parameter of the prior.

From (3.6), the conditional densities are easily obtained due to the conjugacy property:

$$f(\Phi|\tau, \mathbf{Z}) \sim \mathit{Normal}\left[\widehat{\Phi}_b, \tau(\mathbf{V})^{-1}\right] \tag{3.12}$$

$$f(\tau|\Phi, \mathbf{Z}) \sim \mathit{Gamma}\left[\frac{N + 2\alpha}{2}, D + \frac{1}{2}(\Phi - \widehat{\Phi}_b)'\mathbf{V}(\Phi - \widehat{\Phi}_b)\right] \tag{3.13}$$

The densities in (3.12) and (3.13) are in a suitable form to introduce the main steps of the Gibbs-Sampler algorithm (Tweedie, 1998; Gamerman, 1997; Gelman et al., 1995; Casella and George, 1992; Geman and Geman, 1984):

- **Step 1**. Assign initial values to the parameters: $\Phi^{(0)}$ and $\tau^{(0)}$.
- **Step 2**. Obtain a new observation $\Phi^{(j+1)}$ from the conditional density in (3.12):
  $$\Phi^{(j+1)} \sim f(\Phi|\tau^{(j)}, \mathbf{Z})$$
- **Step 3**. Obtain a new observation $\tau^{(j+1)}$ from the conditional density in (3.13):
  $$\tau^{(j+1)} \sim f(\tau|\Phi^{(j+1)}, \mathbf{Z})$$
- **Step 4.** Stop if the convergence of the Markov chain has been detected. Otherwise, do $j \longleftarrow j + 1$ and return to **Step 2**.

After a sufficiently large number of iterations, the chain of observations $(\Phi^{(j)}, \tau^{(j)})$ converges and it can be treated as a sample from the joint posterior density $f(\Phi, \tau|\mathbf{Z})$. The convergence assessment of MCMC algorithms is discussed in Subsection 3.3.

3.2 Informative Prior t-Gamma

Unfortunately, conjugate priors do not exist for all likelihoods. For the general case of non-conjugacy, the posterior and its conditionals cannot be sampled directly. It occurs, for instance, if we choose for Φ a p-dimensional, t-Student prior, with ν degrees of freedom, location parameter $\boldsymbol{\mu}$ and precision matrix $\mathbf{P}$; and a Gamma density for the parameter τ. Then we have the t-Gamma prior given by:

$$p(\Theta) = p_1(\Phi)p_2(\tau) \tag{3.14}$$

so that

$$\begin{aligned} p_1(\Phi) &\propto \left[1 + \frac{(\Phi - \mu)' \mathbf{P} (\Phi - \mu)}{\nu}\right]^{-\frac{(\nu+p)}{2}} \\ (\Phi) &\sim t\text{-}Student(\mu, \mathbf{P}) \end{aligned} \tag{3.15}$$

$$\begin{aligned} p_2(\tau) &\propto \tau^{\alpha-1} \exp\{-\beta\tau\} \\ (\tau) &\sim Gamma(\alpha, \beta) \end{aligned} \tag{3.16}$$

In this case, the joint posterior is given by:

$$f(\Phi, \tau|\mathbf{Z}) \propto \tau^{\frac{N-p+2\alpha}{2}-1} \exp\left\{-\tau\left[\beta + \frac{B(\Phi)}{2}\right]\right\} p_1(\Phi) \tag{3.17}$$

in which

$$B(\Phi) = (\Phi - \widehat{\Phi})'\mathbf{X}'\mathbf{X}(\Phi - \widehat{\Phi}) + (\mathbf{Z} - \widehat{\mathbf{Z}})'(\mathbf{Z} - \widehat{\mathbf{Z}}) \tag{3.18}$$

Thus, from (3.17), the conditional densities are given by:

$$f(\Phi|\tau, \mathbf{Z}) \propto \exp\left\{-\frac{\tau}{2}\left[(\Phi - \widehat{\Phi})'\mathbf{X}'\mathbf{X}(\Phi - \widehat{\Phi})\right]\right\} p_1(\Phi) \tag{3.19}$$

$$f(\Phi|\tau, \mathbf{Z}) \propto \Psi(\Phi, \tau)\, p_1(\Phi) \tag{3.20}$$

$$f(\tau|\Phi, \mathbf{Z}) \sim Gamma\left[\frac{N - p + 2\alpha}{2}, \beta + \frac{1}{2}B(\Phi)\right] \tag{3.21}$$

Note that the conditional density $f(\Phi|\tau, \mathbf{Z})$ in (3.19) cannot be sampled directly. To overcome this difficulty, we can use the Metropolis-Hastings (MH) algorithm (Tweedie, 1998; Gamerman, 1997; Gelman et al., 1995; Chib and Greenberg, 1995). Roughly speaking, the underlying idea of the MH algorithm is to separate the posterior in (3.19) into two portions: one from which a random sample can be easily obtained, and one to perform an acceptance-rejection test on that sample. In this paper, the density $p_1(\Phi)$ in (3.15) will provide the sample, while the function $\Psi(\Phi, \tau)$ will be used in the acceptance-rejection test. The main steps of the MH algorithm are briefly described next:

- **Step 1**. Assign initial values to the parameters: $\Phi^{(0)}$ and $\tau^{(0)}$.
- **Step 2**. Obtain an observation Φ^* from the prior density in (3.15): $\Phi^* \sim p_1(\Phi)$
- **Step 3**. Calculate the probability of acceptance for the new observation Φ^* using $\Psi(\Phi, \tau)$ in (3.19):

$$\delta(\Phi^*, \Phi^{(j)}) = \begin{cases} \min\left\{1, \dfrac{\Psi(\Phi^*, \tau^{(j)})}{\Psi(\Phi^{(j)}, \tau^{(j)})}\right\} & \text{if} \quad \Psi(\Phi^{(j)}, \tau^{(j)}) \neq 0 \\ 1 & \text{Otherwise} \end{cases}$$

- **Step 4**. Obtain u from Uniform$[0, 1]$. Do $\Phi^{(j)} = \Phi^{(j+1)}$ if $u > \delta(\Phi^*, \Phi^{(j)})$. Otherwise, do $\Phi^{(j+1)} = \Phi^*$.
- **Step 5**. Obtain a new observation for τ from the density in (3.16): $\tau^{(j+1)} \sim f(\tau|\Phi^{(j+1)}, \mathbf{Z})$
- **Step 6**. Stop if the convergence of the Markov chain has been detected. Otherwise, do $j \longleftarrow j + 1$ and return to **Step 2**.

3.3 Simulation Procedures

The following procedures are carried out to obtain the final samples $\{\Theta^{(j)}\} = \{\phi_1^{(j)}, \ldots, \phi_p^{(j)}; \tau^{(j)}\}$, $j = 1, \ldots, m$, for the parameters via MCMC simulation as described in the two previous subsections. (1) M independent chains are simultaneously simulated for m iterations and for different initial values of the parameters. That is, we have a matrix for each parameter $\{\theta_{ij}\}$, $i = 1, \ldots, M$; $j = 1, \ldots, m$, in which the row index i indicates the chain and the column index j indicates the corresponding iteration within chain i. (2) The first half of observations for each chain is discarded to avoid *biased estimates* due to influence of initial values. (3) The convergence of the chains is verified by the Gelman-Rubin method (Gelman et al., 1995; Brooks and Gelman, 1998). (4) After convergence, the M chains are concatenated to form a single chain of length $Mm/2$.

4 MODEL SELECTION

Once the parameters for several AR(p) models have been calculated, the next step is to use the *parsimony principle* (Box et al., 1994) to choose the most suitable value for p. According to this principle, the model should have as few parameters as possible, consistent with the aim of capturing the major features of the data. The Akaike's information criterion (AIC) (Akaike, 1974) and the Schwarz's Bayesian information criterion (BIC) (Schwarz, 1978) are widely used to select the model. They are given by:

$$AIC = \ln(\hat{\tau}^{-1}) + \frac{2W}{N} \tag{4.1}$$

$$BIC = \ln(\hat{\tau}^{-1}) + \frac{W \ln(N)}{N} \tag{4.2}$$

where $W = p + 1$ and N is the length of the time series. These criteria may be minimized over choices of W to form a tradeoff between the fit of the model to the time series (which lowers $\hat{\tau}^{-1}$) and the model's complexity, which is measured by W. We choose as the model's order the value of p (or W, equivalently) that minimizes (4.1) or (4.2).

An alternative is based on posterior predictive distributions (Chang, 1995). This approach uses the *ordered predictive density* (OPD) for the observation Z_{N+k} given $Z_{N+k-1} = (Z_1, \ldots, Z_N, \ldots, Z_{N+k-1})'$, Φ and τ:

$$OPD(M_p, k) = f(Z_{N+k}|Z_{N+k-1}) \tag{4.3}$$

$$OPD(M_p, k) = \int_\Phi \int_\tau \tau^{1/2} \exp\left\{-\frac{\tau}{2}\left(Z_{N+k} - \sum_{i=1}^{p} \phi_i Z_{N+k-i}\right)^2\right\} f(\Phi, \tau|\mathbf{Z}) d\Phi d\tau \tag{4.4}$$

where $f(\Phi, \tau|\mathbf{Z})$ is the joint posterior density for the parameters Φ and τ, and M_p is the AR(p) model being evaluated. Equation (4.4) can be computed using the observations $\Phi^{(j)}$ and $\tau^{(j)}$ generated via MCMC simulation. In this case, the estimated ordered predictive density $\widehat{OPD}(M_p, k) = \widehat{f}(Z_{N+k}|Z_{N+k-1})$ is computed as follows:

$$\widehat{OPD}(M_p, k) = \frac{1}{m} \sum_{j=1}^{m} (\tau^{(j)})^{1/2} \exp\left\{-\frac{\tau^{(j)}}{2}\left(Z_{N+k} - \sum_{i=1}^{p} \phi_i^{(j)} Z_{N+k-i}\right)^2\right\} \tag{4.5}$$

Then, we compute $\widehat{OPD}(M_p) = \prod_{k=1}^{K} \widehat{OPD}(M_p, k)$, $K \gg 1$, for each model M_p, and choose that order p for which $\widehat{OPD}(M_p)$ is the maximum value.

Other possibility is to select the model via *Bayes factor* (BF), defined as $BF_{ij} = p(\mathbf{Z}|M_i)/p(\mathbf{Z}|M_j)$ (Kass and Raftery, 1995), where $p(\mathbf{Z}|M_p)$ is the normalization factor in (2.7). We can write $p(\mathbf{Z}|M_p)$ as follows:

$$p(\mathbf{Z}|M_p) = \int_\Phi \int_\tau L(\mathbf{Z}|\Phi, \tau, M_p) p(\Phi, \tau|M_p) d\Phi d\tau \tag{4.6}$$

where $p(\Phi, \tau | M_p)$ is the prior density corresponding to the model M_p and $L(\mathbf{Z}|\Phi, \tau, M_p)$ is the likelihood function. It is worth noting that $p(\mathbf{Z}|M_p)$ can also be estimated through the observations $\Phi^{(j)}$ and $\tau^{(j)}$ generated via MCMC simulation as follows (Raftery, 1996):

$$\hat{p}(\mathbf{Z}|M_p) = \frac{1}{m}\sum_{j=1}^{m} L(\mathbf{Z}|\Phi^{(j)}, \tau^{(j)}, M_p) \tag{4.7}$$

We choose the model M_j instead of model M_i if $BF_{ij} < 1$.

5 K-STEP AHEAD FORECASTING VIA MCMC

Let $\mathbf{Z}_f = (Z_{(N+1)}, Z_{(N+2)}, \ldots, Z_{(N+k)})'$ be the k values of the time series to be predicted and $\mathbf{Z}_p = (Z_{(N-(p-1)}, Z_{(N-(p-2)}, \ldots, Z_N)'$ the past p observations of the same series. It can be shown that $\mathbf{Z}_f$ is a vector of random variables that follows a multivariate Normal distribution with vector of means $E(\mathbf{Z}_f|\mathbf{Z}_p)$ and covariance matrix $Var(\mathbf{Z}_f|\Phi, \tau, \mathbf{Z}_p)]$ (de Alba, 1993):

$$E(\mathbf{Z}_f|\Phi, \tau, \mathbf{Z}_p) = -\mathbf{A}^{-1}(\Phi)\mathbf{B}(\Phi)\mathbf{Z}_p \tag{5.1}$$

$$Var(\mathbf{Z}_f|\Phi, \tau, \mathbf{Z}_p) = \text{diag}\left\{\tau^{-1}[\mathbf{A}^{-1}(\Phi)][\mathbf{A}^{-1}(\Phi)]'\right\} \tag{5.2}$$

where the matrices $\mathbf{A}(\Phi)$ and $\mathbf{B}(\Phi)$ are defined as follows:

$$\mathbf{A}(\Phi) = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_1 & 1 & 0 & \cdots & 0 \\ -\phi_2 & -\phi_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\phi_{k-1} & -\phi_{k-2} & -\phi_{k-3} & \cdots & 1 \end{pmatrix}_{k\times k} \tag{5.3}$$

$$\mathbf{B}(\Phi) = \begin{pmatrix} -\phi_p & -\phi_{p-1} & -\phi_{p-2} & \cdots & -\phi_1 \\ 0 & -\phi_p & -\phi_{p-1} & \cdots & -\phi_2 \\ 0 & 0 & -\phi_p & \cdots & -\phi_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & -\phi_{k-3} & \cdots & -\phi_k \end{pmatrix}_{k\times p} \tag{5.4}$$

In this paper, we extend the theoretical work of de Alba (1993) by estimating (5.1) and (5.2) through the use of the MCMC samples, as follows:

$$\widehat{E}(\mathbf{Z}_f|\Phi, \tau, \mathbf{Z}_p) = \frac{1}{m}\sum_{j=1}^{m} -\mathbf{A}^{-1}(\Phi^{(j)})\mathbf{B}(\Phi^{(j)})\mathbf{Z}_p \tag{5.5}$$

$$\begin{aligned}\widehat{Var}(\mathbf{Z}_f|\Phi, \tau, \mathbf{Z}_p) = \text{diag}\Bigg[&\frac{1}{m}\sum_{j=1}^{m}\frac{1}{\tau^{(j)}}\left[\mathbf{A}^{-1}(\Phi^{(j)})\right]\left[\mathbf{A}^{-1}(\Phi^{(j)})\right]' + \\ &+ \frac{1}{m}\sum_{j=1}^{m}\left(\mathbf{A}^{-1}(\Phi^{(j)})\mathbf{B}(\Phi^{(j)})\mathbf{Z}_p\right)^2 + \\ &+ \left(\frac{1}{m}\sum_{j=1}^{m}\mathbf{A}^{-1}(\Phi^{(j)})\mathbf{B}(\Phi^{(j)})\mathbf{Z}_p\right)^2\Bigg]\end{aligned} \tag{5.6}$$

Equations (5.5) and (5.6) are, respectively, the Monte Carlo estimates of the forecasted values and their variance. It is worth emphasizing that a clear advantage of Bayesian modeling is that forecasting uncertainty can be easily quantified by constructing a credible interval for the parameters from (5.5) and (5.6). In the next section, we apply the Bayesian and MCMC techniques presented so far to the monthly streamflow forecasting problem.

6 CASE STUDY – THE FURNAS RESERVOIR

It is considered a time series of monthly mean streamflow from the Furnas reservoir which belongs to the Brazilian hydroelectric system. Its dam is located in the mid-course of the Grande river near the rapids known as the Corredeiras das Furnas between the towns of São João da Barra and São João Batista do Glória, in Minas Gerais state. Its construction began in July 1958, with the first unit starting up operations in September 1963 and the sixth - the last unit planned - in July 1965. In the early 1970s, a start was made on expanding this facility with its seventh and eighth units, totaling 1,216 MW, ranking this project among the largest in Latin America.

The privileged location of this power plant (500 kilometers from Rio de Janeiro, 400 kilometers from São Paulo and 300 kilometers from Belo Horizonte) helped prevent a massive black-out in the mid-1960s in Brazil, avoiding rationing and power cuts to the nation's industrial sector. When construction began, its planned capacity was equivalent to one-third of the total installed capacity in Brazil. In addition to being a benchmark in major hydro-power facilities in Brazil, the FURNAS power plant also resulted in the regulation of the Rio Grande river and the construction of a further eight power plants, making full use of over 6,000 MW installed.

The recordings cover a period of sixty years ranging from January 1930 to December 1990. All equations and procedures in this paper were implemented in the script language of Matlab (Hanselman and Littlefield, 1998). The actual observations of monthly mean streamflow are a seasonal time series denoted by $Y_{t(r,m)}$, with $t(r,m) = 12(r-1)+m$, where $r = 1, \ldots, 60$, is the number of years and $m = 1$, $2, \ldots, 12$, is the number of months. We applied a logarithmic transformation to each value of $Y_{t(r,m)}$ and computed the sample mean, μ_m, and the standard deviation, σ_m, of the streamflow values for each month after the transformation. Then the transformed time series is normalized as follows:

$$Z_{t(r,m)} = \frac{\ln Y_{t(r,m)} - \mu_m}{\sigma_m} \tag{6.1}$$

where $\ln(x)$ is the natural logarithm of x. Fig. 1 shows the Furnas time series before and after the transformation in (6.1) and the corresponding histograms. From this figure, we have observed that the transformed time series has an approximately Normal distribution with zero mean and constant variance. An AR(p) model is then fitted to the time series Z_t using the maximum likelihood and the proposed Bayesian approaches.

Fig. 2 shows the autocorrelation function (ACF) and the partial autocorrelation function (PACF) for the transformed series. By analyzing the ACF and the PACF, we can have a first idea of the model (AR(p), MA(q) or ARMA(p,q)) to be fitted to the time series. The order p or q of the model can also be inferred from Fig. 2. Since the ACF has an exponential-like decay, and the partial autocorrelation coefficients cut off after lag k = 2, we can select an AR(2) model to be fitted to the Furnas time series. The dotted lines represent the $\pm 2/\sqrt{n}$ standard error limits consistent with testing the null hypothesis that partial autocorrelation coefficients are zero.

The choice of an AR(2) model can be confirmed by the AIC and BIC criteria, as well as by those using MCMC samples like the predictive density and Bayes factor criteria (see Table 1).

Method	p=1	p=2	p=3
AIC	-1.0350	**-1.0467**	-1.0456
BIC	-1.0243	**-1.0255**	-1.0141
OPD (K=20)	2.4215	**6.4706**	1.8854
Bayes Factor	0.0107	**0.5581**	0.1636

Table 1: Model selection via several criteria.

Note that the values obtained for $p = 2$ and $p = 3$ using the AIC method are very similar, indicating that they provide almost the same result for model fitting (i.e. small $\hat{\tau}$), but the AR(2) is preferred because it has fewer parameters to be determined. For the BIC method, the interpretation of the result is a bit different. In this case, the values for $p = 1$ and $p = 2$ are very similar, but the model with more parameters, i.e. AR(2), is preferred because it provides a slightly lower $\hat{\tau}$. Thus, it is sometimes not

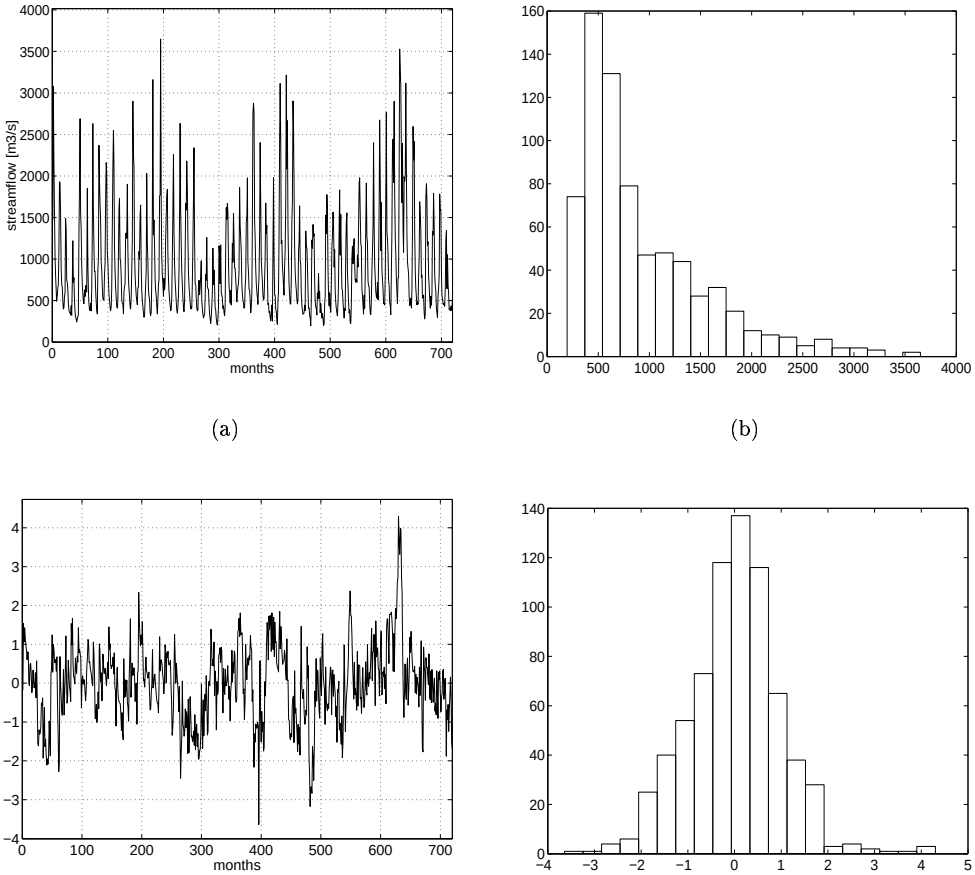


Figure 1: Furnas monthly streamflow (m^3/s) time series. (a)-(b) Before and (c)-(d) after the transformation in (6.1).

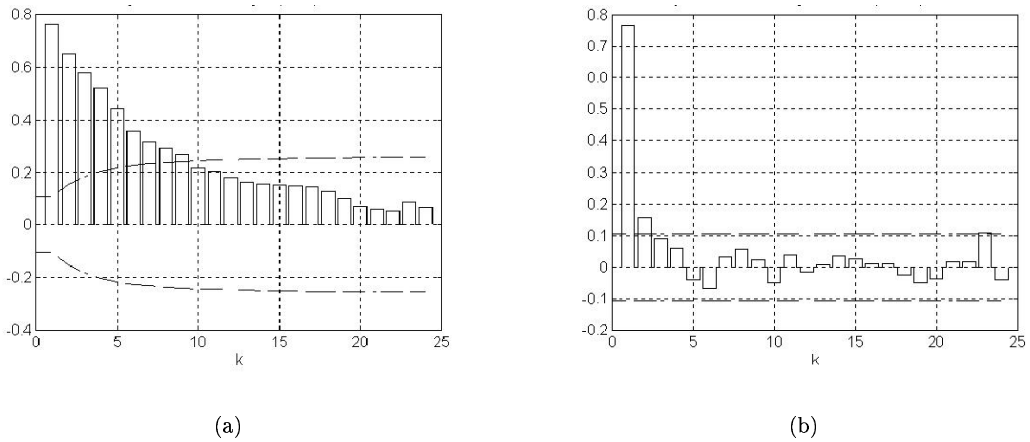


Figure 2: (a) ACF and (b) PACF for the transformed Furnas time series.

very straightforward to choose th order p of the model based on such very similar values provided by the AIC and BIC methods. Unlike AIC and BIC methods, the values provided by the OPD and the Bayes factor methods are more sharp, emphasizing clearly the difference among the different orders and suggesting the use of $p = 2$. For instance, the Bayes factor ratios $BF_{12} = 0.0107/0.5581 = 0.02 < 1$ and $BF_{32} = 0.1636/0.5581 = 0.29 < 1$ indicate that the AR(2) model should be used.

Fig. 3 show the histograms of the samples generated via MCMC simulation for the parameters ϕ_1, ϕ_2 and τ using the Normal-Gamma and t-Gamma prior densities. The subscript b indicates a Bayesian estimate. It is worth noting that the symmetric form of the non-standard posterior density for the case of a t-Gamma prior density could only be inferred through MCMC simulation.

Table 2 shows the estimates and their variances for the parameters of the AR(2) model fitted to the Furnas time series. MLE denotes the maximum likelihood estimates, NGP represents the Bayesian estimates obtained using the Normal-Gamma prior and the Gibbs-Sampling algorithm, and TGP symbolizes the Bayesian estimates obtained using the t-Gamma prior and the Metropolis-Hastings algorithm. The hyperparameters for the Normal-Gamma prior are $\mu = (0,0)'$, $\mathbf{P} = 75I_2$, $\alpha = 2$, and $\beta = 2$; while those for the Normal-Gamma prior are $\mu = \hat{\phi}_{MLE}$, $\mathbf{P} = 75I_2$, $\alpha = 2$, and $\beta = 2$.

Method	ϕ_1	var(ϕ_1)	ϕ_2	var(ϕ_2)	τ	var(τ)
MLE	0.6478	0.0027	0.1623	0.0027	2.0285	—
NGP	0.6435	0.0026	0.1644	0.0028	1.9987	0.0217
TGP	0.6480	0.0025	0.1619	0.0027	1.9488	0.0317

Table 2: Point estimates and their variances.

In Table 3, the estimation uncertainty for the estimates shown in Table 2 is quantified by the credible intervals. The second line in this table represents the lower limit of the interval, while the third line represents the upper limit.

ϕ_1(NGP)	ϕ_1(TGP)	ϕ_2(NGP)	ϕ_2(TGP)	τ(NGP)	τ(TGP)
0.5481	0.5525	0.0624	0.0590	1.7230	1.6711
0.7428	0.7518	0.2661	0.2630	2.2854	2.2641

Table 3: Credible intervals 95% for the estimates in Table 2.

The next simulations compare the predictive ability of the MLE method with the proposed Bayesian one. Three criteria are used to evaluate the forecasted Y_t^p and the actually observed Y_t^o values of mean streamflow, namely:

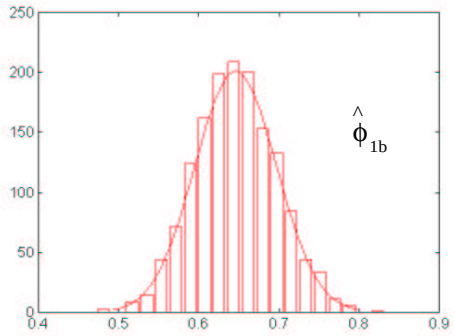
$$\textit{Mean Squared Error} : MSE = \frac{1}{k}\sum_{t=1}^{k}(Y_t^o - Y_t^p)^2 \tag{6.2}$$

$$\textit{Mean Absolute Deviation} : MAD = \frac{1}{k}\sum_{t=1}^{k}|Y_t^o - Y_t^p| \tag{6.3}$$

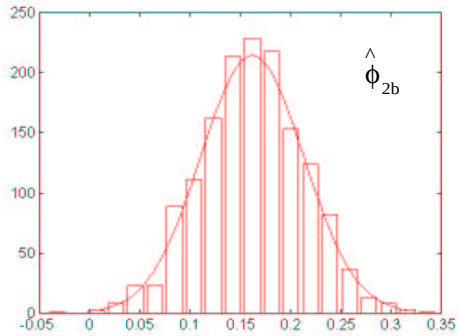
$$\textit{Mean Absolute Percentual Error} : MAPE = \frac{100}{k}\sum_{t=1}^{k}\frac{|Y_t^o - Y_t^p|}{Y_t^p} \tag{6.4}$$

Table 4 shows the forecasted values for a period of 12 months ahead obtained for both the classical MLE approach and Bayesian method with a t-Gamma prior density. The forecasting uncertainty is quantified by the credible intervals shown in the fourth column of Table 4. The values in this table are in the original domain of streamflow measurements (and not in that of the transformed series Z_t) obtained through the following inverse transformation:

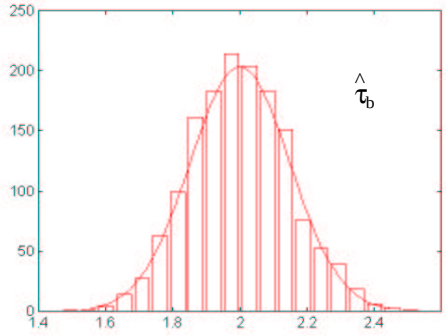
$$E(Y_f|\Phi,\tau) = \exp\left\{\mu_m + \frac{\sigma_m}{2}\left[2E(Z_f|\Phi,\tau,\mathbf{Z}_p) + \sigma_m \operatorname{var}(Z_f|\Phi,\tau,\mathbf{Z}_p)\right]\right\} \tag{6.5}$$



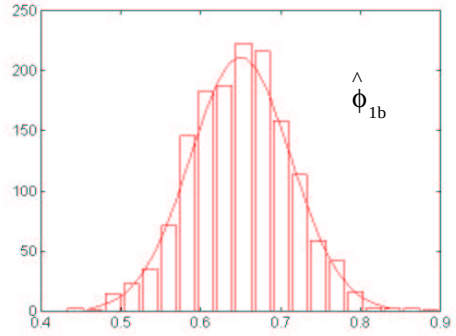
(a)



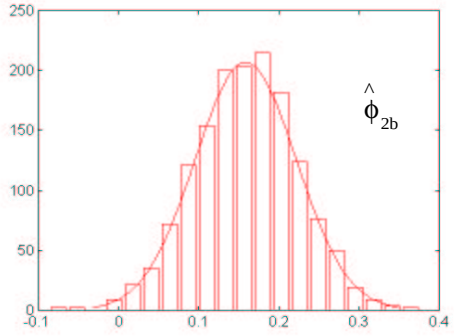
(b)



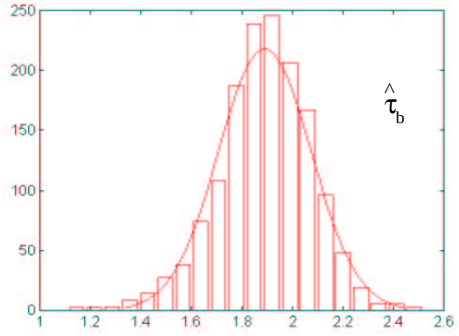
(c)



(d)



(e)



(f)

Figure 3: Histograms for the samples $\phi_1^{(j)}$, $\phi_2^{(j)}$ and $\tau^{(j)}$ generated via MCMC simulation using the Normal-Gamma prior in (a), (b) and (c); and using the the t-Gamma prior in (d), (e) and (f).

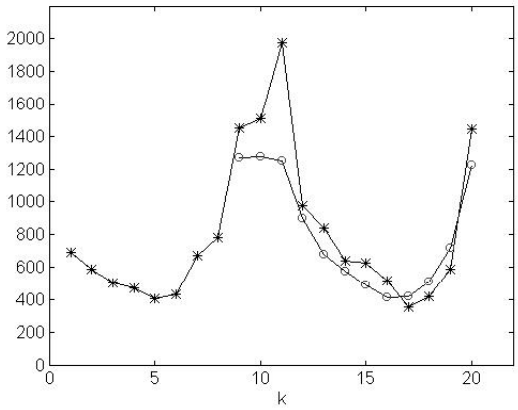


Figure 4: Forecasting 12 months ahead for the Furnas time series. Forecasted values are shown in open circles, while the asterisks denote the actual values.

where $E(Z_f|\Phi,\tau,\mathbf{Z}_p)$ and $\text{var}(Z_f|\Phi,\tau,Z_p)$ are, respectively, the forecasted value and the variance of the forecasting given in (5.5) and (5.6).

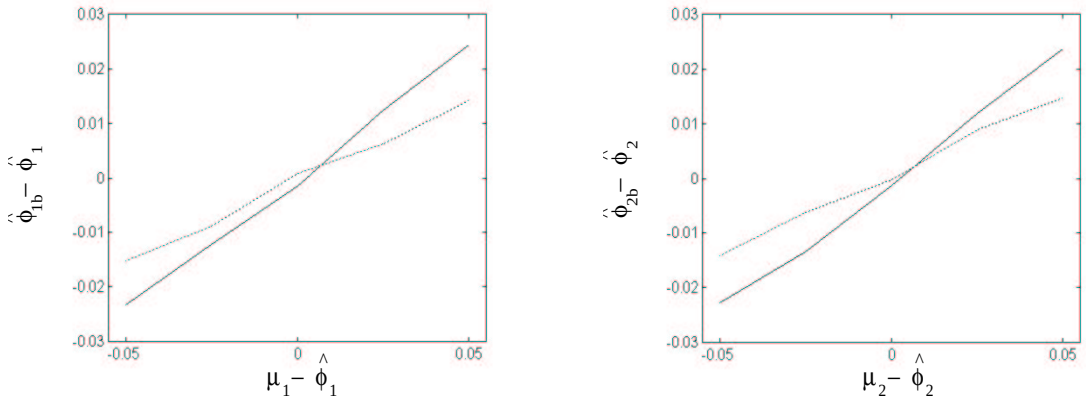
Actual	**Max. Likelihood**	**Bayesian**	**C.I. (95%)**
1450	1264.7	1270.2	[1223.4, 1320.8]
1512	1269.1	1275.4	[1224.8, 1322.7]
1977	1246.4	1253.5	[1202.1, 1300.3]
971	896.2	899.9	[864.4, 930.6]
839	677.5	679.7	[655.1, 700.1]
639	571.2	572.7	[552.2, 589.1]
624	487.4	488.3	[473.9, 499.1]
513	411.4	412.0	[400.5, 420.2]
354	416.3	417.0	[402.4, 426.8]
421	506.4	506.9	[492.9, 515.7]
587	715.0	715.7	[695.2, 727.6]
1447	1224.4	1225.3	[1191.8, 1243.5]
MSE	73.10	72.27	——
MAD	183.28	181.05	——
MRE	18.34%	18.18%	——

Table 4: Forecasting 12 months ahead for Furnas time series.

We can note that the performance of the Bayesian method is slightly better than that of the MLE approach. Fig. 4 plots the forecasted values in Table 4 for the Bayesian approach together with the actual streamflow values for a certain part of the Furnas time series.

6.1 Robustness Analysis of the Bayesian Estimation

The class of autoregressive models is a rather general set of models largely used to represent stationary time series. However, time series often present change points in their dynamic structure, which may have a serious impact on the analysis and lead to misleading conclusions. A change point, which is generally the effect of an external event on the phenomenon of interest, may be represented by a change in the structure of the model or simply by a change of the value of some of the parameters. In particular, we are interested in how the parameter estimates vary with respect to the specification of the hyperparameters of the prior density. Unlike the MLE approach, the Bayesian method allows the study of robustness of the estimator. In the following tests, we evaluate how changes in the value of the location parameter $\boldsymbol{\mu}$



(a) (b)

Figure 5: Robustness analysis using Normal-Gamma prior (solid lines) and the t-Gamma (dotted lines).

of both prior densities influence the corresponding Bayesian point estimates. The other hyperparameters $\mathbf{P}$, α and β are kept constants. The precision matrix $\mathbf{P}$ is diagonal with large elements, i.e., $P_{ii} \gg 1$, $i = 1, \ldots, p$, in order to reduce variability and facilitate the visualization of the results.

Fig. 5 shows the results of the robustness analysis for the Bayesian estimates ϕ_1 and ϕ_2. In this figure, μ_1 and μ_2 are the components of the location parameter $\boldsymbol{\mu}$, $\widehat{\phi}_1$ and $\widehat{\phi}_2$ are the maximum likelihood estimates, and $\widehat{\phi}_{1b}$ and $\widehat{\phi}_{2b}$ are the Bayesian estimates. As expected from (3.11), there is a linear relationship between $\widehat{\Phi}_b - \widehat{\Phi}$ and $\boldsymbol{\mu} - \widehat{\Phi}$ when using the Normal-Gamma prior. The same occurs when using the t-Gamma prior, but that the slope of the curve for the t-Gamma prior is lower than the slope for the Normal-Gamma prior, especially at the central region around zero. This fact indicates that the t-Gamma prior produces Bayesian estimates less sensitive to changes in the location parameter. It is important to emphasize that this type of robustness test is only possible due to MCMC simulation since the posterior density for t-Gamma prior is not standard. Table 5 shows the numerical values corresponding to the robustness test in Fig. 5.

	$\widehat{\phi}_{1b}$	C.I. 95%	$\widehat{\phi}_{2b}$	C.I. 95%
NG	0.5699	[0.5211, 0.6205]	0.0822	[0.0366, 0.1290]
	0.5811	[0.5320, 0.6301]	0.0943	[0.0467, 0.1430]
	0.5932	[0.5416, 0.6473]	0.1062	[0.0571, 0.1592]
	0.6042	[0.5584, 0.6517]	0.1157	[0.0674, 0.1666]
	0.6149	[0.5663, 0.6618]	0.1278	[0.0805, 0.1838]
	0.6288	[0.5763, 0.6746]	0.1410	[0.0948, 0.1882]
	0.6408	[0.5876, 0.6916]	0.1527	[0.1068, 0.2016]
	0.6521	[0.6017, 0.6978]	0.1653	[0.1167, 0.2107]
TG	0.5780	[0.5100, 0.6535]	0.0976	[0.0213, 0.1775]
	0.5959	[0.5261, 0.6713]	0.1028	[0.0333, 0.1747]
	0.6012	[0.5373, 0.6680]	0.1149	[0.0443, 0.1803]
	0.6076	[0.5317, 0.6745]	0.1229	[0.0517, 0.1891]
	0.6172	[0.5476, 0.6816]	0.1289	[0.0555, 0.1967]
	0.6227	[0.5538, 0.6901]	0.1379	[0.0671, 0.2146]
	0.6306	[0.5588, 0.7100]	0.1438	[0.0767, 0.2103]
	0.6434	[0.5774, 0.7076]	0.1495	[0.0785, 0.2247]

Table 5: Robustness test for $\widehat{\phi}_{1b}$ and $\widehat{\phi}_{2b}$.

7 CONCLUSION

This paper dealt with the application of Bayesian statistics and Markov Chain Monte Carlo (MCMC) methods to all the stages of the streamflow forecasting problem. We fitted autoregressive models to the monthly mean streamflow time series of the Furnas reservoir in Brazil, calculated the parameters using the MLE and Bayesian approaches, and studied the forecasting and robustness properties of the Bayesian estimators. In particular, MCMC algorithms have played an essential role when the posterior densities were non-standard, being used to obtain samples from the posterior densities for inference. The proposed procedures are quite general and can easily adapted or extended to other stochastic models, such as MA(q) and ARMA(p, q) models. In the following, we summarize the steps for an efficient forecasting, using AR(p) models, based on Bayesian and MCMC methods.

- Determination of the likelihood function for the data, specification of a prior density for the parameters and the determination of the posterior density via Bayes theorem.
- Initial analysis of the time series trying to observe and eliminate non-stationarity, tendency, seasonality, etc. Determination of the ACF and PACF to infer the AR model and, if possible, its order p.
- Determination of the order p of the model via classical methods such as AIC and BIC, and Bayesian methods such as predictive density and Bayes Factor.
- Parameter estimation via classical methods such as the maximum likelihood, and Bayesian methods together with MCMC simulation.
- Forecasting via classical and Bayesian methods.
- Robustness analysis of the Bayesian estimators via MCMC methods.

The last two procedures are of great importance to the planning and operation studies of a hydroelectric system, and are the main theoretical and practical results of this paper. In particular, forecasting and robustness studies can be used together as the ultimate criteria to select a certain parameter estimation method. For the prior densities studied in this paper, the t-Gamma prior has lead to Bayesian estimates less sensitive to changes in the location parameter than those resulting from the Normal-Gamma prior.

The possibility of evaluating the robustness of the estimator through simulations can justify in part the use of computer-intensive techniques, such as MCMC, in the place of the standard MLE in (2.3) and (2.4). We can conclude that Bayesian statistics together with MCMC algorithms give more flexibility to the process of time series modeling and all its stages such as parameter estimation, model selection, forecasting, and robustness evaluation.

As further work, a comparative study between the proposed Bayesian approach and other common techniques for streamflow forecasting such as neural networks (Atiya et al., 1999) and neuro-fuzzy algorithms (Ballini et al., 1999) is suggested. It can be fruitful to combine these different approaches to time series modeling and evaluate the resulting hybrid method in forecasting streamflow sequences.

References

Akaike, H. (1974). A New look at the statistical identification model. *IEEE Transactions on Automatic Control*, **19**(6), 716–723.

Atiya, A. F., El-Shoura, S.M., Shaheen, S.I. and El-Sherif, M.S. (1999). A Comparison between neural-network forecasting techniques – Case study: river flow forecasting. *IEEE Transactions on Neural Networks*, **10**(2), 402-409.

Awwad, H., Valdes, J., and Restrepo, P. (1994). Streamflow forecasting for Han River Basin, Korea, *Journal of Water Resources Planning and Management*, **120**(5), 651–673.

Ballini, R., Soares, S. and Andrade, M.G. (1999). Seasonal streamflow forecasting via a neural fuzzy system. *Proceedings of the 14th IFAC World Congress*, Beijing, China, 81–86.

Barbieri, M. M. and Conigliani, C. (2000). Bayesian analysis of autoregressive time series with change points. *Journal of the Italian Statistical Association*.

Box, G.E.P, Jenkins, G.M. and Reinsel, G.C. (1994). *Time series analysis: forecasting and control*, 3nd edition, Holden-Day.

Brooks, S.P. (1998). Markov chain Monte Carlo method and its application. *The Statistician*, **47**, 69–100.

Brooks, S.P. and Gelman, A. (1998). Convergence assessment techniques for Markov chain Monte Carlo. *Statistics and Computing*, **8**, 319–335.

Casella, G. and George, E. J. (1992). Explaining the Gibbs Sampler. *The American Statistician*, **46**(3), 167–174.

Chang, H. (1995). Model determination using predictive distributions. *PhD Thesis*, Department of Statistics, University of Connecticut.

Cheng, Y. (1994). Evaluating an autoregressive model for streamflow forecasting. *Proceedings of the 1994 Hydraulic Engineering Conference*, 1105–1109.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hasting algorithm. *The American Statistician*,**49**(4), 327–335.

Choy, S.T.B. and Smith, A.F.M. (1997). On robust analysis of a normal location parameter. *Journal of the Royal Statistical Society-B*, **59**(2), 463–474.

Davis, M.H.A. and Vinter, R. (1985). *Stochastic modelling and control*, Chapman & Hall.

de Alba, E. (1993). Constrained forecasting in autoregressive time series models: A Bayesian analysis. *International Journal of Forecasting*, **9**, 95-108.

Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman & Hall.

Gelman, R., Carlin, J.B., Stern, H.S. and Rubin, D.B, (1995). *Bayesian data analysis*, Chapman & Hall.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

Hanselman, D. and Littlefield, B. (1998). *Mastering MATLAB 5*, Prentice Hall.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors and model uncertainty. *Journal of the American Statistical Association*, **90**, 773–795.

Maceira, M.E.P. and Bezerra C.V. (1997). Stochastic streamflow models for hydroelectric systems, *Proceedings of the 1997 Probabilistic Methods Applied to Power Systems (PMAPS'97)*, Vancouver, BC, Canada.

McCulloch R. E. and Tsay, R. (1993). Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the American Statistical Association*, **88**(2), 968–978.

Pericchi, L.R. and Walley, P. (1991). Robust Bayesian credible intervals and prior ignorance. *International Statistical Review*, **59**, 1–24.

Pericchi, L.R. and Smith, A.F.M. (1992). Exact and approximate posterior moments for a normal location parameter. *Journal of the Royal Statistical Society-B*, **54**(3), 793–804.

Raftery, A. E. (1996). Hypothesis testing and model selection via posterior simulation, In:*Markov Chain Monte Carlo in Practice*, W. R. Gilks, E. Richardson e D. J. Spiegelhalter (eds.), Chapman & Hall.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**(2), 461–464.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, **22**, 1701–1762.

Tweedie, R.L.(1998). Markov chains: structure and applications, *Technical Report*, Department of Statistics, Colorado State University, Available at http://www.stat.colostate.edu/Documents/1998.html