# A Model for Hydro Inflow and Wind Power Capacity for the Brazilian Power Sector

Gilson Matos
gilson.g.matos@ibge.gov.br

Cristiano Fernandes
cris@ele.puc-rio.br

PUC-Rio
Electrical Engineering Department

GAS Workshop

Tenerife 9 - 11th January 2014

# Outline

# Outline

# Introduction

## Motivation

- In Brazil, the peak of wind power production in the Northeast coincides with the dry season of the hydro system in the Southeast (Seasonal complementarity).
- This complementarity between the hydro and wind power systems can be exploited for optimal power generation planning and energy dispatch.
- In practice this would entail the need of jointly forecasting/simulating time series of wind and hydro flow.


Figure: Brazil geographical regions.

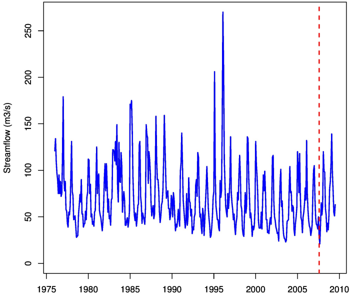



Figure: Monthly streamflow of Paraibuna river (MG), from 1976 to 2009.

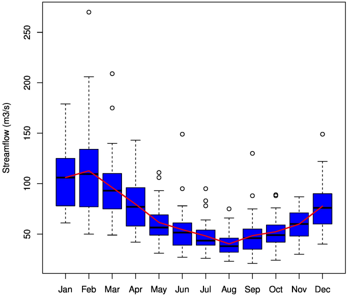



Figure: Boxplots and averages of the monthly streamflow of Paraibuna river.

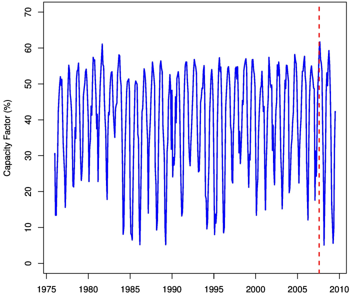



Figure: Monthly capacity factor of a wind farm in Northeast of Brazil, from 1976 to 2009.

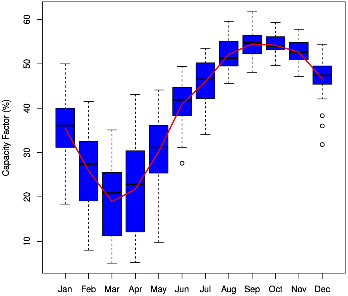



Figure: Boxplots and averages of the monthly capacity factor in Northeast of Brazil.

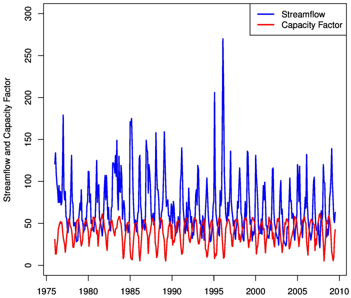



Figure: Streamflow and wind power factor series.

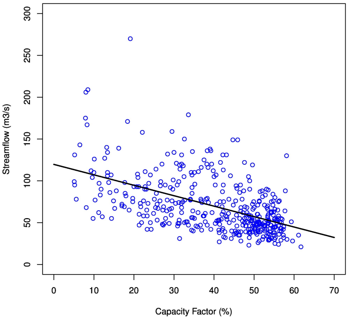



Figure: Scatter plot and regression line.

## Objectives

- To model streamflow and wind power processes, we developed univariate GAS models with predictive gamma and beta densities, respectively.
- GAS structures for the univariate cases: SARIMA and unobserved components.
- Univariate models were used to construct a bivariate GAS model, with marginal gamma and beta distributions.
- The Bivariate GAS model implies negative correlated variables which captures the observed seasonal complementarity between the two series.

## Notation

- $y_t$: variable of interest.
- $f_t$: time-varying vector.
- $x_t$: vector of exogenous variables.
- $\theta$: vector of static parameters.
- $Y^t = (y_1, y_2, ..., y_t)$, $F^t = (f_0, f_1, ..., f_t)$ and $X^t = (x_1, x_2, ..., x_t)$.
- The available information at $t$ is given by $\{f_t, \mathcal{F}_t\}$, where

$$\mathcal{F}_t = \{Y^{t-1}, F^{t-1}, X^t\}, \quad t = 1, 2, ..., n.$$

- Assume that $y_t$ is generated by

$$y_t \sim p(y_t / f_t, \mathcal{F}_t, \theta)$$

- $\nabla_t$ and $s_t$ are the score vector and the weighted score, respectively.

# Outline

## Gamma Model - SARIMA structure for parameter evolution

- Consider $y_{1t} = \lambda_t u_t$, $u_t^{'} s \sim Gamma(\alpha, \alpha^{-1})$ iid.
- Multiplicative error model (MEM): $f_t = \lambda_t$.
- An appropriate link function is given by:

$$f_t = -\left[\sum_{k=1}^{r} \phi_k g(x_{t-k+1})\right] + \ln \lambda_t \quad (1)$$

or

$$\lambda_t = \exp\left[f_t + \sum_{k=1}^{r} \phi_k g(x_{t-k+1})\right] \quad (2)$$

## Gamma Model - SARIMA structure for parameter evolution

$$(y_{1t}/f_t, \mathcal{F}_t, \theta) \sim Gamma\left(\alpha, \alpha^{-1}\lambda_t\right), \qquad \begin{aligned} E\left(y_{1t}/f_t, \mathcal{F}_t, \theta\right) &= \lambda_t \\ Var\left(y_{1t}/f_t, \mathcal{F}_t, \theta\right) &= \lambda_t^2/\alpha \end{aligned} \tag{3}$$

$$s_t = \alpha^d \left(\frac{y_t}{\lambda_t} - 1\right), \quad \text{d} = 0,\ 1/2 \ \text{ or } \ 1 \tag{4}$$

$$f_{t+1} = w + \sum_{i=1}^{p} A_i s_{t-i+1} + \sum_{j=1}^{q} B_j f_{t-j+1} \tag{5}$$

>> Seasonality is captured by making $A_i$ and $B_j$ nonzero, for $i, j$ multiples of 12 and in their vicinity.

## Gamma Model - Unobserved component structure for parameter evolution

$$(y_{1t}/f_t, \mathcal{F}_t, \theta) \sim Gamma\left(\alpha, \alpha^{-1}\lambda_t\right), \qquad \begin{aligned} E\left(y_{1t}/f_t, \mathcal{F}_t, \theta\right) &= \lambda_t \\ Var\left(y_{1t}/f_t, \mathcal{F}_t, \theta\right) &= \lambda_t^2/\alpha \end{aligned} \tag{6}$$

$$s_t = \alpha^d \left(\frac{y_{1t}}{\lambda_t} - 1\right), \quad \text{d} = 0,\ 1/2 \text{ or } \ 1 \tag{7}$$

$$f_{t+1} = w + f_{1,t+1} + f_{2,t+1} + f_{3,t+1} + \sum_{k=1}^{r} \phi_k g(x_{t-k+1}) \tag{8}$$

In this model $\lambda_t = \exp(f_t)$.

The scaled score can be seen as an error component.

## Gamma Model - Unobserved component structure for parameter evolution

**random walk** $$f_{1,t+1} = f_{1,t} + a_1 s_t \tag{9}$$

**seasonality** $$f_{2,t+1} = -\sum_{i=1}^{11} f_{2,t+1-i} + a_2 s_t \tag{10}$$

**autoregressive** $$f_{3,t+1} = \phi f_{3,t} + a_3 s_t \tag{11}$$

## Univariate Gamma Models

- **Pearson residuals**:

$$r_{t/t-1} = s_t \quad (d = 1/2) \tag{12}$$

- Also, from MEM specification ($y_{1t} = \lambda_t u_t$), a natural residual is given by:

$$\begin{aligned} u_t^* &= (u_t - 1) \\ &= s_t \quad (d = 0) \end{aligned} \tag{13}$$

## Beta Model - SARIMA structure for parameter evolution

- Wind power capacity is defined in the range $[0, k]$, for $k \leq 100$. In the application we adopted $k = 70$.
- For this series, we considered a beta distribution, with the time-varying parameter given by:

$$\beta_t = \exp\left[f_t + \sum_{k=1}^{r} \phi_k g(x_{t-k+1})\right] \tag{14}$$

## Beta Model - SARIMA structure for parameter evolution

$$
\begin{aligned}
(y_{2t}/f_t, \mathcal{F}_t, \theta) \sim Beta\left(\beta_t, \alpha\right), \quad E\left(y_{2t}/f_t, \mathcal{F}_t, \theta\right) &= k\frac{\beta_t}{\beta_t + \alpha} \qquad (15)\\
Var\left(y_{2t}/f_t, \mathcal{F}_t, \theta\right) &= k^2 \frac{\beta_t \alpha}{(\beta_t + \alpha)^2(\beta_t + \alpha \ + 1)}
\end{aligned}
$$

$$
s_t = \frac{\ln y_{2t} - [\ln k + \psi_1(\beta_t) - \psi_2(\beta_t + \alpha)]}{\beta_t^{1-2d}[\psi_2(\beta_t) - \psi_2(\beta_t + \alpha)]^{1-d}}, \quad \mathrm{d} = 0,\ 1/2 \ \text{ or } \ 1 \qquad (16)
$$

$$
f_{t+1} = w + \sum_{i=1}^{p} A_i s_{t-i+1} + \sum_{j=1}^{q} B_j f_{t-j+1} \qquad (17)
$$

## Beta Model - Unobserved component structure for parameter evolution

$$(y_{2t}/f_t, \mathcal{F}_t, \theta) \sim Beta\left(\beta_t, \alpha\right), \quad E\left(y_{2t}/f_t, \mathcal{F}_t, \theta\right) = k\frac{\beta_t}{\beta_t + \alpha} \tag{18}$$

$$Var\left(y_{2t}/f_t, \mathcal{F}_t, \theta\right) = k^2 \frac{\beta_t \alpha}{(\beta_t + \alpha)^2(\beta_t + \alpha \ + 1)}$$

$$s_t = \frac{\ln y_{2t} - [\ln k + \psi_1(\beta_t) - \psi_2(\beta_t + \alpha)]}{\beta_t^{1-2d}[\psi_2(\beta_t) - \psi_2(\beta_t + \alpha)]^{1-d}}, \quad \mathrm{d} = 0,\ 1/2 \ \text{ or } \ 1 \tag{19}$$

$$f_{t+1} = w + f_{1,t+1} + f_{2,t+1} + f_{3,t+1} + \sum_{k=1}^{r} \phi_k g(x_{t-k+1}) \tag{20}$$

Components $f_{1,t}$, $f_{2,t}$ and $f_{3,t}$ are the same as in the gamma model.

# Outline

## Bivariate Gamma-Beta Distribution

- We considered a modified version of a bivariate distribution proposed by Nadarajah (2009). Let $u$ and $v$ be gamma variables with:

$$u \sim Gamma(\alpha, \lambda/\alpha), \quad v \sim Gamma(\beta, \lambda/\alpha). \tag{21}$$

- Now, define $y_1$ and $y_2$ as follows:

$$y_1 = u, \quad y_2 = k\frac{v}{u+v} \tag{22}$$

- Thus, $y_1$ and $y_2$ have gamma and beta distributions, respectively. The bivariate $Gamma - Beta(\alpha, \beta, \lambda)$ is easily obtained through a change of variables.

$$p(y_1, y_2) = \frac{k y_1^{\alpha+\beta-1} y_2^{\beta-1} (k - y_2)^{-(\beta+1)} \exp\{-k\alpha y_1/[(k - y_1)\lambda]\}}{(\lambda/\alpha)^{\alpha+\beta}\Gamma(\alpha)\Gamma(\beta)} \tag{23}$$

## Bivariate Gamma-Beta Distribution

- Moments:

$$E(y_1) = \lambda, \quad V(y_1) = \lambda^2/\alpha \tag{24}$$

$$E(y_2) = k\frac{\beta}{\alpha+\beta}, \quad V(y_2) = k^2\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{25}$$

$$Corr(y_1, y_2) = -\left(1+\frac{\alpha+1}{\beta}\right)^{-\frac{1}{2}} \tag{26}$$

## Bivariate Gamma-Beta GAS Model

$$((y_{1t}, y_{2t})/f_t, F_t, \theta) \sim Gamma - Beta(\alpha, \beta_t, \lambda_t) \quad (27)$$

$$s_t = I_{t/t-1}^{d-1} \nabla_t, \quad \mathrm{d} = 0,\ 1/2 \ \text{ or } \ 1 \quad (28)$$

$$f_{t+1} = w + \sum_{i=1}^{p} A_i s_{t-i+1} + \sum_{j=1}^{q} B_j f_{t-j+1} \quad (29)$$

## Bivariate Gamma-Beta GAS Model

- The chosen parameterization is given by:

$$(\beta_t, \lambda_t) = \left( \exp\left[ f_{1,t} + \sum_{k=1}^{r_1} \phi_{1k} g_1(x_{t-k+1}) \right], \exp\left[ f_{2,t} + \sum_{k=1}^{r_2} \phi_{2k} g_2(x_{t-k+1}) \right] \right)' \tag{30}$$

- Score vector and the Fisher Information matrix to $f_t = (f_{1t}, f_{2t})'$:

$$\nabla_t = \begin{bmatrix} \beta_t \left\{ \ln\left[ y_{1t} y_{2t} / (k - y_{2t}) \right] - \ln \lambda_t + \ln \alpha - \psi_1(\beta_t) \right\} \\ k \alpha y_{1t} / [(k - y_{2t}) \lambda_t] - (\alpha + \beta_t) \end{bmatrix} \tag{31}$$

$$I_{t/t-1} = \begin{bmatrix} \beta_t^2 \psi_2(\beta_t) & \beta_t \\ \beta_t & (\alpha + \beta_t) \end{bmatrix} \tag{32}$$

## Bivariate Gamma-Beta GAS Model

- Since the conditional moments of each variable are the same of the univariate models, the dependence structure will come from eq.(29).
- To illustrate, a GAS(1,1) model without exogenous variables has updating equation given by:

$$\begin{pmatrix} f_{1,t+1} \\ f_{2,t+1} \end{pmatrix} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} s_{1,t} \\ s_{2,t} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} f_{1,t} \\ f_{2,t} \end{pmatrix} \quad (33)$$

# Outline

## Application

- Motivation: Seasonal stabilization of energy supply in Brazil by exploiting seasonal complementarity between wind and hydro regimes.
- Univariate and Bivariate GAS models.
- Generation of integrated scenarios including consideration of the dispatch from wind farms in energy planning.

## Data

- Monthly averages of streamflow and wind power capacity factor from January 1976 to July 2009 (403 observations).
- Last 24 observations were separated to evaluate out-of-sample prediction accuracy.
- Covariates: Natural Affluent Energy (ENA) series of the four power grid subsystems of Brazil (lag $r = 2$).

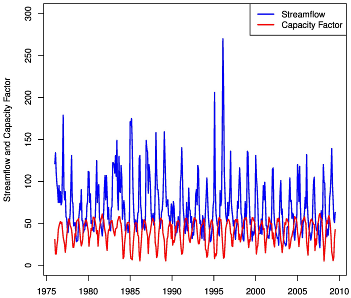



Figure: Streamflow and wind power factor series.

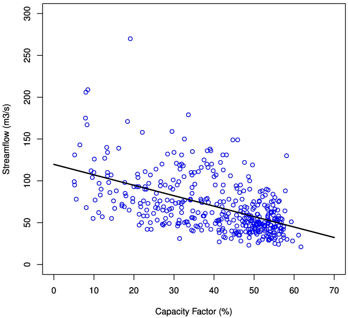



Figure: Scatter plot of the series and regression line.

## Methodology

- Model selection: AIC/BIC.
- Diagnostics: quantile residuals - Ljung-Box (LB) and Jarque-Bera (JB).
- Goodness of fit: RMSE, MAE and

$$MASE = \left(\frac{n-1}{n}\right) \frac{\sum_{t=1}^{n} |\hat{y}_{t/t-1} - y_t|}{\sum_{t=2}^{n} |y_t - y_{t-1}|} \tag{34}$$

$$sMAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|\hat{y}_{t/t-1} - y_t|}{(\hat{y}_{t/t-1} + y_t)} \tag{35}$$

$$pseudo\ R^2 = \left[corr(\hat{y}_{t/t-1}, y_t)\right]^2 \tag{36}$$

## Forecasting

$k$-step ahead forecasts, as given by $E\left(y_{t+k}/f_t, \mathcal{F}_t, \theta\right)$, obtained by Monte Carlo simulation.

**Algorithm:**

- For $k \geq 1$, independently generate $m$ paths, each one with extension of $K$, according to the steps below:
  1. Update $f_{t+k}$ from $\{\mathcal{F}_{t+k-1}\}$ and $f_{t+k-1}$;
  2. Simulate $y_{t+k}$ from the predictive distribution, using $\hat{\theta}$ and the information sets $\{\mathcal{F}_{t+k}\}$ and $f_{t+k}$;
  3. Return to step (a), until $k = K$;
- For each horizon $k = 1, 2, ..., K$ mean, confidence interval and quantiles can be obtained.

## Diagnostics using Quantile Residuals

- Diagnostics using quantile residuals (Kalliovirta, 2009), defined by:

$$R_{t,\theta} = \Phi^{-1}(F(y_t/f_t, \mathcal{F}_t, \theta)) \tag{37}$$

  where $\Phi^{-1}(.)$ is the inverse normal CDF and $F(y_t/f_t, \mathcal{F}_t, \theta)$ is the conditional CDF.

- Observed quantile residuals $r_{t,\hat{\theta}}$ obtained by substituting $\theta$ for $\hat{\theta}$.
- In the multivariate case, the theoretical quantile residuals are defined from the marginal distributions, ie:

$$R_{t,\theta} = \left[R_{t1,\theta}, R_{t2,\theta}, ..., R_{tk,\theta}\right]' \tag{38}$$

# Gamma GAS Models

## Univariate Gamma GAS Models

Table: Model selection for univariate Gamma GAS models.

| Statistic | SARIMA | UCM |
|---|---|---|
| Log-Lik. | -1439,0 | -1442,0 |
| AIC | 2928,1 | 2922,0 |
| BIC | 3025,7 | 2996,2 |

Table: Goodness of fit for univariate Gamma GAS models.

| Period | Measures | SARIMA | UCM |
|---|---|---|---|
| In-sample | RMSE | 16,7 | 16,5 |
| | MAE | 11,6 | 11,5 |
| | MASE | 0,60 | 0,60 |
| | sMAPE | 7,8 | 7,7 |
| | pseudo $R^2$ | 0,76 | 0,77 |
| Out-of-sample$^{(1)}$ | RMSE | 12,9 | 16,3 |
| | MAE | 10,2 | 12,8 |
| | MASE | 0,53 | 0,66 |
| | sMAPE | 7,9 | 9,9 |
| | pseudo $R^2$ | 0,89 | 0,75 |

Note: (1) Predictions $k$-step ahead obtained by simulation.

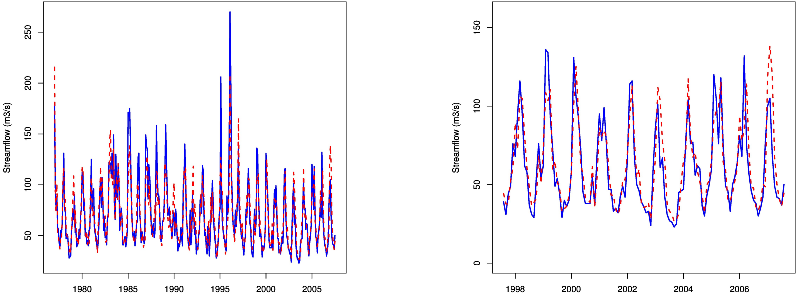



Figure: 1-step ahead for streamflow series (SARIMA evolution).

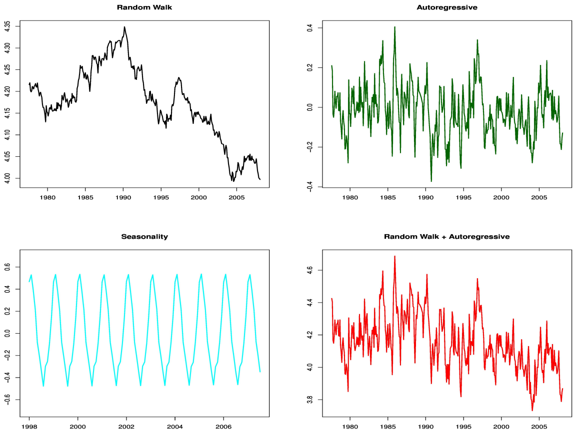



Figure: Random walk, autoregressive and seasonal components for the Gamma GAS model.

Table: Diagnostics (p-value) for the Gamma GAS models.

| Test | SARIMA | UCM |
|---|---|---|
| Autocorrelation (LB) | 0,70 | 0,71 |
| Heteroscedasticity (LB) | 0,01 | 0,04 |
| Normality (JB) | 0,12 | 0,47 |

# Beta GAS Models

## Univariate Beta GAS Models

Table: Model selection for univariate Beta GAS models.

| Statistic | SARIMA | UCM |
|---|---|---|
| Log-lik. | -1074,6 | -1050,8 |
| AIC | 2191,1 | 2129,6 |
| BIC | 2273,1 | 2184,3 |

Table: Goodness of fit for univariate Beta GAS models.

| Period | Measures | SARIMA | UCM |
|---|---|---|---|
| In-sample | RMSE | 5,2 | 4,9 |
| | MAE | 4,0 | 3,6 |
| | MASE | 0,58 | 0,53 |
| | sMAPE | 6,8 | 6,3 |
| | pseudo $R^2$ | 0,86 | 0,87 |
| Out-of-sample$^{(1)}$ | RMSE | 7,7 | 7,9 |
| | MAE | 5,8 | 6,2 |
| | MASE | 0,85 | 0,90 |
| | sMAPE | 12,7 | 13,2 |
| | pseudo $R^2$ | 0,90 | 0,90 |

Note: (1) Predictions $k$-step ahead obtained by simulation.

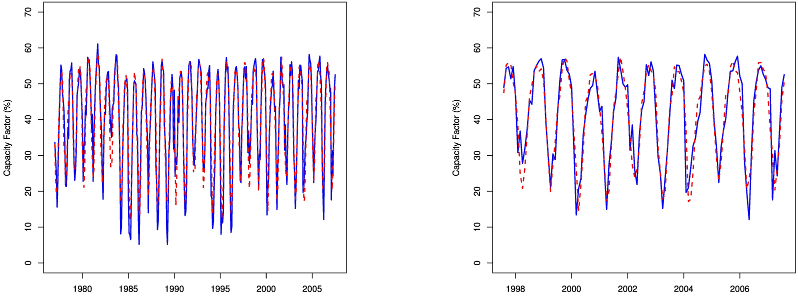



Figure: 1-step ahead for wind power factor series (UC evolution).

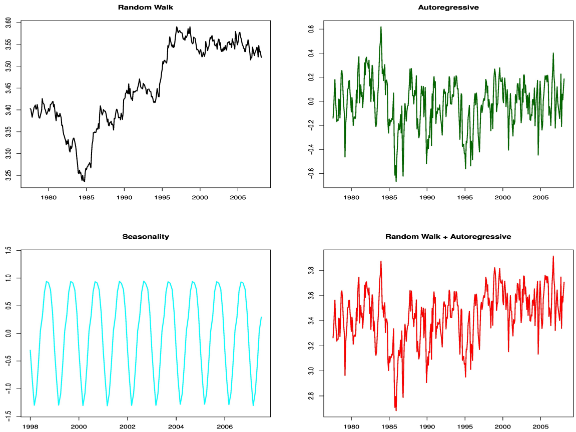



Figure: Random walk, autoregressive and seasonal components for the Beta GAS model.

Table: Diagnostics (p-value) for the Beta GAS models.

| Test | SARIMA | UCM |
|---|---|---|
| Autocorrelation (LB) | 0,54 | 0,52 |
| Heteroscedasticity (LB) | 0,00 | 0,00 |
| Normality (JB) | 0,17 | 0,01 |

# Gamma-Beta GAS Model

## Bivariate Gamma-Beta GAS Model

- Consider the evolution equation of $f_t$:

$$f_{t+1} = w + \sum_{i=1}^{p} A_i s_{t-i+1} + \sum_{j=1}^{q} B_j f_{t-j+1}$$

- We tested some structures (AR; diagonal matrices $A_i$ and $B_j$ - SUTSE; diagonal matrices $A_i$ and $B_j$ with only $B_1$ and $B_{12}$ full);

- Best structure (forecasting/ residual autocorrelation): diagonal matrices and $B_1$ and $B_{12}$ full.

- The chosen structure allows direct communication between the conditional means, capturing short-term and seasonal dependence.

Table: Model selection by type of model.

| **Model** | **Measures** | | |
|---|---|---|---|
| | **Log-Ver.** | **AIC** | **BIC** |
| Bivariate | -2616,1 | 5326,2 | 5509,8 |
| Univariate | -2513,6 | 5119,2 | 5298,8 |

- By univariate model, in this context, we mean the bivariate model assuming independence between streamflow and wind processes.
- In this case, the joint likelihood is obtained multiplying the likelihoods of the univariate models, ie:

$$L(y_1, y_2; \theta, f_0) = L(y_1; \theta_1, f_{10})L(y_1, y_2; \theta_2, f_{20})$$

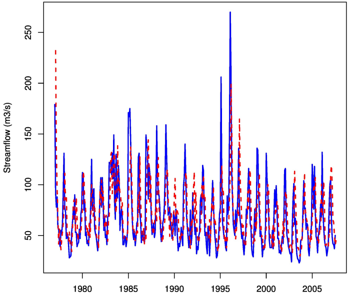
Table: Goodness of fit for bivariate Gamma-Beta GAS models.

| **Period** | **Measures** | **Streamflow** | **Wind** |
|---|---|---|---|
| In-sample | RMSE | 18,2 | 5,1 |
| | MAE | 12,8 | 3,9 |
| | MASE | 0,67 | 0,57 |
| | sMAPE | 8,9 | 6,6 |
| | pseudo $R^2$ | 0,72 | 0,86 |
| Out-of-sample[1] | RMSE | 12,9 | 9,6 |
| | MAE | 10,9 | 7,4 |
| | MASE | 0,57 | 1,08 |
| | sMAPE | 8,9 | 14,9 |
| | pseudo $R^2$ | 0,85 | 0,89 |

Note: (1) Predictions $k$-step ahead obtained by simulation.

Table: Diagnostics (p-value) for the bivariate Gamma-Beta GAS model.

| Test | Streamflow | UCM |
|---|---|---|
| Autocorrelation (LB) | 0,11 | 0,19 |
| Heteroscedasticity (LB) | 0,24 | 0,00 |
| Normality (JB) | 0,10 | 0,66 |



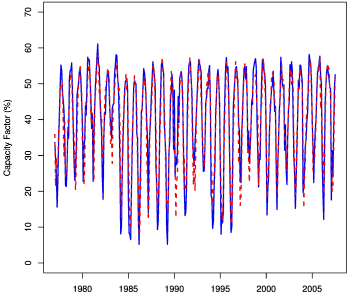




Figure: 1-step ahead for streamflow and wind power series.

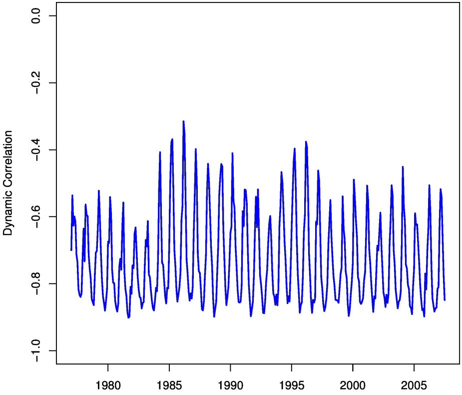



Figure: Conditional correlation.

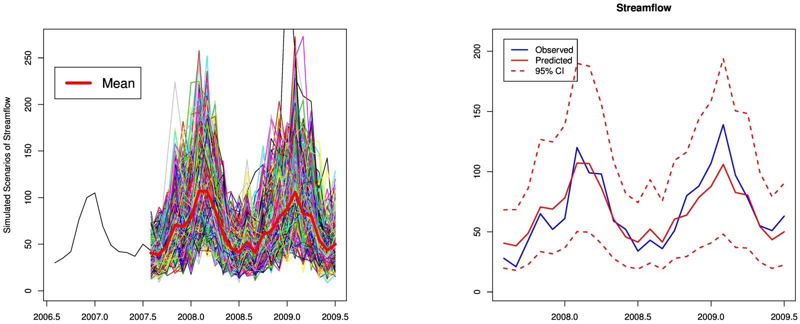



Figure: Streamflow - simulated paths and $k$-step ahead predictions.

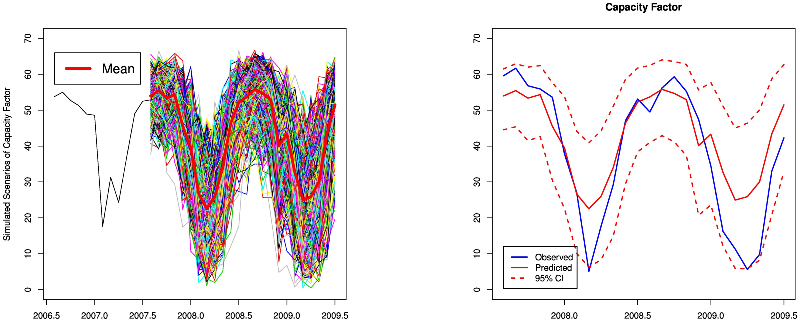



Figure: Power wind factor series - simulated paths and $k$-step ahead predictions.

# Outline

## Conclusion

- We specify and estimate a bivariate Gamma-Beta GAS model based on Nadjahara's distribution.
- The bivariate Gamma-Beta GAS model integrated the univariate models (parameterizations and marginal distributions), besides permitting the joint generation of streamflow and wind power factor scenarios.
- Our model compares favorably to a VARX model fitted to the same bivariate series (Amaral, 2011), with the additional feature that no data transformation is necessary.
- Forecasting obtained by simulation and diagnostics based on quantile residuals.
- Recursion initialization ($f_t$) and maximum likelihood initialization need further thinking.

# Bibliography

AMARAL, B. M. **Modelos VARX para geração de cenários de vento e vazão aplicados à comercialização de energia**. 2011. 155f. Dissertação (Mestrado em Engenharia Elétrica) - PUC-Rio, Rio de Janeiro. 2011.

CREAL, D., KOOPMAN, S. J. and LUCAS, A. Generalized Autoregressive Score Models with Applications. **Journal of Applied Econometrics**, v. 28, n. 5, p. 777 - 795, 2013.

KALLIOVIRTA, L. **Diagnostic Tests Based on Quantile Residuals for Nonlinear Time Series Models**. 2009. 165 f. Thesis (Doctorate in Economics) - University of Helsinki, Helsinki. 2009.

ORD, J. K., KOEHLER, A. B. and SNYDER, R. D. Estimation and prediction for a class of dynamic nonlinear statistical models. **Journal of the American Statistical Association**, v. 92, n. 440, p. 1621-1629, 1997.

NADARAJAH, S. A bivariate distribution with gamma and beta marginals with application to drought data. **Journal of Applied Statistics**, v. 36, n. 3, p. 277-301, 2009.