Joint Deterministic and Stochastic Modelling of Discharge in Large Brazilian Rivers

Prass, Taiane S.[1]
E-mail: taianeprass@gmail.com

Clarke, Robin T.[2]
E-mail: clarke@iph.ufrgs.br

Collischonn, Walter[2]
E-mail: collischonn@iph.ufrgs.br

Lopes, Sílvia R.C.[1]
E-mail: silvia.lopes@ufrgs.br

Universidade Federal do Rio Grande do Sul (UFRGS),
[1] *Instituto de Matemática*
[2] *Instituto de Pesquisas Hidráulicas*
9500, Bento Gonçalves Avenue
91509-900, Porto Alegre, Brazil

Introduction

A cyclical behavior is a common characteristic of most hydrologic time series, and the literature shows that long-range behavior can sometimes be expected (Hosking, 1984; Montanari et al., 1997, 2000; Bisognin and Lopes, 2007; Prass et al., 2011).

To model long memory behavior, Granger and Joyeux (1980) and Hosking (1981) introduced the autoregressive fractionally integrated moving average (ARFIMA or FARIMA) model. To account for the cyclical behavior, Porter-Hudak (1990) introduces the seasonal ARFIMA (SARFIMA) models. The theoretical properties of SARFIMA processes as well as Monte Carlo simulation studies regarding estimation and forecasting on these processes are presented in Bisognin and Lopes (2007, 2009 and 2011).

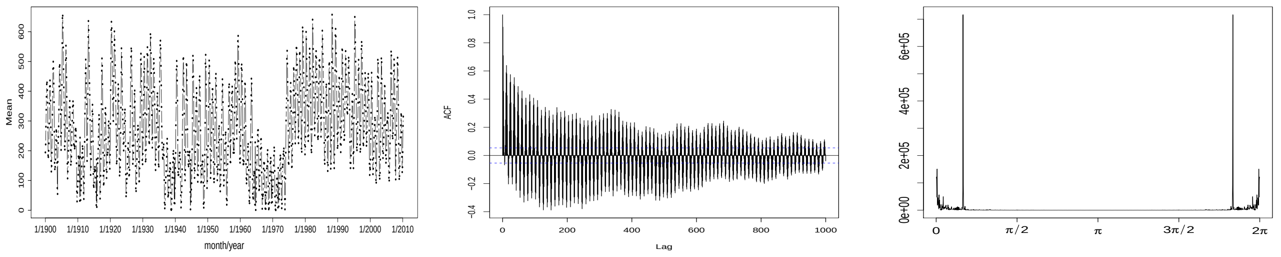
In this work we analyze the mean monthly water-level in the Paraguay River at Ladário and the daily discharge time series for the Amazon River at Óbidos. While a complete SARFIMA model is consider to model the time series of water-level, a harmonic model is used for the time series of discharge.

Paraguay River at Ladário

The Paraguay River is a major river in South America, running through Brazil, Bolivia, Paraguay and Argentina. It flows 2,621 kilometers from its headwaters in the Mato Grosso state (Brazil) to its confluence with the Paraná River north of Corrientes (Argentina). More details on the characteristics of the Paraguay River are given in Prass et al. (2011).

Figure 1 shows, respectively, the time series $\{X_t\}_{t=1}^{n}$ of the mean monthly water-level in the Paraguay River at Ladário, in the period from January 1900 to March 2010 (a total of $n = 1323$ observations), its sample autocorrelation (for lags $h = 0, \cdots, 1000$) and periodogram functions. The decline in water-level over the extended period from 1960 to about 1975, which can be observed in Figure 1, has never been fully explained, but is replicated in other time series of river flows from other part of the la Plata drainage system. From the decay in the sample autocorrelation, from its cyclical behavior and from the peak in the periodogram function, it is evident the presence of seasonal long-memory in this time series. The highest peak in the periodogram function corresponds to the Fourier frequency $\lambda_j = \frac{2\pi j}{1323}$, with $j = 110$, which leads to the conclusion that $s = \frac{n}{j} = 12.027 \approx 12$.

Figure 1: Time Series of the mean monthly water-levels in the Paraguay River at Ladário in the period from January 1900 to March 2010 and its autocorrelation and periodogram functions.



The analysis of the sample autocorrelation and periodogram functions suggests considering a SARFIMA$(p, d, q) \times (P, D, Q)_s$ model, with seasonal period $s = 12$ (see the theoretical properties of SARFIMA processes in Bisognin and Lopes, 2007, 2009, 2011). Thus, one has

$$\phi(\mathcal{B})\Phi(\mathcal{B}^s)(1-\mathcal{B})^d(1-\mathcal{B}^s)^D(X_t - \mu) = \theta(\mathcal{B})\Theta(\mathcal{B}^s)\varepsilon_t, \quad \text{for all } t \in \mathbb{Z},$$

where $\mu \in \mathbb{R}$ is the process mean; $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a white noise process with mean zero and variance σ_ε^2; d and D are, respectively, the nonseasonal and seasonal differencing order parameters (allowed to be fractional); $s \in \mathbb{N}^*$ is the period of the seasonality; $\mathcal{B}$ is the backward shift operator defined by $\mathcal{B}^{sk}(X_t) = X_{t-sk}$, for all $k, s \in \mathbb{N}$; $(1-\mathcal{B})^d$ and $(1-\mathcal{B}^s)^D$ are, respectively, the nonseasonal and the seasonal difference operators, defined through the expression $(1-\mathcal{B}^s)^D = \sum_{k=0}^{\infty} \delta_{D,k}\mathcal{B}^{sk}$, with $\delta_{D,0} = 1$ and $\delta_{D,k} = \frac{\Gamma(k+D)}{\Gamma(-D)\Gamma(k+1)}$, for all $k > 0$, and the operator $(1-\mathcal{B})^d$ is obtained when $D = d$ and $s = 1$; $\phi(\cdot)$ and $\Phi(\cdot)$ are, respectively, the nonseasonal and seasonal autoregressive polynomials, $\theta(\cdot)$ and $\Theta(\cdot)$ are, respectively, the nonseasonal and seasonal moving average polynomials, defined by $\phi(z) = \sum_{k=0}^{p}(-\phi_k)\, z^k$, $\theta(z) = \sum_{k=0}^{q}\theta_k\, z^k$, $\Phi(z) = \sum_{k=0}^{P}(-\Phi_k)\, z^{sk}$ and $\Theta(z) = \sum_{k=0}^{Q}\Theta_k\, z^{sk}$, with $\phi_0 = -1 = \Phi_0$ and $\theta_0 = 1 = \Theta_0$.

The model selection was performed as follows:

1. We use the first $n = 1310$ observations to fit the model and save the last 13 to compare with the out-of sample forecasts.

2. We set $\hat{\mu} = \bar{X}$, where $\bar{X}$ is the sample mean of $\{X_t\}_{t=1}^n$ and we consider all possible models with $p, q \in \{0, \cdots, 4\}$ and $P, Q \in \{0, 1, 2\}$. Given the time series $\{X_t - \hat{\mu}\}_{t=1}^n$, the parameter estimation is carried out by minimizing the function $Q(\cdot)$ (see Prass et al., 2011) which is an approximation of the Gaussian maximum likelihood function in the spectral domain (see Beran, 1994). The spectral density function of a complete SARFIMA process, as well as its asymptotic behavior near the seasonal frequencies is given in Bisognin and Lopes (2009).

3. Once the model is estimated, the residuals $\{\hat{\varepsilon}_t\}_{t=1}^n$ are calculated based on the infinite order autoregressive representation of a SARFIMA process (see Prass et al., 2011). By letting $\hat{X}_t := X_t - \hat{\varepsilon}_t$, for all $t \in \{1, \cdots, n\}$, we obtained the fitted-values or in-sample forecast.

4. For each residuals time series $\{\hat{\varepsilon}_t\}_{t=1}^n$ we verify graphicaly the assumption of non-correlated residuals by ploting the sample autocorrelation function. As measures of in-sample forecasting performance,

we calculate the residuals mean absolute value (mae) and the mean absolute percentage error ($mape$), defined as

$$mae = \frac{1}{n}\sum_{t=1}^{n} |\hat{\varepsilon}_t| \quad \text{and} \quad mape = \frac{1}{n}\sum_{t=1}^{n} \frac{|\hat{\varepsilon}_t|}{|X_t|}, \qquad \text{for all } t \in \{1, \cdots, n\}.$$

5. If more than one model presents uncorrelated residuals, we perform the out-of-sample forecast based on those models. Expression for the h-step ahead forecast and its mean square error are given in Bisognin and Lopes (2011). The forecasting performance is measured by calculating the mean absolute error of forecast (mae_f), defined as

$$mae_f = \frac{1}{n_p}\sum_{h=1}^{n_p} |X_{n+h} - \hat{X}_{n+h}|,$$

   where n is the forecasting origin, $n_p = 1323 - n$ is the total number of predicted values.

6. Among all the models with similar forecasting performances, we choose the more parsimonious one.

7. After selecting the final model, we also estimate the model parameters by considering other two different sub-samples from the data. We fix the starting point as the first observation in the time series and consider as ending points the values $n_1 = 721$ and $n_2 = 913$. These values correspond, respectively, to January 1961 and January 1976 (Prass et al., 2011, consider the values $n_1 = 661$ and $n_2 = 992$). This analysis allow us to observe if the parameters of the model change in the period from January 1961 to December 1975 and also, if the model is able to describe (or predict, when $n = 721$) the time series behavior in this period.

Table 1 presents the estimated parameter values for two SARFIMA$(p, d, q) \times (P, D, Q)_s$ models. To calculate the p-values for the estimated parameters we consider their asymptotic distribution, which is Gaussian. Since this distribution is well known and the p-values can be easily calculated, they are not presented in Table 1. Although there is no previous indication that $d = 0$, we also consider this possibility (Model 2) given that in our first analysis (Model 1) we found $d + D > 0.5$.

Tabel 1: Parameter estimation for the mean monthly water-level in the Paraguay River at Ladário. The value in parenthesis corresponds to the standard error of the estimate.

n	Model	Estimate						
		$\hat{d}$	$\hat{D}$	$\hat{\phi}_1$	$\hat{\theta}_1$	$\hat{\Phi}_1$	$\hat{\mu}$	$\hat{\sigma}^2_\varepsilon$
721	1	0.2681 (0.1340)	0.3566 (0.0329)	0.7909 (0.0890)	0.2703 (0.0608)	-0.2473 (0.0454)	264.1535	1412.1780
	2	-	0.3680 (0.0324)	0.8962 (0.0175)	0.3942 (0.0359)	-0.2493 (0.0454)	264.1535	1434.1800
913	1	0.2152 (0.1164)	0.3529 (0.0294)	0.7982 (0.0764)	0.3033 (0.0526)	0.2299 (0.0408)	243.3060	1450.1480
	2	-	0.3641 (0.0290)	0.8922 (0.0159)	0.3997 (0.0319)	-0.2333 (0.0407)	243.3060	1462.6980
1310	1	0.2423 (0.0960)	0.3957 (0.0244)	0.7614 (0.0702)	0.3480 (0.0391)	0.2536 (0.0335)	273.8886	1332.9840
	2	-	0.4118 (0.0240)	0.8806 (0.0139)	0.4387 (0.0261)	-0.2561 (0.0335)	273.8886	1348.0860

By letting $D(n)$ (equivalently, $d(n), \phi_1(n), \theta_1(n)$ and $\Phi_1(n)$) be the parameter D (respectively, d, ϕ_1, θ_1 and Φ_1) corresponding to the SARFIMA model for $\{X_t\}_{t=1}^{n}$, for any $n \in \{721, 913, 1310\}$, the results on Table 1 can be summarized as follows

1. as expected, when the parameter $d(n)$ is introduced in the model, the estimated values of $\phi_1(n)$ and $\theta_1(n)$ decrease and the other estimates, including $\hat{\sigma}_\varepsilon^2(n)$, remain almost the same as in the case $d(n) = 0$.

2. the p-values of the estimates of $d(n)$, for $n \in \{721, 913, 1310\}$, are, respectively, 0.0454, 0.0645 and 0.0116, which lead to the conclusion that $d(n) \neq 0$ at the 5% significance level.

3. the standard deviation $\sigma_\varepsilon(n) := \sqrt{\sigma_\varepsilon^2(n)}$ values, for all $n \in \{721, 913, 1310\}$, are close together.

4. for both models, the confidence intervals for $\eta(n_1)$ and $\eta(n_2)$, for any $\eta \in \{d, D, \phi_1, \theta_1, \Phi_1\}$ and $n_1, n_2 \in \{721, 913, 1310\}$ is non-empty. Thus, one cannot reject the hypothesis that $\eta(n)$ are all statistically equal, for all $n \in \{721, 913, 1310\}$ and any $\eta \in \{d, D, \phi_1, \theta_1, \Phi_1\}$.

Based on these findings, there is no statistical evidence that the model varies as n increases and the same model could be used if only a few new observations are available.

Table 2 shows the descriptive statistics for the residual $\{\hat{\varepsilon}_t\}_{t=1}^n$ time series of the SARFIMA models adjusted to the mean monthly water-level in the Paraguay River at Ladário, for all $n \in \{721, 913, 1310\}$. From this table one observes that, for each statistic considered there, the estimated values remain almost the same, as n increases. One also observes that the sample mean is close to zero and the residuals distribution is almost symmetric (see also the histograms in Figure 2). This result shows that the hypothesis that $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a stationary process seems to hold. As expected, in both cases, the *mae* decreases as the sample size increases. From the *mae* and *mape* values presented in Table 2 one observes that both models seem to fit the data well. However, the *mae* values for Model 1 are slightly smaller than for Model 2. The graphs with the observed and the fitted values showed that, for $n \in \{913, 1310\}$, both models capture the time series behavior in the period from January 1961 to December 1975 (these graphs are available upon request).

Tabel 2: Descriptive statistics for the residuals of the models adjusted to the mean monthly water-level in the Paraguay River at Ladário.

n	Model	Min	1st Q	Median	Mean	3rd Q	Max	*mae*	*mape*
721	1	-202.5549	-20.3621	-1.4125	0.1119	17.1150	252.0664	26.1346	0.2576
	2	-200.4060	-20.2701	-2.3088	0.0703	16.8290	246.3375	26.7158	0.2552
913	1	-202.6292	-20.2652	-1.9036	-0.1618	16.6212	249.4995	26.0266	0.4168
	2	-200.7576	-20.3567	-2.5577	-0.2341	15.5220	246.6846	26.4339	0.3988
1310	1	-209.9899	-18.6809	-2.3516	0.0748	15.4163	251.2380	24.5228	0.3071
	2	-207.8147	-18.9811	-2.1124	0.1250	15.4565	248.1358	24.8991	0.2958

Figure 2 presents the time series $\{\hat{\varepsilon}_t\}_{t=1}^{1310}$ corresponding to the residuals of Model 1 (the graphs for Model 2 are almost identical and are available upon request). This figure also shows the sample autocorrelation function $\hat{\rho}_\varepsilon(\cdot)$, for $h \in \{0, \cdots, 200\}$, the histogram (in the same graph is also the kernel density function) and the QQ-plot for the residuals time series. The graphs for $n \in \{721, 913\}$ present a similar behavior and are available upon request. The similarity between the graphs in Figure 2 and the graphs for Model 2 is not a surprise, given that the values of the parameters ϕ_1 and θ_1 are smaller when $d \neq 0$ than when $d = 0$. The sample autocorrelation function $\hat{\rho}_\varepsilon(\cdot)$ supports the hypothesis that the $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a white noise process. The histogram and the QQ-plot show that the residuals distribution is almost symmetric and clearly, it is not a Gaussian one. (Similar conclusion is achieved for the other values of n).

Figure 2: Residuals time series $\{\hat{\varepsilon}_t\}_{t=1}^{1310}$ (Model 1), its sample autocorrelation function $\hat{\rho}_\varepsilon(h)$, with $h \in \{0, \cdots, 200\}$, histogram (in the same graph is also the kernel density function) and QQ-plot.

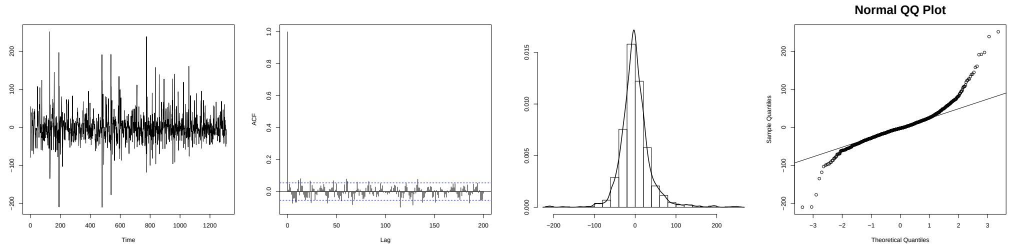
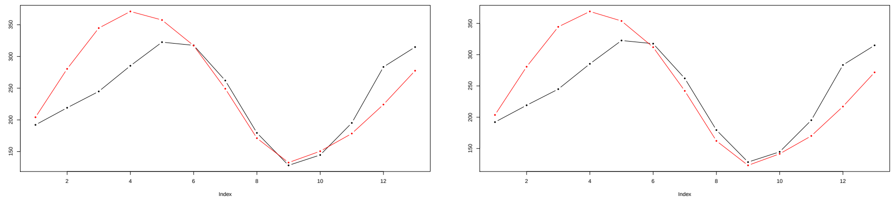


Figure 3 shows the graphs of the observed values X_{1310+h} (in black) and the corresponding h-step ahead forecast value (in red), for $h \in \{1, \cdots, 13\}$, obtained from Model 1 and 2, respectively. The graphs, considering the forecasting origin $n \in \{721, 913\}$ are available upon request. For $n \in \{721, 913\}$ we observed that, as h increases, the predicted values converge to a curve that oscillates around the mean. This fact was expected given the theoretical properties of the h-step ahead forecast, discussed in Bisognin and Lopes (2011). Also, for $n = 721$, none of the models was able to predict the fall that occurred in the period from January 1961 to December 1975. The mae_f values for $n \in \{721, 913, 1310\}$ are, respectively, 109.6520, 123.6648 and 33.775, for Model 1 and 109.2761, 121.6991 and 36.4261, for Model 2. Thus, although the differences in the mae_f values are small, the Model 2 performed better than Model 1, when $n \in \{721, 913\}$.

Figure 3: Observed values X_{1310+h} (in black) and the corresponding h-step ahead forecast value (in red), for $h \in \{1, \cdots, 13\}$, obtained from Model 1 and 2, respectively.



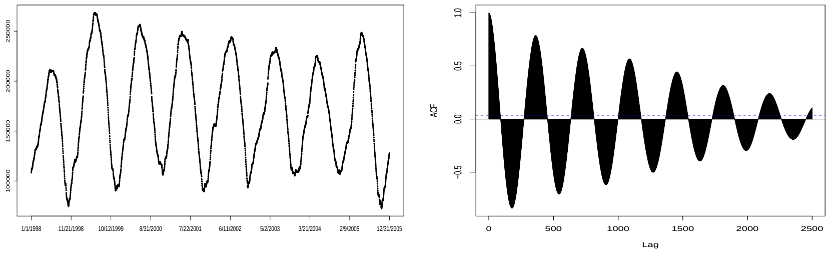
Amazon River at Óbidos

The Amazon is recognized as the world's largest river by volume, but has generally been regarded as second in length to the River Nile. The Amazon River has an average discharge greater than the next seven largest rivers combined together (not including Madeira and Negro Rivers, which are tributaries of the Amazon) and it accounts for approximately one-fifth of the world's total river flow.

This section presents a comparison between deterministic and stochastic modelling of Amazonian discharges. The goal in this work is to assess the prediction performance for each modelling approach. While the stochastic approach only considers the historical time series of discharges, the deterministic model describes the causative relationships between river discharge and precipitation (rainfall), evaporation, and characteristics of soil, vegetation, topography and geology. Here, we consider the time series of daily discharges for the Amazon River at Óbidos for the period January 1, 1998 to December 31, 2005 and in particular, the residuals given by the deterministic model MGB-IPH (for a description of this model, see Collischonn et al., 2007 and 2008). This analysis is part of an ongoing project and the stochastic model presented here is not our final model (yet to be determined).

Figure 4 presents, respectively, the Amazon discharges time series $\{Y_t\}_{t=1}^{n}$, with $n = 2922$ observations, where $t = 1$ and $t = n$ correspond, respectively, to January 1, 1998 and December 31, 2005 and its sample autocorrelation function (sample ACF). This graphs suggest that the time series presents a seasonal behavior with period $s = 365$ (one year). The slowly and steady decay of the sample ACF indicates non-stationarity. This hypothesis is confirmed by the so-called Phillips-Perron test, where we obtained the test statistic equals to -1.7992, with p-value = 0.6633.

Figure 4:* *Time series $\{Y_t\}_{t=1}^{2922}$ representing the Amazon River daily discharge in the period from January 1, 1998 to December 31, 2005 and its sample autocorrelation function.



The steps for model selection in the deterministic approach are as follows:

1. To fit the MGB-IPH model, the record of discharge is divided approximately into two halves. One half is used for model calibration (not necessarily the first half of the observation), and the other half for model verification. The model is fitted using the calibration data (i.e., the record of discharge, together with the records of daily rainfall and evaporation for the same period). Parameters are estimated by an iterative procedure analogous to the way in which genetic information is passed between generations to improve fitness, and measures of goodness of fit are calculated for this calibration period.

2. One of the most commonly used criteria for model selection is the Nash-Sutcliffe efficiency E, which is analogous to a coefficient of determination, and it is given by

$$E = 1 - \frac{\sum_{t=1}^{n}(X_t - \hat{X}_t)^2}{\sum_{t=1}^{n}(X_t - \bar{X})^2},$$

   where n is the discharge record length. The closer E is to 1, the better the fit. Another measure of goodness of fit which gives more weight to lower discharges is to calculate E using $\ln(X_t)$ instead of X_t. The mean absolute error value (mae) is also used.

3. The true measure of the model's ability to describe how a river basin responds to rainfall is obtained from the data retained for model validation. Estimates of model parameters calculated using the calibration data are now used to estimate the time series of river discharge, using records of rainfall and evaporation for the validation period. Goodness-of-fit measures are again calculated and if (for example) the Nash-Sutcliffe E is close to one for the validation period, the model is regarded as very satisfactory.

Figure 5: Deterministic Analysis: at the left hand side the observed time series $\{Y_t\}_{t=1}^{2922}$ (in black) and the fitted values $\{\hat{Y}_t\}_{t=1}^{2922}$ (in red) and at the right hand side the residuals time series.

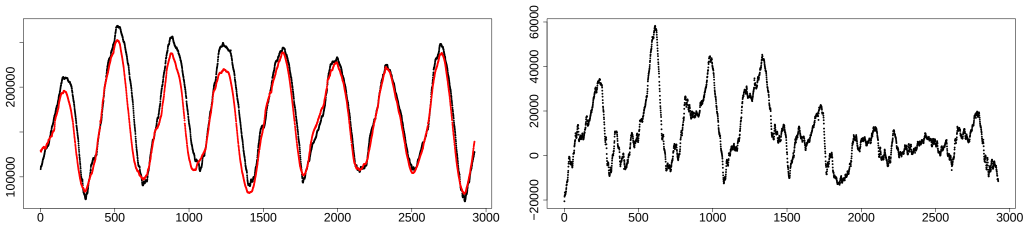


Figure 5 shows, respectively, the graph of the observed time series $\{Y_t\}_{t=1}^{2922}$ (in black) and the fitted values $\{\hat{Y}_t\}_{t=1}^{2922}$ (in red) obtained from the MGB-IPH model and the residuals time series. From this figure, the model seems to fit the data relatively well in the second half of the time series. However, for the first half of the data (validation sample) the model does not present the same performance. Moreover, one observes that the residuals for the first half of the data still present a seasonal behavior. The *mae* and *mape* for the first half of the data are, respectively, 20913.95 and 0.1248. For the second half of the data the *mae* and *mape* are, respectively, 7015.19. and 0.1139.

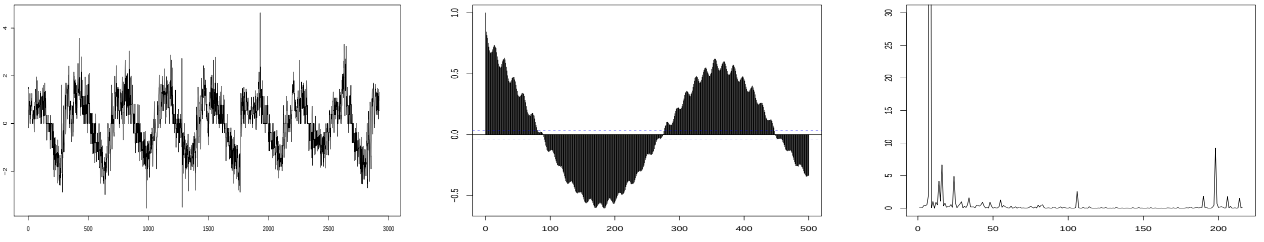
The steps in the stochastic modeling approach are as follows:

1. given the magnitude of the data, we consider a rescaled version of the original time series, given by $X_t = 10^{-3} \times Y_t$, for all $t \in \{1, \cdots, 2922\}$, where $\{Y_t\}_{t=1}^{2922}$ is the original time series.

2. The first 2981 values of the time series $\{X_t\}_{t=1}^{2922}$ are used to fit the model and we save the last 31 values to assess the out-of sample performance of the selected model.

3. Non-stationarity is removed by applying the first difference operator $(1 - \mathcal{B})$ to the time series $\{X_t\}_{t=1}^{n}$. The resulting time series is denoted by $\{U_t\}_{t=2}^{2981}$. This time series, its sample autocorrelation (ACF) and periodogram functions are presented in Figure 6. While the sample ACF indicates a periodic behavior with at least two seasonal periods, the periodogram function shows high peaks at several frequencies. Based on these facts a harmonic model seems more appropriated for this data set.

Figure 6: Time series $\{U_t\}_{t=2}^{2981}$, where $U_t = (1 - \mathcal{B})X_t$ and $\{X_t\}_{t=1}^{2922}$ are the rescaled discharges, its sample autocorrelation function at lags $h = 0, \cdots, 500$ and the periodogram function at the frequencies $\lambda_j = 2\pi j/2980$, for $j = 1, \cdots, 216$.



4. SARFIMA models do not fit the data well for any $p, q \in \{0, \cdots, 4\}$ and $P, Q \in \{0, 1, 2\}$. Therefore, we consider only harmonic models, given by

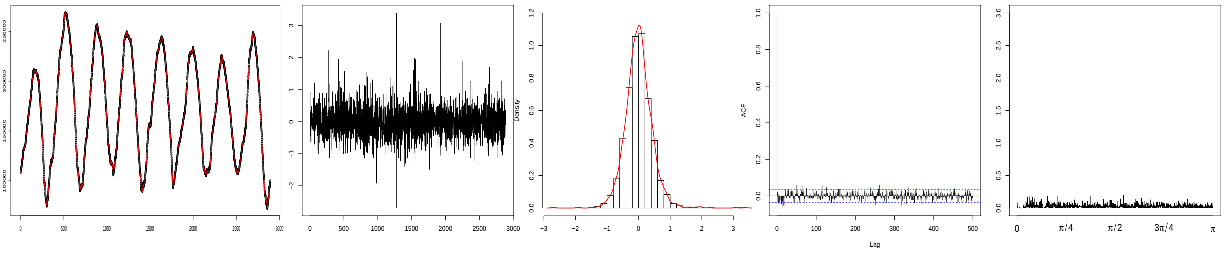
$$(1) \qquad U_t = \mu + \sum_{k=1}^{m} [A_k \cos(\omega_k t) + B_k \sin(\omega_k t)] + \varepsilon_t, \quad \text{for all } t \in \mathbb{Z},$$

   where $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is a white noise process. These processes and the optimization procedure to adjust this model is described with details in Bloomfield (1976).

5. We start the model selection by setting m as the number of frequencies λ_j for which $I_n(\lambda_j) > 0.2$, where $I_n(\cdot)$ is the periodogram function. We also set those values of λ_j as starting values of ω_k, for $k = 1, \cdots, m$, in the optimization algorithm. We increase the number of frequencies in the model until we obtain non-correlated residuals. The final model is composed by $m = 62$ frequencies (the table with the coefficients for the final model is available upon request). The authors are still working in a method to detect which coefficients can be eliminated from this model to make it more realistic.

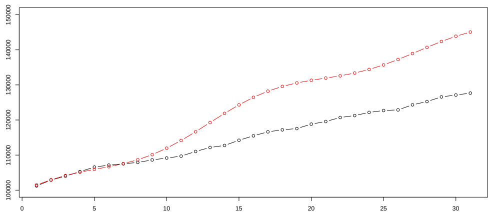
6. The fitted values $\{\hat{X}_t\}_{t=2}^{2981}$ and the residuals $\{\hat{\varepsilon}_t\}_{t=2}^{2981}$ are obtained through expression (1) upon replacing the model parameters by their estimated values. The fitted-values are then rescaled to have the same magnitude as the original data. The resulting time series is denoted by $\{\hat{Y}_t\}_{t=2}^{2981}$. Figure 7 shows, respectively, the observed (in black) and the fitted values (in red) and the residuals time series, histogram (and kernel density function), sample autocorrelation and periodogram functions. From these graphs one concludes that the model seems to fit the data well. Moreover, the residuals seem to satisfy the model's assumption and, although the distribution of the residuals seems to be symmetric, it is not Gaussian.

Figure 7: Stochastic Analysis: observed time series $\{Y_t\}_{t=1}^{2891}$ (in black) and the fitted values $\{\hat{Y}_t\}_{t=1}^{2891}$ (in red) and the residuals time series, histogram (and kernel density function), sample autocorrelation function and periodogram function.



7. The h-step ahead forecast is obtained through expression (1) upon replacing the model parameters by their estimated values and by setting $\varepsilon_{2981+h} = 0$, for $h \in \{1, \cdots, 31\}$. The final values are then rescaled to have the same magnitude as the original data. Figure 8 presents the h-step ahead forecast for $h = 1, \cdots, 31$, obtained from the harmonic model. The mae_f for this model is 8379.452 and the mean percentage error of prediction is 0.0698.

Figure 8: Graphs of the h-step ahead forecast $\{\hat{Y}_{t+h}\}_{h=1}^{31}$, for $t = 2981$ (in red) and the observed values $\{Y_{t+h}\}_{t=1}^{31}$ (in black).



Conclusions

Here we consider a stochastic model that allows for long memory and periodic behavior for the time series of water-level at a gauging site on the Paraguay River, a major tributary of the la Plata drainage system. We also present a comparison between deterministic and stochastic modelling of Amazonian discharges. This analysis is part of an ongoing project and our goal is to assess the prediction performance for each modelling approach. While the stochastic approach only considers the historical time series of discharge, the deterministic model describes how river discharge is determined by precipitation, evaporation and drainage basin characteristics of soil, geology and vegetation.

For the time series of water-level we compare two SARFIMA models. While in Model 1 the parameter d was estimated (we found $d \neq 0$ at a 5% significance level), in Model 2 we fixed $d = 0$.

Different sample sizes were considered to test if the coefficients of the model remain the same over the time. We found no evidence that the parameters change as the sample size increases. In-sample and out-of-sample forecasting performances of both, Model 1 and Model 2, were very similar and there is not enough evidence to decide which model is better.

A harmonic model was considered for the time series of Amazonian discharges. Given the high number of frequencies (62 frequencies) needed to provide a good fit (in terms of uncorrelated residuals), we conclude that further investigation is needed and perhaps a new class of models have to be considered. The deterministic model seems to provide a good fit to the data, in terms of mean absolute error. However, the residuals of this model still present a seasonal behavior. Thus, our next step is to consider a stochastic model for the residuals time series.

REFERENCES (RÉFERENCES)

Beran, J. (1994). *Statistics for Long-Memory Processes*, Chapman & Hall, New York.

Bisognin, C. and S.R.C. Lopes (2007). "Estimating and Forecasting the Long-Memory Parameter in the Presence of Periodicity". *Journal of Forecasting*, Vol. **26**, 405-427.

Bisognin, C. and S.R.C. Lopes (2009). "Properties of seasonal long memory processes". *Mathematical and Computer Modelling*, Vol. **49**, 1837-1851.

Bisognin, C. and S.R.C. Lopes (2011). "Estimation and Forecasting in Seasonal Long Memory Processes". Submitted.

Bloomfield, P. (1976). *Fourier Analysis of Time Series: An Introduction.* New York: John Wiley & Sons.

Collischonn, W., D. Allasia, B.C. da Silva and C.E.M. Tucci (2007). "The MGB-IPH model for large-scale rainfall-runoff modelling / Le modèle MGB-IPH pour la modélisation pluie-débit à grande échelle". *Hydrological Sciences Journal*, Vol. **52**(5), 878-895.

Collischonn, B., Collischonn, W. and C.E.M. Tucci (2008). "Daily hydrological modeling in the Amazon basin using TRMM rainfall estimates". *Journal of Hydrology*, Vol. **360**(1-4), 207-216.

Granger, C.W.J. and R. Joyeux (1980). "An Introduction to Long Memory Time Series Models and Fractional Differencing". *Journal of Time Series Analysis*, Vol. **1**(1), 15-29.

Hosking, J.R.M. (1981). "Fractional Differencing". *Biometrika*, Vol. **68**(1), 165-176.

Hosking, J.R.M. (1984). "Modelling Persistence in Hydrological Time Series Using Fractional Differencing". *Water Resources Research*, Vol. **20**(12), 1898-1908.

Montanari, A., R. Rosso and M.S. Taqqu (1997). "Fractionally differenced ARIMA models applied to hydrologic time series: identification, estimation, and simulation". *Water Resources Research*, Vol. **33**(5), 1035-1044.

Montanari, A., R. Rosso and M.S. Taqqu (2000). "A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan". *Water Resources Research*, Vol. **36**(5), 1249-1259.

Porter-Hudak, S. (1990). "An Application of the Seasonal Fractionally Differenced Model to the Monetary Aggregates". *Journal of the American Statistical Association*, Vol. **85**(410), 338-344.

Prass, T.S, J.M, Bravo, R.T. Clarke, W. Collischonn and S.R.C. Lopes (2011). "Use of a Seasonal Fractionally-Differenced Model for Forecasting Mean Monthly Water-Level in the Paraguay River, Brazil". Submitted.

RÉSUMÉ (ABSTRACT)

Although the Nile is the world's longest river, the Amazon has the highest mean discharge. Here we present a comparison between deterministic and stochastic modelling of Amazonian discharges. The goal in this work is to assess the prediction performance for each modelling approach. While *the stochastic approach only considers the historical time series of discharge, the deterministic model describes how river discharge is determined by precipitation, evaporation and drainage basin characteristics of soil, geology and vegetation. A stochastic model that allows for long memory and periodic behavior is also explored for water-level at a gauging site on the Paraguay River, a major tributary of the la Plata drainage system.*

Keywords: *Stochastic and deterministic models, long-memory, seasonality.*