



Seasonal Hydrological and Meteorological Time Series

Edilberto Cepeda Cuervo¹, Jorge Alberto Achcar², Marinho G. Andrade³.

¹ Statistics department Science Faculty Universidad Nacional de Colombia, Bogotá, Colombia email: ecepedac@unal.edu.co

² Social Medicine Department FMRP, Universidade de São Paulo Ribeirão Preto, S.P., Brazil

³ Applied Mathematics and Statistics Department ICMC, Universidade de São Paulo Sao Carlos, S.P., Brazil

ABSTRACT

Time series models are often used in hydrology and meteorology studies to model streamflows series in order to make forecasting and generate synthetic series which are inputs for the analysis of complex water resources systems. In this paper we introduce a new modeling approach for hydrologic and meteorological time series assuming a continuous distribution for the data, where both the conditional mean and conditional variance parameters are modeled. Bayesian methods using standard MCMC (Markov Chain Monte Carlo Methods) are used to simulate samples for the joint posterior distribution of interest. Two applications to real data sets illustrate the proposed methodology, assuming that the observations come from a normal, a gamma or a beta distribution. A first example is given by a time series of monthly averages of natural streamflows, measured in the year period ranging from 1931 to 2010 in Furnas hydroelectric dam, Brazil. A second example is given with a time series of 313 air humidity data measured in a weather station of Rio Claro, a Brazilian city located in southeastern of Brazil. These applications motivate us to introduce new classes of models to analyze hydrological and meteorological time series.

Keywords: Hydrology time series, Meteorological time series, Conditional regression models, Bayesian analysis, MCMC methods.

Series de Tiempo hidrológicas y meteorológicas estacionales

RESUMEN

Los modelos de series de tiempo se usan a menudo en estudios de hidrología y meteorología para modelar series de flujos a fin de hacer pronósticos y generar series sintéticas que son insumos para el análisis de sistemas complejos de recursos hídricos. En este artículo presentamos un nuevo enfoque de modelado para series de tiempo hidrológicas y meteorológicas asumiendo una distribución continua para los datos, donde se modelan los parámetros tanto de la media condicional como de la varianza condicional. Métodos bayesianos estándares que usan MCMC (Markov Chain Monte Carlo) son usados para simular muestras de la distribución a posteriori conjunta de interés. Dos aplicaciones a conjuntos de datos reales ilustran la metodología propuesta, asumiendo que las observaciones provienen de una distribución normal, gamma o beta. Un primer ejemplo está dado por una serie temporal de promedios mensuales de los caudales naturales, medidos en el período anual que va de 1931 a 2010 en la presa hidroeléctrica de Furnas, Brasil. Un segundo ejemplo considera una serie temporal de 313 datos de humedad del aire medidos en una estación meteorológica de Río Claro, una ciudad brasileña ubicada en el sureste de Brasil. Estas aplicaciones nos motivan a introducir nuevas clases de modelos para analizar series de tiempo hidrológicas y meteorológicas.

Palabras clave: Series de tiempo hidrológicas; Series de tiempo meteorológicas; Modelos de regresión condicional; Análisis Bayesiano; Métodos MCMC.

Record

Manuscript received: 09/06/2017

Accepted for publication: 27/04/2018

How to cite item

Cepeda-Cuervo, E., Achcar, J. A., & Andrade, M. G. (2018). Seasonal Hydrological and Meteorological Time Series. *Earth Sciences Research Journal*, 22(2), 83-90.

DOI: <http://dx.doi.org/10.15446/esrj.v22n2.65577>

1. Introduction

Time series models are often used in hydrology studies to model streamflow series in order to make predictions and to generate synthetic series which are inputs for the analysis of complex water resources systems (see, for example, Salas et al., 1980, 1982; Hosking, 1984; Hipel & McLeod, 1994; Montanari et al., 1997; Hasebe et al., 2000). In many studies, hydrologists also use time series data to display the amount of rainfall that has fallen in a region for the past day, year or a period of 10 years (see for example, Guimaraes & Santos, 2011, and Lee & Lee, 2000).

Modeling hydrological variability is very important in the planning and management of water resources. Many aspects of the hydrologic cycle could be described by time series data. Researchers, usually use time series data to evaluate the resources of a water basin. Important variables related to streamflow and watershed describes streamflow properties such as monthly flows or streamflow parameters. Assuming a specified time series model, we usually could estimate the streamflow parameters using a classical or a Bayesian inference approach.

Different time series models as ARMA and higher orders of MA models have been used by some authors when considering hydrologic regionalization of watersheds (see for example, Chiang et al., 2002 a, b; Wang et al., 2015). Valipour et al. (2013) studied annual runoff time series in Dahuofang reservoir, in northeast China, using autoregressive integrated moving average (ARIMA) models coupled with ensemble empirical mode decomposition (Wu, Z., & Huang, N. E., 2009). Spectral analysis and forecasting of hydrological time series has also been considered (Marques et al., 2006), including geostatistical applications (Robin et al., 1993) and climate change investigations (Lall and Mann, 1995).

Considering hydrological time series, the monthly streamflow series typically have a periodic behavior in the mean and variance and in general, periodic autoregressive models are used in the analysis of the data (see, for example, Modal & Wasimi, 2006). In this situation, usually it is assumed that the series flow has a normal or log-normal distribution (see for example, Tesfaye et al., 2006; Wang et al., 2009).

The behavior of meteorological time series is similar to the behavior of hydrological time series. In this case, given that the relative air humidity is a random variable with values given in the open interval $(0, 1)$, we could assume a beta distribution to analyze the data. Another possibility to analyze the data set is to consider a transformation of the data and to assume a normal distribution for the transformed data. As a special case, we could assume a logistic transformation.

In this paper, a more general model is considered in the analysis of the hydrological or meteorological time series conditional to the historical available information: it is assumed that the data is generated from a normal, a gamma or a beta distribution, with conditional mean and variance, given respectively, by $E[Y_t|Y_{t-1}]$ and $V[Y_t|Y_{t-1}]$, where the index t is related to time. Thus a general model is proposed to analyze hydrological or meteorological time series, assuming that the observations are generated from a continuous biparametric exponential family of distributions.

As an illustration and motivation for the use of the proposed models, we first consider as a first example, a data set consisting of the time series of monthly averages of natural streamflows, measured in the year period ranging from 1931 to 2010, in Furnas hydroelectric dam, located in southeastern Brazil. This time series is shown in Figure 1. From this Figure, we observe that the streamflow series have a periodic behavior in the mean and variance and in this case, general periodic autoregressive models are usually assumed in the analysis of the time series data (Modal & Wasimi, 2006).

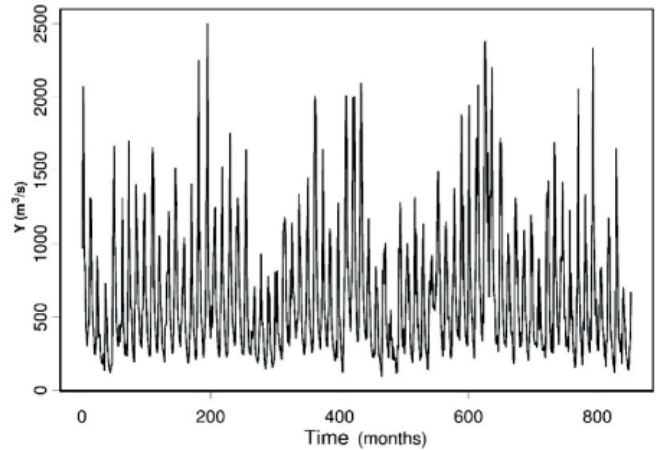


Figure 1. Time series of monthly averages of natural streamflows.

To take into account the heteroscedasticity in the time series of streamflows showed in this first example, we propose a new model which assumes seasonal and autoregressive terms in the modeling of the mean and variance parameters. In this way, we propose a periodic and heteroscedastic model, in which the variance also presents an autoregressive structure.

As a second example, we consider another time series with behavior similar to the assumed hydrological time series presented in the first example, consisting of a meteorological time series given by weekly averages of air humidity, measured in the city of Rio Claro, São Paulo state, Brazil (Figure 2). In this case, given that the relative air humidity is a random variable with values in the open interval $(0, 1)$, it is assumed a beta distribution in the analysis of the data. In this case, we propose a jointly modeling for the mean and for the variance of the data considering beta regression models, including seasonal and autoregressive terms in both regression models

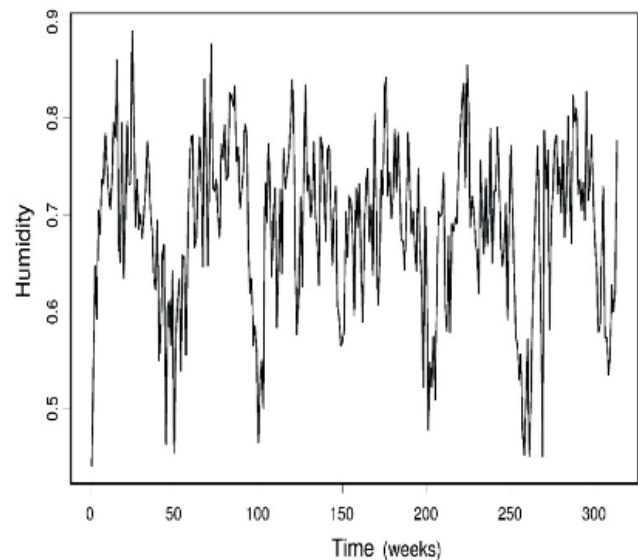


Figure 2. Air humidity time series data.

The paper is structured as follows: in section 2, it is introduced the seasonality analysis of the time series. In section 3, it is proposed seasonal autoregressive models. In section 4, it is presented the results of the analysis of the hydrological time series obtained using the proposed models assuming normal and gamma distributions. In section 5, it is presented the results of the analysis of air humidity time series. Finally, in section 6 it is presented some conclusions and future research topics.

2. A period model

In this section, we introduce a new modeling approach including seasonality terms in the model which better describes time series of monthly averages of natural stream flows, denoted by Y_t . In the first case, we consider the time series data related to the monthly averages of natural streamflows in Furnas hydroelectric dam (Figure 1). A preliminary spectral analysis is developed for this time series to determine the time periods to be considered in the mean and variance model formulations. Thus, if in the spectral analysis, the number of observations is $T = 2q + 1$, where q is a positive integer number, the Fourier time series model given by:

$$y_t = \alpha_0 + \sum_{i=1}^q \{ \alpha_{1,i} \cos(2\pi f_i t) + \alpha_{2,i} \sin(2\pi f_i t) \} + e_t, \quad (1)$$

is to be fitted by the data, where $f_i = i/T$ is the i th harmonic of the fundamental frequency $1/T$ and $\alpha_{1,i}$ and $\alpha_{2,i}$, $i = 1, \dots, q$, are the related coefficients and e_t is an error term assumed to have the first and second moments given respectively by $E(e_t) = 0$, $E(e_t^2) = \sigma^2$ and to be uncorrelated, that is, $E(e_t e_{t+k}) = 0$ for $k \neq 0$.

Considering the n observations of the time series, the least square estimates of the coefficients α_0 and $(\alpha_{1,i}, \alpha_{2,i})$, $i = 1, \dots, q$, are obtained from the equations:

$$\hat{\alpha}_0 = \frac{1}{T} \sum_{t=1}^T y_t \quad (2)$$

$$\hat{\alpha}_{1,i} = \frac{2}{T} \sum_{t=1}^T y_t \cos(2\pi f_i t) \quad (3)$$

$$\hat{\alpha}_{2,i} = \frac{2}{T} \sum_{t=1}^T y_t \sin(2\pi f_i t) \quad (4)$$

Based on these estimates, the intensity of each frequency is calculated by:

$$I(f_i) = \frac{T}{2} (\hat{\alpha}_{1,i}^2 + \hat{\alpha}_{2,i}^2) \quad (5)$$

Observe that the highest frequency is 0.5 cycle per month (time interval) since the smallest period is 2 months. This preliminary analysis of the intensities of frequencies allows us to reduce the number of harmonics considered in the time series included in the model, considering only those frequencies that have higher intensities (Marques et al., 2006). For a series consisting of 80 years of observations (that is 960 months) we would have 480 harmonics, but considering this preliminary analysis, we could reduce the number of harmonics to three or four. Usually, in the case of monthly time series, these periods are given by periods of 6 and 12 months. These periods will be discussed in Section 4, where the specification of the parameters related to the periods is made simultaneously with the specification of the parameters for the autoregressive models

assumed for the mean and variance in the time series formulations to be introduced in Section 3. If some of these parameters show to be not significant after a preliminary statistical analysis, the terms related to these parameters are deleted from the model to develop a further analysis. The final decision on which harmonics should be included in the model must be made considering some criteria for model selection or hypothesis test on the fitted coefficients $\alpha_{1,i}$ and $\alpha_{2,i}$.

3. The proposed seasonal autoregressive model

In hydrological or meteorological time series, we usually assume that the observations of interest are generated from a conditional continuous probability distribution function. As special cases, we could assume that the observations are generated from standard normal, gamma, beta or exponential conditional density functions, denoted by $f(y_t | H_{t-1})$, $t = 1, 2, \dots, T$, where H_{t-1} is the available information up to time $t-1$ and thus Y_t has conditional means and variances given respectively by $\mu_t = E(Y_t | H_{t-1})$ and $h_t = Var(Y_t | H_{t-1})$, following the models:

$$\mu_t = \alpha_0 + \sum_{i=1}^q \{ \alpha_{1,i} \cos(2\pi f_i t) + \alpha_{2,i} \sin(2\pi f_i t) \} + \sum_{i=1}^p \phi_i y_{t-i} \quad (6)$$

$$\log(h_t) = \lambda_0 + \sum_{i=1}^q \{ \lambda_{1,i} \cos(2\pi f_i t) + \lambda_{2,i} \sin(2\pi f_i t) \} + \sum_{i=1}^p \theta_i y_{t-i} \quad (7)$$

where $\beta = \{ \alpha_0, \alpha_{1,1}, \dots, \alpha_{1,q}, \alpha_{2,1}, \dots, \alpha_{2,q}, \phi_1, \dots, \phi_p \}$, is the vector of parameters for the mean model and $\gamma = \{ \lambda_0, \lambda_{1,1}, \dots, \lambda_{1,q}, \lambda_{2,1}, \dots, \lambda_{2,q}, \theta_1, \dots, \theta_p \}$ is the vector of the parameters for the variance model, $f_i = i/T$ the i th harmonic of the fundamental frequency $1/T$. In this paper, the parameters of the proposed models are estimated under a Bayesian approach.

In order to illustrate the proposed methodology, we also include the regression equations relating the mean and variance parameters to the assumed covariates assuming gamma and beta distributions for the data. In this way, we have, the following modeling steps:

1. If Y_t , $t = 1, 2, \dots, T$, follows a gamma conditional distribution $G(p_t, q_t)$, where $G(p, q)$ denotes a gamma distribution with mean p/q and variance p/q^2 , the conditional mean and variance are related to the original parameters by the equations $\mu_t = p_t/q_t$ and $h_t = p_t/q_t^2$.
2. If Y_t , $t = 1, 2, \dots, T$, follows a beta distribution function $B(p_t, q_t)$, we consider a reparameterization of the beta distribution density as a function of the mean and precision, $\phi t = p_t + q_t$, which results to be appropriate in order to define the joint mean and precision beta regression models as introduced by Cepeda (2001). This reparameterization, where $\phi = p + q$, $p = \mu\phi$ and $q = \phi(1 - \mu)$, has been extensively used in the literature following the joint modeling approach for the mean and precision beta parameters introduced by Cepeda (2001) and Cepeda and Gamerman (2005), under a Bayesian approach. It is important to point out that Ferrari and Cribari-Neto (2004) also introduced a beta modeling approach for the mean but assuming constant precision parameters, under a classical approach. In all of these cases, ϕ can be interpreted as a precision parameter in the sense that, for fixed values of μ , larger values of ϕ correspond to smaller values for the variance of Y . This interpretation could be not so simple. In this paper, we use the mean and variance reparameterization of the probability beta distribution in the definition of joint mean and variance beta regression models (Cepeda-Cuervo, 2015), taking into account that $\mu(1 - \mu) > \sigma^2$; in this way, samples of the joint posterior distribution for the regression parameters should be simulated in the subspace of parameters that satisfy this property. Although this reparameterization results in a

complex expression for the beta distribution, it leads to a best and more easily interpretation for the statistical analysis results in the applications.

$$p_t = \frac{(1 - \mu_t)\mu_t^2 - \mu_t\sigma_t^2}{\sigma_t^2} \quad (8)$$

$$q_t = \frac{(1 - \mu_t)[\mu_t - \mu_t^2 - \sigma_t^2]}{\sigma_t^2} \quad (9)$$

In this reparameterization, where the autoregressive seasonal beta regression models have the conditional mean and variance model given by the equations (6) and (7).

Special cases of this model could be easily obtained from this general model. A first model is given by a seasonal regression mean model, with mean given by (6) and autoregressive variance not considering the presence of seasonal terms. A second model, also a seasonal mean model, is given assuming the mean given by (6) and a seasonal variance model not considering the presence of autoregressive terms. A third model, is given by an autoregressive mean model and a variance model, not considering the presence of seasonal terms in the mean and in the variance. A fourth model, is given by an autoregressive model, with constant variance.

4. Hydrological time series

In this section we consider an analysis of the Furnas dam hydroelectric hydrological time series dataset, introduced in Section 1, assuming seasonal autoregressive conditional heteroscedastic models.

The first step in the proposed analysis is to determine the period for the harmonics of higher intensity in the spectral analysis of the streamflows data. In this way, we note in Figure 3 that the harmonics of higher intensity (I) corresponds to the cycle ($1/f_i$) of 6 and 12 months, given that the periodogram shows that this time series contains two cosine-sine peaks at these frequencies. There are other very small peaks in this periodogram, possibly caused by noise components. Thus, the seasonal terms to be included in the mean equation model of the streamflow series are given by $\cos(2\pi t/6)$, $\sin(2\pi t/6)$, $\cos(2\pi t/12)$ and $\sin(2\pi t/12)$. The periodogram of a time series can be obtained using the R-function periodogram (x, method,..) of the library "GeneCycle" (Ahdesmaki et al., 2012). Other statistical software can be used to estimate the periodogram of a seasonal time series, for example, MathLab or the statistical software Excel-XLStat.

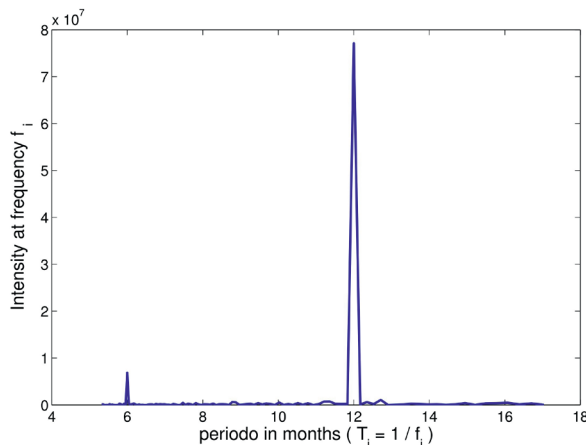


Figure 3. Periodogram of natural streamflows time series.

Many autoregressive models could be assumed to analyze this data set. As special cases, we will assume, in section 4.1, a seasonal conditional normal distribution for the data and, in section 4.2, a seasonal conditional gamma distribution for the data. In order to apply the Bayesian methodology, independent normal prior distributions $N(0,10^k)$, with $k=2$, are assumed for the regression parameters associated with the seasonal terms. For the other parameters of the model, we assume independent normal prior distributions $N(0,10^k)$, with $k=5$. It is important, to point out that non-informative prior could be used (see, for example, Geman, 2006), but in our case, the obtained results are very similar.

Since we are using the free available software WinBugs to simulate samples for the joint posterior distribution of interest in all assumed models, which only requires the introduction of the joint distribution for the data and the prior distributions for all parameters of the model, we will not present in this paper the required full conditional posterior distributions for all parameters needed in the Gibbs sampling or the Metropolis-Hastings algorithms.

4.1. Normal seasonal time series

In this section, we present the results of the Bayesian analysis for the time series of monthly averages of natural streamflows, measured in the period ranging from 1931 to 2010, considering the dataset related to a hydroelectric dam introduced in Section 1, assuming jointly modeling for the mean and variance, assuming a normal model for the data. From the model given by equations (6) and (7), many autoregressive models were considered in the analysis of the data, and the model with smallest DIC (deviance information criterion, introduced by Spiegelhalter et al, 2002) value was given by the heteroscedastic normal regression model with conditional mean and variance structures given respectively by,

$$\mu_t = \alpha_0 + \alpha_{1,1} \cos(2\pi t/12) + \alpha_{1,2} \cos(2\pi t/6) + \alpha_{2,1} \sin(2\pi t/6) + \phi_1 y_{t-1} + \phi_2 y_{t-2} \quad (10)$$

$$\log(h_t^2) = \lambda_0 + \lambda_{1,1} \cos(2\pi t/12) + \theta_1 y_{t-1} + \theta_3 y_{t-3} \quad (11)$$

Samples of the joint posterior distribution of interest were simulated using standard MCMC (Markov Chain Monte Carlo) methods and the free available WinBugssoftware (Spiegelhalter et al, 2003). In each of the cases, many samples were generated starting from different initial values. All of them showed the same behavior, after a small burn-in period consisting of 3,000 or 5,000 generated samples. Convergence of the simulation algorithm was observed from trace plots of the generated Gibbs samples.

For the model given by equations (10) and (11), the value of the logarithm of the likelihood function evaluated at the obtained estimates for the parameters of the model was given by $-2\log L=10479.200$ and the obtained DIC value was $DIC=10876.500$. Monte Carlo estimates of the posterior means for each parameter based on the generated Gibbs samples and their respective standard deviations are given in Tables 1 and 2, including both autoregressive and seasonal terms in the conditional mean and variance terms.

Table 1. Normal model: Bayesian estimates for the mean parameters.

Parameter	α_0	$\alpha_{1,1}$	$\alpha_{1,2}$	$\alpha_{2,1}$	ϕ_1	ϕ_2
Mean	77.44 (6.409)	162.7 (9.71)	62.1 (5.557)	40.45 (5.597)	0.674 (0.03023)	0.124 (0.02258)

Table 2. Bayesian estimates of the variance parameters.

Parameter	α_0	$\lambda_{1,1}$	θ_1	θ_3
Mean	7.926 (0.1563)	1.743 (0.1351)	0.002357 (2.097E-4)	5.559E-4 (2.452E-4)

As an illustration of the performance of the proposed model, Figure 4 shows the agreement between monthly averages of natural streamflows and the fitted mean estimate. At the same time, Figure 5, depicting of the fitted squared root of the expected volatility shows periods of high volatility around of the months 200, 400, 600 and 800, as observed in the data behavior.

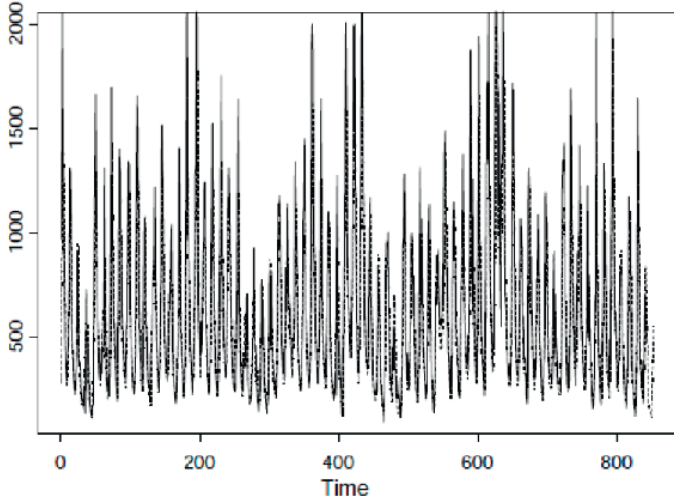


Figure 4. Monthly averages of natural streamflow and normal fitted mean estimates.

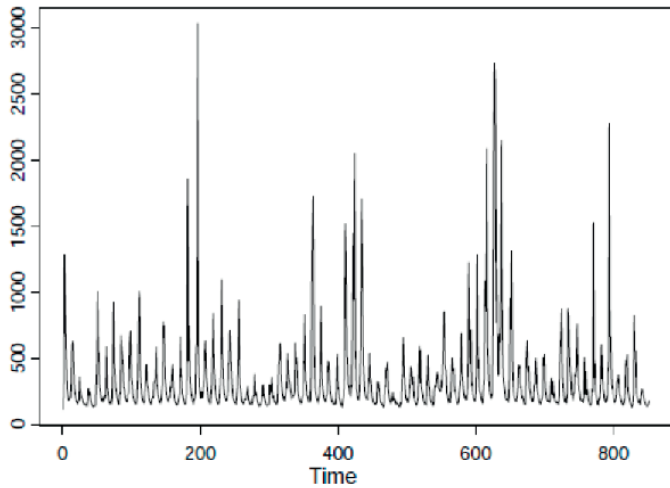


Figure 5. Normal fitted squared root of the expected volatility.

4.2 Gamma seasonal time series

In this section we present the results of the Bayesian analysis for the time series for monthly averages of natural streamflows, measured in the period 1931 to 2010, in Furnas hydroelectric dam, assuming joint modeling for the mean and variance autoregressive gamma models, that is, we assume that the observations of the interest are generated from a conditional gamma density function given by $f(y_t | H_{t-1})$, where H_{t-1} is the information up to time $t - 1$ and Y_t has conditional mean and conditional variance given respectively by $\mu_t = E(Y_t | H_{t-1})$ and $\sigma_t^2 = Var(Y_t | H_{t-1})$, and defined by (10) and (11). For this model, the logarithm of the likelihood function evaluated at the estimates for the parameters of interest is given by $-2\log L = 10649$ and the DIC value is given by 10862.300. Using the DIC criterion to discriminate the two models (normal seasonal time series and gamma seasonal time series), we observe better fit of the data for the gamma seasonal time series model, since we have smaller DIC value for this model.

The Bayesian estimates of the posterior means for the parameters together with the corresponding standard deviations are given in Tables 3 and 4.

Table 3. Gamma model: Bayesian estimates for the mean parameters.

Parameter	α_0	$\alpha_{1,1}$	$\alpha_{1,2}$	$\alpha_{2,1}$	ϕ_1	ϕ_2
Mean	78.74 (7.126)	176.6 (8.615)	60.2 (5.645)	41.59 (5.567)	0.7239 (0.03278)	0.1181 (0.0238)

Table 4. Gamma model: Bayesian estimates for the variance parameters.

Parameter	λ_0	$\lambda_{1,1}$	θ_1	θ_3
Mean	7.942 (0.1394)	1.595 (0.1156)	0.002513 (1.829E-4)	5.467E - 4 (2.151E-4)

Although the mean parameter estimates given in Tables 1 and 3, and variance parameter estimates given in Tables 2 and 4, show some agreement between conditional normal and conditional gamma regression parameter estimates, the DIC value of the conditional gamma model is smaller than that of the conditional normal heteroscedastic models, showing that the class of gamma models is better to fit the monthly averages of natural streamflow data.

In order to determine the forecasting performance of the proposed Gamma seasonal model, we fit this model to the first 910 observed values of the Furnas dam hydrological time series, in order to predict the temporal behavior over the 50 last months. The parameter estimates agree with those reported in tables 3 and 4. Figure 6 shows that this model has good performance in predicting monthly averages of natural streamflows. This figure shows the 90% prediction interval for the variable of interest in these last weeks in dashed lines, and the observed values of natural streamflows in black points.

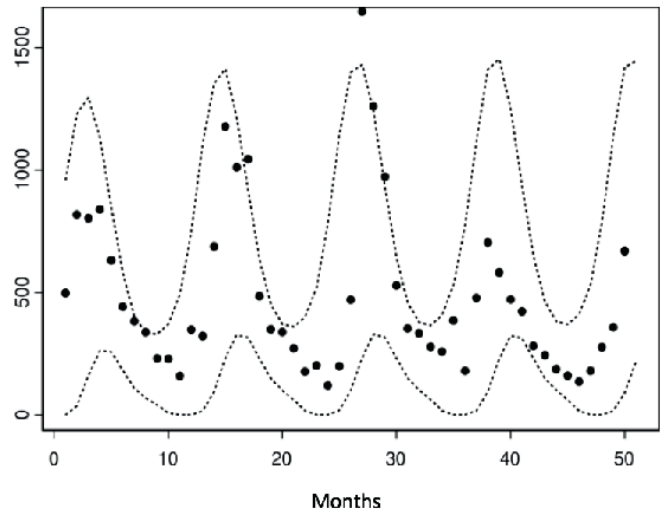


Figure 6. Monthly averages of natural streamflow.

5. Beta mean and variance seasonal models applied to air humidity time series

In this section it is assumed that the time series data are generated from a beta distribution, $B(p_i, q_i)$, with conditional mean and variance given, respectively, by (6) and (7). To illustrate the application of the proposed model to be fitted by the data set, we consider the time series of weekly averages air humidity, measured in Rio Claro city, located in southeastern of Brazil, from 18/10/2002 to 08/10/2008. This time series, introduced in section 1, is shown in Figure 2.

As in Section 4, the first step in the proposed analysis is to determine the period of harmonics of higher intensity in the spectral analysis of the humidity data. From the corresponding periodogram, Figure 7, we note that the

harmonics of higher intensity corresponds to the cycle ($1/f$) of 26 and 52 days. As in the analysis of streamflow data, the periodogram shows that this time series contains two cosine-sine peaks at these frequencies, two larger peaks, and other very small peaks, possibly caused by noise components. Thus, the seasonal term to be included in the mean equation model of the streamflow series are given by: $\cos(2\pi t/26)$, $\sin(2\pi t/26)$, $\cos(2\pi t/52)$ and $\sin(2\pi t/52)$.

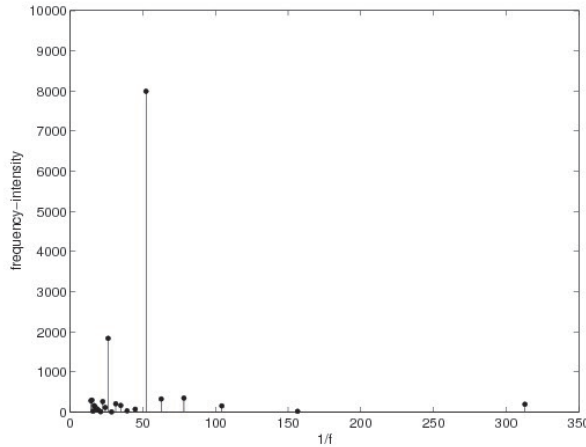


Figure 7. Periodogram of humidity Times series.

5.1. Seasonal mean and conditional variance models

In this section, double seasonal beta regression models are proposed to analyze the air humidity time series data introduced in Section 1. In this way, we conclude that the best beta seasonal mean and variance model is the one where the mean and dispersion regression structures are given by:

$$\text{logit}(\mu_t) = \alpha_0 + \alpha_{1,1} \cos(2\pi t \setminus 52) + \alpha_{1,2} \cos(2\pi t \setminus 26) + \alpha_{2,1} \text{sen}(2\pi t \setminus 52) + \alpha_{2,2} \text{sen}(2\pi t \setminus 26) + \phi_1 \text{logit}(y_{t-1}) \quad (12)$$

$$\log(h_t) = \lambda_0 + \lambda_{1,0} \cos(2\pi t \setminus 52) + \lambda_{2,1} \text{sen}(2\pi t \setminus 52) + \lambda_{2,2} \text{sen}(2\pi t \setminus 26) \quad (13)$$

Assuming independent normal prior distributions $N(0,100)$ for the regression parameters α_0 and $\alpha_{1,i}$, for $i=1,2$ and for λ_0 and $\lambda_{2,i}$, for $i=1,2$. To fit the model 20,000 samples of the joint posterior distribution for the parameters of the model were also generated using the MCMC methods and the WinBugs software (Spiegelhalter et al, 2003). Monte Carlo estimates for the posterior means of each parameter were obtained from the final simulated Gibbs sample, after an initial burn-in sample period of 2000 samples. This ‘‘burn-in sample’’ was discarded to eliminate the effect of the initial values in the iterative procedure. After this ‘‘burn-in sample’’ period we simulated another 20,000 Gibbs samples choosing every 20 iteration to get approximately non-correlated samples, which gives a final sample of size 1,000 used to get the posterior summaries of interest. The posterior summaries of interest are given in Table 5, for the mean regression parameters, and in Table 6, for the variance regression parameters.

Table 5. Beta model: Bayesian estimates for mean parameters.

Parameter	α_0	$\alpha_{1,1}$	$\alpha_{2,1}$	$\alpha_{1,2}$	$\alpha_{2,2}$	ϕ_1
Mean	0.5845 (0.0469)	-0.2091 (0.03039)	0.09499 (0.02583)	-0.06574 (0.02488)	0.08705 (0.02481)	0.2741 (0.05449)

Table 6. Beta model: Bayesian estimates for the variance parameter.

Parameter	λ_0	$\lambda_{1,1}$	$\lambda_{2,1}$	$\lambda_{1,2}$
Mean	-5.521 (0.07953)	0.2095 (0.1161)	0.2327 (0.1195)	-0.213 (0.1173)

The logarithm of the likelihood function evaluated at the obtained estimates for the parameters of the model is given by $\log L = -853.082$ and the DIC criterion has a value equals to -825.525 . In Figure 8, we observe a good agreement between data and the fitted mean, showing the good performance of the proposed model to analyze this data set. Figure 9, reveals a good agreement for the variances, where smaller means are accompanied by smaller variance. This behavior also was observed in the original time series. That is, we conclude that this model fits by the time series data very well.

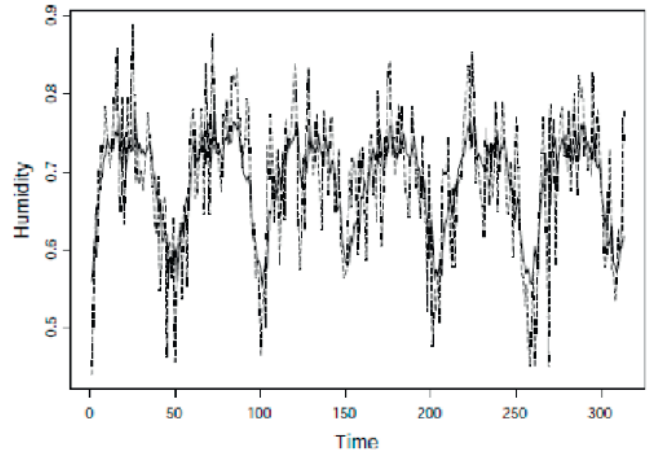


Figure 8. Air humidity time series data (continuous line) and fitted mean (diseased line) for double seasonal model.

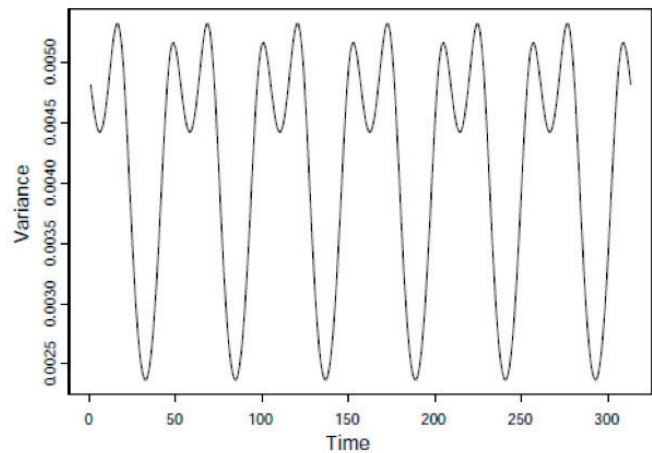


Figure 9. Fitted conditional variance time series.

To determine the forecasting performance of the proposed models, we fit the seasonal beta autoregressive model to the first 282 observed values of the air humidity time series data, in order to predict the time behavior in the 30 last weeks. Figure 10 shows the 90% prediction interval for these last weeks, in dotted lines and the observed values of natural streamflows in black points.

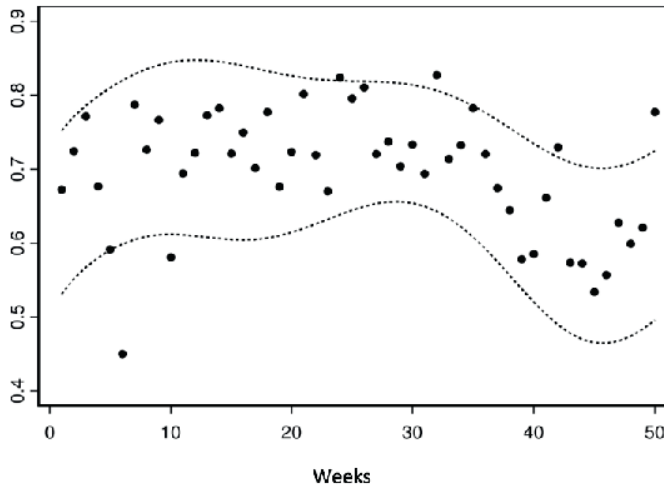


Figure 10. Forecasting humidity time series.

6. Concluding Remarks

In this paper, we introduce a new class of time series models assuming continuous random variables within the exponential family applied to hydrological and meteorological data. Special cases of this proposed methodology were considered assuming, normal, gamma and beta distributions. These new models could be of great interest to analyze hydrological and meteorological data, since the original data usually could not be fitted assuming the standard modeling approach for the means of the data or assuming standard distribution assumptions for time series data. Other important practical point: in practical work, usually it is required to model simultaneously the mean and the variance depending on a vector of parameters to get better predictions and to discover the real behavior of a hydrological or a meteorological time series. With this modeling approach, we could get better inferences as it was observed in the illustrations introduced in this paper. This approach usually is not considered by hydrological or meteorological researchers using standard available software for hydrology or meteorology time series. In this way, the use of Bayesian methods is a good alternative to get accurate inferences and predictions, especially using MCMC methods, since these inference methods are not based on asymptotical results as its common with standard existing classical approaches based on maximum likelihood estimation procedures. Other important point: under the use of a Bayesian approach, we could consider informative prior distributions when is available prior opinion of experts in hydrology and meteorology. This means, better predictions. It is also important, to point out that the computational work needed in the simulation of samples of the joint posterior distribution to get the posterior summaries of interest is greatly simplified using standard existing softwares like the WinBugs software. The proposed methodology was illustrated considering two Brazilian data sets: a hydro- logical time series and a meteorological time series. A final point of relevance for the results of this paper: new statistical models for time series to model climatic variables leading to better inferences and predictions based on data from hydrology and meteorology can be of great practical interest, especially with the large climate changes that have been observed in the world in recent decades.

To determine the performance of the proposed models, an extensive comparison with other models proposed in the literature should be developed. However, in order to compare the performance of the gamma seasonal models with in a model without a seasonal component, we fit the autoregressive model Ar(3), for which the DIC value was 11990, bigger than the DIC value of the fitted gamma seasonal model. The sum of square residuals obtained from applying the AR(3) model is also bigger than that for the proposed models. For the AR(3) model, the sum of square residuals is 65679347 and for the proposed model it is 52976950. Finally, we fit the Ar(3) model to the humidity data, for which the DIC value was -774.550, bigger than -825.525, the BIC value of the fitted beta seasonal model. The sum of square residuals obtained from applying

the AR(3) model to the humidity data is 17.1565, larger than the DIC value (1.3786) of the proposed models.

References

- Ahdesmaki M, Fokianos K, Strimmer K (2012) Package ‘GeneCycle’. R-project website. URL: <http://cran.r-project.org/web/packages/GeneCycle/GeneCycle.pdf>. Accessed 2017 Dec 5.
- Cepeda-Cuervo, E. (2001). *Variability Modeling in Generalized Linear Models*, Unpublished Ph.D. Thesis. Mathematics Institute, Universidade Federal do Rio de Janeiro. http://www.docentes.unal.edu.co/ecepedac/docs/MODELAGEMDA_VARIABILIDADE.pdf
- Cepeda-Cuervo, E., Andrade, M. G., & Achcar, J. A. (2012). *A seasonal and heteroscedastic gamma model for hydrological time series: A Bayesian approach*. In: AIP Conference Proceedings, Vol. 1490, p. 97.
- Cepeda C. E. & Gamerman D. (2005). Bayesian methodology for modeling parameters in the two parameter exponential family. *Estatística*, 57, 93-105.
- Chiang, S. M., Tsay, T. K. & Nix, S. J. (2002a). Hydrologic regionalization of watersheds. I: Methodology development. *Journal of Water Resources Planning and Management*, 128(1), 3-11.
- Chiang, S. M., Tsay, T. K. & Nix, S. J. (2002b). Hydrologic regionalization of watersheds. II: Applications. *Journal of Water Resources Planning and Management*, 128(1), 12- 20.
- Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modeling rates and proportions. *Journal of Applied Statistics*, 31, 799-815.
- Fleming, S. W., Marsh Lavenue, A., Aly, A. H., & Adams, A. (2002). Practical applications of spectral analysis to hydrologic time series. *Hydrological processes*, 16(2), 565-574.
- Guimaraes, R. & Santos, E. G. (2011). Principles of stochastic generation of hydrologic time series for reservoir planning and design: A case study. *Journal of Hydrologic Engineering*. In Press.
- Hasebe, M., Dandou, T., Kumekawa, T. & Neijou, S. (2000). *Time series analysis of monthly rainfall, mean air temperature and carbon dioxide*. In: W. Z. Y. & S. X. Hu (eds.) Proceedings of the eighth International Symposium on Stochastic Hydraulics, 533-537. Beijing, China.
- Hipel, K. W. & McLeod, A. E. (1994). *Time series modeling of water resources and environmental systems*. Elsevier, Amsterdam, The Netherlands.
- Hosking, J. R. M. (1984). Modeling persistence in hydrological time series using fractional differencing. *Water Resources Research*, 20(12), 1898-1908.
- Lee, J. Y. & Lee, K. K. (2000). Use of hydrologic time series data for identification of recharge mechanism in a fractured bedrock aquifer system. *Journal of Hydrology*, 229, 190-201.
- Lall U, Mann M. 1995. The Great Salt Lake: a barometer of low frequency climatic variability. *Water Resources Research* 31(10), 2503–2515.
- Marques, C. A. F., Ferreira, J. A., Rocha, A., Castanheira, J. M., Melo-Goncalves, P., Vaz, N. & Dias, J. M. (2006). Singular spectrum analysis and forecasting of hydrological time series. *Physics and Chemistry of the Earth, Parts A/B/C*, 31(18), 1172-1179.
- Modal, M. S., & Wasimi, S. A. (2006). Generating and forecasting monthly flows of the Ganges river with par model. *Journal of Hydrology*, 323(1-4), 41-66.
- Montanari, A., Rosso, R. & Taquq, M. S. (1997). Fractionally differenced arima models applied to hydrologic time series: Identification, estimation and simulation. *Water Resources Research*, 33(1-4), 1035-1044.
- Robin MJL, Gutjahr AL, Sudicky EA, Wilson JL. 1993. Crosscorrelated random field generation with the direct Fourier transform method. *Water Resources Research* 29(7): 2385–2397.
- Salas, J. D., Delleur, J. W., Yevjevich, V. & Lane, W. L. (1980). *Applied*

- modeling of hydrologic time series*. Water Resources Publications, Littleton, USA.
- Salas, J. D., Boes, D. C. & Smith, R. A. (1982). Estimation of ARMA models with seasonal parameters. *Water Resources Research*, 18(4), 1006-1010.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, 64(4), 583-639.
- Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W. R. (2003). *Win-BUGS User Manual (version 1.4)*. MRC Biostatistics Unit, Cambridge, U.K.
- Tesfaye, Y. G., Meerschaert, M. M. & Anderson, P. L. (2006). Identification of periodic autoregressive moving average models and their application to the modeling of river flows. *Water Resources Research*, 42(W01419), 1-11.
- Wang, Q. J., Robertson, D. E. & Chiew, F. H. S. (2009). A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resources Research*, 45(W05407), 1-18.
- Wang, W. C., Chau, K. W., Xu, D. M., & Chen, X. Y. (2015). Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resources Management*, 29(8), 2655-2675.
- Wu, Z., & Huang, N. E. (2009). Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis*, 1(01), 1-41.
- Valipour, M., Banihabib, M. E., & Behbahani, S. M. R. (2013). Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *Journal of hydrology*, 476, 433-441.