Projeto de Graduação

DEPARTAMENTO DE ENGENHARIA ELÉTRICA

June 26, 2017

A STUDY ON THE IMPACT OF EL NIÑO SOUTHERN OSCILLATION ON HYDRO POWER GENERATION IN BRAZIL

Raphael Augusto Proença Rosa Saavedra

PUC RIO

Projeto de Graduação

DEE DEPARTAMENTO DE ENGENHARIA ELÉTRICA

A STUDY ON THE IMPACT OF EL NIÑO SOUTHERN OSCILLATION ON HYDRO POWER GENERATION IN BRAZIL

Student: **Raphael Augusto Proença Rosa Saavedra**

Advisor: **Alexandre Street de Aguiar**

Co-advisor: **Cristiano Augusto C. Fernandes**

Work presented as partial requirement to the conclusion of the Electrical Engineering major in the Pontifical University of Rio de Janeiro, Rio de Janeiro, Brazil.

Acknowledgements

First and foremost, I'd like to thank my parents for always doing their best to raise me as well as they could. This work wouldn't be possible without the excellent upbringing I had, thanks to their efforts.

I'd like to thank my professor and academic advisor, Alexandre Street, for his always uplifting talks and constant orientations, and for giving me the opportunity to work with him and his other students in LAMPS, where I discovered the true nature of research, and where I made many friends and colleagues.

I'd also like to thank my co-advisor, Cristiano Fernandes, for his invaluable academic insights, for always being present and available to listen to my questions, and for doing his best to answer them.

Thank you to every LAMPS member for making the workplace such a pleasant environment, and for the daily conversations and discussions about both academic and non-academic affairs. Thank you to André Lawson, Alexandre Velloso, Henrique Helfer, Marcelo Ruas, Mario Souto, Thuener Silva, and especially to Bruno Fanzeres for all the talks we had throughout the years and all the guidance you gave me.

Thank you to my friends from Colégio Santo Agostinho for all the daily talks and laughs. Thank you to João Vitor Penna Reis, Pedro Henrique Soares, Pedro Roque, Tomás Rolim, Victor Macêdo, and especially to Thiago Milagres, with whom I studied side by side throughout all of the Electrical Engineering course.

Thank you to my friends Adriano Fonsêca, Arthur Pacheco, Giovanni Vieira, Guilherme Grittz, Igor Strozzi, Lucas Guiselini, and Marcus Cintra for the daily laughs and entertaining philosophical discussions.

I'd also like to thank Maria Carolina Marinho for the incredible company throughout this semester. Thank you for all the love and support.

Finally, I'd like to thank PUC-Rio for providing an excellent environment. The university has practically been my second home during the past years and it has been very pleasant and rewarding.

Abstract

Power plants that utilize renewable resources to produce electrical energy, such as hydroelectric power plants, comprehend the majority of the Brazilian energy matrix. However, their generations have underlying uncertainties associated with hydrological factors, such as precipitation. An adequate description of these future uncertainties is necessary for a thorough analysis of the risks that are inherent to the production of hydro power plants. Among these uncertainties, we highlight the possibility that large scale climatic phenomena, specifically the El Niño Southern Oscillation (ENSO), might considerably affect hydro power generation. Motivated by this hypothesis, we utilize Generalized Autoregressive Score (GAS) models in order to estimate the impact of ENSO on the Affluent Natural Energy (ANE), which serves as an indicator of monthly hydro power generation in each one of Brazil's main regions. Based on simulations, our model indicates an up to 11% ENSO effect on the ANE, depending on the analyzed Brazilian region.

Keywords: El Niño Southern Oscillation, Dynamic Score Models, Hydro Power Generation, Affluent Natural Energy

ESTUDO DO IMPACTO DO EL NIÑO OSCILAÇÃO SUL NA GERAÇÃO HÍDRICA NO BRASIL

Resumo

Usinas que utilizam recursos renováveis para produzir energia elétrica, como as grandes centrais hídricas e pequenas centrais hidrelétricas (PCHs), compreendem a maior parte da matriz elétrica brasileira, mas suas gerações possuem incertezas associadas a fatores hidrológicos. A descrição adequada dessas incertezas futuras é de fundamental importância para uma análise criteriosa dos riscos inerentes à produção de fontes renováveis. Dentre essas incertezas, destacamos a possibilidade de que fenômenos climáticos de larga escala, em especial o El Niño Oscilação Sul (ENOS), afetem consideravelmente a geração hídrica. Motivados por essa hipótese, utilizamos modelos GAS (Generalized Autoregressive Score) com o objetivo de estimar o impacto do ENOS na Energia Natural Afluente, que serve como indicador da capacidade de geração hídrica mensal em cada uma das regiões do Brasil. Baseado em simulações, nosso modelo indica um efeito de até 11% do ENSO nas ENAs, dependendo da região brasileira analisada.

Palavras-chave: El Niño Oscilação Sul, Modelos de Score Dinâmico, Geração Hídrica, Energia Natural Afluente

Summary

List of Figures

List of Tables

1 Introduction

Power plants that utilize renewable resources to produce electrical energy, such as hydroelectric power plants, comprehend the majority of the Brazilian energy matrix. However, their generations have underlying uncertainties associated with hydrological factors, such as precipitation. The benefits related to these resources are numerous and include the diversification of the energy matrix and the reduction of the emission of greenhouse gases emitted by thermal power plants. Nonetheless, the insertion of this kind of energy in the electrical system must be done with caution due to its intermittent nature.

From a commercial standpoint, the generation uncertainty creates a financial risk for generation companies, especially if the contract is set on produced energy, which is a common practice in the Free Trading Environment (FTE) [2, 3]. From a strategic planning standpoint, this uncertainty also creates challenges, ranging from its incorporation in the operation planning to the daily prediction which is essential for real time operation. Even though we possess a long term history for conventional hydro power generation, we are not immune to upsets, as seen in 2014 and 2015, when water inflow models systematically failed [4, 5].

All these factors make it clear that an adequate description of the uncertainties that are inherent to the production of hydro power plants is necessary for a thorough analysis of the risks and benefits this kind of energy offers. Among these uncertainties, we highlight the possibility that large scale climatic phenomena, specifically the El Niño Southern Oscillation (ENSO), might considerably affect hydro power generation. The impact of these phenomena has not yet been thoroughly studied in Brazil. The pioneer work of Nascimento et al. [6] investigated the effect of ENSO on wind speed in Northeastern Brazil, using non-Gaussian time series models. Among the results, the Gamma GAS model pointed a change of 3.98% in wind speed for a 1°C change in sea surface temperature.

Motivated by these preliminary results, in this work we investigate with further details the impact of climatic phenomena, specifically ENSO, on hydro power generation in Brazil. We use Generalized Autoregressive Score (GAS) models, proposed by Creal et al. [7, 8], adjusted to historical time series of Affluent Natural Energy (ANE), which serve as an indicator of monthly hydro power generation in each one of Brazil's main regions. We include in the model exogenous variables associated with ENSO historical data in order to estimate its effects and relevance.

2 El Niño Southern Oscillation

a The ENSO phenomenon

The term "teleconnection" is used in meteorology to designate a causal connection or correlation between meteorological or other environmental phenomena that occur a long distance apart. The Southern Oscillation was one of the first teleconnections to be documented, having been discovered in 1932 by Bliss & Walker [9]. It was observed that an increase of atmospheric pressure in the Pacific Ocean would occur paired with a decrease of atmospheric pressure in the Indian Ocean.

The El Niño phenomenon is one of the most widely known teleconnections in the world and impacts the pressure, the wind and the sea surface temperature of different places around the globe. One of El Niño's main symptoms is the unusual warming of the sea surface in a known region in the middle of the Pacific [1]. The La Niña phenomenon can be defined as the cold phase of El Niño, with the sea surface temperature of the same region becoming cooler than its historic average. Later on, the El Niño started to be analyzed paired with the Southern Oscillation, originating the name ENSO.

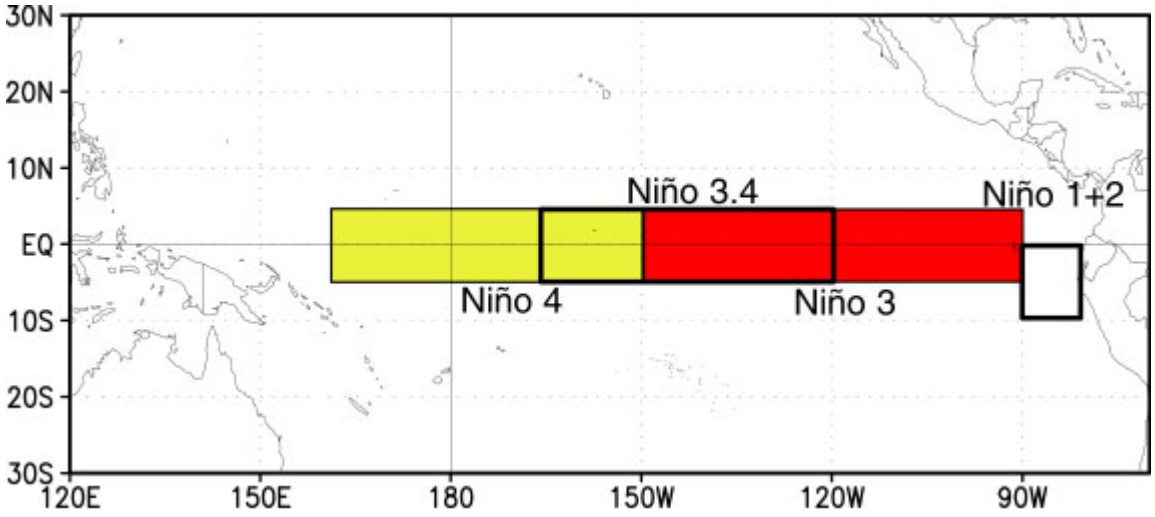

Figure 1: The Niño region in the Pacific Ocean.

The Niño region is divided in subregions such as Niño 3 and Niño 4. Each one has its particular properties, which go beyond the scope of this work. Meteorology institutes such as the National Oceanic and Atmospheric Administration (NOAA) [1] have developed methods for identifying and indicating the nature and intensity of ENSO episodes. These methods result in several indicators, such as the Oceanic Niño Index (ONI), Southern Oscillation Index (SOI) and Multivariate ENSO Index (MEI) [10].

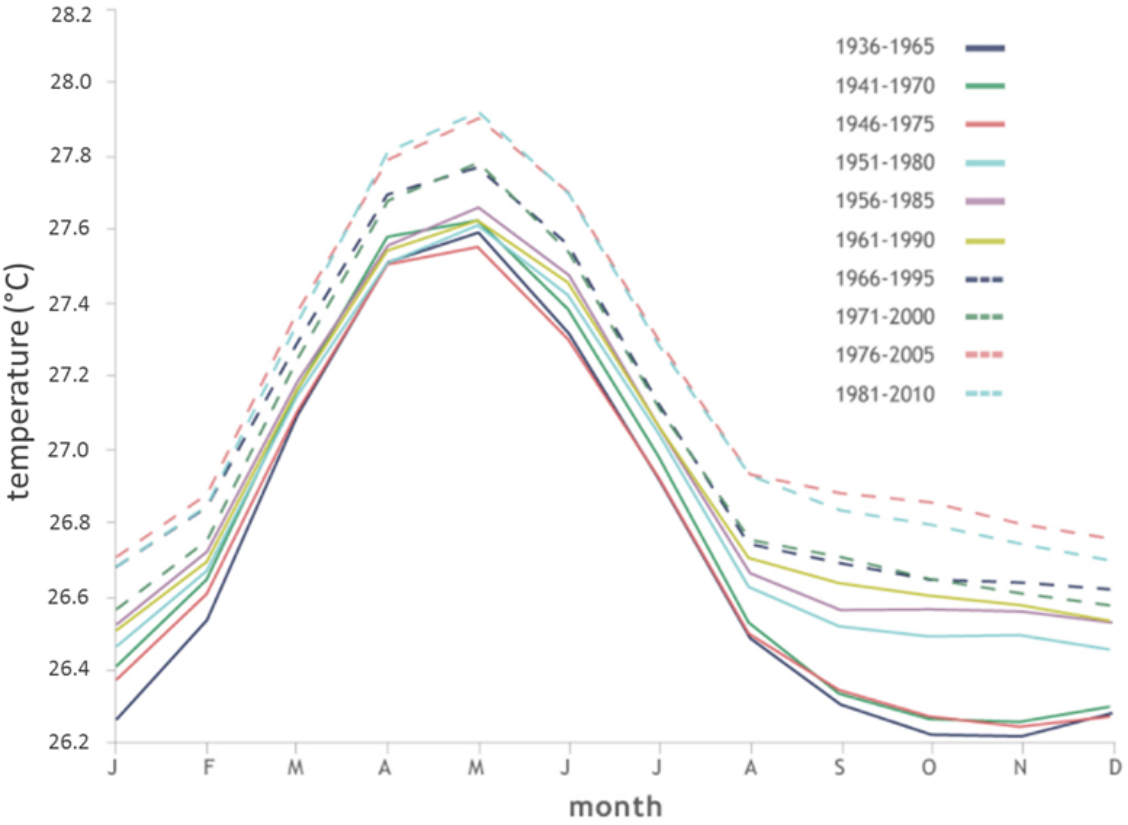

Figure 2: Average surface temperature in the central tropical Pacific. Source: NOAA website [1] and the author.

b Utilized data

In order to consistently estimate the effect of ENSO in hydro power generation, we need long term historical data related to both.

To represent ENSO, there are several different indicators, most of them publicly available at the NOAA website [1]. The main ENSO indicators are: Oceanic Niño Index (ONI), Southern Oscillation Index (SOI), Multivariate ENSO Index (MEI) and Sea Surface Temperature anomalies (SST), with ONI being the index we utilized in this study. ONI is a monthly index related to the Niño 3.4 region (see Figure 1). It takes into account a 30-year sea surface temperature average and verifies deviations from that average, in such a way that long term temperature changes are computed (see Figure 2). ONI values are positive for El Niño (warmer water) and negative for La Niña (cooler water), and can be classified in intensity as weak (absolute values between 0.5 and 0.9), moderate (absolute values between 1.0 and 1.4) and strong (absolute values above 1.5). Values between -0.5 and 0.5 are considered neutral. At the time of this study, ONI was available from 1950 to 2016.

To represent hydro power generation, we utilize Affluent Natural Energy (ANE) data, publicly available at the National Electrical System Operator (ONS) website [11]. ANE is the energy that can be produced from natural affluent water inflows in reservoirs, in average Mega Watt (aMW). It is also split into four different time series, one for each region in Brazil's electrical system: Southeast, South, Northeast and North. At the time of this study, the ANE series were available from 1931 to 2015. It is important to highlight that ANE has a different behavior in every region and therefore our case studies are done individually for each region. The ANE historical behavior for each region can be seen in Figure 3.

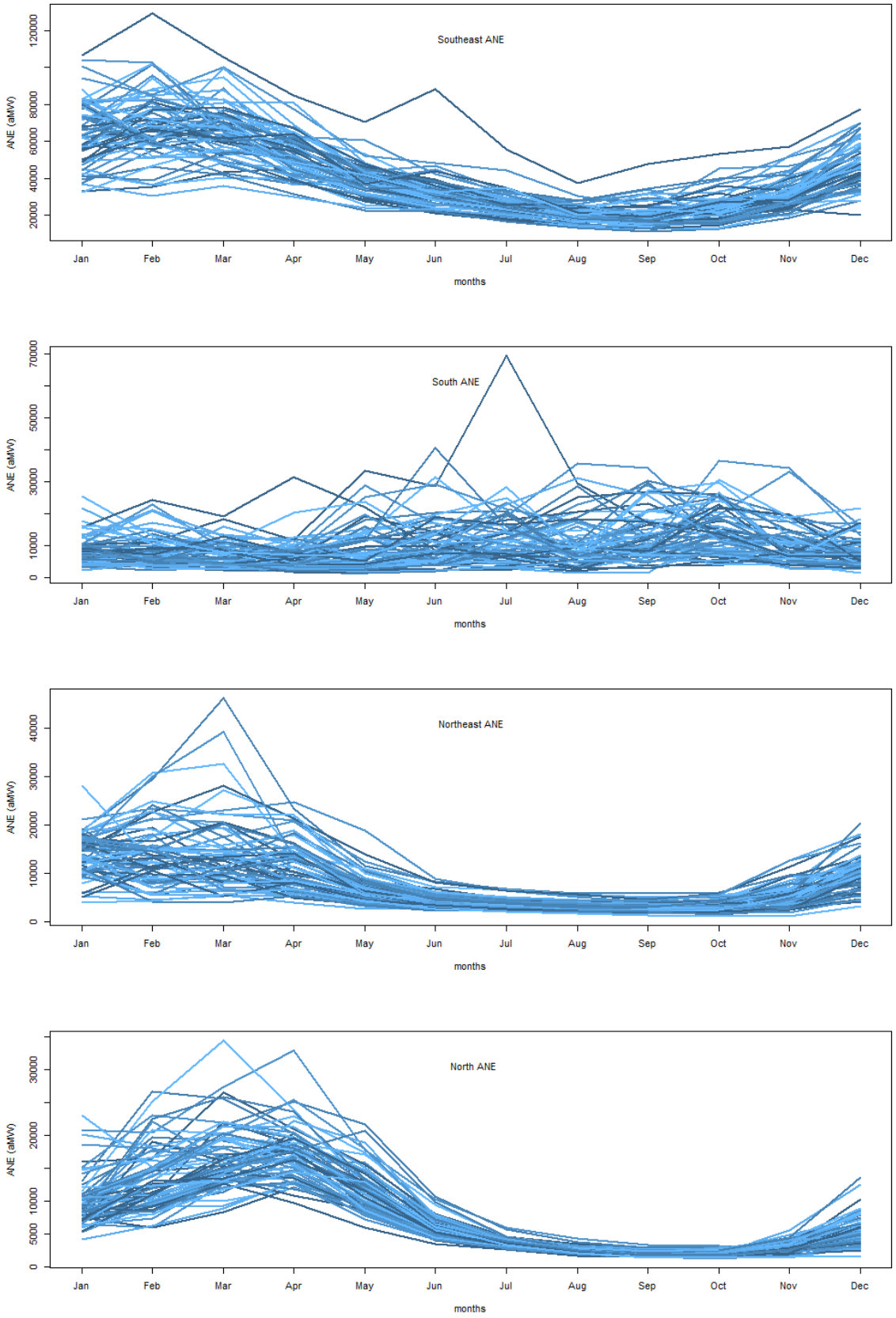

Figure 3: Historical ANE data for each one of Brazil's four regions, for years 1951-2015.

3 GAS models

We now provide a detailed explanation of GAS models, the class of time series models that will be used in this study.

Generalized Autoregressive Score models, or simply GAS models, represent a class of observation driven time series models with time-varying parameters proposed in [7] and further detailed in [8]. GAS models have the scaled score of the likelihood function as their driving mechanism, which allows for a consistent framework for introducing time-varying parameters in a wide class of non-linear models. The GAS model has some well-known models as particular cases, such as the GARCH model.

The GAS formulation introduces a function that relates the time-varying parameters of the series' conditional distribution to lagged values of themselves and of the conditional score, as well as exogenous variables. The GAS model framework is presented in the sequel.

Let $n \times 1$ vector y_t denote the dependent variable of interest, f_t the time-varying parameter vector, x_t a vector of exogenous variables and θ a vector of static parameters. Define $Y^t = \{y_1, \ldots, y_t\}$, $F^t = \{f_0, f_1, \ldots, f_t\}$ and $X^t = \{x_1, \ldots, x_t\}$. The available information set at time t consists of $\{f_t, \mathcal{F}_t\}$, where

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

We assume that y_t is generated by the observation density

$$y_t \sim p(y_t | f_t, \mathcal{F}_t; \theta) \tag{2}$$

Furthermore, we assume that the mechanism for updating the time-varying parameter f_t is given by the following autoregressive updating equation, known as a GAS(p,q) mechanism:

$$f_{t+1} = \omega + \sum_{i=1}^{p} A_i s_{t-i+1} + \sum_{j=1}^{q} B_i f_{t-j+1} \tag{3}$$

where ω is a vector of constants, coefficient matrices A_i and B_j have appropriate dimensions for $i = 1, ..., p$ and $j = 1, ..., q$. The unknown coefficients in Eq. (3) are functions of θ: $\omega = \omega(\theta)$, $A_i = A_i(\theta)$ and $B_j = B_j(\theta)$ for $i = 1, ..., p$ and $j = 1, ..., q$. Finally, s_t is the scaled score function and is defined as:

$$s_t = S_t . \nabla_t, \qquad \nabla_t = \frac{\partial \ln p(y_t | f_t, \mathcal{F}_t; \theta)}{\partial f_t}, \qquad S_t = S(t, f_t, \mathcal{F}_t; \theta) \tag{4}$$

where $S(\cdot)$ is an arbitrary scaling function, generally defined as a function of Fisher information:

$$S_t = \mathcal{I}_{t|t-1}^{-d}, \qquad \mathcal{I}_{t|t-1} = E_{t-1}[\nabla_t \nabla_t'] \tag{5}$$

where $d = 0$, $\frac{1}{2}$ or 1 and E_{t-1} denotes an expectation with respect to $p(y_t | f_t, \mathcal{F}_t; \theta)$.

The advantages of GAS models include: the ability to model non-linearity, non-stationarity and non-Gaussianity; the fact that the evaluation of the likelihood function has a closed form; and the fact that it outputs a complete conditional probability density, which facilitates probabilistic interpretations and simulations.

Nevertheless, the likelihood function is generally non-linear and in many cases non-convex, which means optimality is not always guaranteed. In the study presented here, the likelihood function is not only non-convex, but also extremely ill-behaved, which forced us to develop a methodology involving multiple iterations and numerical methods such as BFGS and Nelder-Mead in order to estimate the maxima as consistently as possible.

4 Lognormal GAS model with time-varying location parameter

We utilize the Lognormal distribution because it is an adequate way of modeling water inflow time series, as done in the NEWAVE model [12] and seen in [13, 14]. The Lognormal probability density function for different values of the scale parameter σ can be seen in Figure 4.

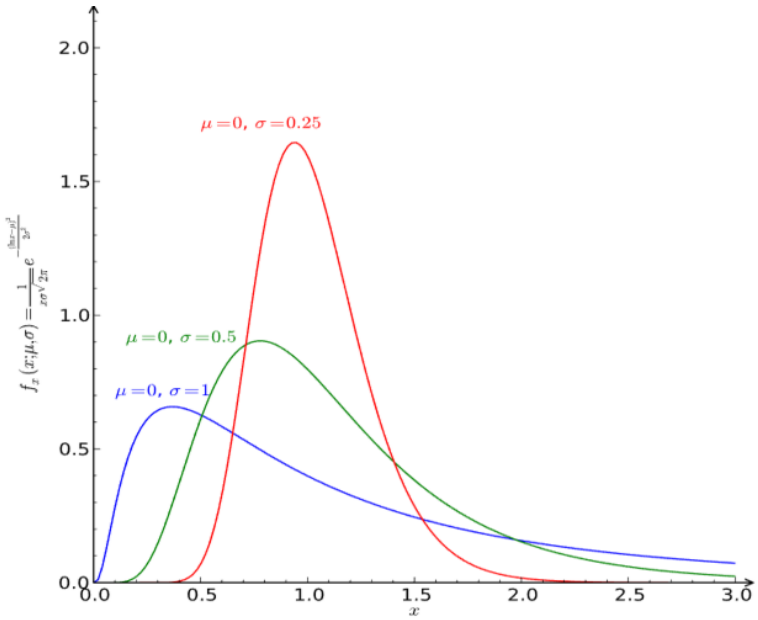

Figure 4: Lognormal density function for different values of the scale parameter σ.

The Lognormal conditional probability density, expectation and variance are as follows:

$$p(y_t|f_t, \mathcal{F}_t; \theta) = \frac{1}{y_t\sigma\sqrt{2\pi}} \exp\left(\frac{-(\ln y_t - \mu_t)^2}{2\sigma^2}\right) \tag{6}$$

$$E[y_t|f_t, \mathcal{F}_t; \theta] = \exp\left(\mu_t + \frac{\sigma^2}{2}\right) \tag{7}$$

$$V[y_t|f_t, \mathcal{F}_t; \theta] = \exp(\sigma^2 - 1) \cdot \exp(2\mu_t + \sigma^2) \tag{8}$$

We consider the location parameter (μ) as our time-varying parameter because of the seasonal and uncertain nature of the series. The parametrization of our choice and the derivation of the score are as follows:

$$f_t^{(1)} = \mu_t \tag{9}$$

$$\begin{aligned}
\nabla_t &= \frac{\partial}{\partial \mu_t}\Big(-\ln(y_t\sigma\sqrt{2\pi}) - \frac{(\ln y_t - \mu_t)^2}{2\sigma^2}\Big) \\
&= \frac{\partial}{\partial \mu_t}\Big(-\frac{(\ln y_t)^2 - 2\mu_t \ln y_t + \mu_t^2}{2\sigma^2}\Big) \\
&= -\frac{1}{2\sigma^2}\frac{\partial}{\partial \mu_t}(-2\mu_t \ln y_t + \mu_t^2) \\
&= -\frac{1}{2\sigma^2}(-2\ln y_t + 2\mu_t) \\
&= \frac{\ln y_t - \mu_t}{\sigma^2}
\end{aligned} \tag{10}$$

We can thus derive the Fisher information from the score:

$$
\begin{aligned}
\mathcal{I}_{t|t-1} &= -E_{t-1}[\nabla_t \nabla_t'] \\
&= -E_{t-1}\Big[\frac{\partial^2 \ln p(y_t|f_t, \mathcal{F}_t; \theta)}{\partial f_{t_1}^2}\Big] \\
&= -E_{t-1}\Big[\frac{\partial}{\partial \mu_t}\Big(-\frac{\ln y_t - \mu_t}{\sigma^2}\Big)\Big] \qquad (11) \\
&= E_{t-1}\Big[\frac{1}{\sigma^2}\Big] \qquad (12) \\
&= \frac{1}{\sigma^2}
\end{aligned}
$$

In this study, we utilize a GAS model with lags 1, 2, 3, 11 and 12, as well as an exogenous variable associated with ENSO. Lags 1, 2 and 3 intend to capture the short-term autoregressive nature of the series, while lags 11 and 12 intend to capture its seasonal behavior.

The underlying scale parameter σ is one of the constant parameters that are estimated through the method presented in Section 5. Using the derivations presented above and $d = 1$, we have the following score and updating mechanism for the time-varying parameter:

$$
\begin{aligned}
s_t &= \mathcal{I}_{t|t-1}^{-1} \nabla_t \\
&= \Big(\frac{1}{\sigma^2}\Big)^{-1} \cdot \frac{\ln y_t - \mu_t}{\sigma^2} \qquad (13) \\
&= \ln y_t - \mu_t
\end{aligned}
$$

$$
f_{t+1} = \omega + A_1 s_t + A_2 s_{t-1} + A_3 s_{t-2} + A_4 s_{t-10} + A_5 s_{t-11} + B_1 f_t + B_2 f_{t-1} + B_3 f_{t-2} + B_4 f_{t-10} + B_5 f_{t-11} + \phi x_{t-\ell} \qquad (14)
$$

where x_t is the ENSO indicator at time t (in our case study, we use ONI) and l is the lag applied to the exogenous variable (in our case study, we use $\ell = 0$).

5 Methodology

a Parameter estimation

The GAS constant parameters – in our case, A_i's, B_i's, ω and ϕ – are estimated via maximum likelihood estimation (MLE). The likelihood function is given by

$$\mathcal{L}(\theta; y_1, ..., y_n) = p(y_1, ..., y_n; \theta) = \prod_{i=t}^{n} p(y_t | f_t, \mathcal{F}_t; \theta) \tag{15}$$

where p denotes the probability density function. Given that dealing with a sum is more convenient than dealing with a product, we utilize the log-likelihood:

$$\ell(\theta; y_1, ..., y_n) = \ln \mathcal{L}(\theta; y_1, ..., y_n) = \sum_{t=1}^{n} \ln p(y_t | f_t, \mathcal{F}_t; \theta) \tag{16}$$

Therefore, our MLE problem is

$$\theta^* = \arg\max \ell(\theta; y_1, ..., y_n) \tag{17}$$

Non-linear optimization methods such as BFGS and Nelder-Mead can be employed to solve this problem. However, as mentioned in Section 3, in our particular case this function is not only non-convex but also extremely ill-behaved, which means optimality cannot be guaranteed. Even though this is certainly an imperfection of the model and should be addressed, our main goal is to accurately estimate a model for our time series and use it to simulate future scenarios. Global optimality, though desired, is not necessary as long as we verify adequate in-sample and out-of-sample results.

In order to estimate the maxima as consistently as possible, we utilize a simple algorithm that runs an arbitrary number of rounds with random initial conditions in each round and returns the parameters with highest likelihood. This allows us to inspect a wider range of results and discard solutions that are local maxima with lower likelihood values. The second part of the algorithm consists of checking if the covariance matrix of the parameters is positive semi-definite. If not, that solution is discarded and the round with the next highest likelihood is checked. The algorithm is as follows:

Algorithm 1 Maximum likelihood parameters estimation

Let r be the arbitrary number of rounds for the algorithm and p the number of parameters;
Let $\text{BFGS}(\ell, \theta, y)$ be the BFGS function call that maximizes log-likelihood function ℓ with initial parameters θ for series y and returns the maximum log-likelihood and its associated parameters;
Let H_i be the numerically differentiated Hessian matrix returned by the BFGS function and W_i be the covariance matrix of the parameters for round i;
for $i = 1$ to r **do**
  $\theta^{(i)} \leftarrow p \times 1$ random vector with elements $x_j \sim N(0, 0.5)$, $j = 1, ..., p$
  $(\ell_{(i)}, \hat{\theta}^{(i)}, H_{(i)}) \leftarrow \text{BFGS}(\ell, \theta^{(i)}, y)$
end for
$i^* \leftarrow \arg\max\{\ell_{(i)}\}$
$W_{(i^*)} \leftarrow H^{-1}_{(i^*)}$
while $W_{(i^*)}$ is not positive semi-definite **do**
  $i^* \leftarrow$ The next i with the highest likelihood;
  $W_{(i^*)} \leftarrow H^{-1}_{(i^*)}$
end while
$\hat{\theta} \leftarrow \theta^{(i^*)}$
return $\hat{\theta}$

b Goodness-of-fit statistics and model diagnostics

After estimating the parameters through Algorithm 1, we generate a time series using our GAS model defined in Section 4. This is done by running the GAS model and obtaining a series of f_t and s_t, for $t = 1, ..., n$, and then using these values to construct the conditional expectation series by using Eq. (7). Finally, we utilize the original series and the generated series to run diagnostics and calculate goodness-of-fit statistics.

The following goodness-of-fit statistics were employed: coefficient of determination (Pseudo-R^2), mean average error (MAE) and symmetric mean absolute percentage error (sMAPE) [15, 16].

$$\text{Pseudo-}R^2 = \big[\text{corr}(\hat{y}_t, y_t)\big]^2 = \big[\text{corr}\big(E[y_t|f_t, \mathcal{F}_t; \theta], y_t\big)\big]^2 \tag{18}$$

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|\hat{y}_t - y_t| = \frac{1}{n}\sum_{t=1}^{n}|E[y_t|f_t, \mathcal{F}_t; \theta] - y_t| \tag{19}$$

$$\text{sMAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\frac{|\hat{y}_t - y_t|}{\hat{y}_t + y_t} = \frac{100\%}{n}\sum_{t=1}^{n}\frac{|E[y_t|f_t, \mathcal{F}_t; \theta] - y_t|}{E[y_t|f_t, \mathcal{F}_t; \theta] + y_t} \tag{20}$$

The autocorrelation function (ACF), partial autocorrelation function (PACF) and Jarque-Bera test [17, 18] were also estimated to investigate the behavior of the quantile residuals, which under correct model specification, should show no serial correlation and be normally distributed [19].

c Monte Carlo simulation

After estimating the parameters and generating the GAS time series, we now have a series f_t that are actually location parameters for the Lognormal distribution. Given that σ is a constant parameter estimated through Algorithm 1, we have the necessary tools to perform Monte Carlo simulation [20]. In order to do that, we just need to input the ENSO variables for the desired period of time.

This is a powerful tool for measuring ENSO effect, since we can conduct conditional simulations by utilizing arbitrary ENSO values, assuming strong El Niño or La Niña periods, and compare both results. This methodology is further discussed in Section 5.d.

d Confidence bands

Given that we estimated $\hat{\theta}$, the MLE estimator of θ, we can obtain values of f_t, $t = 1, ..., n$ by using Eq. (14). However, f_t has sources of uncertainty that should be taken into account, namely parameter uncertainty – since we do not know the true parameter vector θ, just the MLE estimate $\hat{\theta}$ –, filtering uncertainty – since we consider Eq. (14) as the updating equation, but we do not know the true time-varying parameter process – and model uncertainty – since we consider $p(\cdot)$ in Eq. (2) but we do not know the true data generation process for y_t.

We are going to focus on parameter uncertainty and model it by generating confidence bands, as described in [21]. These confidence bands can be constructed by using the cumulative delta method or by Monte Carlo simulation. Given that the GAS model outputs probability distributions, it is more convenient to construct simulation-based bands.

Let $\hat{\mathbf{W}}$ be the covariance matrix of the parameters. By operating under the assumption of asymptotic distribution for the MLE estimates

$$\hat{\theta}^{(i)} \sim N(\hat{\theta}, \hat{\mathbf{W}}), \qquad i = 1, ..., m \tag{21}$$

and by using Monte Carlo simulation, we draw parameter values $\theta^{(i)}$ from the distribution of $\hat{\theta}$ for $i = 1, ..., m$, where m represents the number of simulated scenarios. Then, for each $\theta^{(i)}$, we can determine the sequence $\hat{f}_t^{(i)}$ through Eq. (14). With a collection $\hat{f}_t^{(i)}$, we can construct our time series for each realization, and therefore have m scenarios. Finally, by choosing a significance level α, we have the $\alpha/2$ and $1 - \alpha/2$ quantile series as our inferior and superior confidence bands, respectively.

The confidence bands, in addition to modeling parameter uncertainty, also act as a tool that can be used to verify the statistical significance of the ENSO effect, by utilizing the previously mentioned conditional simulations. If the El Niño-conditioned and La Niña-conditioned confidence bands are completely disjointed, i.e. the inferior band of one simulation is above the superior band of the other, this indicates a significant ENSO effect in the generation.

6 Results

For the case study, the following data was utilized:

- Historical ANE monthly series for each one of Brazil's four regions, used as our dependent variable of interest;
- Historical ONI monthly series, serving as an ENSO indicator, used as an exogenous explanatory variable.

The utilized data spans from 1951 to 2015, comprehending 65 years or 780 monthly observations.

a Estimation

Below are the results for each one of the studied regions, following the methodology presented in Section 5, with Algorithm 1 running 100 rounds. The values for the Jarque-Bera and Ljung-Box tests represent the correspondent p-values, and (r) indicates the test is done on the quantile residuals while (r^2) indicates it is done on the squared quantile residuals.

		SE	S	NE	N
Parameter estimate	$\hat{\phi}$	0.005968	0.1357	0.0009361	-0.01236
Goodness-of-fit statistics	Pseudo-R^2	0.8382	0.3661	0.7537	0.8803
	MAE (aMW)	5773	3914	1700	1288
	sMAPE	6.74%	19.78%	9.50%	6.93%
Model diagnostics	Jarque-Bera (r)	0.0846	0.0360	$< 10^{-15}$	$< 10^{-15}$
	Ljung-Box (r)	0.169	0.723	0.0155	$3.31 \cdot 10^{-4}$
	Ljung-Box (r^2)	0.00112	0.139	$< 10^{-15}$	10^{-14}

Table 1: Model's result summary.

b Simulation

The simulation was conducted with 10,000 scenarios. In order to compare the El Niño and La Niña outcomes, we arbitrarily considered the final 18 months of the series as either strong El Niño, with ONI values equal to $+2$, or strong La Niña, with ONI values equal to -2.

In the simulation graphs, the vertical line indicates the starting point of the artificial El Niño and La Niña periods, which is why the confidence bands are become distinct after it. The bands represent the 5% and 95% quantiles for the El Niño and La Niña conditional simulations. The results can be seen in the following pages.

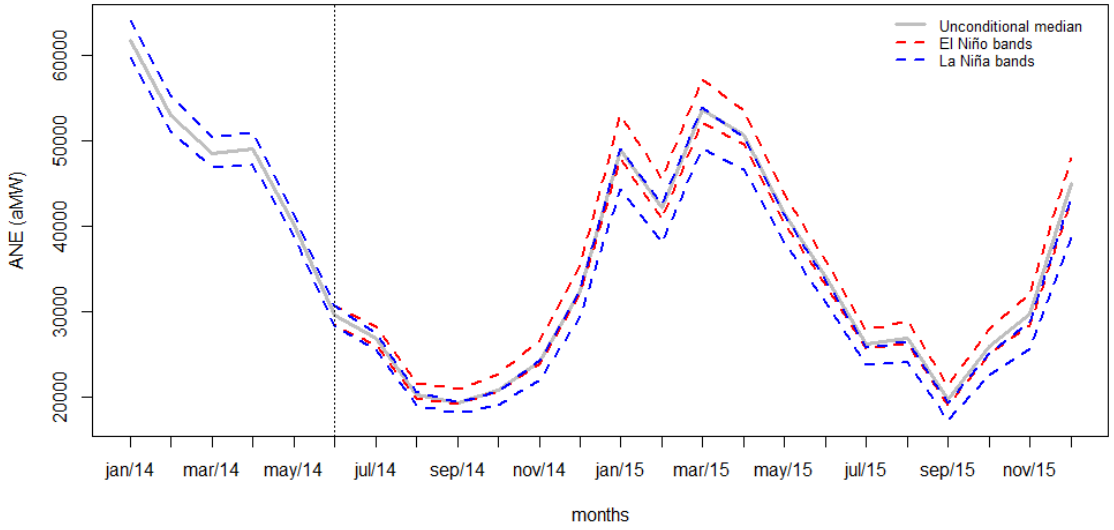

Figure 5: ANE simulation conditioned to El Niño and La Niña for the Southeast region.

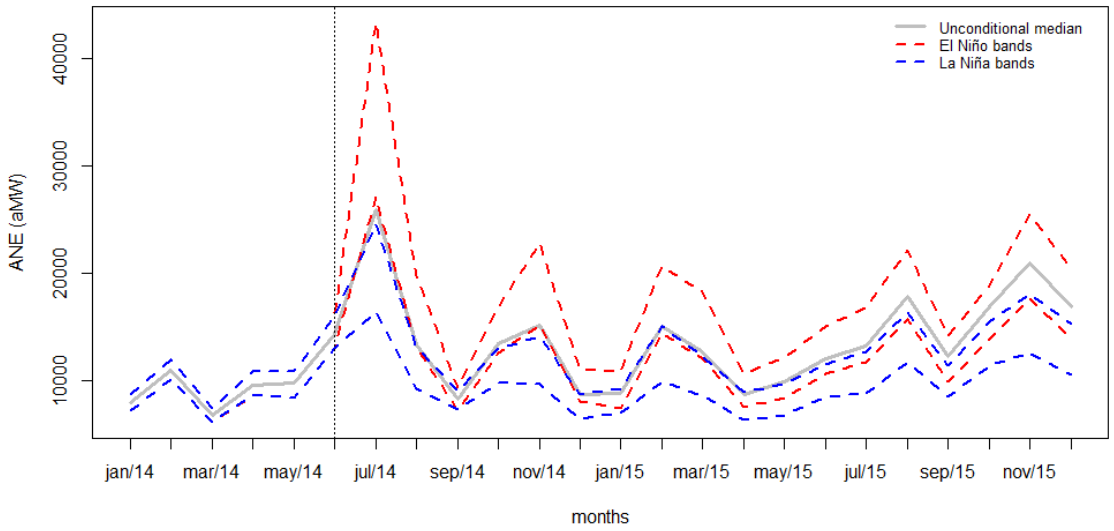

Figure 6: ANE simulation conditioned to El Niño and La Niña for the South region.

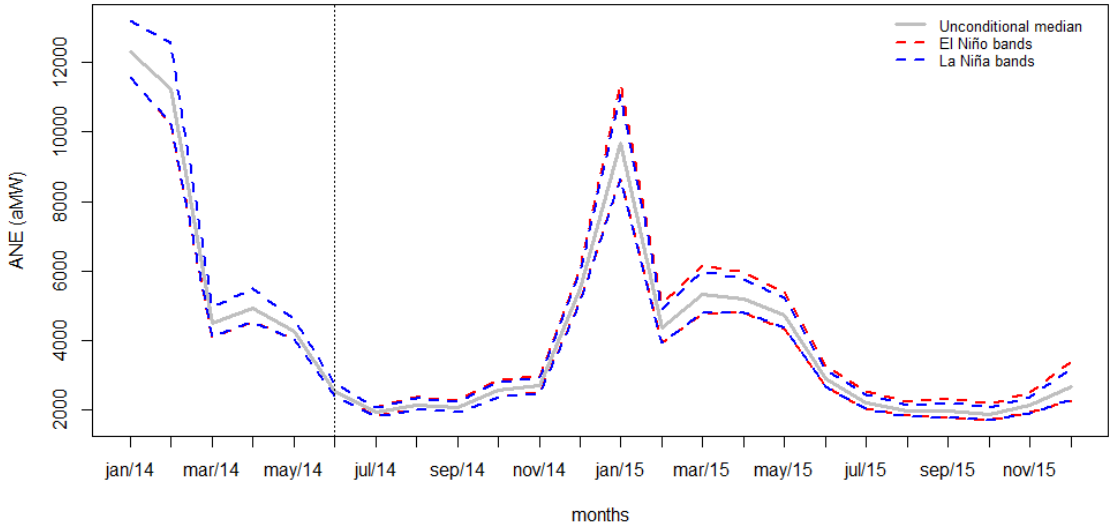

Figure 7: ANE simulation conditioned to El Niño and La Niña for the Northeast region.

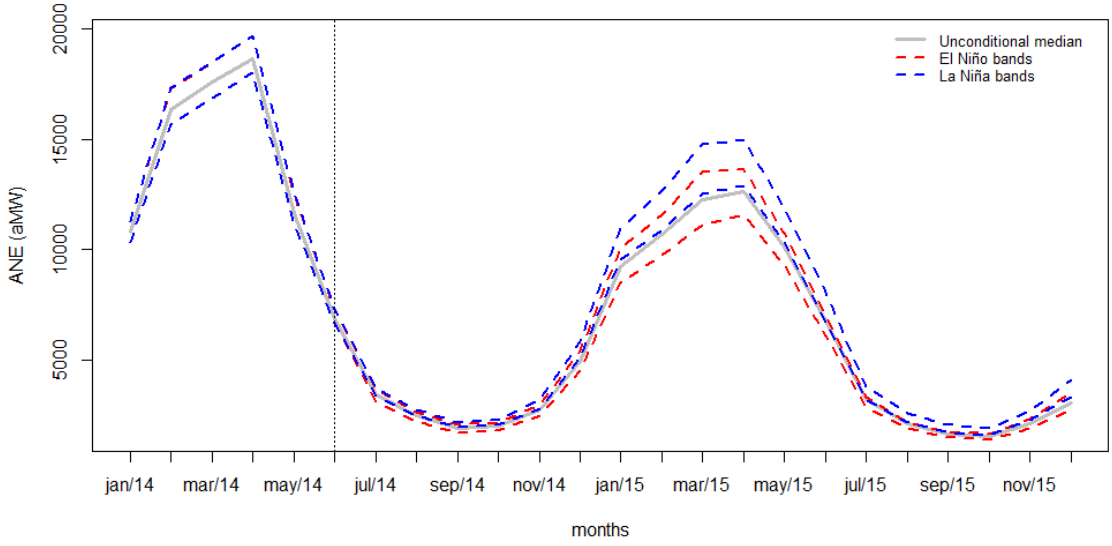

Figure 8: ANE simulation conditioned to El Niño and La Niña for the North region.

c Result summary

As evidenced by Figure 3 and by the results, the ANEs of each region have very distinct natures. The model diagnostics were mostly appropriate for the Southeast and South regions, while the Jarque-Bera and Ljung-Box tests were rejected for the Northeast and North regions. The ACF and PACF were acceptable for every region (see Appendix), which indicates the lags successfully captured the behavior of the series. On the other hand, Gaussianity was rejected for the residuals of South, Northeast and North regions at the 95% significance level through the Jarque-Bera test. This is probably due to outliers that should be treated through dummies. Figures 9 through 20, available in the Appendix, include the autocorrelation function, partial autocorrelation function

and QQ-plot results for each region. It is also evident by the goodness-of-fit statistics that the model fitting had different levels of success among regions.

For the Southeast region, a Pseudo-R^2 of 0.83 and sMAPE of 6.74% indicate a good in-sample fitting. That, allied to the good residuals results, suggests that our Lognormal GAS model is adequate for modeling the ANE of this region. The parameter simulation indicates a considerable ENSO effect, with the El Niño lower band being very close to the La Niña upper band.

For the South region, we had a very high parameter estimate, with $\phi = 0.1357$, indicating a strong ENSO effect. However, it was the region with the worst goodness-of-fit statistics by far, which is to be expected since Southern Brazil ANEs have a very different pattern when compared to the other regions, with less seasonality and a more random behavior. Due to this irregular behavior, it is the hardest region to adequately model. Since the in-sample fitting was not good, the ENSO effect estimate cannot be considered meaningful.

For the Northeast region, a Pseudo-R^2 of 0.75 and sMAPE of 9.50% also indicate a good in-sample fitting. The ENSO parameter estimate was $\phi = 0.0009361$, indicating a minimal ENSO effect on ANE. This is visible in Figure 7, where the confidence bands for El Niño and La Niña are almost identical. Finally, the QQ-plot and Jarque-Bera test indicate non-Gaussianity for the quantile residuals.

For the North region, we had very good in-sample fitting, with an Pseudo-R^2 of 0.88 and sMAPE of 6.93%. Similarly to the Northeast region, the QQ-plot and Jarque-Bera point towards non-Gaussianity. Interestingly, in the North region, the ENSO effect is reversed when compared to the other regions, with La Niña positively affecting ANE.

El Niño & La Niña comparison	SE	S	NE	N
Percentage difference	$+7.34\%$	$+32.40\%$	$+1.11\%$	-11.63%
Absolute difference	43097 MW	70206 MW	710 MW	-12625 MW

Table 2: Comparison between El Niño-conditioned and La Niña-conditioned scenarios.

Finally, Table 2 shows the percentage and absolute differences for the El Niño and La Niña-conditioned simulated scenarios. Positive values indicate El Niño-conditioned ANEs are higher and negative values indicate La Niña-conditioned ANEs are higher. The results indicate significant ENSO effect in every region except the Northeast, though the unsatisfactory fitting for the South region discredits its results. Nonetheless, we have found the Southeast and North regions to be affected by ENSO in non-negligible manner with acceptable goodness-of-fit statistics and model diagnostics.

7 Conclusions & future work

This work allowed us to conclude that the Lognormal GAS model is a viable and adequate method for modeling ANE and possibly other water inflow-based metrics. However, since each one of Brazil's four regions has a different nature and seasonal behavior, the model should be adjusted accordingly. The inappropriate fitting for the South region was to be expected by looking at the historical ANE series (see Figure 3), which has a very distinct distribution from the other regions, having a more random behavior.

One strong assumption we made throughout this work was that ENSO effect is constant over time, since we used only a constant parameter to model it. There are reasons to believe that the ENSO effect is time-variant, as discussed in [22–24], which means a single constant parameter is not an adequate way to model ENSO. Further studies should be done by using different time windows and different parameters for each season.

The biggest concern about our results is the fact that the optimality of the parameters is not guaranteed in our maximum likelihood estimation. By using Algorithm 1 with a large number of rounds, we can filter out local maxima results, but as the number of rounds gets very large, the computational time does the same.

Out-of-sample studies should be done to further investigate the goodness-of-fit of the model, since using only in-sample metrics can be misleading at times.

Finally, in this study, we utilized a Lognormal GAS model with one time-varying parameter – the location parameter. That means we assumed a constant scale parameter σ. However, this simplification might hurt the model since, in some regions, the variance is seasonal. It is visible in Figure 3 that the variance is consistently smaller in the second semester when compared to the beginning of the year. Therefore, a Lognormal GAS model with two time-varying parameters will probably be better suited to model time series such as these.

A Appendix

a Southeast

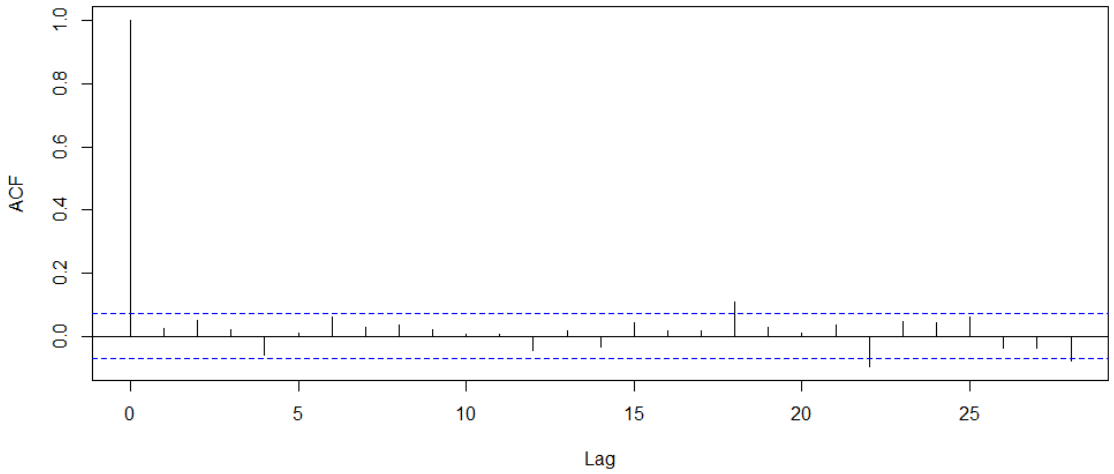

Figure 9: Autocorrelation function for the Southeast region.

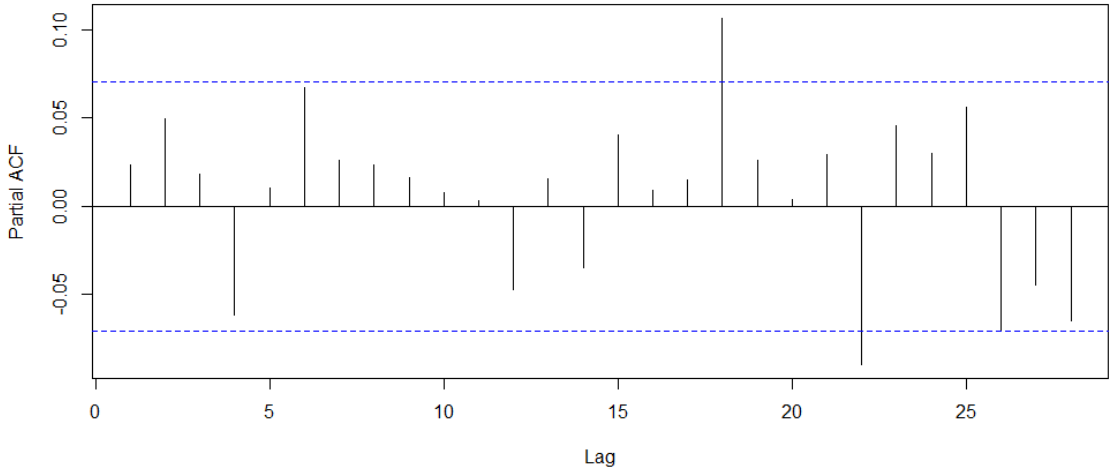

Figure 10: Partial autocorrelation function for the Southeast region.

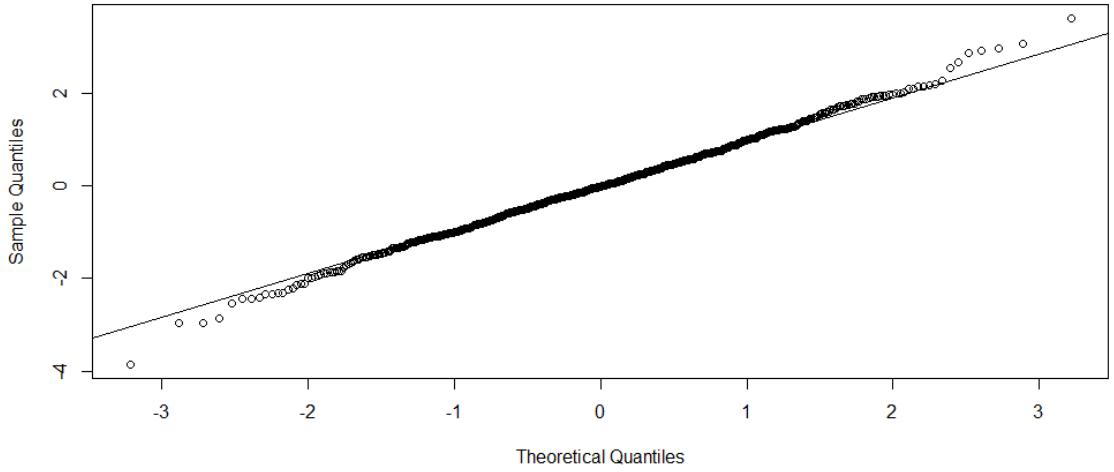

Figure 11: QQ-plot for the Southeast region.

b South

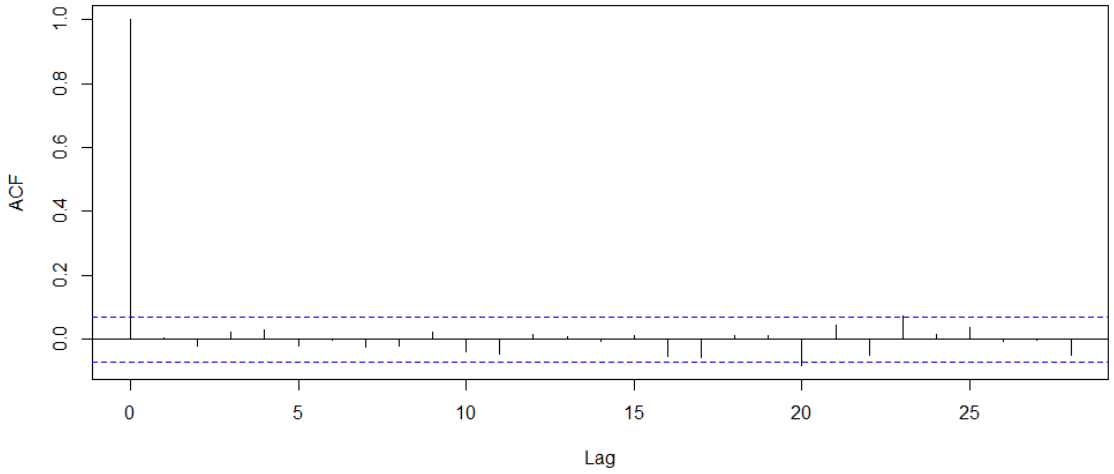

Figure 12: Autocorrelation function for the South region.

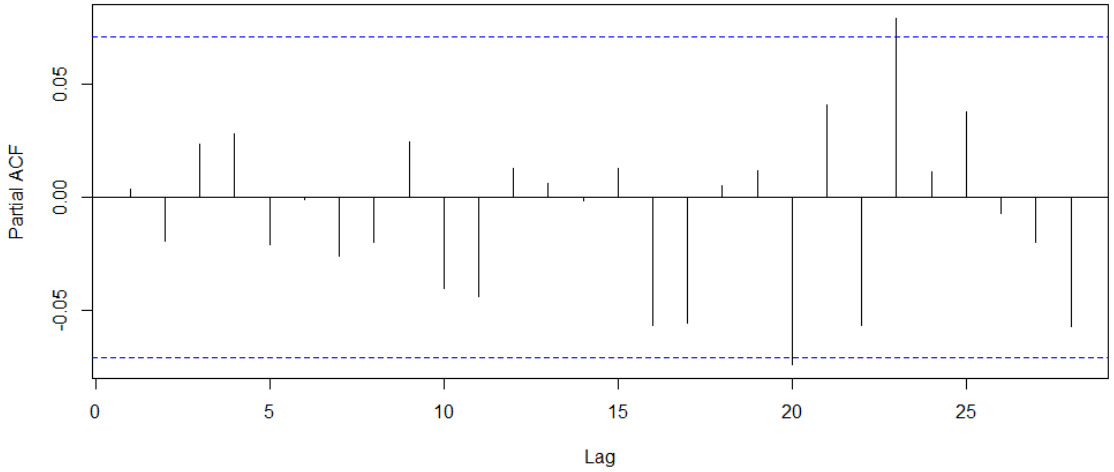

Figure 13: Partial autocorrelation function for the South region.

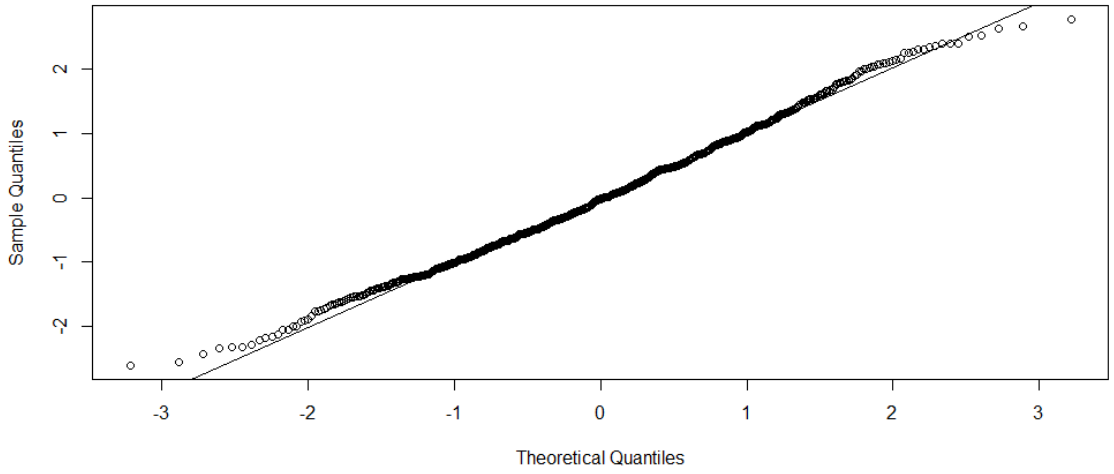

Figure 14: QQ-plot for the South region.

c Northeast

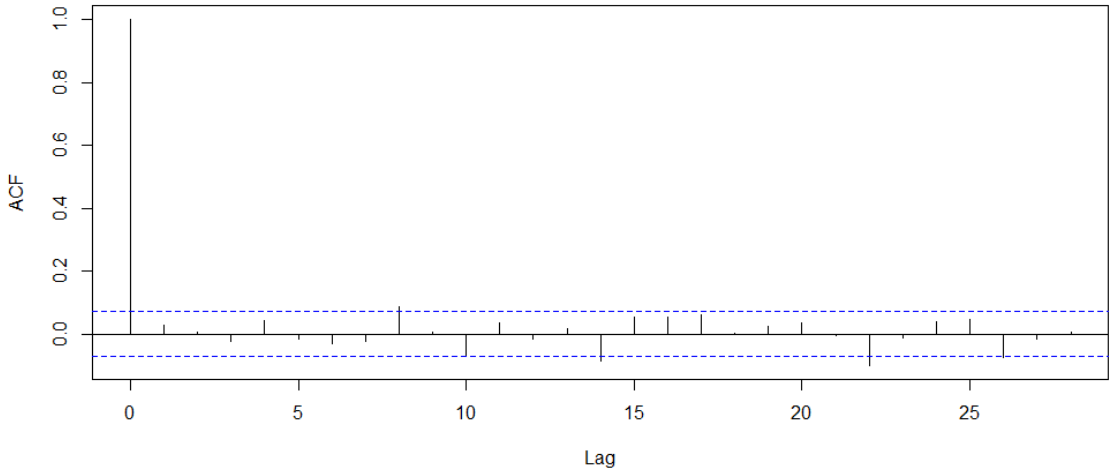

Figure 15: Autocorrelation function for the Northeast region.

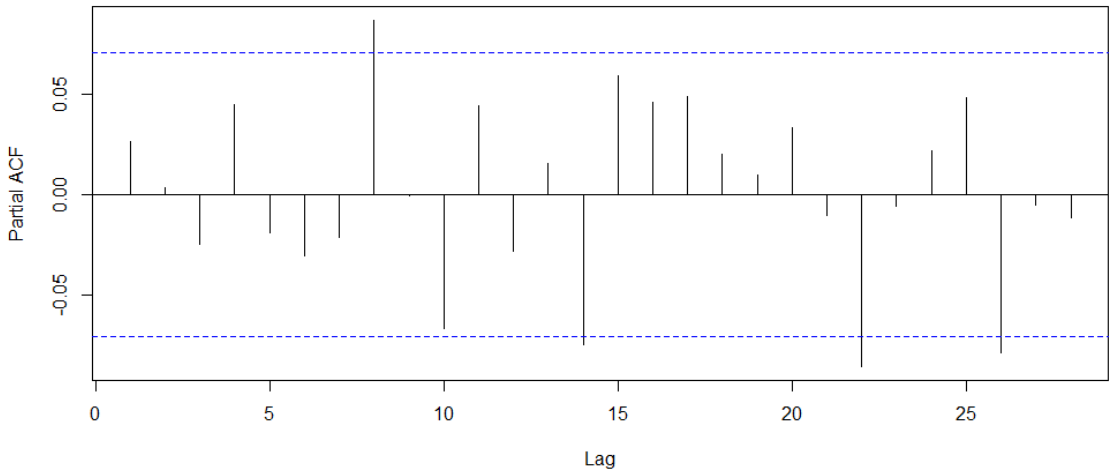

Figure 16: Partial autocorrelation function for the Northeast region.

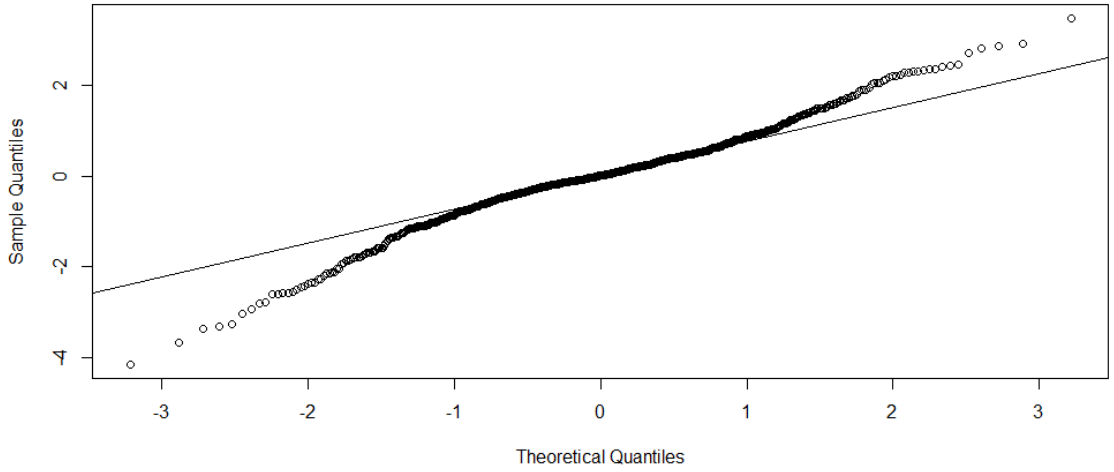

Figure 17: QQ-plot for the Northeast region.

d North

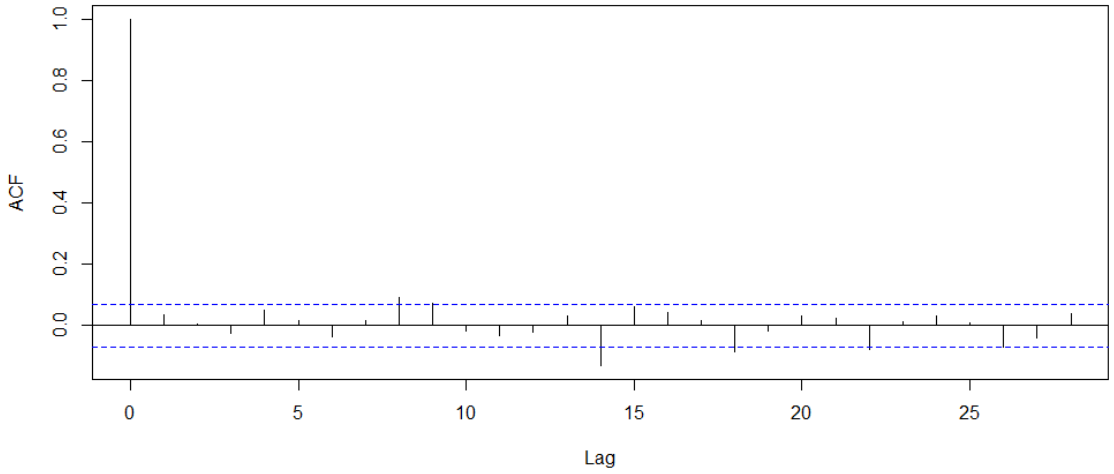

Figure 18: Autocorrelation function for the North region.

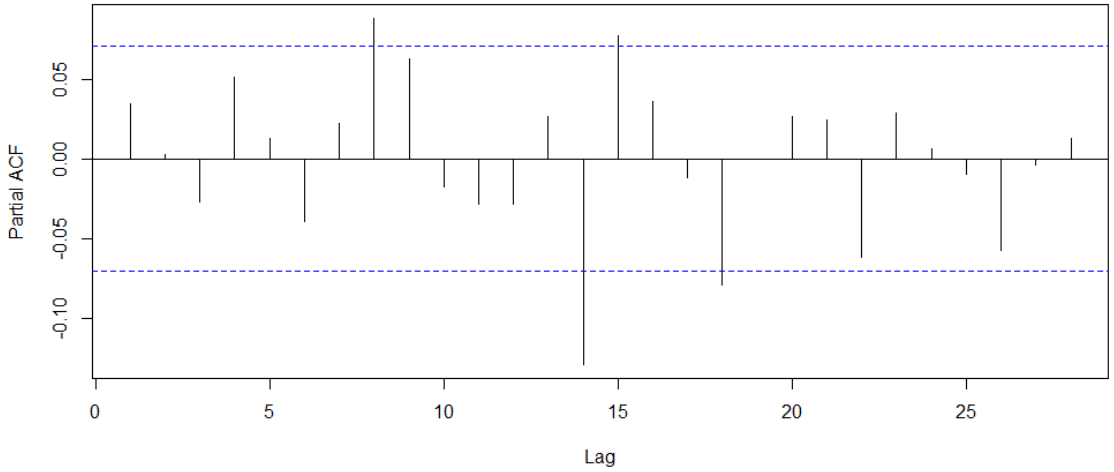

Figure 19: Partial autocorrelation function for the North region.

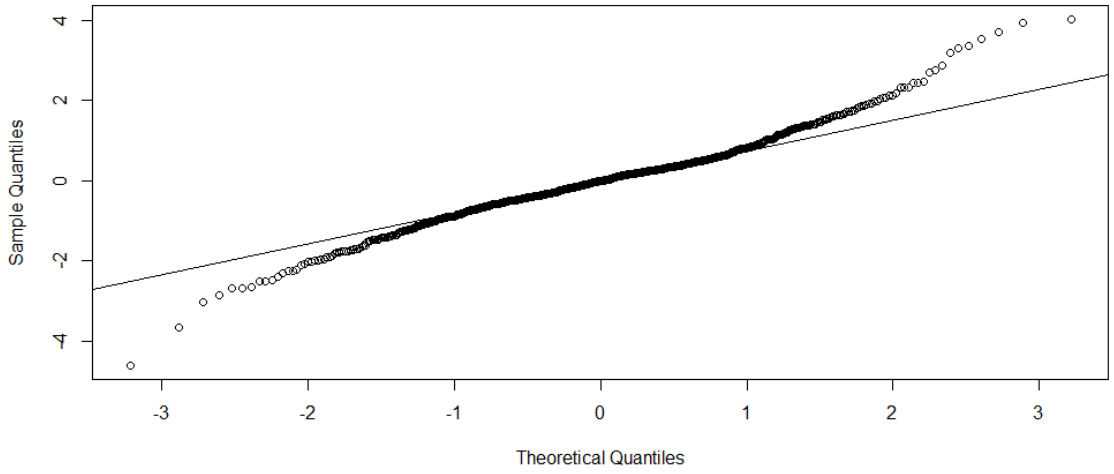

Figure 20: QQ-plot for the North region.

References

[1] www.cpc.ncep.noaa.gov.

[2] A. A. de Almeida, "Política tarifária e comercialização de energia elétrica no ambiente de contratação livre," *Simpósio Brasileiro de Sistemas Elétricos*, 2006.

[3] L. Freire, "Modelo de comercialização de energia renovável no ambiente de contratação livre via teoria de jogos cooperativos," Master's thesis, PUC-Rio, 2013.

[4] PSR Inc., "Vazão do São Francisco: tema que continua tirando nosso sono." *PSR Energy Report, Junho de 2016*, 2016.

[5] S. R. dos Santos Gonçalves and R. F. Calili, "Análise das séries sintéticas de energia natural afluente bruta da região nordeste do brasil."

[6] C. N. Nascimento, C. A. Fernandes, G. B. França, and G. Matos, "Estimação do Impacto do El Niño/La Niña na Intensidade dos Ventos do Nordeste Brasileiro." *Anuário do Instituto de Geociências*, 2014.

[7] D. Creal, S. J. Koopman, and A. Lucas, "Generalized Autoregressive Score Models with Applications." *Tinbergen Institute Discussion Paper*, 2008.

[8] ——, "Generalized Autoregressive Score Models with Applications." *Journal of Applied Econometrics*, vol. 28, 2013.

[9] E. W. Bliss and G. T. Walker, "World Weather V." *Memoirs of the Royal Meteorological Society*, vol. 4, 1932.

[10] www.ncdc.noaa.gov/teleconnections/enso.

[11] www.ons.org.br.

[12] M. E. P. Maceira, V. Duarte, D. D. J. Penna, L. A. M. Moraes, and A. C. G. Melo, "Ten years of application of stochastic dual dynamic programming in official and agent studies in brazil - description of the newave program," *16th PSCC, Glasgow, Scotland*, 2008.

[13] T. H. de Mello, V. L. de Matos, and E. C. Finardi, "Sampling strategies and stopping criteria for stochastic dual dynamic programming: a case study in long-term hydrothermal scheduling," *Energy Systems*, vol. 2, 2011.

[14] R. Sacchi, M. C. Ozturk, J. C. Principe, A. F. M. Carneiro, and I. N. da Silva, "Water inflow forecasting using the echo state network: A brazilian case study," in *Neural Networks, 2007. IJCNN 2007. International Joint Conference on*. IEEE, 2007.

[15] J. M. Wooldridge, *Introductory econometrics: A modern approach*. Nelson Education, 2015.

[16] S. Makridakis, "Accuracy measures: theoretical and practical concerns," *International Journal of Forecasting*, vol. 9, no. 4, 1993.

[17] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[18] C. M. Jarque and A. K. Bera, "Efficient tests for normality, homoscedasticity and serial independence of regression residuals," *Economics letters*, vol. 6, no. 3, 1980.

[19] L. Kalliovirta, "Diagnostic tests based on quantile residuals for nonlinear time series models," Ph.D. dissertation, University of Helsinki, 2009.

[20] C. Z. Mooney, *Monte Carlo simulation*. Sage Publications, 1997, vol. 116.

[21] F. Blasques, S. J. Koopman, K. Łasak, and A. Lucas, "In-sample confidence bands and out-of-sample forecast bands for time-varying parameters in observation-driven models," *International Journal of Forecasting*, vol. 32, no. 3, pp. 875–887, 2016.

[22] A. Grimm, S. Ferraz, and J. Gomes, "Precipitation Anomalies in Southern Brazil Associated with El Niño and La Niña Events." *Journal of Climate*, vol. 11, 1997.

[23] A. Grimm, V. Barros, and M. Doyle, "Climate Variability in Southern South America Associated with El Niño and La Niña Events." *Journal of Climate*, vol. 13, 1999.

[24] A. Grimm, "The El Niño Impact on the Summer Monsoon in Brazil: Regional Processes versus Remote Influences." *Journal of Climate*, vol. 16, 2002.