

Comparative Causality Analyses between Hydrological Natural Inflow and Climate Variables in Brazil

Xu Huang, Paula Medina Macaira, Hossein Hassani*, Fernando Luiz Cyrino Oliveira and Gurjeet Dhesi

**corresponding author: hassani.stat@gmail.com*

Abstract

Numbers of studies have proved the significant influence of climate variables on hydrological series. Considering the pivotal role of the hydroelectric power plants play in the electricity production in Brazil this paper considers the natural hydrological inflow data from 15 major basins and 8 climate variables containing 7 El Niño Southern Oscillation proxies and the sunspot numbers. The causal relationships between hydrological natural inflows and climate variables are investigated by adopting and comparing 5 different causality detection methods (Granger Causality test, Frequency Domain Causality test, Convergent Cross Mapping Causality test, Single Spectrum Analysis (SSA) Causality test and Periodic Autoregressive Model Causality test) that cover both well established and novel empirical approaches. **Both time domain and frequency domain causality tests gain valid evidences of unidirectional causality for a group of series; CCM achieved unidirectional causality for 18% of pairs and overwhelmingly indicated the opposite direction of causality; a mixture of results are concluded by SSA causality test; PAR based causality test obtained six unidirectional causality, but only one is really significant.**

Keywords: Hydrological Natural Inflow; Climate Variables; Causality Detection; Granger Causality; Frequency Domain Causality; Convergent Cross Mapping; Single Spectrum Analysis; Periodic Autoregressive Model.

23 Nomenclature

<i>CCM</i>	Convergent Cross Mapping.
<i>EDM</i>	Empirical Dynamic Modeling.
<i>ENSO</i>	El Niño-Southern Oscillation.
<i>GC</i>	Granger Causality Test.
<i>NGDC</i>	National Geophysical Data Center.
<i>NOAA</i>	National Oceanic and Atmospheric Administration.
24 <i>ONI</i>	Oceanic Niño Index.
<i>PAR</i>	Periodic Autoregressive Model.
<i>PARX</i>	Periodic Autoregressive Model with One Exogenous Variable.
<i>RMSE</i>	Root Mean Square Error.
<i>SOI</i>	Southern Oscillation Index.
<i>SSA</i>	Singular Spectrum Analysis.
<i>SST</i>	Sea Surface Temperature.

25 1 Introduction

26 In Brazil there are 1268 hydroelectric power plants in operation, corresponding to 65% of total
27 installed capacity and responsible for 73% of electricity production in 2016 [1]. This kind of power
28 plant produces electricity by harnessing a river hydraulic potential so the electricity generation
29 depends directly on hydrological regimes.

30 Since the 90s there are several studies showing that not only there is **an** influence of climate
31 variables like El Niño-Southern Oscillation (ENSO) on hydrological series [2–6, 8], but also that
32 when correlation is taken into account there is improvement in the forecasting/modelling exercise
33 of inflow time series [9–13], for instance, storm tides data at the Baltic Sea in [14] and streamflow
34 data of the East River basin of China in [15] by adopting the significant Hurst exponent [16],
35 **which has also been applied in birth time series [17]. Another recent research considered Hurst**
36 **exponent in analyzing hydrogeological series can be found in [18].**

37 This paper aims to establish comprehensive causality analyses between natural inflow and
38 climate variables in Brazil by embracing and comparing both well established and advanced
39 causality detection methods, **including time domain Granger causality (GC) test [19], frequency**
40 **domain causality test [20], Convergent Cross Mapping (CCM) [21], Singular Spectrum Analysis**
41 **(SSA) based causality test [22] and the Periodic Autoregressive model (PAR) based causality**
42 **test [23, 24].**

43 Most of the works previously cited study the influence of ENSO events using the Sea Surface
44 Temperature (SST) variable for the Northeast region of Brazil, ignoring others geographic regions
45 and also other variables that possibly indicate a proxy for ENSO. In this paper, all the fifteen
46 Brazil major basins are considered to test the causality with more than seven ENSO proxies
47 and the Sunspot climate event.

48 The remainder of this paper is organized such that the background of this study is presented
49 in Section 2; the causality detection techniques adopted in this paper are briefly summarized in

50 Section 3; Section 4 introduces the data and summarizes the descriptive statistics along with
51 correlation analyses; the detailed causality test results by different methods are listed in Section
52 5; the paper concludes in Section 6 with proposals of future research.

53 2 Background

54 It is possible to find several studies that identify the influence of ENSO events in the Brazilian
55 river basins, but none of them apply any type of causality test. Amarasekera et al. [2] concludes
56 that the annual discharges of the Amazon river is weakly and negatively correlated with the
57 equatorial Pacific Sea Surface Temperature (SST) anomaly, while the the Paraná river shows
58 a strong and positive correlation. Dettinger & Diaz [3] uses El Niño variations to characterize
59 geographic differences in the seasonality and year-to-year variability of stream flow from several
60 sites around the world, and shows that the Amazon basin is drier-than-normal in El Niño
61 years accordingly to **Southern Oscillation Index (SOI)** and North Atlantic Oscillation (NAO)
62 [25] index. Foley et al. [4] shows that during the El Niño there is a decrease in the Amazon
63 and Tocantins river discharge, and the opposite during the La Niña. Berri et al. [5] presents
64 that exactly the opposite happens in Paraná river, i.e., during El Niño the average inflow are
65 always larger than those observed during La Niña events. Garcia & Mechoso [6] concludes
66 that the Amazon, Tocantins, San Francisco, Paraguay, Paraná river streamflows shows El Niño-
67 like periodicities. Soares et al. [8] notice that the sub-basins of the southern Brazilian regions
68 showed positive variations in water production during El Niño, while the Amazon basin showed
69 no response.

70 Souza Filho [9] shows that the correlation between climate and hydrological variables is
71 beneficial for the prediction of reservoirs inflows in Ceará. Cardoso & Silva Dias [10] use the
72 SST index to show that there is improvement in the reservoir inflow forecasting of Paraná River.
73 Lima & Damien [11] apply dynamic linear models to predict the inflow of the Brazil fifteen main
74 basins using precipitation and an El Niño index. Maçaira et al. [12] developed a causal PAR
75 model to estimate the influence between several El Niño indices and the inflow time series of
76 some Brazilian locations. Silveira et al. [13] propose the Periodic Autoregressive model with
77 one **exogenous** variable (PARX) to simultaneously predict all natural inflows of the National
78 Interconnected System.

79 Apart from the significance of studying the causal links between natural inflow and climate
80 factors, this paper has adopted 5 different causality detection techniques covering both well
81 established and advanced time series analysis methods (note that the detailed introduction of
82 these methods are summarized in section 3). It worth to be noted as another contribution of
83 this paper that it comprehensively investigates the causal relationship with the most up to date
84 time series analysis techniques to our knowledge.

85 The well established and widely applied GC approach enables researchers to evaluate de-
86 pendence relationship, mostly linear, among factors in a complex system. It brings insights on
87 whether the changes of one factor have relationship with the changes of another factor in the

88 current sequence or after specific lag of time. However, it assumes linearity and separability for
 89 the selected variables in the model and the nonlinear applicability is limited. The frequency do-
 90 main causality test extends the GC approach to identify the causality for each frequency instead
 91 of a single statistics for the whole time series, whilst the restricted assumptions and nonlinear
 92 applicability maintain. In addition, by adopting the advanced time series analysis techniques
 93 like SSA and CCM, this paper also explores the causality detection from the aspect of nonlin-
 94 earity and other complex dynamics. These advanced non-parametric techniques are relatively
 95 new and have no assumptions of linear or restricted nonlinear model. They are designed to be
 96 widely applicable and assumption free with straightforward way of thinking and implementing.
 97 By adopting these advanced methods, this paper seeks to further distinguish possible causal
 98 relationships that the empirical tests cannot achieve or fall short at. In general, to the best of
 99 our knowledge, this paper is the first attempt that applies and compares all these five causality
 100 detection methods to date. Moreover, for most of the advanced methods, it is also the initial
 101 implementation study on the natural inflow and climate variables in Brazil.

102 **3 Causality Detection Methods**

103 **3.1 Time Domain Granger Causality Test**

104 GC test [19] is the most generally accepted and significant method for causality analyses in
 105 various disciplines. The regression formulation of granger causality states that vector X_i is the
 106 cause of vector Y_i if the past values of X_i are helpful in predicting the future value of Y_i , two
 107 regressions are considered as follows:

$$Y_i = \sum_{t=1}^T \alpha_t Y_{i-t} + \varepsilon_{1i}, \quad (1)$$

$$Y_i = \sum_{t=1}^T \alpha_t Y_{i-t} + \sum_{t=1}^T \beta_t X_{i-t} + \varepsilon_{2i}, \quad (2)$$

108 where $i = 1, 2, \dots, N$ (N is the number of observations), T is the maximal time lag, α and β are
 109 vectors of coefficients, ε is the error term. The first regression is the model that predicts Y_i by
 110 using the history of Y_i only, while the second regression involves both X_i and Y_i . Therefore, the
 111 conclusion of existing causality is conducted if the second model is a significantly better model
 112 than the first one.

113 **3.2 Frequency Domain Causality Test**

114 The frequency domain causality test is the extension of time domain GC test that identifies the
 115 causality between different variables for each frequency. In order to briefly introduce the testing
 116 methodology, we mainly follow [20, 26].

117 It is assumed that two dimensional vector containing X_i and Y_i (where $i = 1, 2, \dots, N$ and N
 118 is the number of observations) with a finite-order Vector Auto-regression Model representative

119 of order p ,

$$\Theta(R) \begin{pmatrix} Y_i \\ X_i \end{pmatrix} = \begin{pmatrix} \Theta_{11}(R) & \Theta_{12}(R) \\ \Theta_{21}(R) & \Theta_{22}(R) \end{pmatrix} \begin{pmatrix} Y_i \\ X_i \end{pmatrix} + \mathcal{E}_i, \quad (3)$$

120 where $\Theta(R) = I - \Theta_1 R - \dots - \Theta_p R^p$ is a 2×2 lag polynomial and $\Theta_1, \dots, \Theta_p$ are 2×2 autoregressive
 121 parameter matrices, with $R^k X_i = X_{i-k}$ and $R^k Y_i = Y_{i-k}$. The error vector \mathcal{E} is white noise
 122 with zero mean, and $E(\mathcal{E}_i \mathcal{E}_i') = \mathbf{Z}$, where \mathbf{Z} is positive definite matrix. The moving average
 123 representative of the system is

$$\begin{pmatrix} Y_i \\ X_i \end{pmatrix} = \Psi(R) \eta_i = \begin{pmatrix} \Psi_{11}(R) & \Psi_{12}(R) \\ \Psi_{21}(R) & \Psi_{22}(R) \end{pmatrix} \begin{pmatrix} \eta_{1i} \\ \eta_{2i} \end{pmatrix}, \quad (4)$$

124 with $\Psi(R) = \Theta(R)^{-1} \mathbf{G}^{-1}$ and \mathbf{G} is the lower triangular matrix of the Cholesky decomposition
 125 $\mathbf{G}' \mathbf{G} = \mathbf{Z}^{-1}$, such that $E(\eta_t \eta_t') = I$ and $\eta_i = \mathbf{G} \mathcal{E}_i$. The causality test developed in [26] can be
 126 written as:

$$C_{X \Rightarrow Y}(\gamma) = \log \left[1 + \frac{|\Psi_{12}(e^{-i\gamma})|^2}{|\Psi_{11}(e^{-i\gamma})|^2} \right]. \quad (5)$$

127 However, according to this framework, no Granger causality from X_i to Y_i at frequency γ
 128 corresponds to the condition $|\Psi_{12}(e^{-i\gamma})| = 0$, this condition leads to

$$|\Theta_{12}(e^{-i\gamma})| = |\sum_{k=1}^p \Theta_{k,12} \cos(k\gamma) - i \sum_{k=1}^p \Theta_{k,12} \sin(k\gamma)| = 0, \quad (6)$$

129 where $\Theta_{k,1,2}$ is the $(1, 2)$ th element of Θ_k , such that a sufficient set of conditions for no causality
 130 is given by [20]

$$\begin{aligned} \sum_{k=1}^p \Theta_{k,1,2} \cos(k\gamma) &= 0 \\ \sum_{k=1}^p \Theta_{k,1,2} \sin(k\gamma) &= 0 \end{aligned} \quad (7)$$

131 Hence, the null hypothesis of no Granger causality at frequency γ can be tested by using a
 132 standard F-test for the linear restrictions (7), which follows an $F(2, B - 2p)$ distribution, for
 133 every γ between 0 and π , with B begin the number of observations in the series.

134 3.3 Convergent Cross Mapping

135 CCM is firstly introduced in [21] that aimed at detecting the causation among time series
 136 and provide a better understanding of the dynamical systems that have not been covered by
 137 other well established methods like GC. CCM has proven to be an advanced non-parametric
 138 technique for distinguishing causations in a dynamical system that contains complex interactions
 139 in ecosystems and climate studies [21, 27], more details can be found in [28, 29]. Some significant
 140 rationales of embracing this advanced technique include: CCM is non-parametric approach
 141 with no restrictions of assumptions for parametric methods; CCM can distinguish statistically
 142 significant causality by considering only two key variables instead of building a complex model
 143 by incorporating many possible influential variables based on regression modelling; CCM has
 144 remarkable sensitivity at detecting causal links within complex systems whilst not being limited

145 to linearity or nonlinearity; the calculation itself is efficient and comparatively straight forward.
 146 CCM is briefly introduced below by mainly following [21].

147 Assume there are two variables X_i and Y_i , for which X_i has a causal effect on Y_i . CCM
 148 test will test the causation by evaluating whether the historical record of Y_i can be used to get
 149 reliable estimates of X_i . Given a library set of n points (not necessarily to be the total number of
 150 observations N of two variables) and here set $i = 1, 2, \dots, n$, the lagged coordinates are adopted
 151 to generate an E -dimensional embedding state space [30,31], in which the points are the library
 152 vector X_i and prediction vector Y_i

$$X_i : \{x_i, x_{i-1}, x_{i-2}, \dots, x_{i-(E-1)}\}, \quad (8)$$

$$Y_i : \{y_i, y_{i-1}, y_{i-2}, \dots, y_{i-(E-1)}\}, \quad (9)$$

153 The $E+1$ neighbors of Y_i from the library set X_i will be selected, which actually form the smallest
 154 simplex that contains Y_i as an interior point. Accordingly, the forecast is then conducted by this
 155 process, which is the nearest-neighbour forecasting algorithm of simplex projection [31]. The
 156 optimal E will be evaluated and selected based on the forward performances of these nearby
 157 points in an embedding state space.

158 Therefore, by adopting the essential concept of Empirical Dynamic Modeling (EDM) and
 159 generalized Takens' Theorem [30], two manifolds are constructed based on the lagged coordinates
 160 of the two variables under evaluation, which are the attractor manifold M_Y constructed by Y_i
 161 and respectively, the manifold M_X by X_i . The causation will then be identified accordingly
 162 if the nearby points on M_Y can be employed for reconstructing observed X_i . Note that the
 163 correlation coefficient ρ is used for the estimates of cross map skill due to its widely acceptance
 164 and understanding, additionally, leave-one-out cross-validation is considered a more conservative
 165 method and adopted for all evaluations in CCM.

166 3.4 Singular Spectrum Analysis based Causality Test

167 As GC formalized the causality concept and claimed causality if the elimination of one variable
 168 from a system is harmful for explaining the other variable. Similarity, as can be seen in Figure 1,
 169 the SSA based causality analysis is obtained by comparing the forecast values obtained by the
 170 univariate procedure—SSA and multivariate process— multivariate SSA (MSSA). Consequently,
 171 if the forecasting errors using MSSA are significantly smaller than those of univariate SSA, it is
 172 concluded that there is a causal relationship detected between these series. As a nonparametric
 173 technique, the SSA causality test is able to capture possible nonlinearities using a data-driven
 174 approach without specifying any known functional nonlinear model to the relationship, which in
 175 turn, could be incorrectly specified in the first place. Detailed introduction is presented below
 176 which mainly follow [22, 32], where also summarize the details of SSA and MSSA formulation
 177 and forecasting algorithms.

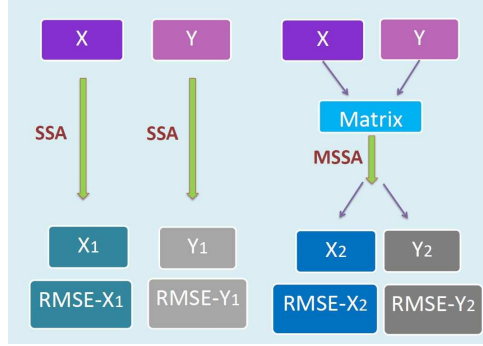


Figure 1: Flowchart of Cause Detection based on SSA Forecasting Accuracy.

178 Let us consider the procedure for constructing vectors of forecasting error for out-of-sample
179 tests in a two variable case X_N and Y_N by both univariate and multivariate SSA techniques
180 respectively. Firstly, the series $X_N = (x_1, \dots, x_N)$ is divided into two separate subseries X_R
181 and X_F that satisfy $X_N = (X_R, X_F)$, where $X_R = (x_1, \dots, x_R)$ and $X_F = (x_{R+1}, \dots, x_N)$. Same
182 procedure is also conducted for Y_N . The subseries X_R and Y_R are used in the reconstruction step
183 to provide the noise-free series \tilde{X}_R and \tilde{Y}_R . The noise-free series are then used for forecasting the
184 subseries X_F and Y_F with the help of the forecasting algorithms (see Appendix A) of SSA and
185 MSSA respectively. For variable X_N , two different forecasting values of $\hat{X}_F = (\hat{x}_{R+1}, \dots, \hat{x}_N)$
186 by SSA and MSSA are then used for computing the forecasting errors accordingly, which will
187 be the same process in terms of variable Y_N . Therefore, in a multivariate system like this, the
188 vectors of forecasts obtained can be used in computing the forecasting accuracy and therefore
189 conducting the causality analysis between the two variables.

190 The length of out-of-sample does not have specific limitation, generally considering the sim-
191 ulation scenario, the length of time series for reconstruction will take 2/3 of the whole series
192 and the rest 1/3 is considered as out-of-sample for constructing forecasting error. The separate
193 point to define the out-of-sample size for different series can be chosen respectively, whilst it
194 is important that when it goes to comparing the performances of different techniques based on
195 constructed forecasting error of one specific series, the sizes of reconstruction and out-of-sample
196 for all techniques should be identical. In addition, the choices of window length L and the
197 referring options of numbers of eigenvalues r should also be carefully evaluated in practice of
198 SSA causality test respectively. Considering this as the first attempt of application, also in
199 order to conduct the most accurate results, all the possibilities of L and its referring choices of r
200 should be applied for both univariate SSA and MSSA processes, then the optimal ones with best
201 performance of forecasting will be chosen to construct the finally cause detection procedure.

202 Consequently, define the criterion $F_{X|Y} = \Delta X_{F|Y} / \Delta X_F$ corresponding to the forecast of the
203 series X_N in the presence of the series Y_N . Specifically, if $F_{X|Y}$ is small, then having information
204 obtained from the series Y can help to achieve better forecasts of the series X . If $F_{X|Y} < 1$, it is
205 concluded that the information provided by the series Y can be regarded as useful or supportive
206 for forecasting the series X . Alternatively, if the values of $F_{X|Y} \geq 1$, then either there is no

207 detectable causality between X and Y or the performance of the univariate SSA is better than of
 208 the MSSA (this may happen, for example, when the series Y has structural breaks misdirecting
 209 the forecasts of X).

210 3.5 Periodic Autoregressive Model based Causality Test

211 To perform monthly forecasts and simulation of hydrological series, the classical PAR model has
 212 been widely used [23]. This type of model adjusts the series using the estimated parameters of
 213 the historical data [33], and does not consider any exogenous information that could affect the
 214 hydrological regimes in equation (10). To consider any exogenous variable in the PAR model,
 215 there is the Periodic Autoregressive model with one exogenous variable (PARX) as presented in
 216 equation (11). PAR models fit for each season an autoregressive term being able to capture the
 217 monthly variability of hydrological regimes, this is the main reason for its success for this type
 218 of data. The mathematical details of PAR and PARX can be found in [12, 13, 24].

$$\left(\frac{Y_i - \mu_m}{\sigma_m}\right) = \sum_{t=1}^{p_m} \varphi_t^m \left(\frac{Y_{i-t} - \mu_{m-t}}{\sigma_{m-t}}\right) + \varepsilon_t, \quad (10)$$

219

$$\left(\frac{Y_i - \mu_m}{\sigma_m}\right) = \sum_{t=1}^{p_m} \varphi_t^m \left(\frac{Y_{i-t} - \mu_{m-t}}{\sigma_{m-t}}\right) + \sum_{t=0}^{v_m} \vartheta_t^m X_{i-t} + \varepsilon_t, \quad (11)$$

220 Where μ_m and σ_m are the average and the standard deviation of season m , respectively; φ_t^m
 221 is the t -th autoregressive coefficient of season m , p_m is the order of the autoregressive operator
 222 of season m . In (11), X_i is the predictor variable, ϑ_t^m is the autoregressive coefficient and v_m is
 223 the order of the autoregressive operator of season m for the predictor variable.

224 Similar to the SSA based Causality Test, it was developed the PAR based Causality Test
 225 that compares the forecasts values obtained by the univariate process PAR and the PARX. If
 226 the forecasting errors using PARX are significantly smaller than those of PAR, it is conclude
 227 that there is a causality detected among the variables.

228 4 Data

229 4.1 The Natural Inflow Series in Brazil

230 According to the Brazilian Electricity Regulatory Agency (ANEEL) there are fifteen major
 231 river basins in Brazil, with an installed capacity of approximately 90 GigaWatts [GW] in 2016,
 232 representing 66% of the total installed capacity in the country (Figure 2). The Parana river
 233 basin has the highest hydroelectric potential, around 43 GW, which represents 48% of total
 234 hydroelectric capacity. It can be further subdivided into six minor basins based on its major
 235 rivers - Paranaíba, Grande, Tiete, Parana, Paranapanema and Iguacu.

Figure 2: Major rivers basins in Brazil.(Source: [11])

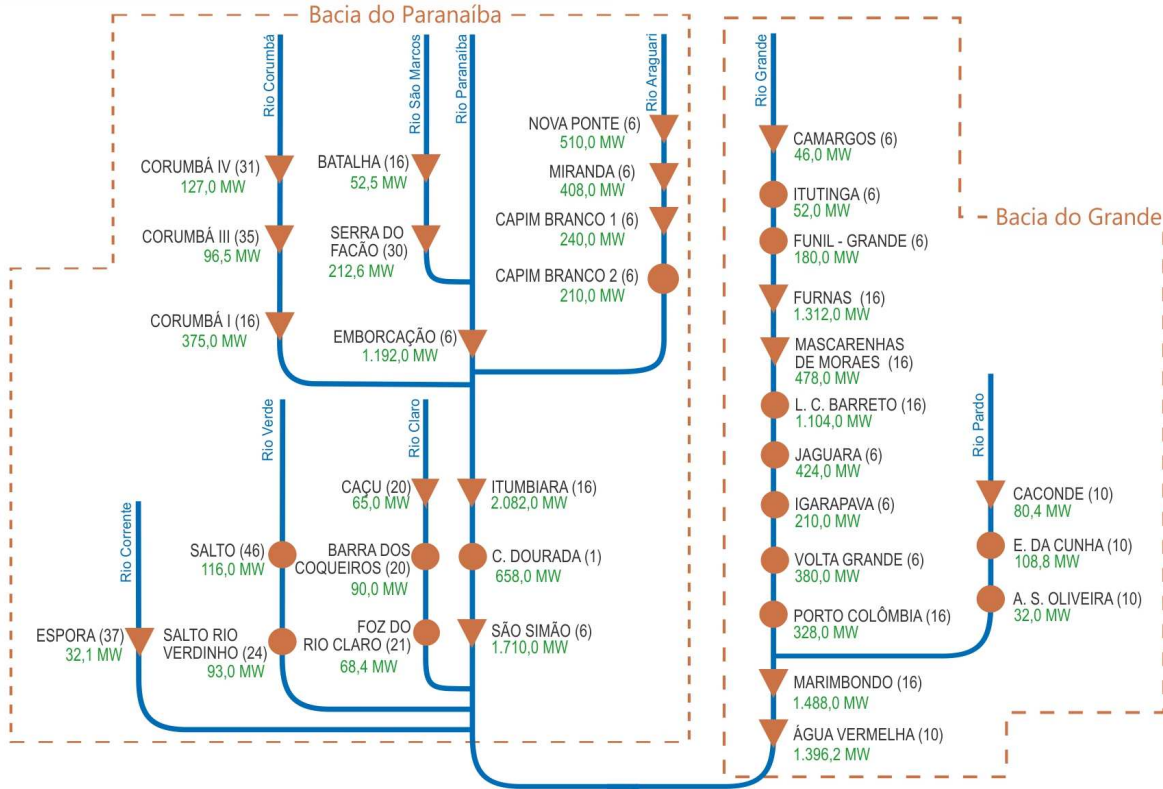


236 The historical data available is the natural inflow¹ for each generator, on a monthly basis,
237 starting in January 1931 and ending in December 2015, measured in cubic meters per second
238 [m^3/s]. For generators built after 1931, the National Electric System Operator performs a back-
239 ward forecasting in order to standardize the records for the hydrothermal dispatch optimization
240 process.

241 In the major rivers there are around 164 hydroelectric power plants currently in operation
242 [34], and these plants operate in a cascade scheme, see in Figure 3 this cascade scheme for
243 Paranaíba and Grande basins with 19 generators with reservoirs, represented by triangles, and
244 15 generators with no reservoir (circles). This way decisions taken at the upstream reservoirs
245 will impact the inflow of the downstream reservoirs.

¹The natural inflow is the average incoming water per unit of time at each generator's reservoir from affluent rivers, lakes and its own drainage area.

Figure 3: Example of a cascade scheme.(Source: Adapted from [1])



246 Since there is a cascade scheme, the natural inflow of each generator has to be calculated
 247 based on the concept of incremental inflows. For exemplifications reason, assume that Camargos
 248 is Generator number 1 and Itutinga is Generator number 2 in Figure 3. If Generator 1 comes
 249 first in the cascade, the incremental inflow will be equal to its natural inflow. But, if Generator
 250 2 has 1 upstream, so its incremental inflow will be given by the difference between its natural
 251 inflow and the natural inflow of Generator 1. The generators will be grouped by basin creating
 252 an equivalent generator with natural inflow equal to the sum of the incremental inflows of all
 253 reservoirs belonging to the basin (Figure 4). It is of note that all natural inflow data analyzed
 254 in the following sections have been adjusted accordingly considering the cascade scheme.

255 4.2 Climate Variables

256 The climate variables were selected trough a literature search. The selected variables are related
 257 to El Niño and the Sunspots numbers; the variables representing El Niño/La Niña phenomenon
 258 are: Southern Oscillation Index (SOI), Equatorial SOI, Niño variations and Oceanic Niño Index
 259 (ONI).

260 The SOI is calculated based on the difference between the atmospheric pressure at sea level in
 261 the regions of Tahiti (in the Western Pacific) and Darwin (Australia, Western Pacific) [35]. The
 262 Equatorial SOI measures the average difference of atmospheric pressure at sea level between two
 263 regions centered on the equator: Indonesia and East Pacific. The range to indicate the presence



Figure 4: The natural inflow series in Brazil.

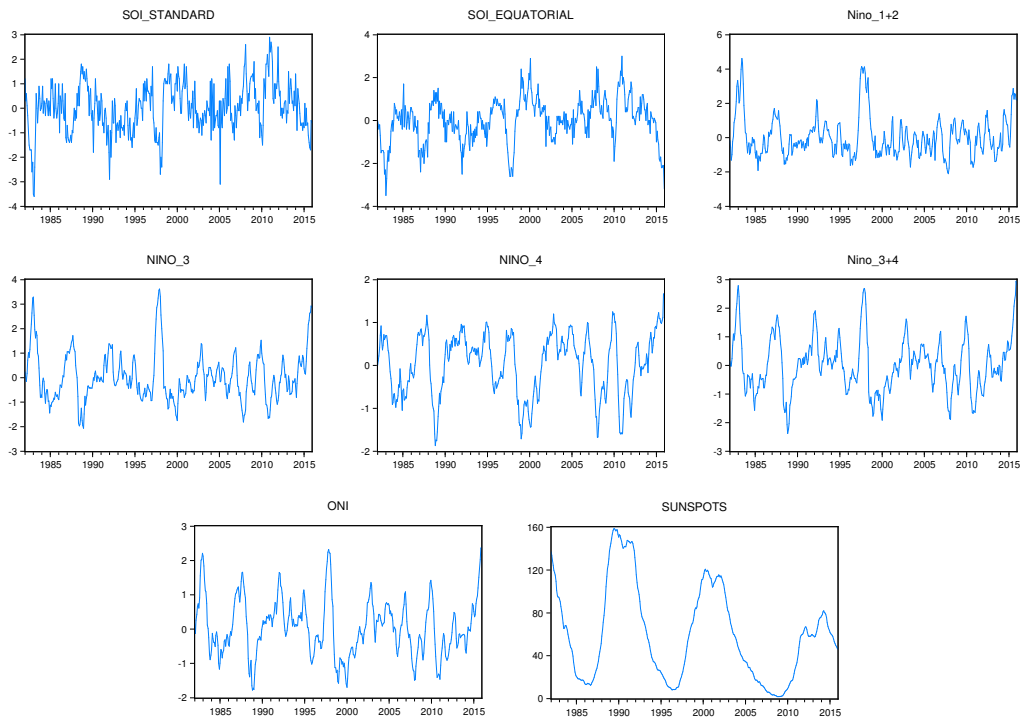


Figure 5: The climate variables in Brazil.

264 or absence of El Niño/La Niña is the same for both the SOI index and Equatorial SOI. It is also
 265 of note that the influence of El Niño in North America indicate a nearly 30 year long cycle due
 266 to the different geographical zone [36]. Consecutive periods of negative figures indicate El Niño
 267 phenomenon occurrence; meanwhile consecutive positive figures denote the presence of La Niña
 268 and values close to zero indicate a normal situation, where none of the two phenomena occur.
 269 The official historical monthly series of these indices are provided by the National Oceanic and
 270 Atmospheric Administration (NOAA).

271 The Sea Surface Temperature (SST) is the water temperature close to the ocean’s surface.
 272 The SST anomaly, that is, the temperature variation by month, is a proxy for El Niño and La
 273 Niña. Thus, this index is used to classify and quantify such phenomena in four Niño regions:
 274 Niño 1+2, Niño 3, Niño 4 and Niño 3.4, defined as follows by NOAA in 2014. Through the
 275 location of the Niño regions it is possible to conclude that regions Niño 1+2 and Niño 3 better
 276 identify temperature anomalies for the Eastern Pacific Ocean sea surface and region Niño 4 for
 277 the Western Pacific. The Niño 3.4 region is centralized in the Pacific, which allows a better
 278 understanding of anomalies across it. Therefore, currently the Niño 3.4 region is the official
 279 measure used to represent SST. However, depending on the study, other regions may be a better
 280 alternative. The threshold for the normal state of this index is between $-0.5^{\circ}C$ and $+0.5^{\circ}C$. The
 281 criteria commonly used to define an El Niño phenomenon consists of five consecutive averages
 282 of SST anomalies above $+0.5^{\circ}C$. Similarly, for La Niña, this criterion remains, but now the

283 SST anomaly should be below $-0.5^{\circ}C$. The monthly time series for all regions are provided by
284 NOAA.

285 The ONI measures the average sea surface temperature anomalies for the region Niño 3.4,
286 removing the existing warming trend on it. According to the NOAA website, multiple centered
287 30-year based periods are adopted for obtaining ONI values of five successive years. For instance,
288 the 1956-1960 ONI values are based on the 1941-1970 period, while 1936-1965 base period
289 produces the 1950-1955 ONI values. The El Niño and La Niña are indicated in the same
290 manner as the SST index, the time series is monthly and is provided by NOAA.

291 Sunspots comprehend solar surface regions of high magnetic field, which have considerably
292 lower temperature than its surroundings and thus appear as a dark area. The magnetic flux
293 amount on the sun surface varies over eleven year periods, known as sunspot and solar cycles.
294 During this cycle there is a minimum and a maximum magnetic flux, which is not only difficult
295 to identify the sunspots and but also they appear almost all the time. The cycle reaches its
296 maximum approximately every eleven years, therefore the observed cycle duration corresponds
297 to eleven years.

298 The number of sunspots calculation is accomplished with the Relative Index American num-
299 ber of sunspots. This index indicates the solar phenomenon occurrence taking into account their
300 relationship with the Earth, including geomagnetic variations and ionosphere effects. The Solar
301 Division from American Association of Variable Star Observers coordinates the data collection
302 program and the analysis of this phenomenon. Thus, the National Geophysical Data Center
303 (NGDC), provides the historical data from the number of sunspots per month since 1749.

304 Considering the availability of all series and since the SST is only available after 1982, the
305 data used for this paper are at monthly frequency from **January 1982 to December 2015**.
306 A brief summary table is listed below in Table 1 and the descriptive statistics can be found in
307 Table 2.

Table 1: Summary of tested series.

	Abbreviation	Variable
Natural Inflow Series	AMZ	Amazon
	EAT	East Atlantic
	GRA	Grande
	IGU	Iguacu
	P1	Paranaiba
	P2	Parapanema
	P3	Parana
	P4	Paraguay
	P5	Parnaiba
	SAT	South Atlantic
	SEAT	Southeast Atlantic
	SF	Sao Francisco
	TIE	Tiete
	TOC	Tocantins
URU	Uruguay	
Climate Variables	SOI _{St}	Southern Oscillation Index Standard
	SOI _{Eq}	The Equatorial Southern Oscillation Index
	NN12	Sea Surface Temperature of Niño 1+2 Region
	NN3	Sea Surface Temperature of Niño 3 Region
	NN4	Sea Surface Temperature of Niño 4 Region
	NN34	Sea Surface Temperature of Niño 3.4 Region
	ONI	The Oceanic Niño Index
	SS	Sunspots Number

Table 2: Descriptive statistics of data.

Series	Mean	Median	Maximum	Minimum	Std. Dev.	Skewness	Kurtosis	Obs
AMZ	33553.30	27753.5	95088	5600	21564.07	0.47	1.97	408
EAT	622.75	356	6690	45	746.51	3.67	22.33	408
GRA	2155.66	1721	7938	364	1323.23	1.36	4.79	408
IGU	1802.17	1414	11670	206	1368.56	2.12	10.77	408
P4	425.12	352	1124	212	178.78	1.07	3.42	408
P3	2587.07	2309	8911	130	1368.81	1.35	5.69	408
P1	3094.21	2496	11025	705	1796.74	1.29	4.53	408
P2	2857.46	2504	16004	699	1521.65	3.12	20.81	408
P5	438.33	366.5	1668	178	232.30	1.70	6.71	408
SF	3152.20	2229	15360	406	2481.43	1.55	5.49	408
SAT	868.39	689.5	4524	110	616.66	1.64	7.01	408
SEAT	1535.49	1226	6862	319	975.27	1.56	5.97	408
TIE	1787.83	1549.5	5519	548	849.09	1.64	6.03	408
TOC	13064.57	8229	45317	1772	10943.19	0.93	2.81	408
URU	1939.78	1472	11834	262	1472.04	2.00	9.32	408
SOI _{St}	0.03	0	2.9	-3.6	1.01	-0.21	3.60	408
SOI _{Eq}	0.02	0.1	3	-3.5	1.03	-0.37	3.51	408
NN12	0.08	-0.18	4.62	-2.1	1.21	1.32	5.00	408
NN3	0.05	-0.13	3.62	-2.07	0.99	0.95	4.46	408
NN34	0.03	0.005	2.95	-2.38	0.97	0.41	3.25	408
NN4	0.04	0.19	1.67	-1.87	0.72	-0.50	2.49	408
ONI	0.05	-0.01	2.37	-1.78	0.84	0.36	3.00	408
SS	60.12	56.6	158.5	1.7	44.31	0.57	2.24	408

308 4.3 Correlation Analysis

309 Prior to the comparison of causality analyses by different methods, the correlation analysis
310 results are here summarized in Table 3 and Table 4. Note that the results are Pearson correlation
311 coefficients respectively considering the empirical status of Pearson approach and ** indicates

312 significance at the 0.01 level whilst * reflects the 0.05 level.

313 As can be seen in Table 3, the correlation between natural inflow and climate variables are
 314 overwhelmingly weak, except a few weak correlations detected among NN12, NN3, P2, P3, SAT
 315 and URU. The correlations between the climate series are also evaluated and summarized in
 316 Table 4. Similar conclusions are obtained as expected following the results in [12]: SOI indices
 317 hold negative correlation with the others, whilst the El Niño and ONI series indicate strong
 318 positive values.

Table 3: Correlation between natural inflow and climate variables.

	AMZ	EAT	GRA	IGU	P1	P2	P3	P4	P5	SAT	SEAT	SF	TIE	TOC	URU
SOI_{St}	0.06	-0.06	0.00	-0.07	-0.03	-0.09	-0.09	0.09	0.09	-0.13**	0.01	-0.03	0.01	0.06	-0.11*
SOI_{Eq}	0.03	-0.04	-0.05	-0.22**	-0.05	-0.24**	-0.23**	0.06	0.09	-0.27**	0.01	-0.04	-0.06	0.05	-0.26**
NN12	-0.03	-0.05	0.08	0.37**	0.04	0.36**	0.38**	-0.03	-0.13**	0.33**	-0.03	0.02	0.14**	-0.07	0.37**
NN3	-0.06	-0.03	0.03	0.28**	0.04	0.26**	0.31**	-0.08	-0.14**	0.32**	-0.05	-0.01	0.09	-0.09	0.33**
NN4	-0.03	0.02	-0.06	0.11*	0.01	0.07	0.19**	-0.05	-0.08	0.16**	-0.07	-0.02	-0.02	-0.03	0.16**
NN34	-0.06	0.01	0.01	0.20**	0.03	0.18**	0.25**	-0.08	-0.12*	0.26**	-0.05	-0.01	0.05	-0.07	0.25**
ONI	-0.07	0.01	0.01	0.19**	0.04	0.18**	0.23**	-0.09	-0.11*	0.28**	-0.04	0.01	0.04	-0.07	0.27**
SS	0.00	0.10*	-0.02	0.04	0.02	0.06	0.02	0.05	0.00	0.01	-0.06	0.03	0.05	0.05	0.00

Table 4: Correlation between climate variables.

	SOI _{St}	SOI _{Eq}	NN12	NN3	NN4	NN34	ONI	SS
SOI_{St}	1.00							
SOI_{Eq}	0.80**	1.00						
NN12	-0.47**	-0.65**	1.00					
NN3	-0.67**	-0.83**	0.82**	1.00				
NN4	-0.69**	-0.75**	0.41**	0.73**	1.00			
NN34	-0.75**	-0.85**	0.64**	0.94**	0.88**	1.00		
ONI	-0.74**	-0.85**	0.63**	0.92**	0.88**	0.99**	1.00	
SS	-0.02	-0.04	-0.02	-0.02	-0.03	-0.03	-0.03	1.00

319 5 Causality Analyses Comparison

320 The causality detections between natural inflow and climate variables in Brazil are here evaluated
 321 and compared by implying different causality detection methods summarized in section 2. It is
 322 of note that all the results were obtained using R.

323 5.1 Time Domain Granger Causality Test

324 Given the significant and empirical role of GC causality test, the GC test results are firstly con-
 325 ducted and summarized as follows in Table 5. It is of note that the preconditions of time domain
 326 GC test are satisfied for all tests across various combinations of variables and the corresponding
 327 optimal lag is determined respectively by a group of information criteria. Specifically, the re-
 328 sults that are highlighted in red represent that the valid evidence is obtained for unidirectional
 329 causality from corresponding climate variable to the natural inflow. Note that these valid cases
 330 have no conflicts of causality for the reverse direction and all shows significance level less than
 331 10%.

332 It is observed that the GC test shows relatively promising performance for NN12 and ONI
 333 across climate variables, for AMZ, URU and SAT among all natural inflow series. However,

334 there are general misleading results of the reverse direction and many cases of mutual directional
335 causality with high significant levels.

336 **5.2 Frequency Domain Causality Test**

337 The frequency domain test extends the GC test and further investigates into the causal links
338 by each particular frequency. Note that the preconditions are stratified and the optimal lag-
339 structures are maintained for all tests. As can be seen in Table 6, the valid cases are again
340 highlighted in red, which indicates unidirectional causality from climate variable to natural
341 inflow without the evidence of causality for the other direction.

342 In general, NN34 and ONI obtain overall valid evidences of unidirectional causality without
343 misleading results, whilst only AMZ out of all the natural inflow series shows identical valid
344 results with NN12, NN4 and NN34. Even P2, P3, IGU, URU and SAT indicate a few valid
345 unidirectional causality cases, however, it is not consistent and solid enough considering the
346 amount of misleading results showing causality in reverse direction or mutual direction.

Table 5: Time domain GC test results.

	SOLSt	SOLSt	SOLSt	NN12	NN3	NN4	NN34	ONI	SS
	lag	resault	lag	lag	lag	lag	lag	lag	lag
	resault	resault	resault	resault	resault	resault	resault	resault	resault
GRA	→		→	→	→	→	→	→	→
	←		←	←	←	←	←	←	←
P1	→		→	→	→	→	→	→	→
	←	6 YES*	←	←	←	←	←	←	←
P2	→	4 YES*	→	→	→	→	→	→	→
	←		←	←	←	←	←	←	←
P3	→	10 YES*	→	→	→	→	→	→	→
	←		←	←	←	←	←	←	←
SEAT	→	6 YES*	→	→	→	→	→	→	→
	←		←	←	←	←	←	←	←
SF	→		→	→	→	→	→	→	→
	←		←	←	←	←	←	←	←
P4	→		→	→	→	→	→	→	→
	←	12 YES*	←	←	←	←	←	←	←
AMZ	→	12 YES**	→	→	→	→	→	→	→
	←		←	←	←	←	←	←	←
TOC	→	12 YES*	→	→	→	→	→	→	→
	←	6 YES***	←	←	←	←	←	←	←
EAT	→		→	→	→	→	→	→	→
	←		←	←	←	←	←	←	←
TIE	→	12 YES*	→	→	→	→	→	→	→
	←		←	←	←	←	←	←	←
IGU	→	11 YES**	→	→	→	→	→	→	→
	←		←	←	←	←	←	←	←
URU	→	11 YES**	→	→	→	→	→	→	→
	←		←	←	←	←	←	←	←
SAT	→	11 YES**	→	→	→	→	→	→	→
	←		←	←	←	←	←	←	←
P5	→		→	→	→	→	→	→	→
	←		←	←	←	←	←	←	←

←: causality from climate variable to natural inflow.

→: causality from natural inflow to climate variable.

empty cell: no causality detected.

*, **, ***: indicate 10%, 5% and 1% significant level respectively.

Table 6: Frequency domain causality test results.

	SOI.St freq	Y/N	SOI.Eq freq	Y/N	NNI2 freq	Y/N	NN3 freq	Y/N	NN4 freq	Y/N	NN34 freq	Y/N	ONI freq	Y/N	SS freq	Y/N
GRA	→ ←		→ ←		→ ←		→ ←		→ ←		→ ←		→ ←		→ ←	
P1	→ ←		→ ←		→ ←		→ ←		→ ←		→ ←		→ ←		→ ←	
P2	→ ←	Y	→ ←		→ ←	Y	→ ←	>10.5	→ ←	all	→ ←	Y	→ ←	>12.6	→ ←	2-2.5
P3	→ ←		→ ←		→ ←	Y	→ ←	>7.9	→ ←	>12.6 >15.7	→ ←	Y	→ ←	>9	→ ←	
SEAT	→ ←	Y	→ ←		→ ←		→ ←		→ ←		→ ←		→ ←		→ ←	
SF	→ ←		→ ←		→ ←		→ ←		→ ←		→ ←		→ ←		→ ←	
P4	→ ←		→ ←		→ ←		→ ←		→ ←		→ ←		→ ←		→ ←	
AMZ	→ ←		→ ←		→ ←		→ ←		→ ←	3.1-6.3	→ ←	Y	→ ←		→ ←	
TOC	→ ←	Y	→ ←		→ ←	Y	→ ←		→ ←		→ ←		→ ←		→ ←	
EAT	→ ←	Y	→ ←		→ ←	Y	→ ←		→ ←	4.2 4.2-12.6	→ ←		→ ←		→ ←	
TIE	→ ←		→ ←	Y	→ ←		→ ←		→ ←		→ ←		→ ←		→ ←	2-2.5
IGU	→ ←		→ ←		→ ←	Y	→ ←	4.2-6.3 >7.9	→ ←	>12.6	→ ←	Y	→ ←	>10.5	→ ←	2-3.1
URU	→ ←		→ ←		→ ←	Y	→ ←	>6.3 >7.9	→ ←	>31.4	→ ←	Y	→ ←	>7.9	→ ←	2-5.2
SAT	→ ←		→ ←		→ ←	Y	→ ←	>12.6 >7.9	→ ←	>62.8	→ ←	Y	→ ←	>7.9	→ ←	
P5	→ ←	all Y	→ ←	all Y	→ ←	2.6-4.5	→ ←		→ ←	>3.1	→ ←	Y	→ ←		→ ←	

←: causality from climate variable to natural inflow.

→: causality from natural inflow to climate variable.

empty cell: no causality detected.

*, **, ***: indicate 10%, 5% and 1% significant level respectively.

347 5.3 CCM

348 The CCM causality test has the significant advantage of no prior linear model assumptions are
349 made and this technique is designed for better understanding of causal relationships in complex
350 dynamical system. The results of CCM tests are briefly summarized in Table 7, Table 8 and
351 Table 9 and organized by each pair of tested variables. Moreover, the time lag has been involved
352 to the evaluation, where lag 1 to 6 are considered covering 6 months of lag effect. It is of note
353 that all test results are obtained by the optimal embedding dimension respectively, which is
354 determined by the nearest neighbor forecasting performance using simplex projection and leave-
355 one-out cross validation is applied for the best choice on library size with optimal performance.

356 The results overwhelmingly indicate causality from natural inflow to climate variable², whilst
357 only 18% of the pairs get positive evidence on unidirectional causality from climate variable to
358 natural inflow. However, even among those 18% pairs, there are misleading results of no clear
359 causality detected for specific time lag options. In general, SAT and SF along with NN3 and
360 NN4 obtain relatively more positive results.

²This is possibly due to the oversensitivity of CCM on noise, however, it is of note that the cross mapping skills of both directions are significantly high, indicating the strong link between natural inflow and climate variables.

Table 7: CCM causality test results (1).

	SOI.St		SOI.Eq		NN12		NN3		NN4		NN34		ONI		SS			
	Lag	dir.	Lag	dir.	Lag	dir.	Lag	dir.	Lag	dir.	Lag	dir.	Lag	dir.	Lag	dir.		
GRA	1	5	→	1	6	→	1	2	→	1	2	→	1	2	→	1	2	→
	2	7	→	2	6	→	2	2	→	2	2	→	2	2	→	2	2	→
	3	5	→	3	5	→	3	2	→	3	2	→	3	2	→	3	2	→
	4	5	→	4	7	→	4	2	→	4	2	→	4	2	→	4	2	→
	5	4	→	5	6	→	5	2	NO	5	2	NO	5	2	NO	5	2	→
	6	8	→	6	5	→	6	8	→	6	2	←	6	2	NO	6	2	→
P1	1	5	→	1	6	→	1	2	→	1	2	→	1	2	→	1	2	→
	2	7	→	2	6	→	2	2	→	2	2	→	2	2	→	2	2	→
	3	5	→	3	5	→	3	3	→	3	2	→	3	2	→	3	2	→
	4	5	→	4	7	→	4	2	←	4	2	→	4	2	→	4	2	→
	5	4	→	5	6	→	5	2	NO	5	2	→	5	2	→	5	2	→
	6	8	→	6	5	→	6	8	→	6	2	→	6	2	→	6	2	→
P2	1	5	→	1	6	→	1	2	→	1	2	NO	1	2	→	1	2	→
	2	7	→	2	6	→	2	2	→	2	2	→	2	2	→	2	2	→
	3	5	→	3	5	→	3	3	→	3	2	→	3	2	→	3	2	→
	4	5	→	4	7	→	4	2	NO	4	2	→	4	2	→	4	2	→
	5	4	→	5	6	→	5	2	NO	5	2	→	5	2	→	5	2	→
	6	3	→	6	5	NO	6	8	→	6	2	→	6	2	→	6	2	→
P3	1	5	→	1	6	→	1	2	→	1	2	→	1	2	→	1	2	→
	2	7	→	2	6	NO	2	2	→	2	2	→	2	2	→	2	2	→
	3	5	→	3	5	→	3	3	→	3	2	→	3	2	→	3	2	→
	4	5	→	4	7	→	4	2	→	4	2	→	4	2	→	4	2	→
	5	4	→	5	6	→	5	2	→	5	2	→	5	2	→	5	2	→
	6	8	→	6	5	→	6	8	→	6	2	→	6	2	→	6	2	→
SEAT	1	5	→	1	6	→	1	2	→	1	2	→	1	2	→	1	2	→
	2	7	→	2	6	→	2	2	→	2	2	→	2	2	→	2	2	→
	3	5	→	3	5	→	3	3	→	3	2	→	3	2	→	3	2	→
	4	5	→	4	7	→	4	2	→	4	2	→	4	2	→	4	2	→
	5	4	→	5	6	→	5	2	→	5	2	→	5	2	→	5	2	→
	6	8	→	6	5	→	6	8	→	6	2	→	6	2	→	6	2	→

←: causality from climate variable to natural inflow.

→: causality from natural inflow to climate variable.

NO: no causality detected.

Table 8: CCM causality test results (2).

	SOI_St		SOI_Eq		NN12		NN3		NN4		NN34		ONI		SS	
	Lag	dir.	Lag	dir.	Lag	dir.	Lag	dir.	Lag	dir.	Lag	dir.	Lag	dir.	Lag	dir.
SF	1	5 →	1	6 →	1	2 →	1	2 →	1	2 →	1	2 →	1	2 →	1	2 →
	2	7 →	2	6 →	2	2 →	2	2 →	2	2 →	2	3 →	2	2 →	2	2 →
	3	5 →	3	5 →	3	3 →	3	2 →	3	2 →	3	2 NO	3	3 →	3	2 →
	4	5 →	4	7 →	4	2 →	4	2 →	4	2 →	4	2 NO	4	2 NO	4	2 →
	5	4 →	5	6 →	5	2 →	5	2 →	5	2 →	5	2 NO	5	3 →	5	2 →
	6	8 →	6	5 →	6	8 →	6	2 →	6	2 →	6	2 NO	6	2 →	6	2 →
P4	1	5 →	1	6 →	1	2 →	1	2 →	1	2 →	1	2 →	1	2 →	1	2 →
	2	7 →	2	6 →	2	2 →	2	2 →	2	2 →	2	3 →	2	2 →	2	2 →
	3	5 →	3	5 →	3	3 →	3	2 →	3	2 →	3	2 →	3	3 →	3	2 →
	4	5 →	4	7 →	4	2 →	4	2 →	4	2 →	4	2 →	4	2 →	4	2 →
	5	4 →	5	6 →	5	2 →	5	2 →	5	2 →	5	2 →	5	3 →	5	2 →
	6	8 →	6	5 →	6	8 →	6	2 →	6	2 →	6	2 →	6	2 →	6	2 →
AMZ	1	5 →	1	6 →	1	2 →	1	2 →	1	2 →	1	2 →	1	2 →	1	2 →
	2	7 →	2	6 →	2	2 →	2	2 →	2	2 →	2	3 →	2	2 →	2	2 →
	3	5 →	3	5 →	3	3 →	3	2 →	3	2 →	3	2 →	3	3 →	3	2 →
	4	5 →	4	7 →	4	2 →	4	2 →	4	2 →	4	2 →	4	2 →	4	2 →
	5	4 →	5	6 →	5	2 →	5	2 →	5	2 →	5	2 →	5	3 →	5	2 →
	6	8 →	6	5 →	6	8 →	6	2 →	6	2 →	6	2 →	6	2 →	6	2 →
TOC	1	5 →	1	6 →	1	2 →	1	2 →	1	2 →	1	2 →	1	2 →	1	2 →
	2	7 →	2	6 →	2	2 →	2	2 →	2	2 →	2	3 →	2	2 →	2	2 →
	3	5 →	3	5 →	3	3 →	3	2 →	3	2 →	3	2 →	3	3 →	3	2 →
	4	5 →	4	7 →	4	2 →	4	2 →	4	2 →	4	2 →	4	2 →	4	2 →
	5	4 →	5	6 →	5	2 →	5	2 →	5	2 →	5	2 →	5	3 →	5	2 →
	6	8 →	6	5 →	6	8 →	6	2 →	6	2 →	6	2 →	6	2 →	6	2 →
EAT	1	5 →	1	6 →	1	2 NO	1	2 →	1	2 →	1	2 →	1	2 →	1	2 →
	2	7 →	2	6 →	2	2 →	2	2 →	2	2 →	2	3 →	2	2 →	2	2 →
	3	5 →	3	5 →	3	3 →	3	2 →	3	2 →	3	2 →	3	3 →	3	2 →
	4	5 NO	4	7 →	4	2 NO	4	2 →	4	2 →	4	2 →	4	2 →	4	2 →
	5	4 →	5	6 →	5	2 NO	5	2 →	5	2 →	5	2 →	5	3 →	5	2 →
	6	8 →	6	5 →	6	8 →	6	2 NO	6	2 →	6	2 →	6	2 →	6	2 →

←: causality from climate variable to natural inflow.

→: causality from natural inflow to climate variable.

NO: no causality detected.

Table 9: CCM causality test results (3).

	SOISt		SOI_Eq		NN12		NN3		NN4		NN34		ONI		SS			
	Lag	dir.	Lag	dir.	Lag	dir.	Lag	dir.	Lag	dir.	Lag	dir.	Lag	dir.	Lag	dir.		
TIE	1	5	→	1	6	→	1	2	→	1	2	→	1	2	→	1	2	→
	2	7	→	2	6	→	2	2	→	2	2	→	2	2	→	2	2	→
	3	5	→	3	5	→	3	2	→	3	2	→	3	2	→	3	2	→
	4	5	→	4	7	→	4	2	→	4	2	→	4	2	→	4	2	→
	5	4	→	5	6	→	5	2	→	5	2	→	5	2	→	5	2	→
	6	8	→	6	5	→	6	8	→	6	2	→	6	2	→	6	2	→
IGU	1	5	→	1	6	→	1	2	→	1	2	→	1	2	→	1	2	→
	2	7	→	2	6	→	2	2	NO	2	2	→	2	2	→	2	2	→
	3	5	→	3	5	→	3	3	→	3	2	→	3	2	→	3	3	→
	4	5	→	4	7	→	4	2	NO	4	2	→	4	2	→	4	2	→
	5	4	→	5	6	→	5	2	→	5	2	→	5	2	→	5	3	→
	6	8	→	6	5	→	6	8	→	6	2	→	6	2	→	6	2	→
URU	1	5	→	1	6	→	1	2	→	1	2	→	1	2	→	1	2	→
	2	7	→	2	6	→	2	2	→	2	2	→	2	2	→	2	2	→
	3	5	→	3	5	→	3	3	→	3	2	→	3	2	→	3	3	→
	4	5	→	4	7	→	4	2	→	4	2	NO	4	2	→	4	2	→
	5	4	→	5	6	→	5	2	→	5	2	→	5	2	→	5	3	→
	6	8	→	6	5	→	6	8	→	6	2	→	6	2	→	6	2	→
SAT	1	5	→	1	6	→	1	2	→	1	2	→	1	2	→	1	2	→
	2	7	→	2	6	→	2	2	→	2	2	→	2	2	→	2	2	→
	3	5	→	3	5	→	3	3	→	3	2	→	3	2	→	3	3	→
	4	5	→	4	7	→	4	2	→	4	2	→	4	2	→	4	2	→
	5	4	→	5	6	→	5	2	→	5	2	→	5	2	NO	5	3	→
	6	8	→	6	5	→	6	8	→	6	2	→	6	2	→	6	2	→
P5	1	5	→	1	6	→	1	2	→	1	2	→	1	2	→	1	2	→
	2	7	→	2	6	→	2	2	→	2	2	→	2	2	→	2	2	→
	3	5	→	3	5	→	3	3	→	3	2	→	3	2	→	3	3	→
	4	5	→	4	7	→	4	2	→	4	2	→	4	2	→	4	2	→
	5	4	→	5	6	→	5	2	→	5	2	→	5	2	→	5	3	→
	6	8	→	6	5	→	6	8	→	6	2	→	6	2	→	6	2	→

←: causality from climate variable to natural inflow.

→: causality from natural inflow to climate variable.

NO: no causality detected.

361 5.4 SSA based Causality Test

362 Follow the brief introduction of SSA based causality test in section 3, the test results of natural
363 inflow and climate variable in Brazil are summarized in Table 10³. It is of note that both recur-
364 rent and vector forecasting algorithms are evaluated respectively; the out-of-sample is defined as
365 the last 1/3 of the total observation for both SSA and MSSA forecastings; the root mean square
366 error (RMSE) of forecasting for SSA and MSSA are the optimal outcome obtained respectively
367 with the optimal window length (L) and numbers of eigenvalues (r) that are also listed in the
368 table; causality is detected if the corresponding F statistics is smaller than 1 and the significant
369 level of causality increases while the value of F statistics decreases.

370 In general, the results are again a mixture of different unidirectional causality, mutual direc-
371 tional causality and no causality, and no significant pattern can be identified, except that NN34
372 and NN4 work slightly better among all series. Moreover, the F statistics are very close to 1,
373 which indicate that the forecasting of MSSA by involving the other variable is improved by a
374 very limited amount comparing to the performance of univariate SSA.

³It is of note that the listed pairs are part of all combinations that cover almost all tested series and types of results. The complete details of these results are available upon request from the authors.

Table 10: SSA based causality test results.

	SSA				MSSA				SSA Causality			
	Rec		Vec		Rec		Vec		Rec	Vec	Direction	Decision
	L,r	RMSE	L,r	RMSE	L,r	RMSE	L,r	RMSE	F stat	F stat		
ONI	3,2	0.21	16,14	0.20	9,6	0.20	3,2	0.21	0.92	1.09	URU caus ONI	Wrong
URU	12,3	1349.63	3,1	1357.71	2,1	1426.68	2,1	1426.68	1.06	1.05	ONI caus URU	
NN34	3,2	0.28	3,2	0.28	3,2	0.28	3,2	0.29	1.02	1.02	SAT caus NN34	No
SAT	20,5	541.68	20,10	544.38	11,3	556.97	12,3	559.69	1.03	1.03	NN34 caus SAT	
SS	16,14	1.00	16,11	1.22	8,6	1.19	8,6	1.47	1.20	1.21	P4 caus SS	Yes
P4	19,4	100.14	17,5	106.68	19,4	85.11	15,4	91.61	0.85	0.86	SS caus P4	
SOISt	2,1	0.74	3,1	0.73	2,1	0.69	2,1	0.69	0.93	0.94	SAT caus SOISt	Wrong
SAT	20,5	541.68	20,10	544.38	11,3	556.97	12,3	559.69	1.03	1.03	SOLSt caus SAT	
NN3	3,2	0.34	3,2	0.35	3,2	0.33	3,2	0.34	0.96	0.96	IGU caus NN3	Wrong
IGU	18,1	1292.42	20,1	1282.83	17,1	1338.35	19,1	1315.74	1.04	1.03	NN3 caus IGU	
NN4	6,3	0.20	17,8	0.21	3,2	0.22	3,2	0.22	1.06	1.07	GRA caus NN4	No
GRA	19,5	829.02	13,3	821.12	19,5	844.93	17,5	854.13	1.02	1.04	NN4 caus GRA	
NN12	3,2	0.57	3,2	0.57	8,5	0.49	9,6	0.48	0.86	0.83	SF caus NN12	Mutual
SF	12,5	1529.22	13,3	1509.66	20,3	1395.22	16,6	1425.42	0.91	0.94	NN12 caus SF	
NN34	3,2	0.28	3,2	0.28	3,2	0.28	3,2	0.29	1.01	1.02	P2 caus NN34	Yes
P2	20,7	1223.53	20,7	1224.24	16,7	1111.85	12,4	1088.23	0.91	0.89	NN34 caus P2	
ONI	3,2	0.21	16,14	0.20	5,4	0.19	5,4	0.20	0.88	1.03	P3 caus ONI	Mutual
P3	20,5	1330.56	5,1	1320.23	20,6	1307.70	15,4	1306.01	0.98	0.99	ONI caus P3	
NN34	3,2	0.28	3,2	0.28	3,2	0.28	3,2	0.29	1.01	1.02	AMZ caus NN34	Yes
AMZ	13,6	5052.20	20,12	4762.11	10,7	4904.91	13,9	4678.68	0.97	0.98	NN34 caus AMZ	
NN4	6,3	0.20	17,8	0.21	3,2	0.22	3,2	0.22	1.06	1.06	SAT caus NN4	No
SAT	20,5	541.68	20,10	544.38	11,3	556.97	12,3	559.69	1.03	1.03	NN4 caus SAT	
NN3	3,2	0.34	3,2	0.35	3,2	0.33	3,2	0.34	0.96	0.97	P1 caus NN3	Wrong
P1	20,5	1049.30	13,3	1029.57	20,5	1047.73	20,5	1053.18	1.00	1.02	NN3 caus P1	
SOIEq	2,1	0.52	2,1	0.52	7,3	0.54	7,3	0.55	1.05	1.05	URU caus SOIEq	No
URU	12,3	1349.63	3,1	1357.71	2,1	1426.68	17,10	1385.08	1.06	1.02	SOIEq caus URU	
NN4	6,3	0.20	17,8	0.21	3,2	0.22	8,4	0.23	1.10	1.12	TOC caus NN4	Yes
TOC	11,5	5595.27	20,12	5229.04	15,10	4538.45	15,10	4358.17	0.81	0.83	NN4 caus TOC	
NN12	3,2	0.57	3,2	0.57	10,6	0.49	10,6	0.50	0.86	0.87	TIE caus NN12	Wrong
TIE	16,4	642.53	16,5	633.07	12,6	729.93	19,14	702.37	1.14	1.11	NN12 caus TIE	
SOISt	2,1	0.74	3,1	0.73	8,3	0.69	8,3	0.68	0.93	0.93	IGU caus SOISt	Wrong
IGU	18,1	1292.42	20,1	1282.83	17,1	1338.35	18,1	1315.74	1.04	1.03	SOLSt caus IGU	
ONI	3,2	0.21	16,14	0.20	8,6	0.17	8,6	0.18	0.82	0.92	SF caus ONI	Mutual
SF	12,5	1529.22	13,3	1509.66	20,3	1395.22	16,6	1425.42	0.91	0.94	ONI caus SF	
NN12	3,2	0.57	3,2	0.57	14,9	0.48	14,9	0.48	0.83	0.84	P3 caus NN12	Mutual
P3	20,5	1330.56	5,1	1320.23	20,6	1307.70	15,4	1306.01	0.98	0.99	NN12 caus P3	
SOIEq	2,1	0.52	2,1	0.52	5,2	0.56	5,2	0.55	1.07	1.06	P2 caus SOIEq	Yes
P2	20,7	1223.53	20,7	1224.24	16,7	1111.85	12,4	1088.23	0.91	0.89	SOIEq caus P2	

Yes: only causality from climate variable to natural inflow is detected.
 Wrong: only causality from natural inflow to climate variable is detected.
 No: no causality detected.
 Mutual: mutual directional causality between climate variable and natural inflow.

375 5.5 Periodic Autoregressive Model based Causality Test

376 The PAR causality test results are summarized in Table 11⁴ where both PAR and PARX RMSE
 377 are present when forecasting the last 1/3 of the total observation. Following the procedure
 378 describe in SSA causality test, if the corresponding F statistics is smaller than 1 then there is
 379 causality. When causality is detected in both directions the causality is not computed, and when
 380 the direction of causality is from the natural inflow in the climate variable then is computed as
 381 wrong. The causality is computed in the right decision only in six cases, but the F statistics
 382 is very close to one in most cases, showing that even when causality can be considered, the
 383 improvements of considering a climate variable it's on the edge. The only case where can be
 384 clearly found a causality is between SOI Equatorial and Paranapanema basin.

⁴It is of note that the listed pairs are part of all combinations that cover almost all tested series and types of results. The complete details of these results are available upon request from the authors.

Table 11: PAR based causality test results.

	RMSE		F stat	PAR Causality	
	PAR	PARX		Direction	Decision
GRA	1129.17	1076.456	0.953	NN4T caus GRA	Mutual
NN4	1.128	0.789	0.7	GRA caus NN4	
P2	2056.97	1235.528	0.601	SOLEq caus P2	Yes
SOLEq	1.69	1.754	1.038	No	
P3	1618.171	1369.67	0.846	ONI caus P3	Yes
ONI	1.306	1.955	1.497	No	
P3	1618.171	1587.472	0.981	NN3 caus P3	Yes
NN3	1.56	2.199	1.409	No	
P3	1618.171	1434.836	0.887	NN4 caus P3	Yes
NN4	1.128	1.34	1.188	No	
P3	1618.171	1583.034	0.978	NN34 caus P3	Yes
NN34	1.574	2.143	1.362	No	
AMZ	8519.513	8581.495	1.007	No	Wrong
NN4	1.128	0.91	0.807	AMZ caus NN4	
TOC	6359.823	5489.439	0.863	ONI caus TOC	Mutual
ONI	1.306	1.006	0.77	TOC caus ONI	
EAT	806.574	472.349	0.586	SS caus EAT	Mutual
SS	64.101	24.251	0.378	EAT caus SS	
TIE	827.817	747.594	0.903	NN3 caus TIE	Yes
NN3	1.56	1.624	1.041	No	
IGU	1629.477	1348.128	0.827	ONI caus IGU	Mutual
ONI	1.306	0.847	0.649	IGU caus ONI	
URU	1827.213	1497.444	0.82	SOLSt caus URU	Mutual
SOLSt	1.537	1.152	0.75	URU caus SOLSt	
SAT	712.875	550.122	0.772	NN12 caus SAT	Mutual
NN12	1.606	1.406	0.876	SAT caus NN12T	
P5	193.369	196.162	1.014	No	Wrong
SOLEq	1.69	1.544	0.914	P5 caus SOLEq	

Yes: only causality from climate variable to natural inflow is detected.

Wrong: only causality from natural inflow to climate variable is detected.

No: no causality detected.

Mutual: mutual directional causality between climate variable and natural inflow.

6 Final Discussion and Conclusion

In general, this paper successfully obtains comprehensive investigation of the causality relationship between natural inflow and climate variables in Brazil by analyzing the data of 15 major basins and 8 different climate series. For the first time to the best of our knowledge, it incorporates and compares five different causality detection methods for the causality study on hydrological series. In specific, GC test shows relatively promising performance for AMZ, URU and SAT among all natural inflow series, NN12 and ONI across climate variables; frequency domain causality test indicates generally valid evidences of unidirectional causality for AMZ, NN34 and ONI; CCM overwhelmingly obtains significant unidirectional causality from the opposite direction (natural inflow to climate variables), whilst SAT, SF, NN3 and NN4 relatively give more positive results of the valid direction; SSA based causality test shows that NN34 and NN4 work slightly better, and the forecasting improvements by involving the other variable are generally very limited; PAR based causality test computed six unidirectional causality, but only one is really significant (P2 and SOLEq).

The overall results indicate that there is no single method which stands out and outperforms the others. The conclusions are a mixture of different unidirectional, mutual directional, and no causality. There is no obvious pattern that can be clearly identified across 15 natural inflow series and 8 climate variables. However, it is noticed that the overwhelming evidences of opposite

403 direction of causality are obtained by CCM, which is the most concurrent outcome of all five
404 different tests. It is frankly interesting discovery that is possibly caused by significant noises
405 that generally exist in those series, which will be one of the main focuses for future research.

406 The works presented in the Background section showed improvements when using informa-
407 tion from the climate variables in the inflow prediction procedure, so even if the tests applied
408 here did not present favourable results, a natural continuation of this study will be the appli-
409 cation of different models that incorporate exogenous variables to verify the significance of the
410 climate variables in the prediction of each of the inflow series studied.

411 **References**

- 412 [1] Operador Nacional do Sistema Elétrico (2017). www.ons.com.br.
- 413 [2] Amarasekera, K., Lee, R., Willians, E. & Eltahir, E. (1997) ENSO and natural variability
414 in the flow of tropical rivers. *Journal of Hydrology*, 200, 24–39.
- 415 [3] Dettinger, M. & Diaz, H. (2000). Global characteristics of streamflow seasonality and vari-
416 ability. *Journal of Hydrometeorology*, 1, 289–310.
- 417 [4] Foley, J., Botta, A., Coe, M. & Costa, M. (2002). El Niño-Southern oscillation and the
418 climate, ecosystems and rivers of Amazonia. *Global Biogeochemical Cycles*, 16, 1132.
- 419 [5] Berri, G., Ghietto, M. & García, N. (2002). The Influence of ENSO in the flows of the Upper
420 Paraná River of South America over the Past 100 Years. *Journal of Hydrometeorology*, 3,
421 57–65.
- 422 [6] Garcia, N. & Mechoso, C. (2005). Variability in the discharge of South American rivers and
423 in climate. *Hydrological Sciences Journal*, 3, 459–478.
- 424 [7] Stuck, J., G’untner, A. & Merz, B. (2006). ENSO impact on simulated South American
425 hydro-climatology. *Advances in Geosciences*, 6, 227–236.
- 426 [8] Soares, J., Carriello, F., Ferreira, N. & Rennó, C. (2006). Mapping the hydrologic response
427 of the Brazilian hydrologic regions and their variability associated with El Niño and La
428 Niña. *Revista Ambiente & Água*, 1, 21–36.
- 429 [9] Ausloos, M., & Ivanova, K. (2001). Power-law correlations in the southern-oscillation-index
430 fluctuations characterizing El Niño. *Physical Review E*, 63(4), 047201.
- 431 [10] Cardoso, A. & Silva Dias, P. (2006). The relationship between ENSO and Paraná River
432 flow. *Advances in Geosciences*, 6, 189–193.
- 433 [11] Lima, L., Popova, E. & Damien, P. (2014). Modeling and forecasting of Brazilian reservoir
434 inflows via dynamic linear models. *International Journal of Forecasting*, 30, 464–476.

- 435 [12] Maçaira, P., Cyrino Oliveira, F., Ferreira, P., de Almeida, F. & Souza, R. (2017). Intro-
436 ducing a Causal PAR(p) Model to Evaluate the Influence of Climate Variables in Reservoir
437 Inflows: a Brazilian Case. *Pesquisa Operacional*, 37, 107–128.
- 438 [13] Silveira, C., Alexandre, A., Souza Filho, F., Vasconcelos Junior, F. & Cabral, S. (2017).
439 Monthly streamflow forecast for National Interconnected System (NIS) using Periodic Auto-
440 regressive Endogenous Models (PAR) and Exogenous (PARX) with climate information.
441 *Brazilian Journal of Water Resources*, 22, 1–10.
- 442 [14] Domino, K., Bachowicz, T., & Ciupak, M. (2014). The use of copula functions for predictive
443 analysis of correlations between extreme storm tides. *Physica A: Statistical Mechanics and
444 its Applications*, 413, 489-497.
- 445 [15] Zhang, Q., Xu, C. Y., Yu, Z., Liu, C. L., & Chen, Y. D. (2009). Multifractal analysis
446 of streamflow records of the East River basin (Pearl River), China. *Physica A: Statistical
447 Mechanics and its Applications*, 388(6), 927-934.
- 448 [16] Hurst, H. E. (1951). Long term storage capacity of reservoirs. *ASCE Transactions*, 116(776),
449 770-808.
- 450 [17] Rotundo, G., Ausloos, M., Herteliu, C., & Ileanu, B. (2015). Hurst exponent of very long
451 birth time series in XX century Romania. Social and religious aspects. *Physica A: Statistical
452 Mechanics and its Applications*, 429, 109-117.
- 453 [18] Ausloos, M., Cerqueti, R., & Lupi, C. (2017). Long-range properties and data validity for
454 hydrogeological time series: The case of the Paglia river. *Physica A: Statistical Mechanics
455 and its Applications*, 470, 39-50.
- 456 [19] Granger, C. W. (1969). Investigating causal relations by econometric models and cross-
457 spectral methods. *Econometrica: Journal of the Econometric Society*, 37(3), 424-438.
- 458 [20] Breitung, J. & Candelon, B. (2006). Testing for short- and long-run causality: A frequency-
459 domain approach. *Journal of Econometrics*, 132, 363-378.
- 460 [21] Sugihara, G., May, R., Ye, H., Hsieh, C. H., Deyle, E., Fogarty, M. & Munch, S. (2012).
461 Detecting causality in complex ecosystems. *Science*, 338(6106), 496-500.
- 462 [22] Hassani, H., Huang, X., Gupta, R. & Ghodsi, M. (2016). Does sunspot numbers cause
463 global temperatures? A reconsideration using non-parametric causality tests. *Physica A:
464 Statistical Mechanics and its Applications*, 460, 54–65. Elsevier.
- 465 [23] Terry, L., Pereira, M., Neto, T., Silva, L. & Sales, P. (1986). Coordinating the Energy Gen-
466 eration of the Brazilian National Hydrothermal Electrical Generating System. *Interfaces*,
467 16, 16–38.

- 468 [24] Ursu, E. & Perea, J. (2017). Estimation and identification of periodic autoregressive models
469 with one exogenous variable. *Journal of the Korean Statistical Society*, 46, 629–640.
- 470 [25] Collette, C., & Ausloos, M. (2004). Scaling analysis and evolution equation of the North
471 Atlantic oscillation index fluctuations. *International journal of modern physics C*, 15(10),
472 1353-1366.
- 473 [26] Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time
474 series. *Journal of the American Statistical Association*, 77, 304-324.
- 475 [27] Huang, X., Hassani, H., Ghodsi, M., Mukherjee, Z. & Gupta, R. (2017). Do trend extraction
476 approaches affect causality detection in climate change studies? *Physica A: Statistical
477 Mechanics and its Applications*, 469, 604–624.
- 478 [28] Deyle, E., Fogarty, M., Hsieh, C., Kaufman, L., MacCall, A., Munch, S., Perretti, C., Ye,
479 H. & Sugihara, G. (2013). Predicting climate effects on Pacific sardine. *Proceedings of the
480 National Academy of Sciences*, 110(16), 6430-6435.
- 481 [29] Ye, H., Deyle, E., Gilarranz, L. & Sugihara, G. (2015). Distinguishing time-delayed causal
482 interactions using convergent cross mapping. *Scientific Reports*, 5, 14750.
- 483 [30] Takens, F. (1981). Detecting strange attractors in turbulence *Dynamical Systems and Tur-
484 bulence. Dynamic Systems and Turbulence*, 898, 366-381.
- 485 [31] Sugihara, G. & May, R. (1990). Nonlinear forecasting as a way of distinguishing chaos from
486 measurement error in time series. *Nature*, 344(6268), 734-741.
- 487 [32] Hassani, H., Zhigljavsky, A., Patterson, K. & Soofi, A. (2010). A comprehensive causal-
488 ity test based on the singular spectrum analysis. *Causality in Science*, 379-406. Oxford
489 University Press.
- 490 [33] Maceira, M. & Damázio, J. (2006). Use of the PAR(p) Model in the Stochastic Dual Dy-
491 namic Programming Optimization Scheme Used in the Operation Planning of the Brazilian
492 Hydropower System. *Probability in the Engineering and Informational Sciences*, 20, 143-
493 156.
- 494 [34] Operador Nacional do Sistema Elétrico (2016). Atualização de Séries Históricas de Vazões,
495 período 1931 a 2015.
- 496 [35] Petroni, F., & Ausloos, M. (2008). High frequency intrinsic modes in El Niño/Southern
497 Oscillation Index. *Physica A: Statistical Mechanics and its Applications*, 387(21), 5246-
498 5254.
- 499 [36] Cimino, G., Del Duce, G., Kadonaga, L. K., Rotundo, G., Sisani, A., Stabile, G., ... &
500 Whitaric, M. (1999). Time series analysis of geological data. *Chemical geology*, 161(1-3),
501 253-270.