JOSIELTON DA SILVA SANTOS

IMPACT OF PRECIPITATION EXTREMES ON ENERGY PRODUCTION ACROSS THE SÃO FRANCISCO RIVER BASIN

Dissertation submitted to the Applied Meteorology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

Adviser: Flávio Barbosa Justino

Co-adviser: Jackson Martins Rodrigues

VIÇOSA - MINAS GERAIS
2023

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
S237i
2023

Santos, Josielton da Silva, 1995-
Impact of precipitation extremes on energy production across the São Francisco river basin / Josielton da Silva Santos. – Viçosa, MG, 2023.
1 dissertação eletrônica (73 f.): il. (algumas color.).

Texto em inglês.
Orientador: Flávio Barbosa Justino.
Dissertação (mestrado) - Universidade Federal de Viçosa, Departamento de Engenharia Agrícola, 2023.
Referências bibliográficas: f. 65-73.
DOI: https://doi.org/10.47328/ufvbbt.2023.559
Modo de acesso: World Wide Web.

1. Energia elétrica - Produção - Medição. 2. Energia elétrica - Produção - Aspectos ambientais. 3. Usinas hidrelétricas - São Francisco, Rio, Bacia. 4. Aprendizado do computador. I. Justino, Flávio Barbosa, 1971-. II. Universidade Federal de Viçosa. Departamento de Engenharia Agrícola. Programa de Pós-Graduação em Meteorologia Aplicada. III. Título.

CDD 22. ed. 621.37

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317

JOSIELTON DA SILVA SANTOS

IMPACT OF PRECIPITATION EXTREMES ON ENERGY PRODUCTION ACROSS THE SÃO FRANCISCO RIVER BASIN

Dissertation submitted to the Applied Meteorology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

APPROVED: July 21, 2023.

Assent:

Josielton da Silva Santos
Author

Flávio Barbosa Justino
Adviser

To my wife.

ACKNOWLEDGEMENTS

To God, for without Him, I would be nothing.

To *Nossa Senhora Imaculada Conceição* and Saint Rita of Cascia, for their intercession and protection.

To my wife, Larissa, for giving me much more than I believe I deserve. Even when distance kept us apart, you were always by my side, believing in me and encouraging me to dream and achieve our goals.

To my parents, Thereza and Josué, and my sister, Taiara, for always supporting and encouraging me in all my decisions, no matter what they may be. You celebrated every new step I took and rejoiced in every accomplishment as if they were your own (and they are).

To my mother-in-law, Maria Helena, and my sisters-in-law, Mara and Morgânia, for their constant presence and support.

To my advisor, Flávio Justino, and my co-advisor, Jackson Rodrigues. Your guidance was essential for the development and completion of this work. The journey of my master's degree was filled with learning, often with uncertain and challenging paths. You were like lights and signposts, guiding me in the right direction on this road.

To my classmates, Bruna Rodrigues, Santiago Maria, Juliete Catein, Mierele Castro, Lormido Ernesto, Elaine, Letícia, Thaís and to the other fellow students of the master's program whom I had the pleasure to meet at UFV. You form a support network, always present in every moment and situation.

To my cousins and friends, Glauber, Jerlane, Jhon, João Carlos, Luciana, Paloma, Thalisson, Adelson, Gabriel, Corcino, Bruno, Ianka, and certainly someone whose name I forgot to mention. Sharing this journey with this network was essential in many moments.

To the Federal University of Viçosa, for providing the opportunity to pursue my postgraduate studies and for the infrastructure that enabled the development of this work.

To the professors with whom I had contact at UFV that contributed to carried out the study: Gabrielle Pires, Paulo Hamakawa, Jose Gleriani, Demetrius Silva, Ricardo Amorim, Denis Cunha, and Marcos Heil. Thank you very much.

To Graça Freitas, the secretary of the Graduate Program, who is always willing to assist students and professors.

To Avanisse, my first teacher, who played a fundamental role in my first experiences in the classroom. Those experiences shaped the researcher I am today.

To teachers Renan and Roberto, dear teachers from high school, who will be happy and proud to hear about these results.

To teacher Tatiana Máximo, who guided me through the paths of research and water resources during my undergraduate studies. It was through her that the first opportunities in this field arose, and she remains one of my main references to this day.

To my former colleagues at SEDURBS-SE, who believed in my potential for growth. Especially, Overland Amaral, Ailton Rocha, João Carlos, Ana Paula, Luciano Queiróz, Ingrid, Marta, Kely, and Adnívea. It was in this environment that my knowledge about climate found a path for growth.

This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* – Brasil (CAPES) – Finance Code 001.

Finally, to the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq), for granting me the scholarship during my master's program.

“A wise man should always follow paths treaded by great men and imitate those who have been supreme, so that if his skill does not match theirs, at least he can come close”.
(Maquiavel)

ABSTRACT

SANTOS, Josielton da Silva, M.Sc., Universidade Federal de Viçosa, July, 2023. **Impact of precipitation extremes on energy production across the São Francisco river basin.** Adviser: Flávio Barbosa Justino. Co-adviser: Jackson Martins Rodrigues.

The Brazilian electrical system is predominantly hydrothermal, with hydroelectric power plants (HPP's) dependent on rainfall variability. The São Francisco river basin plays a fundamental role in the country's electricity production with HPP's in the Northeast and Southeast regions. However, climate extremes events have affected the energy production. To manage the use of HPP's to avoid energy shortage during dry periods and the activation of thermal power plants, is a challenge as it increases production costs and may result in water wastage during rainy periods. The main variables influencing operational decisions are Stored Energy (STE) and Affluent Natural Energy (ANE), used to calculate the Marginal Cost of Operation (MCO) and the Settlement Price of Differences (SPD). The current study investigates the relationships between these variables and climate precipitation extremes events in the São Francisco river basin. Spatial distribution and trends of 11 extremes precipitation indices are analyzed. The seasonality, trends, and correlation between the energy variables and the extreme indices are also investigated. Three machine learning algorithms (Random Forest, Artificial Neural Networks, and k-Nearest Neighbors) were applied as regression models to estimate the energy variables (ANE, STE, MCO, and SPD). Correlations between energy variables show the impact of changes in ANE and STE availability in the São Francisco river basin on MCO and SPD, in the Northeast and Southeast/Midwest subsystems. ANE and STE showed downward trends, while MCO and SPD experienced an upward trend. Furthermore, the seasonal behavior throughout the year was demonstrated for STE and ANE, influenced by extreme precipitation rates in different time scales. Trends indicate a reduction in total precipitation (PRCTOT) and the number of wet days (CWD), as well as an increase in the number of dry days in the basin (CDD). Results based on machine learning algorithm indicate that it is reasonable to efficiently estimate ANE and STE using extremes precipitation data. These findings have significant implications for the planning and management of the Brazilian electricity sector, contributing to strategic

decision-making and the formulation of public policies that ensure the country's energy security.

Keywords: Affluent Natural Energy. Stored Energy. Prediction.

RESUMO

SANTOS, Josielton da Silva, M.Sc., Universidade Federal de Viçosa, julho de 2023. **Impactos dos extremos de precipitação na produção de energia ao longo da Bacia do Rio São Francisco**. Orientador: Flávio Barbosa Justino. Coorientador: Jackson Martins Rodrigues.

O sistema elétrico brasileiro é predominantemente hidrotérmico, com usinas hidrelétricas (UHE's) sensíveis à variabilidade das chuvas. A bacia hidrográfica do Rio São Francisco desempenha um papel fundamental na produção de energia elétrica do país, com UHE's no Nordeste e Sudeste. No entanto, eventos climáticos extremos têm prejudicado sua produção energética. Equilibrar o uso das UHE's para evitar escassez de energia durante períodos secos e o acionamento de termelétricas é um desafio, pois aumenta os custos de produção e pode desperdiçar água em períodos chuvosos. As principais variáveis que influenciam as decisões operacionais são a Energia Armazenável (EAR) e a Energia Natural Afluente (ENA), usadas para calcular o Custo Marginal de Operação (CMO) e o Preço de Liquidação das Diferenças (PLD). Neste estudo, analisou-se as relações entre essas variáveis e os extremos climáticos de precipitação na bacia do Rio São Francisco. Séries temporais de 11 índices extremos climáticos de precipitação foram utilizadas para analisar distribuição e tendências. Investigou-se a sazonalidade, tendência e correlação entre as variáveis energéticas e os índices extremos. Foram aplicados três algoritmos de aprendizado de máquina (Floresta Aleatória, Redes Neurais Artificiais e k-vizinhos mais próximos) como modelos regressivos para estimar as variáveis energéticas (ENA, EAR, CMO e PLD). As correlações entre as variáveis energéticas revelam o impacto das variações na disponibilidade de ENA e EAR na Bacia do Rio São Francisco no CMO e PLD nos subsistemas Nordeste e Sudeste/Centro-Oeste. ENA e EAR apresentaram tendências de redução, enquanto CMO e PLD mostraram tendência de aumento. Observou-se comportamento sazonal ao longo do ano para EAR e ENA, influenciados pelos índices extremos de precipitação em diferentes escalas temporais. As tendências indicam redução na precipitação total (PRCTOT) e no número de dias úmidos (CWD), além de aumento no número de dias secos na bacia (CDD). Os resultados mostraram que é possível estimar eficientemente ENA e EAR usando dados de extremos climáticos de precipitação na bacia. Essas descobertas têm implicações significativas para o planejamento e gerenciamento do setor elétrico brasileiro, contribuindo para tomada

de decisões estratégicas e formulação de políticas públicas que garantam a segurança energética do país.

Palavras-chave: Energia Natural Afluente. Energia Armazenada. Predição.

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

ANE – Affluent Natural Energy
ANEEL - National Electric Energy Agency
ANN - Artificial Neural Networks
BES - Brazilian Electric System
BR-DWGD - Brazilian Daily Weather Gridded
CCEE - Board of Electrical Energy Commercialization
CDD - Consecutive Dry Days
CWD - Consecutive Wet Days
DD - Dry Days
ETCCDI - Expert Team on Climate Change Detection and Indices
EWD - Eastward Wave Disturbances
HPP – Hydroelectric Power Plant
IBGE - Brazilian Institute of Geography and Statistics
IPCA - National Consumer Price Index
ITCZ - Intertropical Convergence Zone
KGE - Kling-Gupta Efficiency Coefficient
KNN - k-Nearest Neighbors
LTA - Long-Term Average
MAE - Mean Absolute Error
MAPE - Mean Absolute Percentage Error
MCO - Marginal Cost of Operation
ONS - National Electrical System Operator
PMO - Monthly Operational Plan
PRCDQ - Driest Quartile Precipitation
PRCPTOT - Total Precipitation
PRCWQ - Wettest Quartile Precipitation
R20mm - Number of days with very heavy precipitation
RF - Random Forest
RMSE - Root Mean Squared Error
RX1DAY - Maximum precipitation in one day
RX5DAYS - Maximum precipitation in five consecutive days
SACZ - South Atlantic Convergence Zone

SDII - Simple Daily Intensity Index

SIG - Geographic Information System

SIN - National Interconnected System

SPD - Settlement Price of Differences

STE - Stored Energy

UTCV - Upper Tropospheric Cyclonic Vortices

WD - Wet Days

WMO - World Meteorological Organization

SUMMARY

1. INTRODUCTION

The Brazilian electrical system has a continental extension, and its energy matrix is heterogeneous, predominantly composed by hydroelectric and thermal power sources. The system is largely dominated by hydroelectric power plants (HPP's), which use large reservoirs to regulate seasonal river flows (MENDES, 2019). The operational planning is sensitive to rainfall variability due to its dependence on hydroelectric generation. While water is a low-cost resource, prolonged rainfall scarcity can lead to energy deficits and the need to apply more expensive sources, such as thermal power plants (LUIZ-SILVA *et al.*, 2018).

The centralized operations of HPP reservoirs are carried out by the National Electric System Operator (ONS) within the scope of the Monthly Operational Plan (PMO) of the National Interconnected System (SIN), which enable the maximization of natural flows utilization, reducing water waste, and minimizing costs. These operations aim to achieve an optimized balance between minimum cost and maximum operational security, ensuring full demand supply (MENDES, 2019).

The Brazilian hydroelectric system is characterized by the interconnection between the stages of power generation and transmission (HIDALGO *et al.*, 2020). One prominent watershed is the São Francisco river basin which covers the Northeast and Southeast regions of Brazil. The São Francisco river plays a fundamental role in supplying electrical energy to the Northeast region of the country (CBHSF, 2022). The São Francisco river is crucial in human activities as it borders five Brazilian states. However, there are few studies on the long-term sediment dynamics of the river and its response to climate (MESCOLOTTI *et al.*, 2021).

In the Northeast, Regoto *et al.* (2021) observed a significant reduction in precipitation extremes, resulting in a drier climate and prolonged periods of drought, which is more pronounced during the rainy seasons, intensifying water scarcity-related problems. Although, the water flows can be regulated and an increase in the use of alternative sources is possible, the country has faced a series of crises in the water and energy sectors in recent years (HIDALGO *et al.*, 2020).

During the period from 2012 to 2017, the contribution of HPP's to meet the total electricity demand in the Northeast was, on average, only 31%. Furthermore, in November 2015 and 2017, the reservoirs of the São Francisco river reached the

lowest level since the construction of the dams in 1994, with only 5% of Stored Energy (STE). This situation is directly associated with the occurrence of extreme weather events that result in dry periods in the region (DE JONG *et al.*, 2018).

According to Silveira *et al.* (2016), studies aiming to identify climate variability in the region contribute to the reformulation of water resource management policies and increase the system resilience in the face of climate change challenges. Additionally, Lucas *et al.* (2021) emphasize that the assessment of climate extremes in a specific area is a fundamental tool for water resource management, decision-making in planning, and formulation of environmental protection policies.

2. LITERATURE REVIEW

2.1. Climate extremes events in Brazil

Intense rainfall is gaining prominence in discussions on global climate change. Often, lack of planning and infrastructure contributes to the amplification of their impacts (OLIVEIRA *et al.*, 2020). Brazil faces extreme weather events that result in economic and social losses, and these events are expected to become more frequent and intense throughout the current century. Analyses of historical records reveal an increase in consecutive drought periods and a decrease in consecutive rainfall periods in almost the entire Brazilian territory (AVILA-DIAZ *et al.*, 2020).

Precipitation extremes exhibit heterogeneous pattern in most of the country, without a dominant characteristic across all regions. However, the most affected region is the Northeast, with a tendency towards aridity, and the South, with increased humidity (REGOTO *et al.*, 2021). Analyses of extreme precipitation indices conducted by Silva *et al.* (2022) in a predominantly semi-arid region of Brazil, demonstrated a decrease in rainfall regime and an increase in dry days, especially on an annual scale and during the rainy season. There were also decreasing trends of wet days and increasing dry days, indicating a reduction in total precipitation.

Prolonged drought periods have contributed to intensification of events, such as the water crisis of 2014-2015 in the Southeast region, and recurrent drought episodes in the Northeast. It worth noting that the Northeast of Brazil, especially parts of the São Francisco basins, has experienced an increase in aridity, with the frequency of extreme precipitation events growing since the 1980s (AVILA-DIAZ *et al.*, 2020).

Analyses by Oliveira *et al.* (2020) revealed sub-regions vulnerable to intensified hydrological and geological processes due to human activity and the recurrence of hydrometeorological phenomena in the São Francisco river basin. The frequency of intense rainfall has highlighted the severity of these events in some sub-regions, emphasizing the importance of public policies for prevention and impact reduction. The increase in dry spells periods affects the economy in basins essential for energy production (AVILA-DIAZ *et al.*, 2020).

Thus, analyses of precipitation extremes can be useful in guiding the formulation of strategies for mitigating climate disasters in areas with high vulnerability (SILVA *et al.*, 2022). Certainly, deep investigation on how these

precipitation extremes events are connected with the energy system in terms of the different quantities related such as Stored Energy (STE) and Affluent Natural Energy (ANE) used to calculate the Marginal Cost of Operation (MCO) and the Settlement Price of Differences (SPD) is crucial to characterize the temporal impact of climate on the Brazilian electrical system.

2.2. Brazilian Electrical System

The Brazilian Electrical System (BES) has continental dimension and unique characteristics globally. Energy generation is predominantly hydroelectric, dependent on the hydrological regime and rainfall. However, this scenario is evolving towards a more balanced energy matrix with diversification of energy sources (MENDES, 2019). The total installed capacity in the National Interconnected System (NIS) was 161,526 MW as of 12/31/2018, distributed as follows: 63.7% in HPP's, 14.2% in conventional and nuclear thermal power plants, and 22.1% in small hydroelectric power plants (SHPs), biomass, wind, and solar power plants (DAHER & MARTINEZ, 2019).

The NIS is a unified electrical network composed of power plants, transmission lines, substations, and distribution networks. It consists of four subsystems: South, Southeast/Central-West, Northeast, and most of the North region. The interconnection of these subsystems allows for the transfer of energy, taking advantage of the synergistic benefits and diversity of the hydrological regimes of their respective river basins. This integration ensures a secure and cost-effective supply to the electricity market (ONS, 2023a).

The ONS is the entity responsible for coordinating and controlling the operation of electricity generation and transmission facilities in the NIS, as well as planning the operation of the country's isolated systems (ONS, 2023b). The operational strategy of the NIS as a whole takes precedence over the individual strategies of HPPs, allowing for the dispatch of energy to regions with low-capacity reservoirs and avoiding the use of local thermal power plants (BRÊDA *et al.*, 2021).

2.2.1. Problem of operational decisions

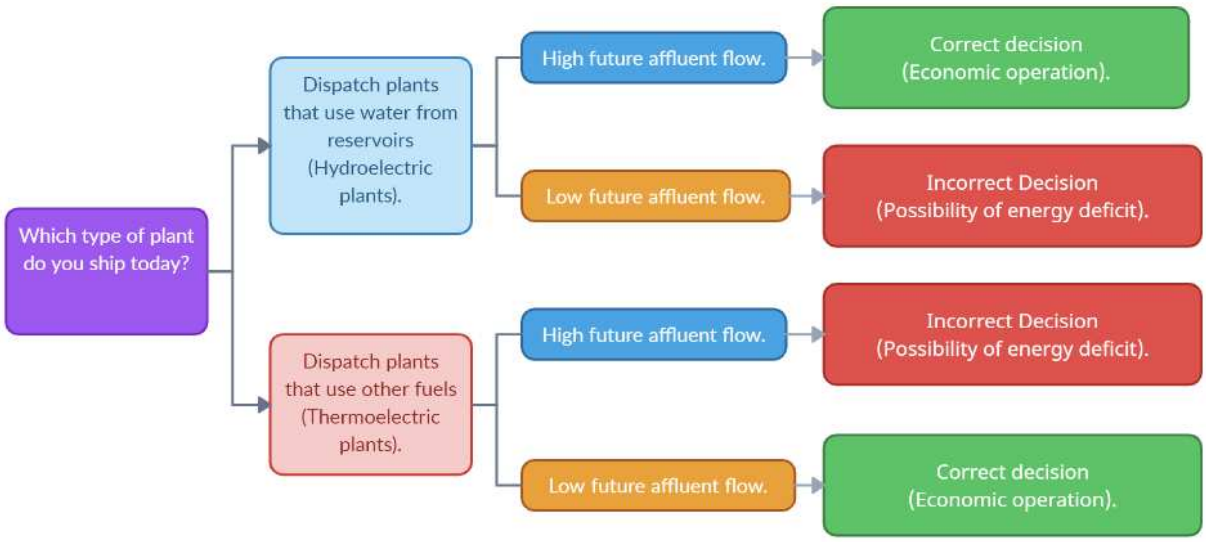
In electrical systems with focus on hydroelectricity, such as the NIS, the main challenge lies in generating electricity to meet market demands during unfavorable hydrological periods (DAHER *et al.*, 2020). An example is the risk of dispatching high-cost thermal generation in the future if there is significant hydroelectric dispatching

before a dry period; similarly, thermal use before a wet period can result in spillage and waste of energy (FERREIRA *et al.*, 2015).

At first glance, it may be thought that water stored in reservoirs has no value, making it advantageous to use this water to meet demand before resorting to thermal power plants with fuel-related costs. However, due to storage limitations and uncertainty about future flow, today's operational decision is related to future operational costs (DAHER *et al.*, 2020).

The dilemma is to balance the intensive use of hydroelectric power to avoid energy scarcity during dry periods, and the frequent activation of thermal power plants, which increases production costs and lead to the waste of water during intense rainy periods (ZAMBON, 2008). The problem of operational decisions, which involves the interdependence between current operational decisions and their future consequences is illustrated in Figure 1.

Figure 1 – Problem of operational decisions



Source: Daher *et al.* (2020)

If reservoir water is currently excessively used and natural inflows are not sufficient to replenish the reservoir level, there will be a high cost of operation in the future due to the need to apply thermal generation to meet energy demands, or even face an energy deficit if the total system generation is insufficient. On the other hand, if the decision is made to conserve reservoir water in the present, by using thermal

generation, concomitant with large volumes of inflow water in the future, this will result in the waste of potential energy from spilled reservoir water (MENDES, 2019).

Thus, in the decision-making process of operating a hydrothermal system, it is crucial to evaluate the immediate advantage of using water compared to the future benefit of its storage (DAHER *et al.*, 2020). Success of the Brazilian electrical system operation depends on climatic and hydrological factors. The ONS (System National Operator), which is responsible for operational decisions, focuses on the most economical energy generation and system security, considering the storage level and water flow in HPP's as the main variables, represented by Stored Energy (STE) and Affluent Natural Energy (ANE), respectively (MENDES, 2019).

The ONS also estimates the Marginal Operating Cost (MCO), which represents the additional cost per unit of energy generated to meet an increase in system demand. The MCO serves as the basis for calculating the Settlement Price for Differences (SPD), whose value is used to assess energy traded in the Short-Term Market. The SPD is calculated and published by the Board of Electrical Energy Commercialization (CCEE) (MEGAWHAT, 2023a).

2.2.2. Affluent Natural Energy (ANE)

The Affluent Natural Energy (ANE) represents the producible energy by the power plant from the natural inflows into the reservoirs, and it is calculated as the product of these inflows with the power plant's productivity at 65% of the useful volumes of its reservoirs (ONS, 2022). According to Medeiros (2022), ANE values can be calculated on a daily, weekly, monthly, or annual basis, as well as by basin and subsystem, using Equations 1 and 2:

$$ANE_{basin} = \sum_{i=1}^{n}[Q_{nat}(i,t).p(i)] \quad (1)$$

$$ANE_{subsystem} = \sum_{j=1}^{m}[Q_{nat}(j,t).p(j)] \quad (2)$$

Where:

t is the time interval adopted for the calculation of ANE;

i is the utilization belonging to the system of developments of the considered basin;

n is the number of developments existing in the system of developments of the considered basin;

Q_{nat} is the natural flow of the development during the considered time interval;

p is the productivity associated with a 65% height drop of the power plant;

j is the development belonging to the development system of the considered subsystem and;

m is the number of developments existing in the development system.

The corresponding values are presented in average MW or as a percentage of the long-term historical average (LTA). The monitoring and forecasting of ANE volumes are carried out in relation to its historical average, known as LTA, verified since 1931 (MEGAWHAT, 2023b).

2.2.3. Stored Energy (STE)

Stored Energy (STE) represents the energy associated with the amount of water available in the reservoir, which can be converted into electrical energy by the power plant itself and at all downstream power plants. The magnitude of STE takes into account the water levels observed in the reservoir on the reference date (ONS, 2022). The STE is the energy generated in the system through the complete draining out of the reservoirs, considering a parallel operation policy and disregarding additional inflows (DAHER *et al.*, 2020).

In other words, according to Zambon (2008), the concept of STE is a mathematical abstraction used in methodologies to simplify the representation of multiple reservoirs to a reduced number of equivalent reservoirs. It converts the volume of water stored in the reservoirs into an amount of energy that would be generated by turbine operation, considering the maintenance of average productivity in all of them. The stored energy in the basinas a whole will be the sum of the energy stored in all the reservoirs that compose it, according to Equation 3.

$$STE = \sum_{i=1}^{n}\left[k \,.\, V_i \sum_{j=J_i}^{m} p_{equiv_i}\right] \quad (3)$$

Where:

k is a constant that depends on the adopted unit system;

n s the number of reservoirs in the basin;

J_i is the set of power plants downstream of reservoir i, including itself;

V_i is the current volume of reservoir i;

p_{equivi} é a is the equivalent productivity of power plant i.

By substituting V_i with $V_{útil}$ in Equation 3, we obtain the maximum STE of the basin.

2.2.4. Marginal Cost of Operation (MOC)

Based on current and future hydrological conditions (forecasts), energy demand, fuel prices, cost of deficit, and the introduction of new generation and transmission equipment, a probabilistic pricing model determines the optimal generation allocation in the Brazilian National Interconnected System (SIN), for the period under analysis. This includes hydraulic and thermal generation in each sub-market. This results in the Marginal Cost of Operation (MCO) for each load level (light, medium, and heavy) and sub-market during the studied period (ABREU, 2012).

The MCO is the value that represents the production cost of the electricity demand that the Brazilian Electrical System requires for the next operating week, always reproducing the lowest cost to meet additional consumption (MENDES, 2019). In other words, the MCO represents the additional cost associated with meeting a marginal increase in load in each region after the consumer demand has been fully achieved (DAHER *et al.*, 2020).

The resource to be used to meet the additional load determines the marginal operating cost (DAHER *et al.*, 2020), as presented in Table 1.

Table 1 – Origin of the MOC

Resource	**MOC**
Turbinable Spill	Null
Thermal Generation	Cost of Thermal Generation
Import from Another Region	MOC of the Exporter
load shedding	cost of rationing
Water Depletion	Water Value

Source: Daher *et al.* (2020).

2.2.5. Settlement Price of Differences (SPD)

The Settlement Price of Differences (SPD) is calculated daily by the Board of Electrical Energy Commercialization (CCEE) for each hour of the following day. The calculation is performed by computational models and is based on the MCO (CCEE, 2022). The SPD indicates the decisive economic signal for the guidelines of the Daily Operation Programming (PDO) and the energy financial market. The SPD is the basis for all financial accounting in the energy sector, making it essential for all distribution, generation, and commercialization agents (MENDES, 2019).

At CCEE, the same models used by ONS to schedule and dispatch the energy generation of the system are applied, with some adjustments to reflect the pricing conditions at CCEE. In the calculation of the SPD, internal transmission constraints of each sub-market and plants under testing are not considered, treating the traded energy as equally available at all consumption points and therefore with a single price within each region (ABREU, 2012).

The SPD varies within a range established by the National Electric Energy Agency (ANEEL), which defines the minimum and maximum potential prices. These limits are reviewed annually (MENDES, 2019), although originating from results of the same probabilistic models and dependent on hydrological variables, the MCO and SPD present differences. The MCO aims to assess the generation cost to meet the load, prioritizing system security and considering constraints such as transmission line limitations, without defined limits. On the other hand, the SPD assesses the discrepancies between resources and requirements, such as sales contracts and load, without considering internal electrical constraints or plants, in the testing phase, but establishes a minimum and maximum value beforehand.

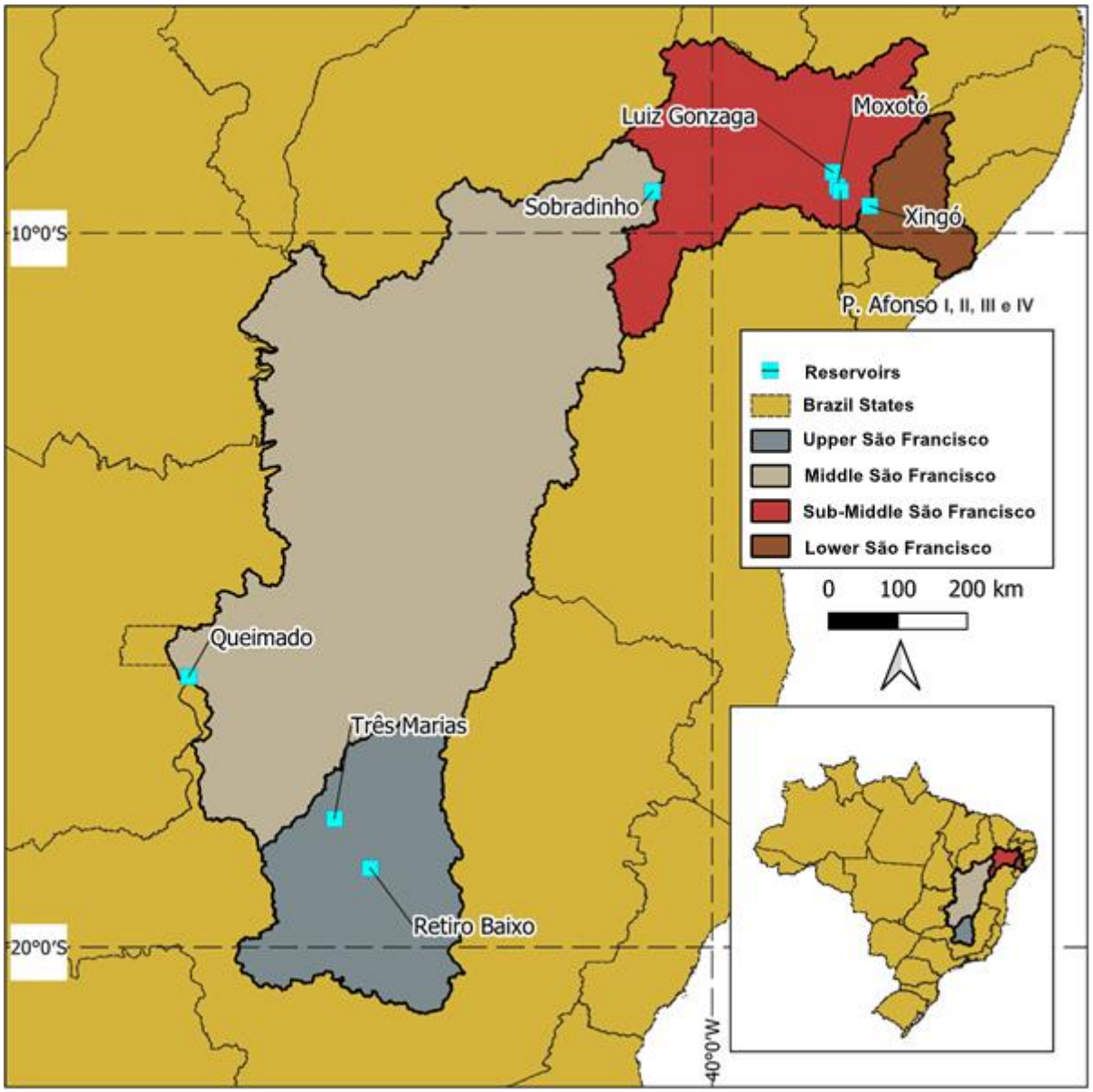
Therefore, this study takes the opportunity to analyze whether extreme rainfall climatic events occurring in the São Francisco river basin may impact on energy variables: Stored Energy (STE), Affluent Natural Energy (ANE), Marginal Operating Cost (MCO), and Settlement Price of Differences (SPD) from 2000 to 2019. Moreover, based on machine learning algorithms it is verified the potential to deliver a conceptual model which can link climate indices and energy variable.

3. MATERIALS AND METHODS

3.1. Study area

The São Francisco river basin (Figure 2) covers 8% of the Brazilian territory with a length of 2,863 km and a drainage area of over 639,219 km². Originating in the state of Minas Gerais, the São Francisco river source at the Serra da Canastra and flows into the Atlantic Ocean, along the borders of the states of Alagoas and Sergipe.

Figure 2 - Hydrographic Regions of the São Francisco river Basin



Source: Santos and Justino (2022).

The Upper São Francisco region comprises the area from the source of the São Francisco river to the city of Pirapora-MG (110,696 km², corresponding to 17% of the basin's surface area). The middle São Francisco extends from Pirapora-MG to

Remanso-BA (322,140 km²; 50% of the basin). The Sub-Middle São Francisco hydrographic region covers the stretch from Remanso-BA to Paulo Afonso-BA (168,528 km²; 26% of the basin). Finally, the lower São Francisco encompasses the stretch from Paulo Afonso-BA to the mouth of the São Francisco river (36,959 km²; 6% of the basin).

Among the main reservoirs in the São Francisco river basin, that are used for flow control and/or hydroelectric power generation, the following stand out: Três Marias, located in the state of Minas Gerais, Sobradinho, Paulo Afonso, and Luiz Gonzaga (Itaparica) in Bahia, and Xingó, situated between the states of Alagoas and Sergipe (CBHSF, 2022). The Paulo Afonso hydroelectric Complex, composed of Paulo Afonso I, II, III, IV, and Apolônio Sales (Moxotó), along with the Xingó hydropower plant, uses run-of-river reservoirs for power generation. In contrast, the other HPP in the São Francisco river basin (Retiro Baixo, Queimado, Luiz Gonzaga, Três Marias, and Sobradinho) operate with flow control reservoirs, which have different characteristics including larger useful volumes than run-of-river reservoirs (ONS, 2023c).

In the Northeast subsystem of the National Interconnected System (SIN), the Paulo Afonso hydroelectric complex, Xingó, Luiz Gonzaga, and Sobradinho hydropower plants are located. On the other hand, the Três Marias, Queimado, and Retiro Baixo hydropower plants, in the São Francisco river Basin, are part of the Southeast/Central-West subsystem (ONS, 2023c).

3.2. Data description

3.2.1. Energy Data

Regarding the production of electrical energy, data on Affluent Natural Energy (ANE), Stored Energy (STE), Marginal Cost of Operation (MCO), and Price Liquidity Differences (SPD) were used. Data on ANE and STE for Brazilian hydro energy basins are available on a daily basis through the database of the ONS, with information from 2000 to the present day. The same database provides MCO values for each operational week by subsystem and load level, as well as the weekly average with elements from the year 2005 up to the period in which this project is being developed. These data can serve as input for energy studies and system operation

cost projections. However, the data are part of a recurrent consistency process and may undergo updates.

SPD data is available for access through the price panel on the CCEE website. They are divided by subsystem (North, Northeast, Southeast/Central-West, and South) and contain historical series of hourly prices, as well as weekly and monthly averages covering the time period from 2001 to the present date.

3.2.2. Climate Data

Two precipitation data sets are used. The first data set is the Brazilian Daily Weather Gridded Data (BR-DWGD) (XAVIER *et al.*, 2022), which consists of daily and monthly meteorological data in grid format, covering the period from January 1, 1961, to July 31, 2020, with a spatial resolution of 0.1º × 0.1º. These data were generated from six distinct interpolation methods using meteorological and rainfall station data (XAVIER *et al.*, 2016) and have been widely utilized in climate studies (ANDRADE *et al.*, 2022; TOMASELLA *et al.*, 2022).

The second data set used is the ERA5-Land, which provides hourly reanalysis data for the land surface on a global scale, covering the period from January 1, 1950, to the present, with a spatial resolution of 0.1º × 0.1º (MUÑOZ-SABATER *et al.*, 2021). Reanalysis combines model data with observations into a comprehensive and consistent data set, using the laws of physics (MUÑOZ-SABATER, 2019). The ERA5-Land dataset has also been used by the scientific community in climate studies in Brazil and worldwide (ARAÚJO *et al.*, 2022; ANDRADE *et al.*, 2022).

3.3. Data pre-processing

3.3.1. Energy Data

For the study of the ANE and STE variables, raw and relative values from the available series were selected. For ANE, the raw value in MW-month (Gross ANE) and the percentage value relative to the long-term average (ANE (%)) were used. For STE, the value verified on the day in MW-month (Gross STE) and the same value in percentage terms (STE (%)) were obtained, considering that the maximum storage capacity is reached when all reservoirs in the basin are full.

The daily ANE and STE series were transformed into monthly series by calculating the average of the daily values for each month.

For the study of the MCO, the weekly average series was chosen, which consists of the average values of the MCO for the light, medium, and heavy load levels. The weekly series was converted into monthly by calculating the average of the weekly values for each month.

The SPD series were acquired in monthly averages. Like the MCO, the SPD does not represent a hydrographic basin but rather an energy subsystem. Therefore, due to the fact that the São Francisco river basin has HPP's in the Northeast and Southeast/Central-West subsystems, two MCO series (MCO-NE and MCO-SE) and two SPD series (SPD-NE and SPD-SE) were studied, representing the average of the two variables in each energy subsystem.

As they are monetary values, the time series of MCO and SPD were adjusted so that, throughout the study period, the currency value was referenced to the value of the Brazilian Real of the last month of the series (December 2022), minimizing the effects of inflation over time. To perform this adjustment, the National Consumer Price Index (IPCA) was used.

The IPCA aims to measure the price variation of a basket of products and services consumed by the population, and its result indicates whether prices have increased or decreased from one month to another (IBGE, 2023).

The MCO and SPD values were adjusted according to the calculation methodology provided by the Brazilian Institute of Geography and Statistics (IBGE), using the "IPCA Calculator" tool. The adjusted value is obtained by multiplying the initial value by the result of dividing the index number of the final month by the index number of the month before the initial month. The result of this division is the factor representing the accumulated variation of the IPCA in the desired period (Equation 4).

$$\mathrm{PC} = P_i * \left(\frac{IPCA_f}{IPCA_i}\right) \quad (4)$$

Where:

PC: Corrected Price;

P_i: Price at start date;

$IPCA_i$: IPCA for the month prior to the initial month and

$IPCA_f$: IPCA for the final (or reference) month.

The monthly average was calculated for all eight energy variables to determining the historical seasonal behavior.

The Table 2 summarizes the energy variables used in this study.

Table 2 – Energy Variables

	Variable	Description	Period
1	Gross ANE	Gross Affluent Natural Energy	2000-2022
2	ANE (%)	Percentage Affluent Natural Energy	2000-2022
3	Gross STE	Gross Stored Energy	2000-2022
4	STE (%)	Percentage Stored Energy	2000-2022
5	MCO-NE	Marginal Cost of Operation of the Northeast Subsystem with corrected values	2005-2022
6	MCO-SE	Marginal Cost of Operating the Southeast/Midwest Subsystem with corrected values	2005-2022
7	SPD-NE	Settlement Price Differences of Northeast Subsystem with corrected values	2005-2022
8	SPD-SE	Settlement Price Differences for Southeast/Center-West Subsystem with corrected values	2005-2022

3.3.2. Climate Data

The daily precipitation series from the BR-DWGD dataset were used. And the hourly data from ERA5-Land were aggregated to generate the daily precipitation series. Both datasets were clipped to the area of the São Francisco river basin in a Geographic Information System (GIS) environment, for the period from 1990 to 2019, representing the most recent 30 common years for both datasets.

3.2.2.1. Precipitation Climate Extremes

In order to analyze the magnitudes and seasonality of rainfall in the basin, maps of annually and monthly average precipitation were generated across the study area. In order to analyze and identify trends in these occurrences, the Expert Team

on Climate Change Detection and Indices (ETCCDI) of the World Meteorological Organization (WMO) has developed 27 indicators based on daily data of maximum and minimum temperature, as well as precipitation (KARL *et al.*, 1999; FRICH *et al.*, 2002). In this study, the indicators listed in Table 3 were selected for the analysis of annual and/or monthly frequency of precipitation extremes. These indicators were calculated using the xclim library (LOGAN *et al.*, 2023).

Table 3 - Precipitation climate extreme index

	Index	Description	Frequency
1	PRCPTOT	Total Precipitation (mm)	Annual and Monthly
2	RX1DAY	Maximum precipitation in one day (mm)	Annual and Monthly
3	RX5DAYS	Maximum precipitation in five consecutive days (mm)	Annual and Monthly
4	SDII	Simple Daily Intensity Index (mm/day)	Annual and Monthly
5	R20mm	Number of days with very heavy precipitation (days)	Annual and Monthly
6	CWD	Consecutive wet days (days)	Annual and Monthly
7	CDD	Consecutive dry days (days)	Annual and Monthly
8	WD	Wet days (days)	Annual and Monthly
9	DD	Dry days (days)	Annual and Monthly
10	PRCWQ	Precipitation in the wettest quartile (mm)	Annual
11	PRCDQ	Precipitation in the driest quartile (mm)	Annual

Days with daily precipitation equal to or greater than 1 mm were considered wet days, while dry days were defined as those with daily precipitation lower than 1 mm. For the 11 indices, the annual averages of the corresponding precipitation extremes were calculated, allowing for the analysis of spatial distribution and trends throughout the São Francisco river basin.

For each of the nine indices calculated on a monthly basis, four additional series were generated, consisting of the sum of extreme values from the previous 3, 6, 12, and 24 months. This provided monthly, quarterly, semi-annual, annual, and biennial accumulated values for each precipitation extreme index.

3.4. Trends

The Mann-Kendall test (MANN, 1945; KENDALL, 1955) and Sen's slope estimation (SEN, 1968; THEIL, 1992) were employed to examine the trends in the monthly series of energy variables and the annual series of extreme climate precipitation indices. The non-parametric Mann-Kendall test is used to assess the

presence of statistically significant trends in a time series, considering the null hypothesis of no increasing or decreasing trends. It does not assume a specific data distribution and was applied with a 95% confidence level, corresponding to a 5% error. It is the most appropriate method for trend analysis in climatological time series (MACHIWAL and JHA, 2008). On the other hand, the Sen's slope test is used to analyze the magnitude of trends. It is a non-parametric method that assumes a linear trend in the series data.

The Sen's slope test is more robust and accurate than a simple linear regression as it is insensitive to outliers and missing information. Both the Mann-Kendall test and the Sen's slope test are widely used in studies to detect trends in time series of climate variables (OLIVEIRA *et al.*, 2016; LUIZ-SILVA *et al.*, 2018).

For conducting the Mann-Kendall and Sen's slope tests, the PyMannKendall library available in the Python language was used. This library enables the application of the tests in a simple and efficient manner, automating the analysis and providing relevant statistical results (HUSSAIN; MAHMUD, 2019).

3.5. Correlations

The correlation coefficient plays the role of indicating the strength and direction of the linear relationship between two random variables. One of the most well-known correlation coefficients is the Pearson coefficient, which measures linear association. According to Shimakura (2006), to assess the degree of association between variables, it is necessary to use the correlation coefficient (*r*). The value of *r* can be interpreted as shown in Table 4.

Let (x_1, x_2, ..., x_n) be the set of measurements for one variable, and let (y_1, y_2, ..., y_n) be the measurements for the other variable. Let $\bar{x}$, $\bar{y}$, S_x and S_y be the sample means and standard deviations of the two data sets. To obtain a measure of the degree of association in the linear relationship between two variables, we use the correlation coefficient (*r*), defined as:

$$r = \frac{S_{xy}}{S_x S_y} \tag{5}$$

Where:

$$S_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1} \tag{6}$$

Table 4 - Interpretation of Pearson's correlation coefficient value (*r*)

***r* (+ or -)**	**Interpretation**
0,00 a 0,19	Very Weak Correlation
0,20 a 0,39	Weak Correlation
0,40 a 0,69	Moderate Correlation
0,70 a 0,89	Strong Correlation
0,90 a 1,00	Very Strong Correlation

Fonte: Shimakura (2006).

The Pearson correlation coefficient was employed to investigate the relationships between energy variables, as well as to identify the association between monthly series of ANE, STE, MCO, and SPD with precipitation extremes indices at different time scales, including monthly, quarterly, semi-annually, annually, and biennially.

3.6. Use of Artificial Intelligence tool for generating predictive regression models

The potential prediction of energy variables based on precipitation extremes may be crucial to the understanding of the relationships between these variables. For this purpose, three machine learning regression methods were used, taking as input the data of precipitation climate extremes with Pearson correlation (*r*), identified as moderate or strong, in order to estimate the respective energy variables on a monthly basis. The methods used are Random Forest (RF), Artificial Neural Networks (ANN), and k-Nearest Neighbors (kNN).

To evaluate the performance of the regression models, the cross-validation technique was utilized, and the following evaluation metrics were applied: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), Kling-Gupta Efficiency (KGE), and Willmott's concordance coefficient *(d)*.

3.6.1. Random Forest (RF)

Random Forest is an estimator that consists of a set of decision trees (BREIMAN, 2001; HASTIE *et al.*, 2001). Each decision tree contains a prediction for the output variable based on a random subset of the input data. For continuous

variables, the estimate is calculated as the average of the values obtained from each tree.

3.6.2. Artificial Neural Networks (ANN)

The Artificial Neural Network algorithm is a method of data processing inspired by the transfer of data in biological neural systems (LEE *et al.*, 2008). The structure of this algorithm includes an input layer, one or more hidden layers, and an output layer. Additionally, each layer is composed of multiple neurons. In an ANN, each connection between neurons has a weight. Each neuron creates an output by passing the sum of the inputs multiplied by the weight of the connection through a transfer function. This output becomes an input for the neurons in the next layer. The output that is closest to the observed data is approximated by adjusting the connection weights (SINGH *et al.*, 2010).

3.6.3. K-Nearest Neighbors (KNN)

The k-Nearest Neighbors algorithm is a non-parametric method that estimates the relationship between inputs and outputs without any predetermined assumptions (ANARAKI *et al.*, 2020). This algorithm searches and selects K feature vectors that have the highest similarity with the historical data. This similarity is calculated based on the Euclidean distance:

$$D_{rt} = \sqrt{\sum_{i=1}^{d} w_i (X_{ir} - X_{it})^2} \quad (7)$$

where D_{rt} is the Euclidean distance between X_r and X_t, X_r is the input related to the predicted data, X_t is the input related to the historical data, d represents the number of inputs, and w expresses the weight for each predicted input (the sum is equal to 1).

After calculating the Euclidean distance, the kernel function estimates the output of the models which is represented as follows (ARAGHINEJAD, 2013):

$$R(D_{rt}) = \frac{\frac{1}{D_{rt}}}{\sum_{j=1}^{k} D_{rt}} \quad (8)$$

where *k* is the number of historical data points with the smallest distance to the predicted data. Finally, the outputs are estimated as:

$$Y_r = \sum_{i=1}^{n} R\left(D_{rj}\right) * Y_j \quad (9)$$

3.6.4. Parameter Selection

The appropriate selection of regression model parameters is crucial for obtaining accurate results. To determine the best hyperparameter values for each regression method, the GridSearchCV algorithm was used.

This algorithm performs an exhaustive search over a predefined grid of hyperparameters, evaluating the model's performance for each parameter combination. By testing all the hyperparameter combinations, the parameter with the highest prediction accuracy on the test set is selected as the best hyperparameter for the model (Ranjan *et al.*, 2019). Table 5 summarizes the chosen parameters for each machine learning algorithm in GridSearchCV*.*

Table 5 - Hyperparameters

Algorithm	**Parameter**	**Options**
RF	*criterion*	*squared_error* ou *absolute_error*
	n_estimators	10, 50, 100, 150 or 200
ANN	*batch_size*	10, 50 or 90
	max_iter	200 or 500
KNN	*n_neighbors*	5, 10, 20 and 30

The parameter "criterion" indicates the criterion by which the RF algorithm will determine the best output, which can be through mean squared error (squared_error) or mean absolute error (absolute_error). The number of decision trees that compose the RF is defined by the parameter "n_estimators".

For the ANN algorithm, the choice of parameters was made for "batch_size," which indicates how many records will be sent to the neural network to update the weights, and "max_iter," which determines the number of iterations until convergence (0.0001) or until reaching that number of iterations.

The parameter "n_neighbors" represents the value of k in Equation (8), which indicates the number of historical data points with the shortest distance to the predicted data points by the KNN algorithm.

3.6.5. Cross-validation

Cross-validation is a statistical method used to evaluate the performance of machine learning models, in which the main goal is to assess the model's ability to generalize to unknown data (CHEN *et al.*, 2021). The cross-validation technique was applied by dividing the data into five subsets, keeping one subset as the test set and the others as the training set. This process is repeated five times, ensuring that each subset is used as the test set at least once. Evaluation metrics are calculated for each round, and the average is taken across the five results. This procedure is repeated 30 times, totaling 150 tests. Results are presented in box plots, showing the average of the evaluation metrics obtained in the 30 repetitions.

3.6.6. Assessment Metrics

To assess the quality of predictions obtained by the regression models five metrics have been used. The Mean Absolute Error (MAE) (Eq. 10) is calculated as the average of the absolute differences between the predicted values and the observed values. It provides a direct measure of the average error of the predictions. The Mean Absolute Percentage Error (MAPE) (Eq. 11) calculates the average of the absolute percentage errors, providing a relative measure of the average error relative to the true values.

The Root Mean Squared Error (RMSE) (Eq. 12) is another widely used metric that calculates the square root of the average of the squared errors between the predictions and the true values. This metric penalizes larger errors more heavily and is sensitive to outliers.

The Kling-Gupta Efficiency (KGE) coefficient (Eq. 13), proposed by Gupta *et al.* (2009) and modified by Kling *et al.* (2012), is a measure of efficiency that compares the variability, trends, and correlation of predictions with the true values. This coefficient ranges from -∞ to 1, where values close to 1 indicate a good fit of predictions to the observed data.

Lastly, the Willmott's concordance coefficient *(d)* (Eq. 16) is used to evaluate the comparative accuracy of estimation values by calculating the discrepancy

between the values of one estimation relative to another (Willmott *et al.*, 1985). The values of this index range from 0 to 1, representing no agreement and perfect agreement, respectively.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - O_i| \quad (10)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|P_i - O_i|}{O_i} \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \quad (12)$$

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \quad (13)$$

$$\beta = \frac{\mu_P}{\mu_O} \quad (14)$$

$$\gamma = \frac{CV_P}{CV_O} \quad (15)$$

$$d = 1 - \left[\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(|P_i - \mu_O| + |O_i - \mu_O|)^2}\right] \quad (16)$$

In the equations shown above, n is the number of data points, μ is the mean value, CV is the coefficient of variation, P is the prediction value, and O is the true value.

These metrics were used to evaluate the performance of each regression model, providing a comprehensive analysis of the quality of the predictions relative to the true values.

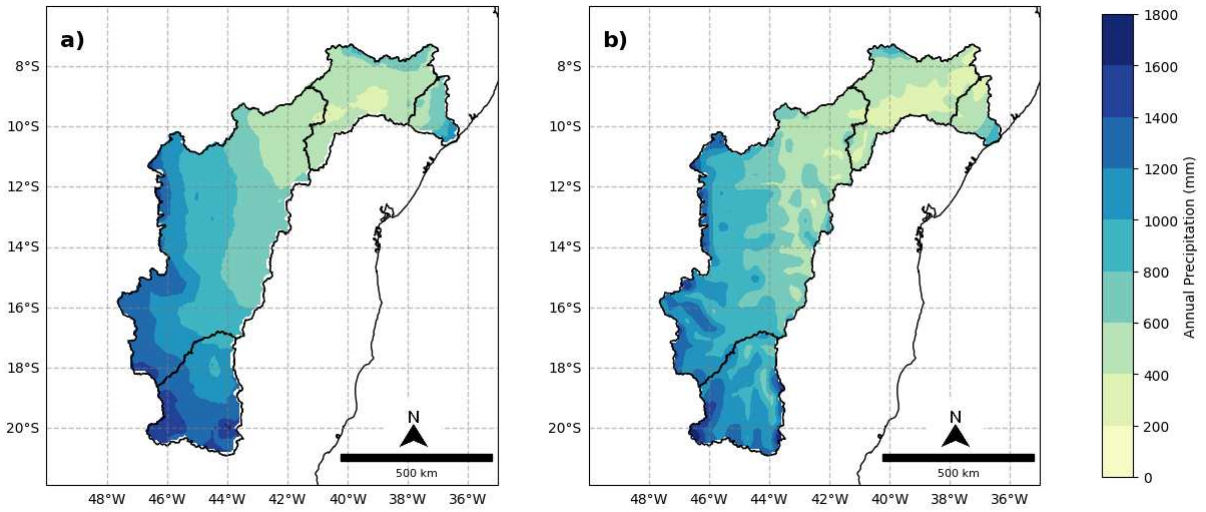
4. RESULTS AND DISCUSSION

4.1. Precipitation Climate Extremes

4.1.1. Characterization of Precipitation

The annual precipitation climatology in the São Francisco river basin can be observed in Figure 3.

Figure 3 – Annual Precipitation Climatology (1990 - 2019)



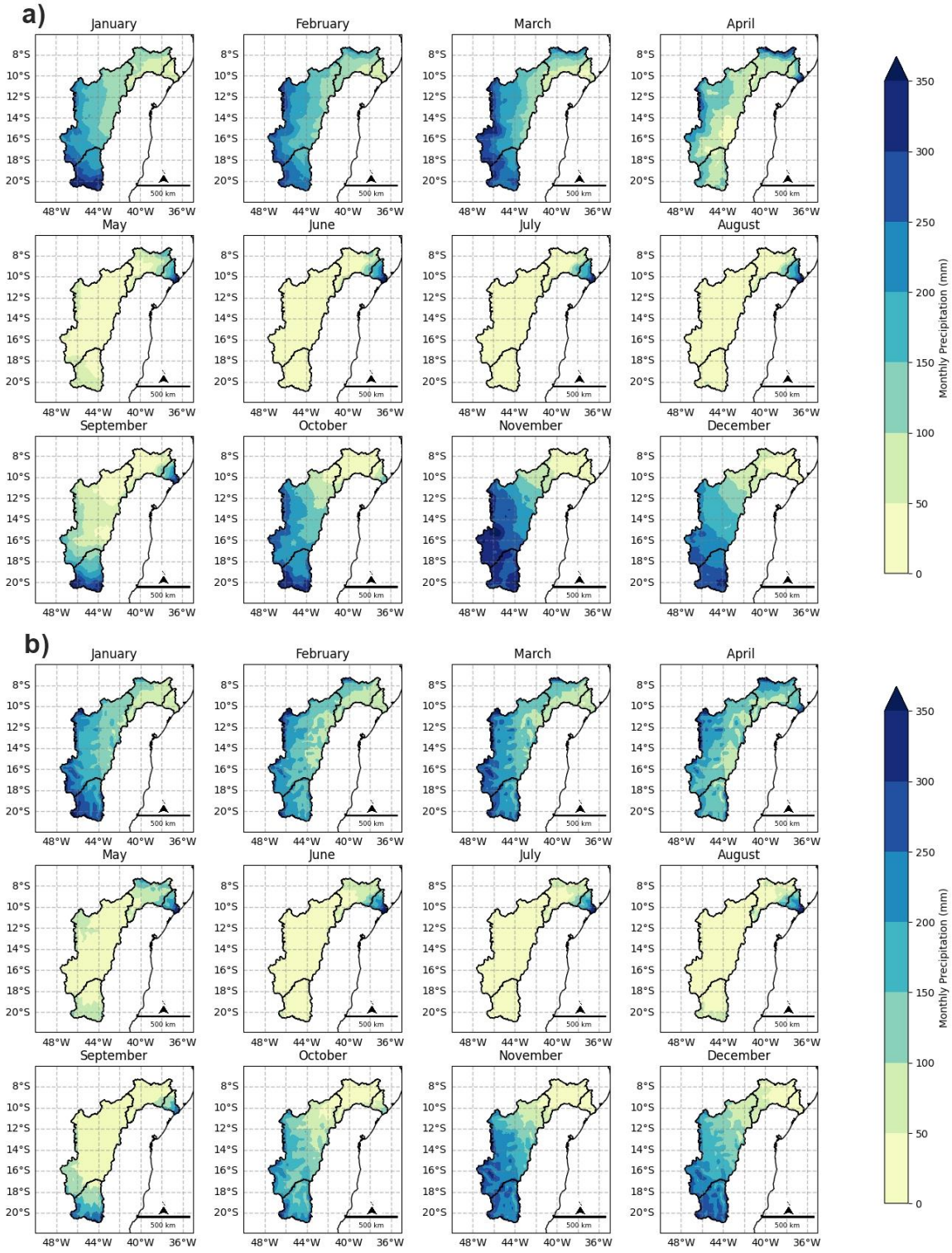
Legend: **a)** BR-DWGD, **b)** ERA5-Land.

It can be observed that, for both analyzed datasets, the average annual total precipitation in the São Francisco river basin generally decreases from the Southeast to the Northeast region. The upper São Francisco hydrographic meso-region records the highest annual volumes, ranging from 1400 to 1800 millimeters, followed by the Middle São Francisco, which presents volumes between 800 and 1600 millimeters distributed over a relatively larger area compared to the other meso-regions. The lower São Francisco, on the other hand, exhibits an annual precipitation by about 600 to 1000 millimeters, while the sub-middle São Francisco is the region with the lowest precipitation, with predominance of only 200 to 600 millimeters annually.

During the period from April to October, it can be observed that the basin as a whole experience the lowest precipitation volume, although the lower São Francisco meso-region registers considerable volumes in a limited area (Figures 4a and 4b). The highest precipitation averages are predominantly recorded between November and March, especially in the upper and middle São Francisco meso-regions.

The significant volume in the lower São Francisco during the dry period occur, according to the study conducted by Gomes *et al.* (2019) due to the occurrence of Easterly Waves Disturbances (EWDs). The authors argue that EWDs play a significant role in the coastal region of Northeast Brazil from april to august, acting in conjunction with other of atmospheric systems such as cold fronts, the Intertropical Convergence Zone (ITCZ), and Upper Tropospheric Cyclonic Vortices (UTCVs). EWDs account for at least 60% of the total rainfall in this region.

Figure 4 – Monthly Precipitation Climatology (1990 – 2019)



Legend: **a)** BR-DWGD, **b)** ERA5-Land.

This spatial distribution of precipitation throughout the year is relevant for understanding the regional climate. Additionally, it is important to highlight that this

distribution is also a result of precipitation extremes. The spatial variability of precipitation among different hydrographic meso-regions, as well as its seasonality, is similar to the results obtained by Rodrigues *et al.* (2021), Souto *et al.* (2019), and Oliveira *et al.* (2022).

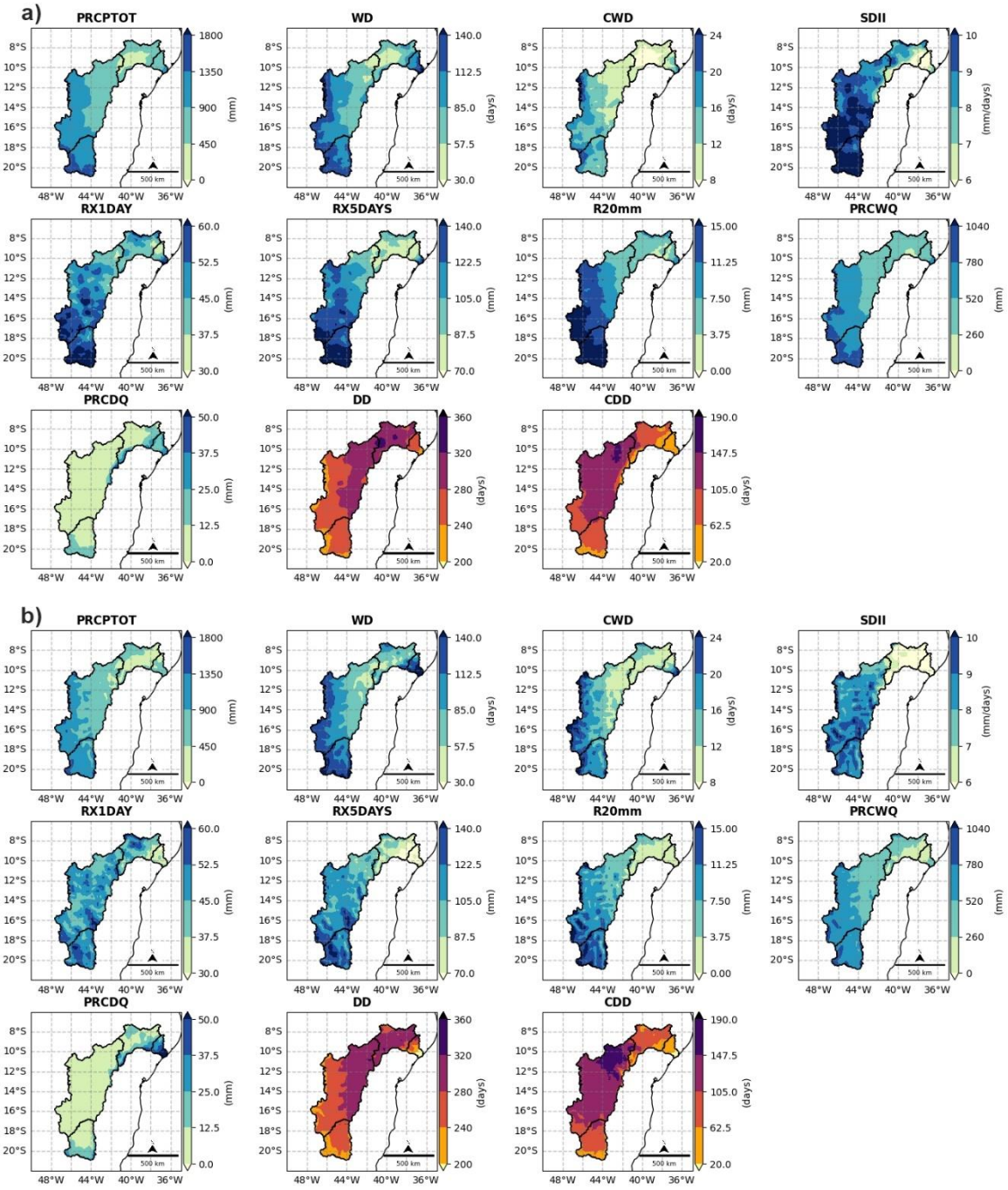
4.1.2. Precipitation Climate Extreme Index

Figure 5 presents the results of two datasets, BR-DWGD and ERA5-Land (Figures 5a and 5b, respectively), which are consistent regarding the spatial distribution of annual mean precipitation extremes in the São Francisco river basin. Among the analyzed variables, the indicators PRCPTOT (Total Precipitation) and PRCWQ (Wettest Quartile Precipitation) stand out. They have the highest values in the upper and middle São Francisco meso-region, indicating that about 60% of the total precipitation, in the rainiest areas occurs in the wettest quarter.

The indicators of wet days (WD) and consecutive wet days (CWD) show higher values in the upper São Francisco meso-region, the western part of the middle São Francisco, and the eastern area of the lower São Francisco. On average, there are more than 85 rainy days per year, with consecutive rainfall events lasting between 16 and 24 days in these regions. On the other hand, a low value of these indices is observed in the sub-middle São Francisco meso-region. It is noticed an average of 30 to 60 rainy days per year, with only 8 to 12 consecutive rainy days in most of this meso-region.

The indices RX5DAYS and R20mm reveal the occurrence of extreme precipitation events in the upper and middle São Francisco meso-regions, as they also have the highest values for these indicators. RX5DAYS measures the maximum intensity of precipitation over five consecutive days and records values above 120mm in these same meso-regions. On the other hand, R20mm, which measures the frequency of days with precipitation exceeding 20mm, indicates the occurrence of more than 11 days with values above this threshold.

Figure 5 – Annual Average of Precipitation Extremes (1990 – 2019)



Legenda: **a)** BR-DWGD, **b)** ERA5-Land.

The variable RX1DAY (Maximum Precipitation in a Single Day) is distributed throughout the basin area, with values exceeding 50mm observed in all meso-regions, with values exceeding 60mm in the upper and middle São Francisco. On the other hand, the SDII indicator (Mean Daily Intensity) shows differences between the datasets, as for BR-DWGD, higher magnitudes (above 10 mm/day) are observed in

the upper and middle São Francisco meso-regions, with values between 8 and 9 millimeters per day in the sub-middle São Francisco meso-region. For the ERA5-Land, the highest SDII values also occur in the upper and middle São Francisco (between 8 and 9 mm/day), with minimum values in the other areas across the basin.

The prominent extreme indices in the lower São Francisco are consistent with a recent study conducted by Morales *et al.* (2023). In this study, the authors observed that the eastern coast of the Northeast region experiences a higher number of extreme precipitation events compared to the semiarid region. Regarding the meso-region most affected by precipitation extremes, there is agreement with the study conducted by Medeiros *et al.* (2022), in which the extreme precipitation events (PRCPTOT, RX1day, RX5day, SDII, R20mm, CWD) are more intense, frequent, and prolonged in the northern, southern, and southeastern regions of Brazil. During the rainy season, the South Atlantic Convergence Zone (SACZ) has a significant influence on the rainfall regime in the southeastern region of Brazil (MARENGO *et al.*, 2015; ROSA *et al.*, 2020; NIELSEN *et al.*, 2019). The study conducted by Aguiar *et al.* (2021) demonstrates that the frequent occurrence of these extreme events is caused by the influence of the SACZ. Their results reveal that the average conditional probability of the occurrence of the SACZ, when there are disasters due to intense or persistent rainfall events in the Southeast, is 48%, and in Minas Gerais, this probability is even higher, reaching 50%.

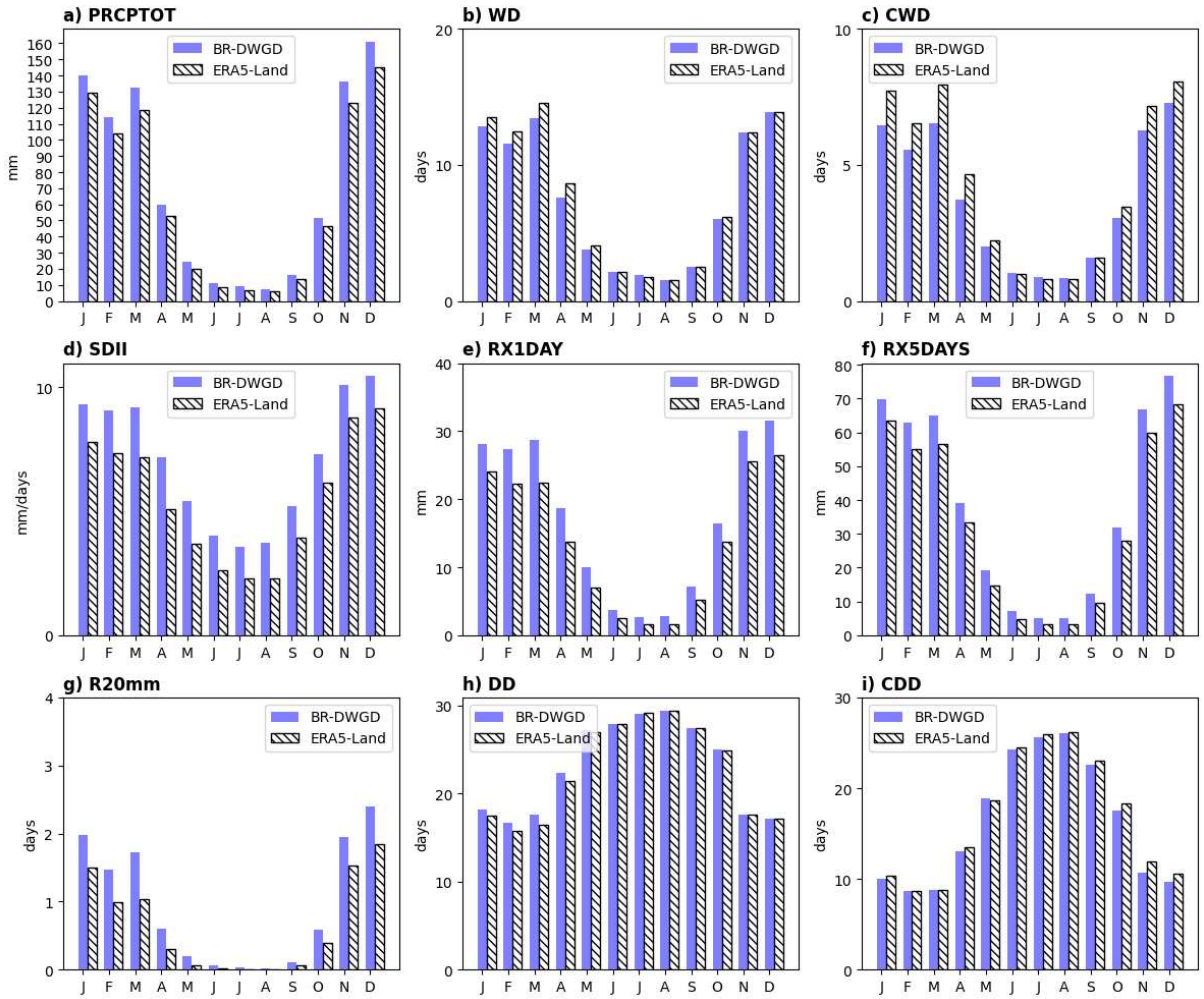
Lastly, it is worth noting analyses of the indicators DD (Dry Days), CDD (Consecutive Dry Days), and PRCDQ (Driest Quartile Precipitation). The highest number of dry days (from 280 to 320 days) is observed throughout the sub-middle São Francisco meso-region and the eastern part of the middle São Francisco, which also records the maximum number of CDD (from 145 to 190 days). These results suggest that, on average, this area goes without precipitation for more than six months, resulting in high vulnerability to drought events.

Predominantly in the lower São Francisco meso-region, precipitation in the driest quarter is higher compared to other meso-regions. These analyses corroborate the study by Medeiros *et al.* (2022), which found lower intensity and frequency of extreme precipitation events in Northeast Brazil. In this region, there is a high number of Consecutive Dry Days (CDD) in the central portion of the Northeast, which has a semiarid climate.

However, this does not mean that there are no occurrences of extreme rainfall in the Northeast, as the RX1DAY and SDII indices stand out. According to the conclusions of Monteiro (2022), in the semiarid area of the Northeast, when atmospheric conditions result from the interaction of two or more meteorological systems, such as UTCV's, EWD, and ITCZ, it is frequent to result in significant rainfall characterizing true occurrences of extreme indices.

These results highlight the spatial heterogeneity of extreme precipitation events in the São Francisco river basin, which requires the adoption of specific strategies for each hydrographic meso-region in order to promote efficient water resources management. It is also important to understand the seasonal behavior of the extreme indices (Figure 6).

Figure 6 – Monthly Average of Extreme Precipitation (1990 – 2019)



Analyses of the distribution of average PRCPTOT values (Figure 6a) throughout the years reveal the presence of a rainy period between November and March, with monthly averages above 100mm which are represented by both datasets. On the other hand, a dry period occurs in the basin from May to September, with monthly total precipitation below 30mm. October and April are transitional months between the rainy and dry periods, with an average of approximately 60mm.

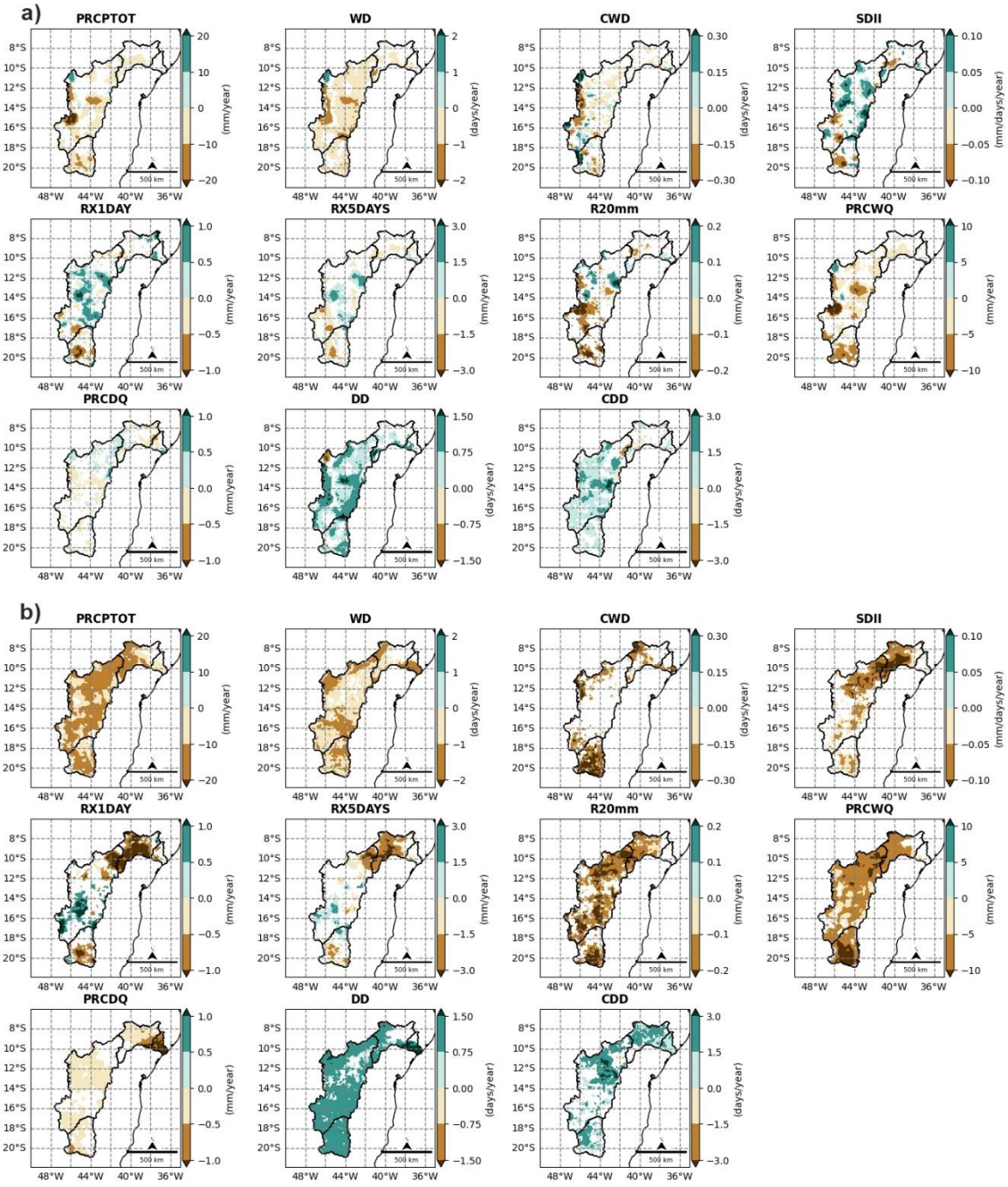
The extreme indices related to heavy rainfall (WD, CWD, SDII, RX1DAY, RX5DAYS, and R20mm) (Figures 6a to 6g), show their maximum during the rainy period, between November and March. This indicates a strong influence of these extreme events on the precipitation regime of the São Francisco river basin. The presence of these extreme events during the rainy season, highlights the influence identified by Aguiar *et al* (2021), as shown by a significant association between natural disasters and the SACZ, in the months of November, December and January. According to the authors, more than half of all SACZ-related events occurred during this period, with respect to others in the Southeast region.

On the other hand, the extreme indices related to drought reach their monthly maximum averages in the months between June and September. During this period, on average, 25 days of each month are characterized as dry (Figure 6h), with at least 20 of these days being consecutively dry (Figure 6i) indicating the dry period in the basin.

4.1.3. Trends in Precipitation Climate Extreme Indexes

Significant trends calculated for the annual extreme precipitation indices over the study period are shown in Figure 7. It is worth noting that only the significant values are highlighted ($p < 0.05$), thus, the entire white area within the basin does not exhibits trends or, if they do, they are not significant. Trends of extreme precipitation indices differ in terms of spatial distribution and magnitude in the two analyzed datasets. This behavior is justified by Regoto *et al* (2021), which states that comparing different datasets and methods can lead to large variations and uncertainties in climate extreme trends. A negative trend is observed throughout the basin for the PRCPTOT, WD, CWD, PRCWQ, and R20mm indices, with larger magnitudes exhibited by the ERA5-Land dataset. It is important to mention that both datasets in most cases deliver similar spatial pattern but with different magnitudes.

Figure 7 – Significant Trends (95%) of Precipitation Climate Extremes (1990 – 2019)



Legend: **a)** BR-DWGD, **b)** ERA5-Land.

These results are similar to those detected by Avila-Diaz *et al* (2020), that found predominantly negative trends in the São Francisco basin for the PRCPTOT, R20mm, and CWD indices, although only CWD showed a significant trend. Downward trends are observed in the sub-middle São Francisco meso-region and the northern portion

of the middle São Francisco for the SDII, RX1DAY, and RX5DAYS indices, especially in the ERA5-Land dataset, diverging from the results obtained by BR-DWGD (Figure 7a), which generally do not indicate significant trends in these areas, and show positive trends in the middle São Francisco for the mentioned indicators.

An increasing trend in DD index is observed throughout the entire area of the São Francisco river basin, with larger magnitudes found in the ERA5-Land, which also revealed positive trends in CDD in all hydrographic meso-region of the basin. On the other hand, the BR-DWGD dataset suggests a relatively milder increasing trend in CDD in the middle São Francisco. Observations of reduction (increase) in intense rainfall events (drought) in the semiarid region of the basin are consistent with the findings by Assis *et al.* (2021). They emphasize the decrease in total precipitation, daily rainfall, maximum volumes in 1 and 5 consecutive days, rainy days, and the number of days with moderate, heavy, and intense rainfall. Positive trends were only observed in consecutive days without rain, reinforcing the negative trend of rainfall and recurrent droughts in northeastern Brazil.

The PRCDQ index shows downward trends throughout the basin, with the greatest magnitude in the lower São Francisco. In the BR-DWGD dataset, these trends are characterized by minimal magnitudes. Climate analyses are essential to understand how climate extremes affect water resource availability for energy generation. Changes in extreme precipitation event trends can directly impact hydroelectric power production. Therefore, efficient energy production, water resource management and sustainable energy supply may benefit for a clear understand of short-scale weather events.

4.2. Relationship between Climate Extremes and Energy Variables

4.2.1. Correlations between Climate Extremes and Energy Variables

The Pearson linear correlation coefficients *(r)* between energy variables and accumulated climate extremes are presented in Table 6. Empty fields in the table indicate that the corresponding correlations did not reach the desired significance level ($p \leq 0.05$). Both datasets used showed r values > |0.70| between Gross Affluent Natural Energy (ANE) and all accumulated extreme indices for six months, except for CDD 6, where the ERA5-Land dataset indicates a moderate correlation. In fact, 1, 3, 6, 12 and 24 indicate that the indices are summed throughout these specific months interval.

Moderate correlations are also observed with indices accumulated over three months. As for ANE (%), strong correlations are observed with CWD 24 from ERA5-Land and R20mm 24 from BR-DWGD, with consensus only for PRCPTOT 24, while correlations are moderate for the other indices accumulated over 24 months, as well as for all indices accumulated over 12 months.

The correlations between Gross Stored Energy (STE) (both gross and percentage) and extreme indices exhibit strong values, particularly highlighting the correlations with RX1DAY 24 and RX5DAYS 24 from the BR-DWGD dataset, as well as with PRCPTOT 24, WD 24, CWD 24, R20mm 24, and DD 24, with agreement between the two data sources. For the other extremes accumulated over 24 months and all accumulated over 12 months, the correlations with stored energy are moderate.

The variables MCO-NE and SPD-NE show moderate correlations with all extreme indices accumulated over 12 and 24 months, except for CDD. It is important to note that MCO-NE revealed weak or no correlations with CWD12 from the BR-DWGD dataset and with SDII 12 in both analyzed datasets.

Moderate correlations were observed between the variables MCO-SE and SPD-SE with extreme indices accumulated over 24 months, except for CWD and CDD. In addition, moderate correlations were found between these variables from the Southeast/Central-West subsystem and PRCPTOT 12 and R20mm 12, as well as between SPD-SE and WD 12, RX1DAY, RX5DAYS, and DD 12. Based on results presented in Table 6, the correlations indicate that energy variables can be affected by climate extreme indices related to precipitation at various time scales.

These previous analyses demonstrated that a gain can be obtained in the interpretation of those correlation. Such as, the more the extremes (PRCPTOT, RX1DAY, RX5DAYS) act in the basin, greater the water availability and, consequently, is cheaper the production costs and selling prices. On the contrary, the smaller the performance of the extremes, the more thermal plants will be activated in the system, thus raising costs and prices.

Table 6 – Correlations between Energy Variables and Precipitation Extremes

	Gross ANE		ANE (%)		Gross STE		STE (%)		MCO - NE*		MCO - SE*		SPD - NE*		SPD - SE*	
	a	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*
PRCPTOT 1	0.17	0.16			-0.22	-0.22	-0.22	-0.22								
PRCPTOT 3	0.66	0.66	0.31	0.31												
PRCPTOT 6	0.78	0.77	0.29	0.30	0.29	0.30	0.29	0.30	-0.21	-0.22			-0.22	-0.21	-0.20	-0.20
PRCPTOT 12	0.38	0.37	0.69	0.68	0.69	0.68	0.69	0.68	-0.55	-0.53	-0.43	-0.42	-0.56	-0.53	-0.51	-0.49
PRCPTOT 24	0.38	0.37	0.72	0.70	0.79	0.78	0.79	0.78	-0.53	-0.52	-0.44	-0.43	-0.54	-0.52	-0.48	-0.47
WD 1	0.25	0.28			-0.22	-0.18	-0.22	-0.18								
WD 3	0.67	0.69	0.29	0.29												
WD 6	0.75	0.74	0.29	0.30	0.33	0.37	0.33	0.37	-0.22	-0.23			-0.22	-0.24	-0.19	-0.20
WD 12	0.35	0.35	0.66	0.65	0.66	0.65	0.66	0.65	-0.49	-0.49	-0.38	-0.38	-0.51	-0.50	-0.45	-0.44
WD 24	0.36	0.35	0.68	0.66	0.76	0.74	0.76	0.74	-0.51	-0.51	-0.42	-0.43	-0.51	-0.52	-0.46	-0.47
CWD 1	0.20	0.23			-0.25	-0.21	-0.25	-0.21								
CWD 3	0.65	0.67	0.26	0.27												
CWD 6	0.74	0.73	0.24	0.26	0.25	0.31	0.25	0.31	-0.16	-0.17			-0.17	-0.17		-0.15
CWD 12	0.33	0.36	0.61	0.67	0.56	0.62	0.56	0.62	-0.39	-0.42	-0.27	-0.31	-0.41	-0.42	-0.35	-0.37
CWD 24	0.35	0.37	0.66	0.70	0.71	0.74	0.71	0.74	-0.42	-0.41	-0.32	-0.32	-0.43	-0.42	-0.37	-0.37
SDII 1	0.19				-0.26	-0.29	-0.26	-0.29								
SDII 3	0.62	0.58	0.23	0.22		-0.16		-0.16								
SDII 6	0.73	0.74	0.22	0.22	0.26	0.20	0.26	0.20	-0.16				-0.18	-0.15	-0.16	
SDII 12	0.24	0.25	0.51	0.51	0.54	0.50	0.54	0.50	-0.39	-0.37	-0.27	-0.25	-0.43	-0.41	-0.38	-0.36
SDII 24	0.35	0.35	0.64	0.66	0.66	0.67	0.66	0.66	-0.47	-0.51	-0.38	-0.42	-0.51	-0.54	-0.47	-0.50
RX1DAY 1	0.24	0.19			-0.24	-0.26	-0.24	-0.26								
RX1DAY 3	0.64	0.61	0.26	0.24												
RX1DAY 6	0.74	0.73	0.25	0.22	0.28	0.22	0.28	0.22	-0.18	-0.18			-0.20	-0.18	-0.18	-0.17
RX1DAY 12	0.31	0.25	0.60	0.47	0.61	0.50	0.61	0.50	-0.48	-0.44	-0.36	-0.34	-0.51	-0.47	-0.46	-0.43
RX1DAY 24	0.36	0.30	0.68	0.57	0.73	0.65	0.73	0.65	-0.56	-0.55	-0.46	-0.47	-0.58	-0.56	-0.53	-0.52
RX5DAYS 1	0.23	0.21			-0.24	-0.24	-0.24	-0.24								
RX5DAYS 3	0.66	0.65	0.29	0.27												
RX5DAYS 6	0.75	0.74	0.26	0.24	0.29	0.27	0.29	0.27	-0.20	-0.19			-0.21	-0.20	-0.19	-0.19
RX5DAYS 12	0.31	0.28	0.59	0.53	0.62	0.56	0.62	0.56	-0.50	-0.47	-0.39	-0.38	-0.53	-0.50	-0.50	-0.48
RX5DAYS 24	0.36	0.33	0.68	0.61	0.72	0.68	0.72	0.68	-0.55	-0.55	-0.46	-0.48	-0.57	-0.57	-0.53	-0.54
R20mm 1					-0.22	-0.23	-0.22	-0.23								
R20mm 3	0.64	0.58	0.33	0.31		-0.17		-0.17								
R20mm 6	0.78	0.76	0.29	0.26	0.27	0.19	0.27	0.19	-0.21	-0.18			-0.21	-0.18	-0.21	-0.19
R20mm 12	0.37	0.35	0.65	0.60	0.67	0.59	0.68	0.59	-0.54	-0.50	-0.43	-0.41	-0.55	-0.51	-0.51	-0.49
R20mm 24	0.38	0.36	0.71	0.68	0.78	0.74	0.78	0.74	-0.52	-0.51	-0.42	-0.42	-0.52	-0.50	-0.47	-0.46
DD 1	-0.27	-0.30			0.21	0.18	0.21	0.18								
DD 3	-0.68	-0.69	-0.27	-0.28												
DD 6	-0.75	-0.73	-0.28	-0.29	-0.33	-0.38	-0.33	-0.38	0.21	0.23			0.22	0.23	0.18	0.20
DD 12	-0.35	-0.34	-0.65	-0.65	-0.66	-0.65	-0.66	-0.65	0.49	0.48	0.37	0.37	0.50	0.50	0.44	0.44
DD 24	-0.36	-0.34	-0.67	-0.66	-0.75	-0.74	-0.75	-0.74	0.50	0.50	0.41	0.42	0.51	0.51	0.45	0.47
CDD 1	-0.37	-0.39			0.22	0.21	0.22	0.21								
CDD 3	-0.67	-0.68	-0.23	-0.23												
CDD 6	-0.71	-0.69	-0.23	-0.22	-0.32	-0.33	-0.32	-0.33	0.17	0.16			0.17	0.17		
CDD 12	-0.26	-0.23	-0.52	-0.45	-0.54	-0.45	-0.54	-0.45	0.31	0.25	0.20	0.16	0.33	0.27	0.26	0.22
CDD 24	-0.29	-0.23	-0.55	-0.45	-0.60	-0.51	-0.60	-0.51	0.32	0.30	0.25	0.25	0.34	0.32	0.31	0.30

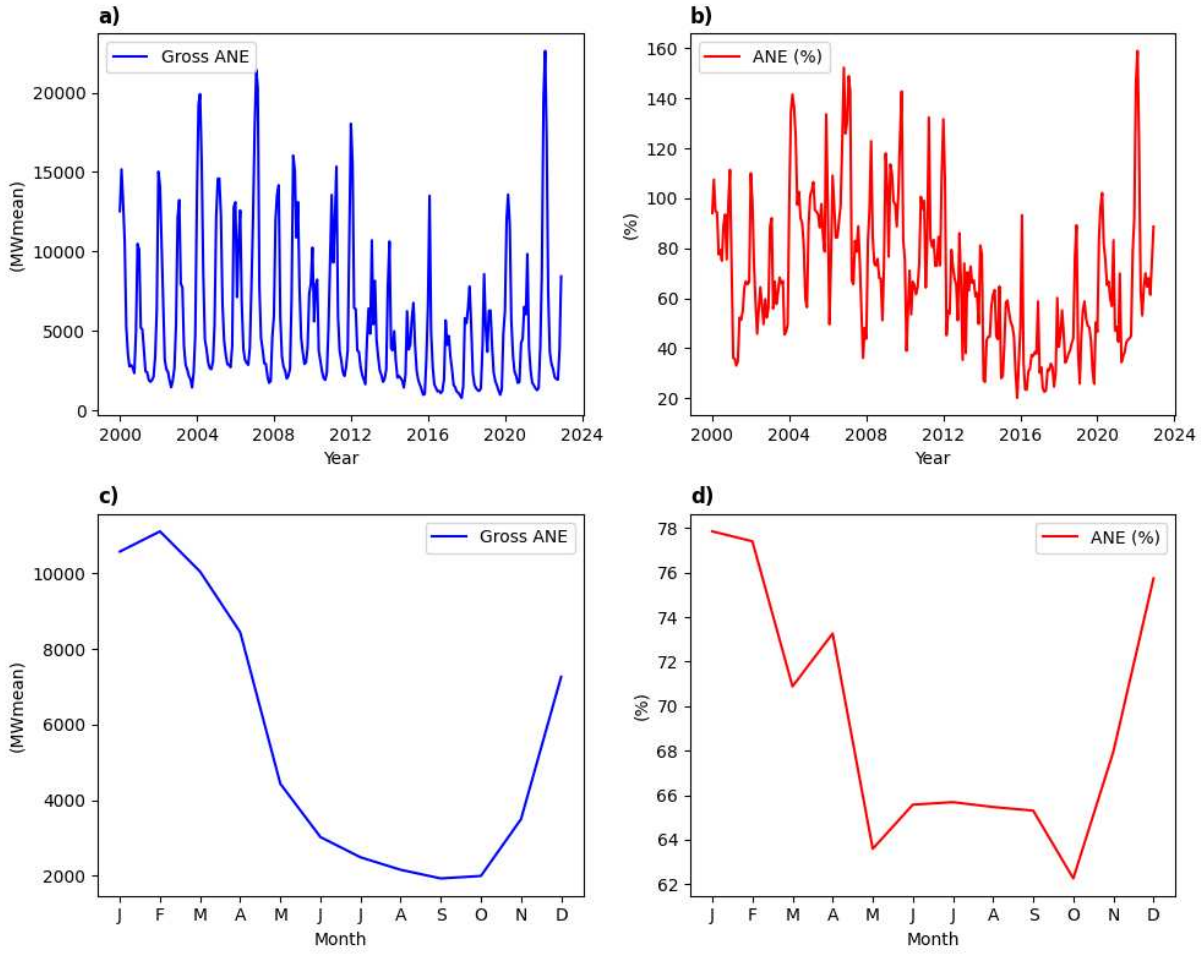
a: BR-DWGD, *b*: Era5-Land

Legend		
	Strong Positive Correlation	$r > 0.7$
	Moderate Positive Correlation	$0.4 > r > 0.7$
	Weak or Null Correlation	$-0.4 > r > 0.4$
	Moderate Negative Correlation	$-0.7 > r > -0.4$
	Strong Negative Correlation	$-0.7 > r$

4.2.2. Affluent Natural Energy (ANE)

In Figure 8, the historical time series and monthly averages of Gross ANE and Percentage of ANE are presented from January 2000 to December 2022. Both time series curves (Figure 8a and b) show a cyclic oscillation, which can be explained by the monthly averages of the historical series (Figure 8c and d) and their relationships with extreme climate precipitation indices.

Figure 8 – Affluent Natural Energy



Legend: **a)** Monthly Time Series of Gross ANE; **b)** Monthly Time Series of ANE (%); **c)** Monthly Average of Gross ANE**; d)** Monthly Average of ANE (%).

The influence of extremes in the São Francisco river basin is observed through the short-term response of Gross Affluent Natural Energy, with stronger correlations occurring in periods of three to six months (Table 6). This relationship reveals a well-defined seasonal behavior, with minimum values recorded in September, the last month of the driest quarter, at approximately 2000 MW.mean. The values of Gross

ANE increase in the following months and peak in February, the fourth month of the rainy season, at over 10000 MW.mean, before decreasing in the subsequent months.

This can be explained by the impact of extremes in the basin, which cause changes in the natural flow of the main river channel in short periods. On the other hand, percentage of ANE, which incorporates long-term averages in its determination, shows long-term responses, ranging from 12 to 24 months.

Seasonality was also observed by Vilar *et al.* (2020), according to the authors, the time series shows a decrease in observed ANE values due to the dry season. Percentage of ANE also exhibits seasonality, but with changes throughout the year. The lowest monthly average value, around 62%, is recorded in October, a transitional month between the dry and rainy periods in the basin. From there, the values gradually increase to reach a maximum of approximately 78% in January, the third month of the rainy season. Subsequently, the values decrease, with increases observed in April and June compared to the previous month.

These increases may be related to the extreme R20mm, which is well correlated with the percentage of ANE, when it is accumulated over 24 months (Table 6). In May, on average, up to two events of intense precipitation can occur in the basin (Figure 6g), influencing the natural river flow in the subsequent month.

The slight increase observed in June is likely associated with the total precipitation in the lower São Francisco hydrographic meso-region and the eastern extreme of the sub-middle São Francisco from May to August (Figure 4). Despite being predominantly dry months in other meso-region of the basin, the Paulo Afonso hydroelectric complex and the Xingó HPP are located in these regions (Figure 2), which use run-of-the-river reservoirs for energy generation, contributing significantly to ANE (%) in the basin.

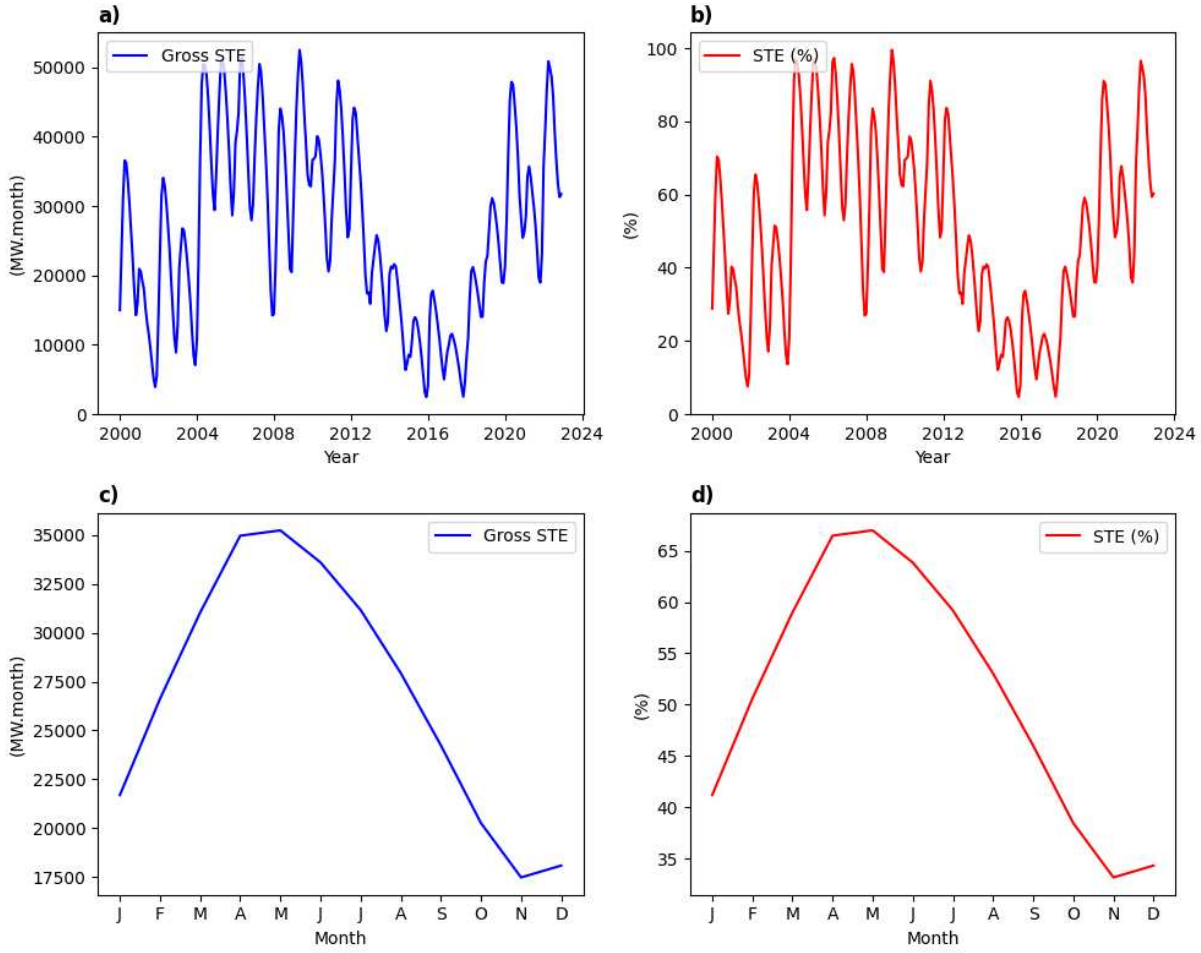
It is also worth noting that ANE (%) is based on a long-term mean (LTA) that has been verified since 1931 and, therefore, may contain climatological patterns not identified in the temporal scope of this study.

4.2.3. Stored Energy (STE)

In Figure 9, the historical time series of Gross Energy Generation (STE) and Percentage of STE, as well as their monthly averages are presented for the January 2000 to December 2022 interval. It is notable that both the time series (Figures 9a and 9b) and the monthly average curves (Figures 9c and 9d) exhibit identical behavior.

This is because, throughout the analyzed period, the maximum STE of the São Francisco river basin remained constant at 52,727 MW.month, according to information from ONS (2023). The cyclic oscillation of the time series can be explained based on their respective monthly averages and their association with extreme precipitation climate indices.

Figure 9 – Stored Energy



Legend: **a)** Monthly Time Series of Gross STE; b) Monthly Time Series of STE (%); c) Monthly Average of Gross STE; d) Monthly Average of STE (%).

The response of STE (gross and percentage) to extremes shows a long-term relationship with precipitation extremes, as stronger correlations occurred in 24-month periods (Table 6). The Gross STE and Percentage of STE (Figures 9c and 9d, respectively) patterns exhibit a clear seasonal cycle. The minimum value is observed in November, at the beginning of the rainy season, at approximately 17,500 MW.month, corresponding to 35% of the maximum energy production capacity of reservoir-based power plant.

Over the following months, this value gradually increases, reaching its maximum in May, during the dry period of the basin, at around 35,000 MW.month, representing approximately 66% of the percentage of STE. This behavior can be attributed to the control of water levels in the reservoirs, which have a considerable storage volume. This provides the system with resilience to variations in water availability over short periods of time.

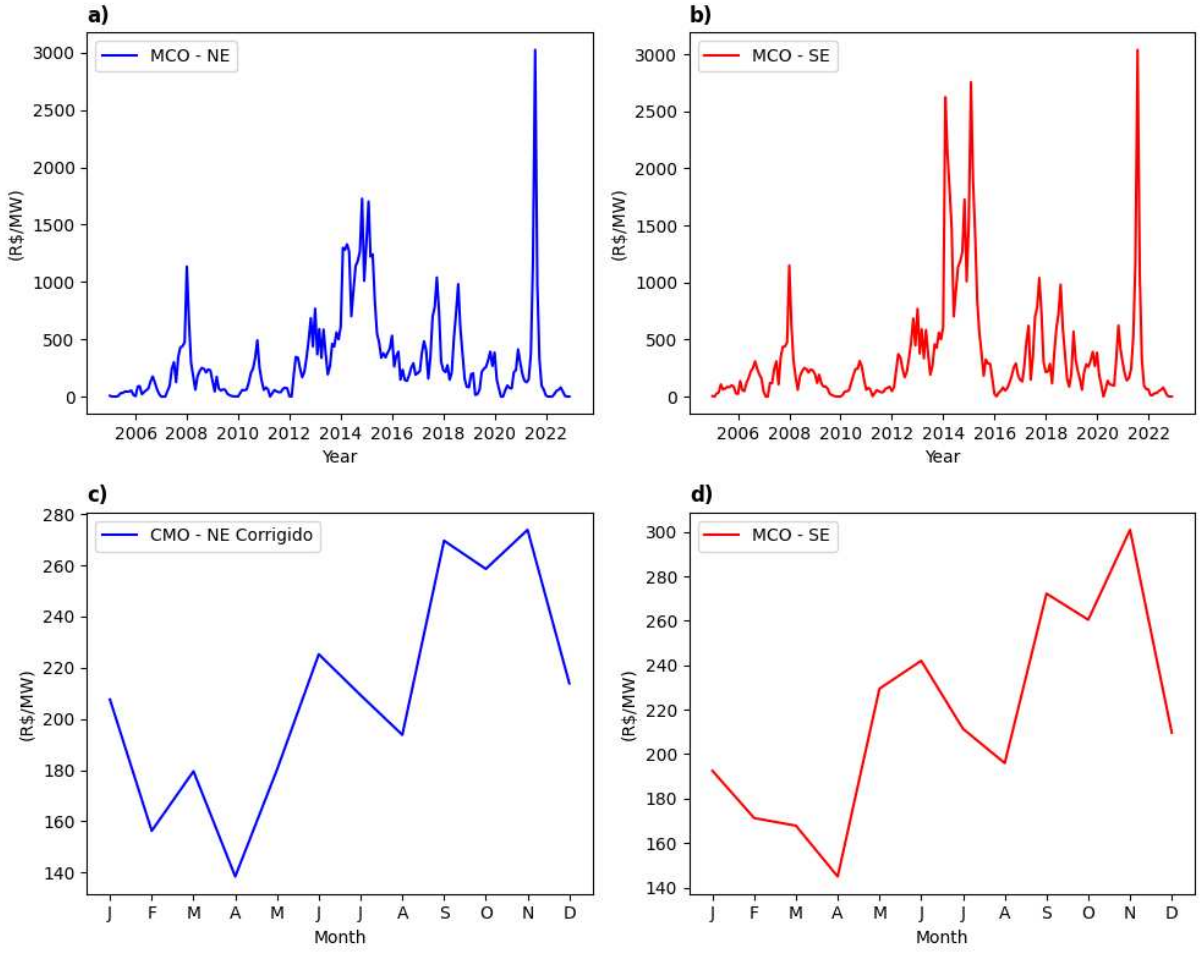
Analyses of the average monthly values of ANE (Figure 8) and STE (Figure 9) reveal a significant difference in magnitude between the two quantities. This difference is mainly due to the fact that, in addition to differences in working volumes, HPP with flow control reservoirs have the ability to store water and therefore generate electrical energy during periods of high demand (STE), while run-of-the-river HPP have their electricity production limited by the available river flow (ANE).

The correlations between the MCO and SPD variables with precipitation extremes in the two analyzed subsystems were moderate (Table 6), mainly because the observed climate extremes in the São Francisco river basin represent only a portion of the water resources that contribute to energy production in each of them. However, moderate correlations are found for the indices accumulated over 12 and 24 months, suggesting that both MCO and SPD also respond in the long term to the occurrence of extremes in the region.

4.2.4. Marginal Cost of Operation (MCO)

In Figure 10, the historical time series of Marginal Cost of Operation (MCO), with monetary values adjusted as of December 2022, are presented. The series cover the period from January 2005 to December 2022. Analyzing the curves of the Marginal Cost of Operation of the Northeast subsystem (MCO-NE) (Figure 10a) and the Southeast/Central-West subsystem (MCO-SE) (Figure 10b), similarities can be observed, which indicate the regional interaction of the electricity sector and the influence of ANE and STE availability in their respective subsystems. It is important to note that only part of this influence comes from the São Francisco river basin.

Figure 10 – Marginal Cost of Operation



Legend: **a)** Monthly Time Series of MCO-NE; **b)** Monthly Time Series of MCO-SE, **c)** Monthly Average of MCO-NE; **b)** Monthly Average of MCO-SE.

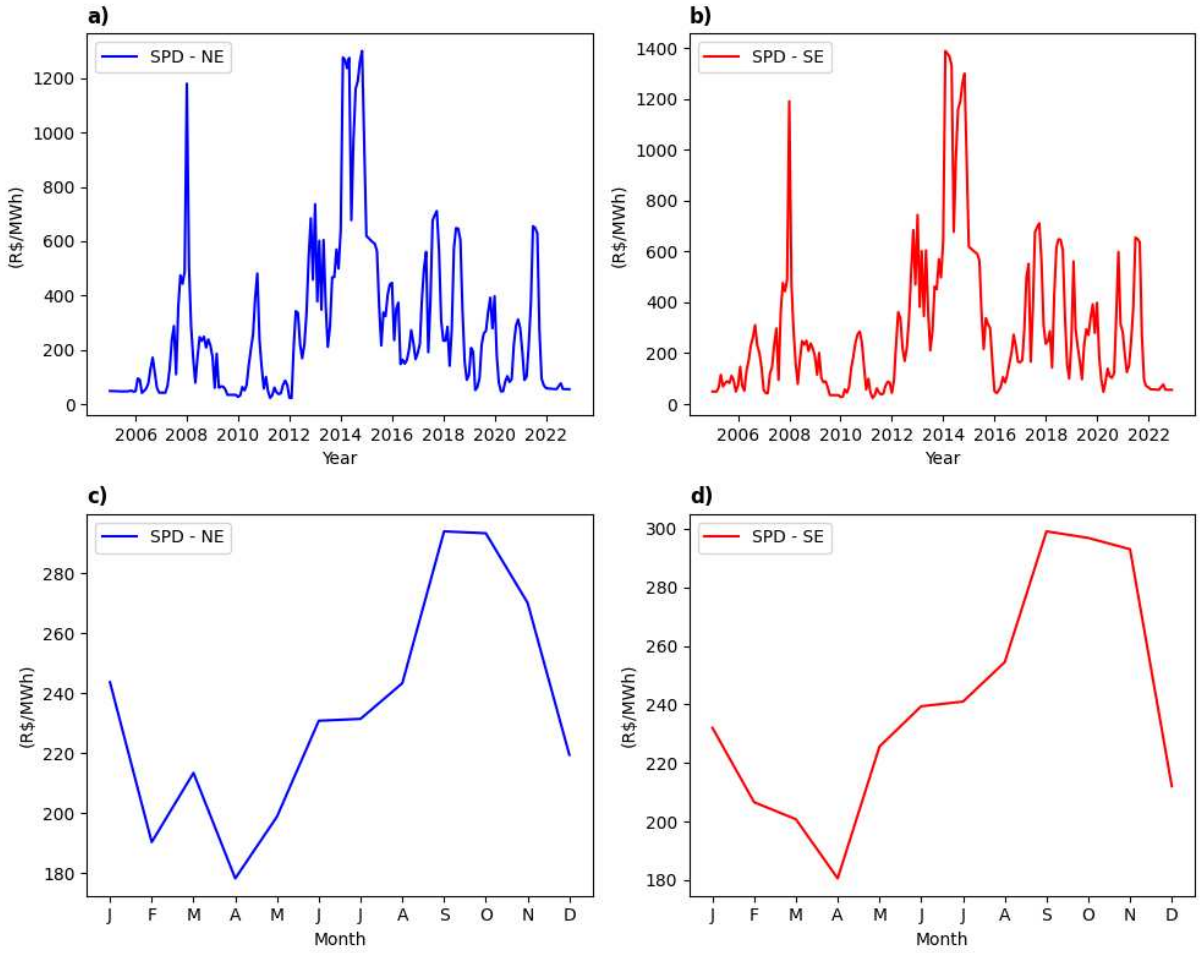
The MCO-NE (Figure 10c) and MCO-SE (Figure 10d) reach their maximum average values in the month of November, during the onset of the rainy season and corresponds to the minimum values of STE. MCO-NE reaches approximately 260 R$/MW, while MCO-SE reaches around 300 R$/MW. On the other hand, the minimum average values occur in April, shortly after the end of the rainy season and one month before the peak of STE. For both subsystems, these minimum values are around 140 R$/MW. These results indicate that MCO can be influenced by precipitation and extreme events that occur during the rainy period of the basin.

It may be noticed that monthly averages of MCO do not follow a continuous trend of growth or decline, showing periods of increase and decrease between the maximum and minimum peaks in both subsystems. This is justified by the different origins of the costs in relation to operational decisions.

4.2.5. Settlement Price of Differences (SPD)

Figure 11 illustrates the historical series of the Settlement Price of Differences (SPD), along with MCO, with monetary values adjusted to December 2022. These series cover the period from January 2005 to December 2022.

Figure 11 – Settlement Price of Differences



Legend: **a)** Monthly Time Series of SPD-NE; **b)** Monthly Time Series of SPD-SE, **c)** Monthly Average of SPD-NE; **b)** Monthly Average of SPD-SE.

Similar to what was observed with MCO, it is evident that the behavior of SPD in the Northeast subsystem (SPD-NE) (Figure 11a) and Southeast/Central-West subsystem (SPD-SE) (Figure 11b) is analogous, demonstrating the interaction between the subsystems and the influence of MCO, which is one of the input data for calculating this price.

It can be observed that the monthly averages of both SPD-NE (Figure 11c) and SPD-SE (Figure 11d) have minimum values in April, which marks the transition

between the rainy and dry periods in the basin. These minimum values are around 180 R$/MW for both subsystems and increase until the month of September, which represents the transition between the dry and rainy periods, reaching the maximum average values in both subsystems, at approximately 290 R$/MW, and then, decreasing in the following months. This distribution, similar to that of MCO, indicates that SPD is influenced by precipitation and extreme events that occur during the rainy period in the basin.

According to Vilar *et al.* (2020), the intense presence of the SACZ phenomenon results in an increase in ANE in the Northeast subsystem, which leads to energy trading at lower prices (lower SPD). On the other hand, when the intensity of this phenomenon is weak, it affects the entire National Interconnected System (SIN), resulting in higher levels of energy trading (higher SPD).

There is an increase in the monthly average of SPD in January compared to December in both subsystems, which can be explained by an increase in load demand in that month due to cultural reasons. Additionally, there is a slight increase in the monthly average of SPD-NE in March compared to the previous month, a behavior strongly influenced by the seasonal pattern of MCO-SE, which exhibits a similar pattern in those months.

4.2.6. Trends of Energy Variables

Results obtained based on the Mann-Kendall trend test for the energy variables are presented in Table 7.

Table 7 - Non-parametric tests for the energy variables

Variable	Trend	*p-value*	Sens's Slope
Gross ANE	decreasing	0.0403	-238.63 MW.month/year
ANE (%)	decreasing	0.0403	-0.45%/year
Gross STE	decreasing	0.0001	-82.02 MW.month/year
STE (%)	decreasing	0.0001	-1.69 %/year
MCO-NE	increasing	0.0111	6.95 R$/MW/year
MCO-SE	increasing	0.0143	5.66 R$/MW/year
SPD-NE	increasing	0.0002	4.23 R$/MW/year
SPD-SE	increasing	0.0001	5.09 R$/MW/year

The ANE and STE deliver decreasing trends over the study period, indicating a reduction in water and electrical energy availability in the São Francisco river basin. On the other hand, the variables SPD and MCO demonstrate increasing trends, indicating higher electricity prices and a greater market demand for the two analyzed subsystems. This suggests the need to invest in management strategies to ensure the long-term sustainability of the electrical system.

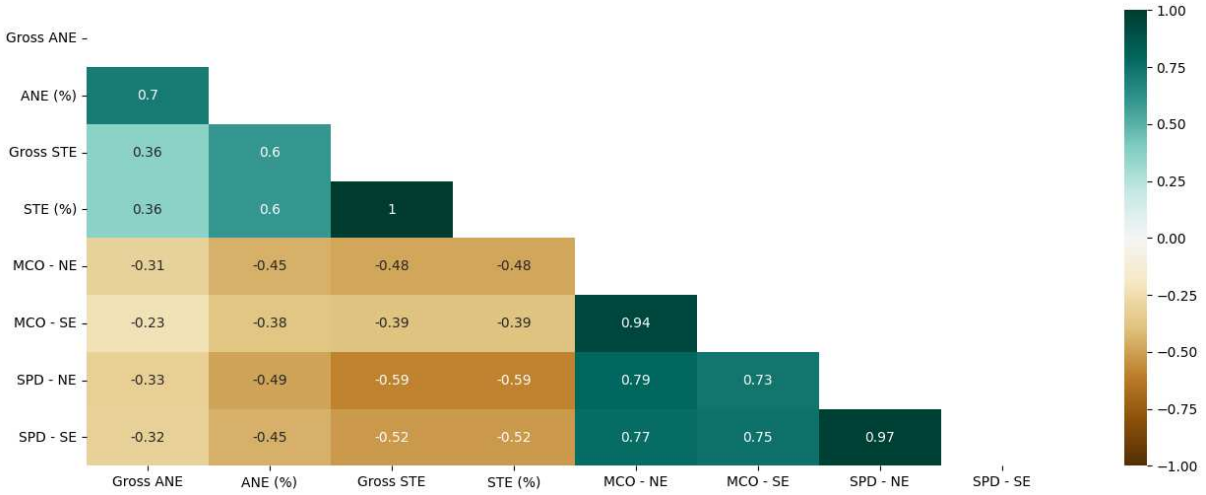
Similarly, Silva *et al.* (2021) observed negative trends when analyzing time series of natural inflows into the Itaparica and Sobradinho reservoirs. The authors attributed these trends to the occurrence of droughts recorded in recent years in the São Francisco river basin.

4.2.7. Correlations between Energy Variables

Significant correlations ($p \leq 0.05$) between energy variables are shown in Figure 12.

Figure 12 shows significant relationships between the energy variables. The perfect correlation between STE and percentage STE (1) occurs due to the absence of changes in the maximum STE of the São Francisco basin over the study period. The strong correlations between Gross ANE and percentage ANE (0.7), as well as between Gross STE and percentage ANE (0.6), are expected since these variables are interdependent and equally influenced by extreme precipitation events throughout the São Francisco basin.

Figure 12 – Pearson correlation (*r*) between energy variables



The strong correlations resulting from the interactions between Marginal Cost of Operation (MCO) and Settlement Price of Differences (SPD) in the two studied subsystems reflect the interaction of these subsystems within the National Interconnected System (SIN). There seems to be a balance regarding the disparity of energy prices between the subsystems, which requires additional information beyond hydrological and climatic factors for a better explanation.

In particular, moderately negative correlations are observed between percentage ANE and SPD-NE (-0.49) and SPD-SE (-0.49), suggesting that the commercialization prices in both subsystems may be influenced by the percentage ANE conditions of the São Francisco basin. The same pattern is observed in relation to percentage ANE and MCO-NE (-0.45), highlighting the importance of the run-of-river hydroelectric complex of Paulo Afonso and Xingó, in the composition of the operational costs of the Northeast subsystem.

Moderately negative correlations are also noteworthy between STE (gross and percentage) and SPD-NE (-0.59) and SPD-SE (-0.52), indicating that the STE of the São Francisco basin plays a significant role in determining the energy commercialization prices in both Northeast and Southeast subsystems. The same pattern is observed for STE (gross and percentage) and MCO-NE (-0.48), which shows the importance of the reservoirs in the São Francisco basin in determining the operational costs of the Northeast subsystem. These correlations indicate that variations in the availability of ANE and STE in the São Francisco basin can directly influence the behavior of operational costs and electricity prices, in the Northeast and Southeast/Central-West subsystems.

Vilar *et al.* (2020) also found these interactions between ANE, STE, and SPD. According to the study, reductions in flows and storage had a direct impact on the price of energy during periods of severe drought in the Northeast. At these times, energy in the NE subsystem was traded at a higher price compared to other subsystems.

4.3. Use of Artificial Intelligence tool for generating predictive regression models

In the following Artificial Intelligence tools are used for generating predictive regression models. Results presented in this section provide a comparative analysis of evaluation metrics for the predictive regression models applied to two datasets: BR-DWGD, represented in blue, and ERA5-Land, represented in green. The evaluated

metrics include Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), Modified Kling-Gupta Efficiency (KGE), and Willmott's concordance index *(d)*. Each metric is presented in six boxplots, representing the Random Forest (RF), Artificial Neural Networks (ANN), and K-Nearest Neighbors (KNN) models. For all models, the input data consisted of accumulated extreme indices with moderate or strong correlation *(r)* with the corresponding energy variable (Table 6). The optimal parameters used for generating the predictive regression models are presented in Table 8.

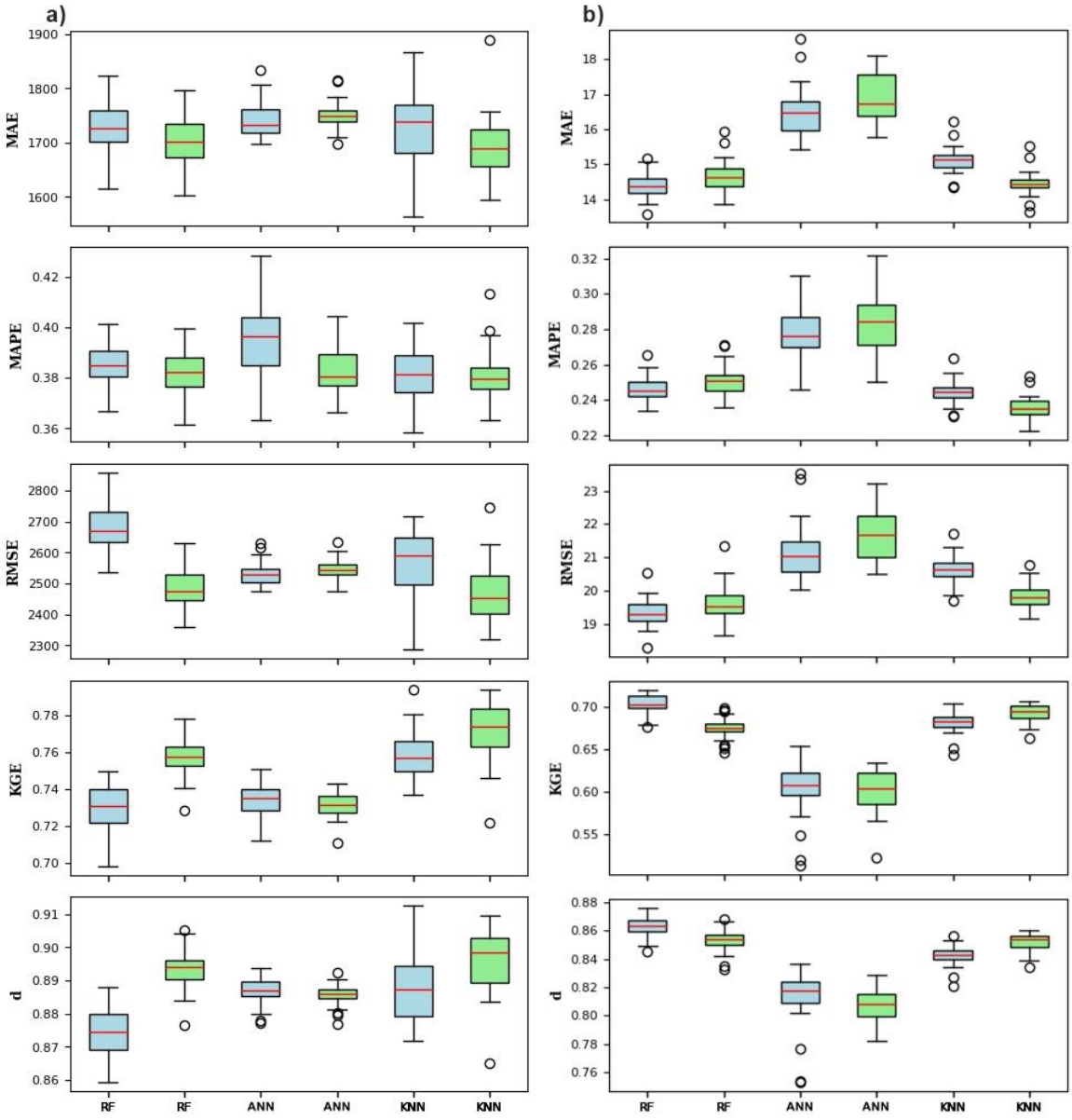
Table 8 - Optimal parameters used in the prediction modeling

		RF		ANN		KNN
		criterion	*n_estimators*	*batch_size*	*max_iter*	*n_neighbors*
Gross ANE	BR-DWGD	*squared_error*	100	50	500	5
	ERA5-Land	*absolute_error*	150	10	500	5
ANE (%)	BR-DWGD	*absolute_error*	100	90	200	30
	ERA5-Land	*absolute_error*	200	50	200	10
Gross STE	BR-DWGD	*absolute_error*	150	10	500	20
	ERA5-Land	*absolute_error*	50	10	500	30
STE (%)	BR-DWGD	*absolute_error*	100	50	200	20
	ERA5-Land	*absolute_error*	100	10	200	30
MCO-NE	BR-DWGD	*absolute_error*	50	90	500	30
	ERA5-Land	*squared_error*	150	50	500	30
MCO-SE	BR-DWGD	*absolute_error*	150	50	200	30
	ERA5-Land	*absolute_error*	200	10	200	30
SPD-NE	BR-DWGD	*absolute_error*	100	90	500	30
	ERA5-Land	*squared_error*	10	90	200	30
SPD-SE	BR-DWGD	*absolute_error*	100	10	200	30
	BR-DWGD	*absolute_error*	200	50	200	30

The optimal parameters were adopted for the regression of the corresponding energy variables. Figure 13 shows the results for Gross ANE and ANE (%). By analyzing Figure 13a, the predictive models for Gross ANE yielded close and satisfactory results for both datasets. The K-Nearest Neighbors (KNN) model demonstrated the best performance in all metrics for the ERA5-Land. For the BR-DWGD dataset, the KNN model also delivers the best performance in terms of KGE, d, and MAPE, but with relatively large dispersion.

Regarding the predictions of ANE (%) (Figure 13b), it can be observed that the results were similar and acceptable for all three models in both datasets. The Random Forest (RF) model showed the best performance with the BR-DWGD data in all metrics. On the other hand, for the ERA5-Land dataset, the performance of the RF model was comparable to that of the KNN model.

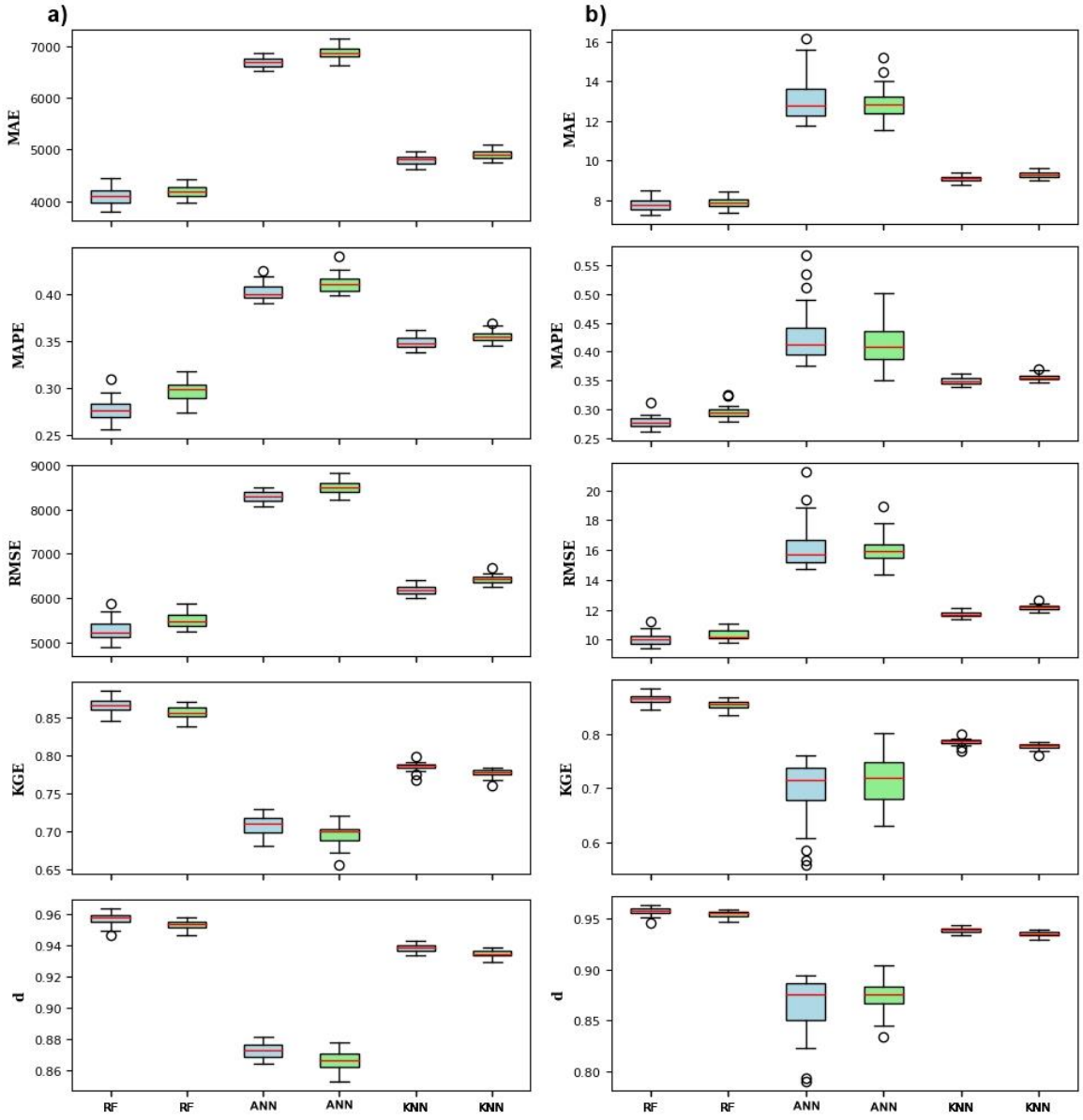
Figure 13 – Regression Model Performance for ANE Prediction



Legend: **a)** Gross ANE, **b)** ANE (%). The blue boxplots represent results from the BR-DWGD dataset, while the green ones depict results from the ERA5-Land dataset.

The estimations of energy accumulation in both gross form (Figure 14a) and percentage form (Figure 14b), also exhibited satisfactory results for the BR-DWGD and ERA5-Land datasets. It can be observed that, except for ANN in STE (%), the dispersion of predicted values is relatively low compared to the ANE estimations.

Figure 14 – Regression Model Performance for STE Prediction



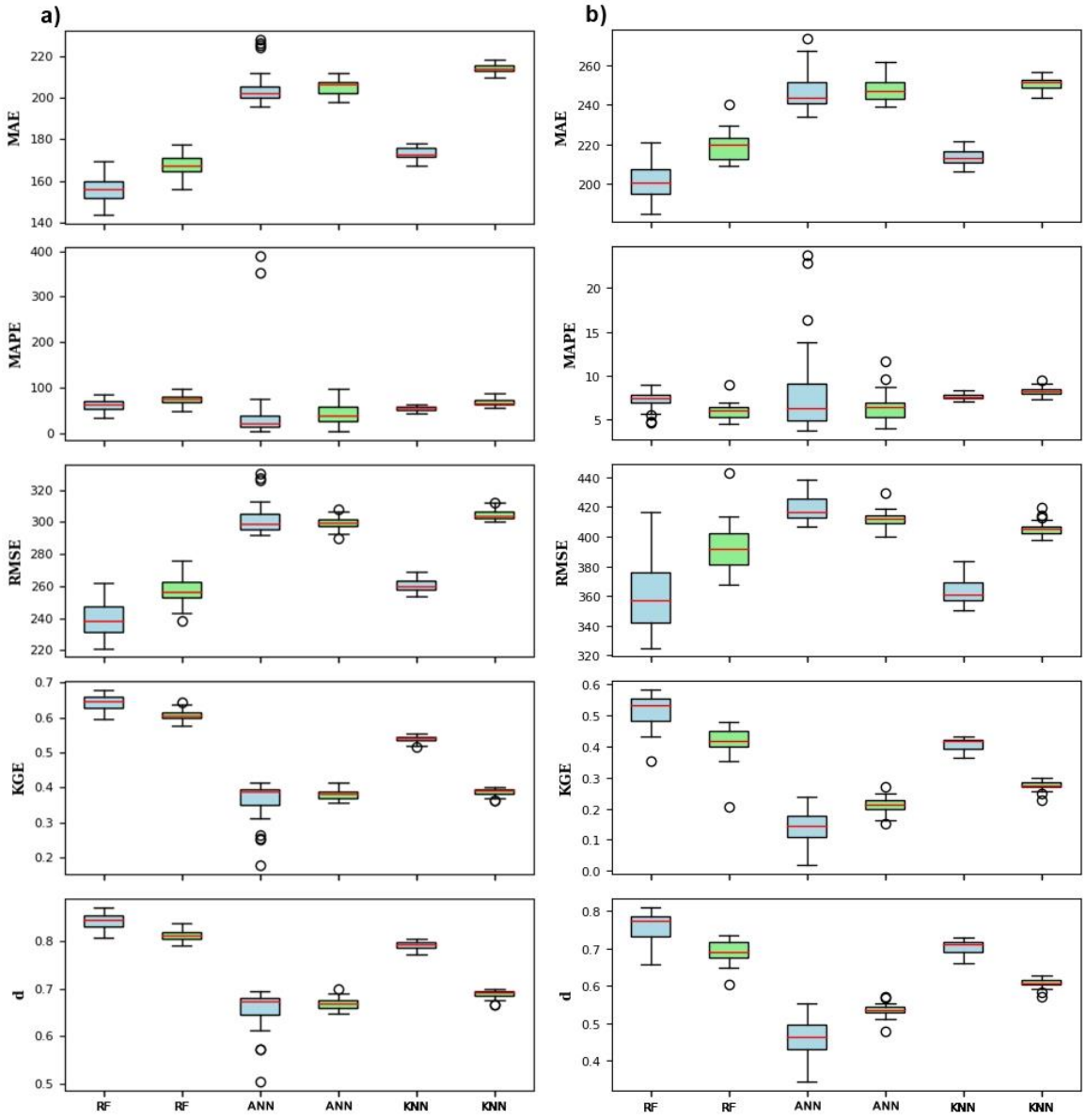
Legend: **a)** Gross STE, **b)** STE (%). The blue boxplots represent results from the BR-DWGD dataset, while the green ones depict results from the ERA5-Land dataset.

The RF model delivers the best performance for STE predictions in all metrics for both datasets. It is noteworthy that, for both gross STE and STE in percentage, the

values of d and KGE are close to optimal values, with a low MAPE ranging between 25% and 30%.

The analysis of Figure 15 reveals that the performance metrics are not satisfactory for the MCO (Marginal Operating Cost) of the two studied subsystems. Despite the RF model presenting acceptable values of KGE and d with the BR-DWGD data, the magnitudes of MAE, MAPE, and RMSE are poor for all other combinations.

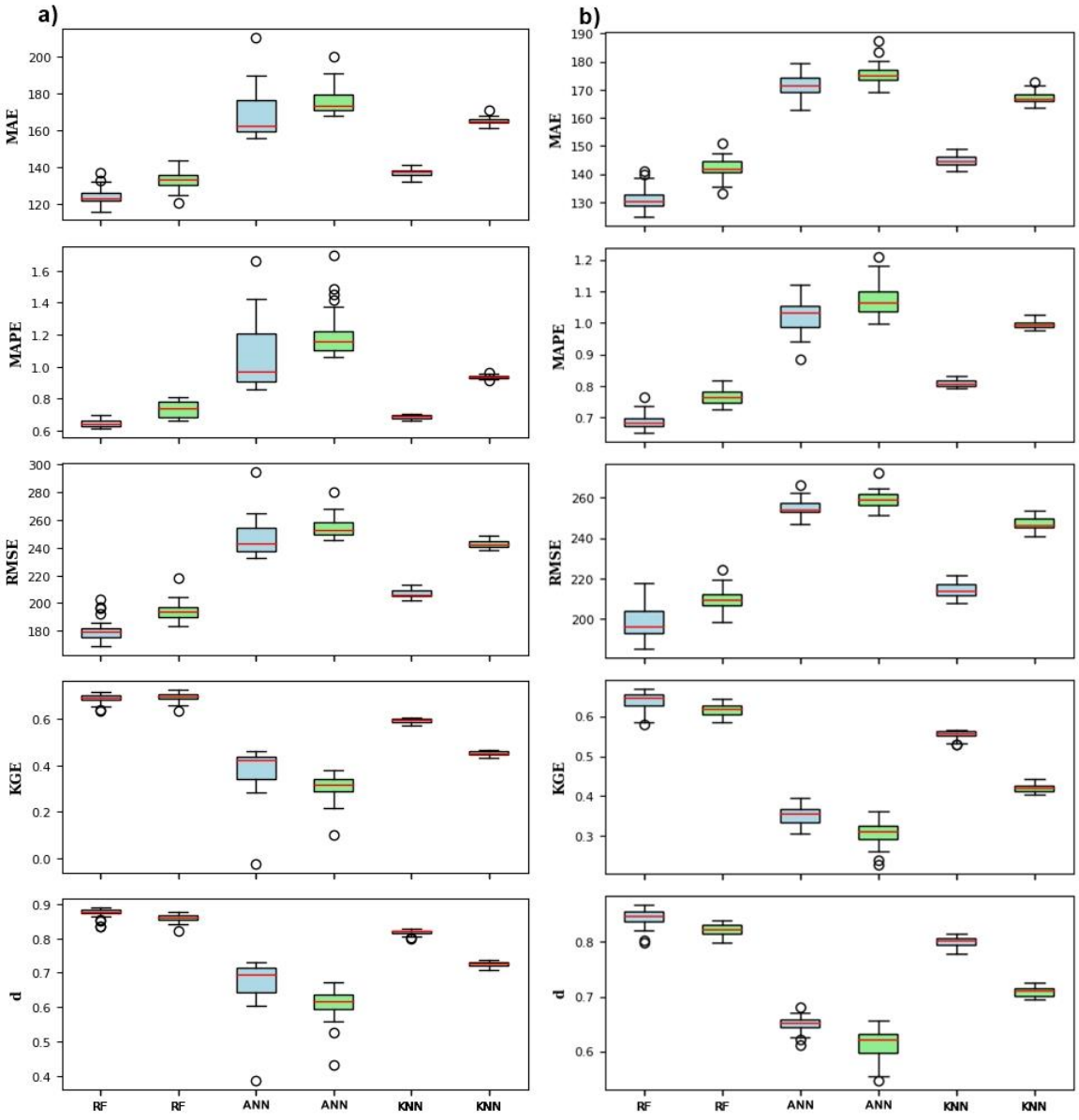
Figure 15 – Regression Model Performance for MOC Prediction



Legend: **a)** MCO-NE, **b)** MCO-SE. The blue boxplots represent results from the BR-DWGD dataset, while the green ones depict results from the ERA5-Land dataset.

The SPD estimation models (Figure 16) show weak performance for the Northeast and Southeast/Central-West subsystems, similar to the MCO prediction results. The RF model with the BR-DWGD dataset presented the best values for KGE and d, but the MAPE value around 70%, combined with large magnitudes in MAE and RMSE, does not allow for considering it as a reasonable performer.

Figure 16 – Regression Model Performance for SPD Prediction



Legend: **a)** SPD-NE, **b)** SPD-SE. The blue boxplots represent results from the BR-DWGD dataset, while the green ones depict results from the ERA5-Land dataset.

The results presented in this section indicate that extreme precipitation events are directly related to both the Energy Stored (STE) in reservoirs and the Affluent Natural Energy (ANE) in the São Francisco river Basin. Furthermore, it is observed that it is feasible to efficiently estimate STE and ANE using only information on extreme precipitation indices in the basin.

This finding has significant relevance for the management of the National Interconnected System (SIN), as Vilar *et al.* (2020) identified considerable differences between the projected values of ANE and STE by the models used by ONS (National Electric System Operator) and the observed values. According to Vilar *et al.* (2020), this indicates the necessity for adjustments in the ONS methodology, especially in severe drought situations. In this context, the estimation methodology proposed in the present study emerges as a promising alternative.

The limitation of MCO and SPD predictions is justified by the various parameters that influence their calculations, including forecasting future hydrological scenarios, the geographic dimensions that delimit each subsystem, as well as variables that go beyond physical and climatic aspects, such as contractual obligations, transmission line availability, limits established by legislation, supply, and demand.

Extreme climate events directly impact hydropower production; however, MCO and SPD values are not always determined solely by these resources. Consequently, estimating MCO and SPD values for these subsystems, based only on extreme precipitation events occurring within the geographic limits of the São Francisco river basin is unfeasible.

A study conducted by Santos *et al.* (2022) proposes a Machine Learning-based approach for predicting time series applied to SPD forecasting. The authors highlight that this line of research is promising, as the obtained projections have a good fit with the actual values. Additionally, they demonstrated that the applied models by ONS have an ‘optimistic’ bias in the long term. Therefore, it is crucial to conduct new studies that aim to anticipate reduce the bias more accurately, including other variables such as ANE and STE, which may reflect potential future water crises.

Understanding the relationships between extreme climate events and energy variables, especially the ability to efficiently estimate ANE and STE, based on information on extreme precipitation events, presents significant potential for future applications in energy production planning in Brazil.

These findings are particularly relevant when considering extreme events predicted for the future, such as prolonged droughts or heavy rainfall, which can have substantial impacts on energy supply. The use of such information can contribute to more resilient and adaptable energy planning, that takes into account challenges arising from extreme climate events.

Although the evaluated models in this study have demonstrated efficiency in estimating STE and ANE based on information on extreme precipitation, it is important to note that the inclusion of other climate variables can further enhance the accuracy of predictive models. In addition to precipitation, factors such as temperature and regional climate patterns play crucial roles in energy production. Therefore, future research exploring the integration of other climate variables into the models, can improve the ability to accurately predict energy availability in situations of extreme climate events.

5. CONCLUSIONS

Analysis of extreme precipitation events in the São Francisco river basin reveals distinct patterns in different regions. The Upper São Francisco meso-region and the western portion of the middle São Francisco are the most impacted by extreme rainfall events, which is related to the rainy season in the basin.

On the other hand, the sub-middle São Francisco and the eastern region of the middle São Francisco are more susceptible to extreme drought events, recording the highest magnitudes throughout the year. Trend tests indicate a reduction in total precipitation and the number of rainy days, as well as an increase in the number of dry days. Seasonal analysis of Energy Stored (STE) and Affluent Natural Energy (ANE) reveals a behavior influenced by precipitation extremes on different time scales. The short-term responses of Gross ANE levels and the long-term responses of ANE (%) and STE (both gross and percentage), are evident in response to extreme climatic events.

In contrast, a well-defined seasonality is not observed for Marginal Operating Cost (MCO) or Settlement Price Differences (SPD). However, the interaction of variables between the Northeast and Southeast/Central-West subsystems is notable, as evidenced by the similar values and strong correlations between them. Changes in the availability of ANE and STE influence the operational costs and electricity trading prices in the studied subsystems. This fact is supported by the observed downward trends in ANE and STE, contrasting with the upward trends in MCO and SPD.

The regression models based on artificial intelligence methods demonstrated efficient performance in estimating ANE and STE based on extreme precipitation events in the basin. These results pave the way for further research and practical applications, including forecasts of future energy conditions based on climate extremes, and the consideration of additional weather variables to enhance the efficiency of the models.

These findings are relevant to the planning and management of the electricity sector, especially in relation to strategic decision-making and the development of public policies aimed at ensuring the energy security of the country.

REFERENCES

ABREU, Wagner Saboia de. *Modelagem e Previsão de Preços à Vista de Energia Elétrica e Aplicações no Contexto de Investimentos sob Incerteza.* 2012. 110 f. Dissertation (Master's degree) - *Curso de Engenharia Generalista, Pontifícia Universidade do Rio de Janeiro* (Puc-Rio), Rio de Janeiro, 2012. Available in: https://web.bndes.gov.br/bib/jspui/bitstream/1408/10494/2/Mestrado_Wagner%20Saboia%20de%20Abreu_P_PO.pdf. Access at: May 24, 2023.

AGUIAR, Louise da Fonseca *et al.* Social and environmental vulnerability in Southeast Brazil associated with the South Atlantic Convergence Zone. Natural Hazards, [S.L.], v. 109, n. 3, p. 2423-2437, 14 jul. 2021. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s11069-021-04926-z.

ANARAKI, Mahdi Valikhan *et al.* Uncertainty Analysis of Climate Change Impacts on Flood Frequency by Using Hybrid Machine Learning Methods. Water Resources Management, [S.L.], v. 35, n. 1, p. 199-223, November 20, 2020. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s11269-020-02719-w.

ANDRADE, João Maria de *et al.* A comprehensive assessment of precipitation products: temporal and spatial analyses over terrestrial biomes in northeastern brazil. Remote Sensing Applications: Society and Environment, [S.L.], v. 28, p. 100842, Nov. 2022. Elsevier BV. http://dx.doi.org/10.1016/j.rsase.2022.100842.

ARAGHINEJAD S (2013) Data-driven modeling: using MATLAB®. In: Water resources and environmental engineering, Springer, Dordrecht, The Netherlands

ARAÚJO, C.S.P. *et al.* Evaluation of air temperature estimated by ERA5-Land reanalysis using surface data in Pernambuco, Brazil. Environ Monit Assess 194, 381 (2022). https://doi.org/10.1007/s10661-022-10047-2.

ASSIS, Janaina Maria Oliveira de *et al.* Analysis of climate indices and impacts on the rainfall regime in the Sub-medium stretch of the São Francisco river Basin, Brazil. Revista Principia - Divulgação Científica e Tecnológica do IFPB, [S.L.], v. 59, n. 4, p. 1475, December 30. 2022. Instituto Federal de

Educação, Ciência e Tecnologia da Paraíba. http://dx.doi.org/10.18265/1517-0306a2021id5570.

AVILA-DIAZ, Alvaro *et al.* Assessing current and future trends of climate extremes across Brazil based on reanalyses and earth system model projections. Climate Dynamics, [S.L.], v. 55, n. 5-6, p. 1403-1426, Jul 15. 2020. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s00382-020-05333-z.

BRÊDA, João Paulo Lyra Fialho *et al.* Modeling coordinated operation of multiple hydropower reservoirs at a continental scale using artificial neural network: the case of brazilian hydropower system. Rbrh, [S.L.], v. 26, p. 1-12, 2021. FapUNIFESP (SciELO). http://dx.doi.org/10.1590/2318-0331.262120210011.

Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

CBHSF - *Comitê da Bacia Hidrográfica do Rio São Francisco. A Bacia*. Minas Gerais, 2022. Available in: https://cbhsaofrancisco.org.br/a-bacia/. Access at: August 30, 2022.

CHEN, Yifu *et al.* A data-driven binary-classification framework for oil fingerprinting analysis. Environmental Research, [S.L.], v. 201, p. 111454, Oct. 2021. Elsevier BV. http://dx.doi.org/10.1016/j.envres.2021.111454.

DAHER, M. J., & MARTINEZ, M. A. Plano da Operação Energética (PEN) (1st ed.). Operador Nacional do Sistema Elétrico (ONS). 2019. Available in: https://www.ons.org.br/AcervoDigitalDocumentosEPublicacoes/PEN_Executivo_2019-2023.pdf. Access at: May 23, 2023.

DAHER, Mário *et al. O SISTEMA INTERLIGADO NACIONAL E O PLANEJAMENTO DA OPERAÇÃO ENERGÉTICA*. In: OLIVEIRA, Francisco José Arteiro de (org.). *O planejamento da operação energética no Sistema Interligado Nacional: conceitos, modelagem matemática, previsão de geração e carga*. São Paulo: Artliber, 2020. p. 85-170. Available in: https://www.ons.org.br/AcervoDigitalDocumentosEPublicacoes/Livros-da-Diretoria-de-Planejamento-na-biblioteca-digital-ONS/LIVRO-O-PLANEJAMENTO-DA-OPERAÇÃO-ENERGÉTICA-NO-SISTEMA%20INTERLIGADO-NACIONAL.pdf. Access at: May 23, 2023.

DE JONG, Pieter *et al.* Hydroelectric production from Brazil's São Francisco river could cease due to climate change and inter-annual variability. Science of

The Total Environment, [S.L.], v. 634, p. 1540-1553, Sep. 2018. Elsevier BV. http://dx.doi.org/10.1016/j.scitotenv.2018.03.256.

FERREIRA, Pedro Guilherme Costa *et al.* The stochastic effects on the Brazilian Electrical Sector. Energy Economics, [S.L.], v. 49, p. 328-335, May 2015. Elsevier BV. http://dx.doi.org/10.1016/j.eneco.2015.03.004.

FRICH, P *et al.* Observed coherent changes in climatic extremes during the second half of the twentieth century. Climate Research, [S.L.], v. 19, p. 193-212, 2002. Inter-Research Science Center. http://dx.doi.org/10.3354/cr019193.

GOMES, Helber B. *et al.* Climatology of easterly wave disturbances over the tropical South Atlantic. Climate Dynamics, [S.L.], v. 53, n. 3-4, p. 1393-1411, Feb 15. 2019. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s00382-019-04667-7.

GUPTA, Hoshin V. *et al.* Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. Journal of Hydrology, [S.L.], v. 377, n. 1-2, p. 80-91, Oct. 2009. Elsevier BV. http://dx.doi.org/10.1016/j.jhydrol.2009.08.003.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). The elements of statistical learning (2nd ed.)., pp. 745. New York: Springer.

HIDALGO, Ieda Geriberto *et al.* Hydropower generation in future climate scenarios. Energy for Sustainable Development, [S.L.], v. 59, p. 180-188, Dec. 2020. Elsevier BV. http://dx.doi.org/10.1016/j.esd.2020.10.007.

HUSSAIN, Md.; MAHMUD, Ishtiak. PyMannKendall: a python package for non parametric mann kendall family of trend tests. Journal of Open Source Software, [S.L.], v. 4, n. 39, p. 1556, Jul 25. 2019. The Open Journal. http://dx.doi.org/10.21105/joss.01556.

IBGE - *Instituto Brasileiro de Geografia e Estatística. Inflação.* Brazil, 2023. Available in: https://www.ibge.gov.br/explica/inflacao.php. Access at: May 18, 2023.

Karl, T.R., Nicholls, N. & Ghazi, A. Clivar/GCOS/WMO Workshop on Indices and Indicators for Climate Extremes Workshop Summary. Climatic Change 42, 3–7 (1999). https://doi.org/10.1023/A:1005491526870

KENDALL, Maurice George. Rank correlation methods. 1948.

KLING, Harald *et al.* Runoff conditions in the upper Danube basinunder an ensemble of climate change scenarios. Journal Of Hydrology, [S.L.], v. 424-

425, p. 264-277, mar. 2012. Elsevier BV. http://dx.doi.org/10.1016/j.jhydrol.2012.01.011.

LEE, Kwan Tun *et al.* Deterministic Insight into ANN Model Performance for Storm Runoff Simulation. Water Resources Management, [S.L.], v. 22, n. 1, p. 67-82, Jan 24. 2007. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s11269-006-9144-x.

LOGAN, Travis *et al.* (2023). Ouranosinc/xclim: v0.40.0 (v0.40.0). Zenodo. https://doi.org/10.5281/zenodo.7535677

LUCAS, Edmundo Wallace Monteiro *et al.* Trends in climate extreme indices assessed in the Xingu river basin- Brazilian Amazon. Weather and Climate Extremes, [S.L.], v. 31, p. 100306, Mar. 2021. Elsevier BV. http://dx.doi.org/10.1016/j.wace.2021.100306.

LUIZ-SILVA, Wanderson *et al.* Climatological and hydrological patterns and verified trends in precipitation and streamflow in the basins of Brazilian hydroelectric plants. Theoretical and Applied Climatology, [S.L.], v. 137, n. 1-2, p. 353-371, Ago 23. 2018. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s00704-018-2600-8.

MACHIWAL, Deepesh; JHA, Madan K.. Comparative evaluation of statistical tests for time series analysis: application to hydrological time series. Hydrological Sciences Journal, [S.L.], v. 53, n. 2, p. 353-366, Apr. 2008. Informa UK Limited. http://dx.doi.org/10.1623/hysj.53.2.353.

MANN, Henry B.. Nonparametric Tests Against Trend. Econometrica, [S.L.], v. 13, n. 3, p. 245, Jul. 1945. JSTOR. http://dx.doi.org/10.2307/1907187.

MARENGO, José A. *et al. A seca e a crise hídrica de 2014-2015 em São Paulo*. Revista Usp, [S.L.], n. 106, p. 31-44, 2 Sep. 2015. *Universidade de São Paulo, Agencia USP de Gestão da Informação Acadêmica* (AGUIA). http://dx.doi.org/10.11606/issn.2316-9036.v0i106p31-44.

MEDEIROS, Felipe Jeferson de et al. Evaluation of extreme precipitation climate indices and their projected changes for Brazil: from cmip3 to cmip6. Weather and Climate Extremes, [S.L.], v. 38, p. 100511, Dec. 2022. Elsevier BV. http://dx.doi.org/10.1016/j.wace.2022.100511.

MEDEIROS, Matheus Sampaio. *PREVISÃO DE VAZÃO E ENERGIA NATURAL AFLUENTE: um estudo de caso avaliando métricas de desempenho*. 2022. 72 f. TCC - *Curso de Engenharia Hidrica, Instituto de Pesquisas Hidráulicas*

da Universidade Federal do Rio Grande do Sul, Porto Alegre, 2022. Available in: https://www.lume.ufrgs.br/bitstream/handle/10183/253374/001158090.pdf?sequence=1. Access at: May 24, 2023.

MEGAWHAT. *Operador Nacional do Sistema Elétrico (ONS)*. São Paulo, 2023a. Available in: https://megawhat.energy/verbetes/28826/operador-nacional-do-sistema-eletrico-ons. Access at: May 23, 2023.

MEGAWHAT. *Energia Natural Afluente (ENA)*. São Paulo, 2023b. Available in: https://megawhat.energy/verbetes/318/energia-natural-afluente-ena. Access at: May 23, 2023.

MENDES, Natalia Dias Sardinha. *Previsão das Vazões Afluentes Diárias por Rede Neural para Projeção de PLD*. 2019. 80 f. Dissertation (Master's degree) - Curso de Programa de Engenharia Elétrica, Coppe, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2019. Available in: https://pantheon.ufrj.br/handle/11422/12287. Access at: May 02, 2022.

MESCOLOTTI, Patricia Colombo *et al.* Fluvial aggradation and incision in the Brazilian tropical semi-arid: climate-controlled landscape evolution of the são francisco river. Quaternary Science Reviews, [S.L.], v. 263, p. 106977, Jul. 2021. Elsevier BV. http://dx.doi.org/10.1016/j.quascirev.2021.106977.

MONTEIRO, Jander Barbosa *et al. A influência de teleconexões e sistemas meteorológicos produtores de precipitação no semiárido nordestino*. Revista Brasileira de Geografia Física, [S.L.], v. 15, n. 1, p. 312, 23 Mar. 2022. Revista Brasileira de Geografia Fisica. http://dx.doi.org/10.26848/rbgf.v15.1.p312-332.

MORALES, Fidel Ernesto Castro *et al.* Spatiotemporal Analysis of Extreme Rainfall Frequency in the Northeast Region of Brazil. Atmosphere, [S.L.], v. 14, n. 3, p. 531, Mar 09. 2023. MDPI AG. http://dx.doi.org/10.3390/atmos14030531.

MUÑOZ SABATER, J. ERA5-Land monthly averaged data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). 2019. doi:10.24381/cds.68d2bb30

MUÑOZ-SABATER, Joaquín *et al.* ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. Earth System Science Data, [S.L.], v. 13, n. 9, p. 4349-4383, 7 set. 2021. Copernicus GmbH. http://dx.doi.org/10.5194/essd-13-4349-2021.

NIELSEN, David Marcolino *et al.* Dynamics-based regression models for the South Atlantic Convergence Zone. Climate Dynamics, [S.L.], v. 52, n. 9-10, p. 5527-5553, Oct 09. 2018. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s00382-018-4460-4.

OLIVEIRA, P. T.; SILVA, C. M. S.; LIMA, K. C.. Climatology and trend analysis of extreme precipitation in subregions of Northeast Brazil. Theoretical and Applied Climatology, [s.l.], v. 130, n. 1-2, p. 77-90, Jul 16. 2016. Available in: https://link.springer.com/article/10.1007%2Fs00704-016-1865-z. http://dx.doi.org/10.1007/s00704-016-1865-z.

OLIVEIRA, Danilo Henrique Morais Castro *et al.* Rainfall and streamflow extreme events in the São Francisco hydrographic region. International Journal Of Climatology, [S.L.], v. 41, n. 2, p. 1279-1291, Dec 02. 2020. Wiley. http://dx.doi.org/10.1002/joc.6807.

OLIVEIRA, Willian dos Santos *et al.* Spatial mapping of annual rainfall in the São Francisco river Basin. Ambiente e Agua - An Interdisciplinary Journal of Applied Science, [S.L.], v. 17, n. 3, p. 1-11, May 31, 2022. Instituto de Pesquisas Ambientais em Bacias Hidrograficas (IPABHi). http://dx.doi.org/10.4136/ambi-agua.2762.

ONS-*Operador Nacional do Sistema Elétrico, Conjunto de dados abertos*. Rio de Janeiro, 2022. Available in: https://dados.ons.org.br/dataset/. Access at: May 02, 2022.

ONS-*Operador Nacional do Sistema Elétrico. O SISTEMA INTERLIGADO NACIONAL*. Rio de Janeiro, 2023a. Available in: https://www.ons.org.br/paginas/sobre-o-sin/o-que-e-o-sin. Access at: May 23, 2023.

ONS-*Operador Nacional do Sistema Elétrico. O que é ONS*. Rio de Janeiro, 2023b. Available in: https://www.ons.org.br/paginas/sobre-o-ons/o-que-e-ons. Access at: May 23, 2023.

ONS-*Operador Nacional do Sistema Elétrico. Reservatórios*. Rio de Janeiro, 2023c. Available in: https://dados.ons.org.br/dataset/reservatorio. Access at: May 23, 2023.

RANJAN, G s K *et al.* K-Nearest Neighbors and Grid Search CV Based Real Time Fault Monitoring System for Industries. 2019 Ieee 5Th International

Conference For Convergence In Technology (I2Ct), [S.L.], p. 1-5, Mar. 2019. IEEE. http://dx.doi.org/10.1109/i2ct45611.2019.9033691.

REGOTO, Pedro *et al.* Observed changes in air temperature and precipitation extremes over Brazil. International Journal of Climatology, [S.L.], v. 41, n. 11, p. 5125-5142, Apr 14. 2021. Wiley. http://dx.doi.org/10.1002/joc.7119.

RODRIGUES, Daniele Tôrres *et al.* Evaluation of the Integrated Multi-SatellitE Retrievals for the Global Precipitation Measurement (IMERG) Product in the São Francisco basin(Brazil). Water, [S.L.], v. 13, n. 19, p. 2714, Oct 01. 2021. MDPI AG. http://dx.doi.org/10.3390/w13192714.

ROSA, Eliana Bertol *et al.* Automated Detection Algorithm for SACZ, Oceanic SACZ, and Their Climatological Features. Frontiers in Environmental Science, [S.L.], v. 8, p. 1-15, Feb 25, 2020. Frontiers Media SA. http://dx.doi.org/10.3389/fenvs.2020.00018.

SANTOS, Cosme Rodolfo Roque dos *et al. Aplicação de aprendizado de máquina para projeção do preço horário de liquidação das diferenças como suporte às estratégias de comercialização de energia elétrica*. Revista Brasileira de Energia, [S.L.], v. 28, n. 1, p. 243-279, Mar 10, 2022. Revista Brasileira de Energia. http://dx.doi.org/10.47168/rbe.v28i1.671.

SANTOS, Josielton da Silva; JUSTINO, Flávio Barbosa. *Análise da relação entre variáveis energéticas e climáticas considerando as diferentes regiões hidrográficas da Bacia do São Francisco*. Congresso Brasileiro de Meteorologia, Bauru, São Paulo, p. 668-680, Dec. 2022. Available in: https://cbmet.com.br/files/anais_cbmet22_texto.pdf. Access at: May 18, 2023.

SEN, Pranab Kumar. Estimates of the Regression Coefficient Based on Kendall's Tau. Journal of The American Statistical Association, [S.L.], v. 63, n. 324, p. 1379-1389, Dec. 1968. Informa UK Limited. http://dx.doi.org/10.1080/01621459.1968.10480934.

Shimakura,S.E.(2006) *Correlação*. In CE003 -Estatística II. Paraná: Dep. de Estatística UFPR:71-78.

SILVA, Marx Vinicius Maciel da *et al.* Projection of Climate Change and Consumptive Demands Projections Impacts on Hydropower Generation in the São Francisco river Basin, Brazil. Water, [S.L.], v. 13, n. 3, p. 332, Jan 29, 2021. MDPI AG. http://dx.doi.org/10.3390/w13030332.

SILVA, Isamara de Mendonça *et al.* Evaluating homogeneity and trends in extreme daily precipitation indices in a semiarid region of Brazil. Frontiers in Earth Science, [S.L.], v. 10, p. 1-20, Dec 20, 2022. Frontiers Media SA. http://dx.doi.org/10.3389/feart.2022.1071128.

SILVEIRA, Cleiton *et al. Mudanças climáticas na bacia do rio São Francisco: uma análise para precipitação e temperatura*. Revista Brasileira de Recursos Hídricos, [S.L.], v. 21, n. 2, p. 416-428, Apr 01, 2016. FapUNIFESP (SciELO). http://dx.doi.org/10.21168/rbrh.v21n2.p416-428.

SINGH, Krishna Kumar *et al.* Estimation of Mean Annual Flood in Indian Catchments Using Backpropagation Neural Network and M5 Model Tree. Water Resources Management, [S.L.], v. 24, n. 10, p. 2007-2019, Dec 02, 2009. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s11269-009-9535-x.

SOUTO, Jefferson *et al.* Performance of Remotely Sensed Soil Moisture for Temporal and Spatial Analysis of Rainfall over São Francisco river Basin, Brazil. Geosciences, [S.L.], v. 9, n. 3, p. 144, Mar 26, 2019. MDPI AG. http://dx.doi.org/10.3390/geosciences9030144.

THEIL, Henri. A Rank-Invariant Method of Linear and Polynomial Regression Analysis. Advanced Studies in Theoretical and Applied Econometrics, [S.L.], p. 345-381, 1992. Springer Netherlands. http://dx.doi.org/10.1007/978-94-011-2546-8_20.

Tomasella, J., Cunha, A.P.M.A., Simões, P.A. *et al.* Assessment of trends, variability and impacts of droughts across Brazil over the period 1980–2019. Nat Hazards (2022). https://doi.org/10.1007/s11069-022-05759-0

VILAR, Rafaella de Araújo Aires *et al. Avaliação do Impacto de Secas Severas no Nordeste Brasileiro na Geração de Energia Elétrica Através do Modelo Newave: projeção das energias afluentes e armazenadas*. Revista Brasileira de Meteorologia, [S.L.], v. 35, n. 1, p. 89-98, Mar. 2020. FapUNIFESP (SciELO). http://dx.doi.org/10.1590/0102-7786351004.

WILLMOTT, C.J. *et al.* Statistics for evaluation and comparison of models. Journal of Geophysical Research, Washington, v.0, n.C5, p.8995-9005, 1985.

XAVIER, Alexandre C. *et al.* Daily gridded meteorological variables in Brazil (1980-2013). International Journal of Climatology, [S.L.], v. 36, n. 6, p. 2644-2659, Oct 08, 2015. Wiley. http://dx.doi.org/10.1002/joc.4518.

XAVIER, Alexandre C. *et al.* New improved Brazilian daily weather gridded data (1961–2020). International Journal Of Climatology, [S.L.], v. 42, n. 16, p. 8390-8404, Jun. 2022. Wiley. http://dx.doi.org/10.1002/joc.7731.

ZAMBOM, Renato Carlos. *Planejamento da Operação de Sistemas Hidrotérmicos de Grande Porte*. 2008. 104 f. Thesis (Doctorate) - Curso de Engenharia Hidráulica, Escola Politécnica da Universidade de São Paulo, São Paulo, 2008. Available in: https://www.teses.usp.br/teses/disponiveis/3/3147/tde-15092008-151643/publico/ZambonRC_2008_Tese.pdf. Access at: May 23, 2023.