



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CAMPUS DE CRATEÚS**  
**CURSO DE ENGENHARIA CIVIL**

**SAMUEL MARTINS PEREIRA**

**PREVISÃO CLIMÁTICA SAZONAL PARA A REGIÃO HIDROGRÁFICA DOS  
SERTÕES DE CRATEÚS UTILIZANDO REGRESSÃO LINEAR MÚLTIPLA.**

**CRATEÚS**

**2022**

SAMUEL MARTINS PEREIRA

PREVISÃO CLIMÁTICA SAZONAL PARA A REGIÃO HIDROGRÁFICA DOS SERTÕES  
DE CRATEÚS UTILIZANDO REGRESSÃO LINEAR MÚLTIPLA

Trabalho de Conclusão de Curso apresentado ao  
Curso de Engenharia Civil da Universidade  
Federal do Ceará, como requisito parcial à  
obtenção do título de bacharel em Engenharia  
Civil.

Orientador: Prof. Dr. Alan Michell Barros  
Alexandre.

CRATEÚS

2022

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

P495p Pereira, Samuel.

Previsão climática sazonal para a região hidrográfica dos Sertões de Crateús utilizando regressão linear múltipla / Samuel Pereira. – 2022.

51 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Crateús, Curso de Engenharia Civil, Crateús, 2022.

Orientação: Prof. Dr. Alan Michell Barros Alexandre.

1. Previsão. 2. Precipitação. 3. Regressão. I. Título.

CDD 620

---

SAMUEL MARTINS PEREIRA

PREVISÃO CLIMÁTICA SAZONAL PARA A REGIÃO HIDROGRÁFICA DOS SERTÕES  
DE CRATEÚS UTILIZANDO REGRESSÃO LINEAR MÚLTIPLA

Trabalho de Conclusão de Curso apresentado ao  
Curso de Engenharia Civil da Universidade  
Federal do Ceará, como requisito parcial à  
obtenção do título de bacharel em Engenharia  
Civil.

Aprovada em: \_\_\_/\_\_\_/\_\_\_\_\_.

BANCA EXAMINADORA

---

Prof. Dr. Alan Michell Barros Alexandre (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Ms. Tatiane Lima Batista  
Universidade Federal do Ceará (UFC)

---

Prof. Ms. Taís Maria Nunes Carvalho  
Universidade Federal do Ceará (UFC)

A Deus.

Aos meus pais, irmãos e amigos.

“A maravilhosa disposição e harmonia do universo só pode ter tido origem segundo o plano de um Ser que tudo sabe e tudo pode. Isso fica sendo a minha última e mais elevada descoberta”.

(Isaac Newton)

## RESUMO

A escassez hídrica é um dos grandes problemas a serem enfrentados no Sertão do Nordeste Brasileiro. Essa região é caracterizada por apresentar baixos índices pluviométricos e uma concentração de precipitação em apenas alguns meses do ano. Além disso, outro fator que interfere negativamente na disponibilidade hídrica da região são a falta de planejamento e a alocação adequada dos recursos hídricos. A realização de previsão de precipitação tem se mostrado como uma alternativa interessante para instruir o poder público na tomada de decisão que amenize os problemas referentes a disponibilidade de água. Assim, este trabalho tem como objetivo gerar modelos de previsão de precipitação para a Região Hidrográfica dos Sertões de Crateús, utilizando a regressão Tradicional, Ridge e Lasso. Os preditores do modelo foram os índices de Temperatura da Superfície do Mar (TSM), uma vez que a precipitação do Nordeste como um todo é grandemente influenciada pelos oceanos. O algoritmo Iterative Input Selection (IIS) foi utilizado para realizar o ranqueamento desses índices, e a partir desse ranqueamento gerou-se todos os modelos, sendo analisado a acurácia dos mesmos. Ademais, para realizar todas as regressões e o ranqueamento dos índices TSM, utilizou-se o programa R-Studio para gerar as rotinas necessárias. A partir dos resultados gerados foi possível avaliar a acurácia de diferentes modelos de previsão de precipitação, identificando aqueles que apresentaram maiores correlações.

**Palavras-chave:** Previsão. Regressão. Precipitação.

## ABSTRACT

Water scarcity is one of the major problems to be faced in the Sertão of the Brazilian Northeast. This region is characterized by low rainfall rates and a concentration of precipitation in only a few months of the year. Besides this, another factor that interferes negatively in the availability of water in the region is the lack of planning and adequate allocation of water resources. Precipitation forecasting has been shown to be an interesting alternative to instruct the public authorities on how to make decisions that mitigate the problems related to water availability. Thus, this work aims to generate precipitation forecast models for the Crateús Hinterlands Hydrographic Region, using Traditional, Ridge and Lasso regression. The model predictors were the Sea Surface Temperature (SST) indices, since the precipitation of the Northeast as a whole is greatly influenced by the oceans. The Iterative Input Selection (IIS) algorithm was used to rank these indices, and from this ranking all models were generated and their accuracy was analyzed. Furthermore, to perform all regressions and the ranking of the TSM indices, the R-Studio program was used to generate the necessary routines. From the results generated, it was possible to evaluate the accuracy of different precipitation forecast models, identifying those that presented the highest correlations.

**Keywords:** Prediction. Regression. Precipitation.

## LISTA DE FIGURAS

Figura 1 – População de cada município que compõe a RHSC .....	20
Figura 2 - Localização das regiões referente a cada índice El Niño.....	21
Figura 3 – Serie temporal da ODP.....	22
Figura 4 – Serie temporal da AMO .....	23
Figura 5 – Fluxograma do Algoritmo IIS.....	24
Figura 6 – Fluxograma do método de pesquisa.....	29
Figura 7 – Regiões hidrográficas do Ceará .....	30
Figura 8– Estratégia metodológica utilizada para determinação da precipitação na RH de Crateús.....	33

## LISTA DE TABELAS

Tabela 1– População de cada município que compõe a RHSC.....	31
Tabela 2 – Estações Pluviométricas utilizadas para geração do Thiessen .....	31
Tabela 3 – Índices TSM utilizados .....	32
Tabela 4 - Classificação dos índices TSM .....	36
Tabela 5 – Resultados dos modelos de previsão utilizando regressão tradicional e como período preditor o trimestre SON. ....	37
Tabela 6 - Resultados dos modelos de previsão utilizando regressão tradicional e como preditor o trimestre JJA. ....	38
Tabela 7 - Resultados dos modelos de previsão utilizando regressão tradicional e como preditor o trimestre MAM. ....	38
Tabela 8 - Resultados dos modelos de previsão utilizando regressão tradicional e como preditor o trimestre DJF. ....	38
Tabela 9 - Resultados dos modelos de previsão utilizando regressão Ridge e como preditor o trimestre SON.....	40
Tabela 10 - Resultados dos modelos de previsão utilizando regressão Ridge e como preditor o trimestre JJA. ....	40
Tabela 11 - Resultados dos modelos de previsão utilizando regressão Ridge e como preditor o trimestre MAM. ....	41
Tabela 12 - Resultados dos modelos de previsão utilizando regressão Ridge e como preditor o trimestre DJF .....	41
Tabela 12 - Resultados dos modelos de previsão utilizando regressão Ridge e como preditor o trimestre DJF .....	42
Tabela 13 - Resultados dos modelos de previsão utilizando regressão Lasso e como preditor o trimestre SON.....	43
Tabela 14 - Resultados dos modelos de previsão utilizando regressão Lasso e como preditor o trimestre JJA. ....	43
Tabela 15 - Resultados dos modelos de previsão utilizando regressão Lasso e como preditor o trimestre MAM. ....	44
Tabela 16 - Resultados dos modelos de previsão utilizando regressão Lasso e como preditor o trimestre DJF. ....	44
Tabela 17 – Comparação dos resultados obtidos de correlação para cada tipo de regressão...	45
Tabela 18 - Comparação dos resultados obtidos de Nash para cada tipo de regressão.....	46



## LISTA DE ABREVIATURAS E SIGLAS

AD	Árvore de Decisão
AMO	Oscilação Multidecadal do Atlântico ( <i>Atlantic Multidecadal Oscillation</i> )
DJFMAM	Período de dezembro a maio
DJF	Período de dezembro a fevereiro
ENOS	El Niño – Oscilação Sul
FMAM	Período de fevereiro a maio
FUNCEME	Fundação Cearense de Meteorologia e Recursos Hídricos
IBGE	Instituto Brasileiro de Geografia e Estatística
MAM	Período de março a maio
NAO	Oscilação do Atlântico Norte ( <i>North Atlantic Oscillation</i> )
NEB	Nordeste Brasileiro
NINO 1.2	Niño 1+2
NINO3	Niño 3
NINO3.4	Niño 3.4
NINO4	Niño 4
NOAA	National Oceanic and Atmospheric Administration
ONI	Niño Oceânico ( <i>Oceanic Niño Index</i> )
PDO	Oscilação Decadal do Pacífico ( <i>Pacific Decadal Oscillation</i> )
RH	Região Hidrográfica
TAD	Dipolo do Atlântico Tropical ( <i>Tropical Atlantic Dipole</i> )
TNA	Atlântico Norte Tropical ( <i>Tropical Northern Atlantic</i> )
TSA	Atlântico Sul Tropical ( <i>Tropical Southern Atlantic</i> )
TSM	Temperatura de Superfície do Mar
VCAN	Vórtices Ciclônicos de Alto Nível
ZCIT	Zona de Convergência Intertropical

## SUMÁRIO

1	INTRODUÇÃO .....	16
1.1	Objetivos.....	17
1.1.1	<i>Objetivo geral</i> .....	17
1.1.2	<i>Objetivos específicos</i> .....	17
2	REFERENCIAL TEÓRICO .....	18
2.1	Caracterização climática.....	18
2.2	Temperatura de Superfície do Mar .....	19
2.2.1	<i>Dipolo do atlântico</i> .....	19
2.2.2	<i>El Niño – Oscilação Sul</i> .....	20
2.2.3	<i>Oscilação Decadal do Pacífico</i> .....	21
2.2.4	<i>Oscilação Multidecadal do Atlântico</i> .....	22
2.2.5	<i>Oscilação do Atlântico Norte</i> .....	23
2.3	Iterative Input Selection .....	23
2.4	Random Forest.....	25
2.5	Regressão Linear Múltipla .....	25
2.5.1	<i>Regressão Tradicional - OLS</i> .....	26
2.5.2	<i>Regressão Lasso e Ridge</i> .....	27
3	METODOLOGIA.....	29
3.1	Delimitação da área de estudo.....	30
3.2	Levantamento dos dados.....	31
3.3	Estratégia metodológica .....	32
3.4	Seleção dos preditores e geração dos modelos de previsão.....	33
4	RESULTADOS E DISCUSSÕES.....	36
4.1	Seleção das variáveis .....	36
4.2	Modelos de previsão .....	37
4.2.1	<i>Regressão Linear Múltipla tradicional</i> .....	37
4.2.2	<i>Regressão Ridge</i> .....	40
4.2.3	<i>Regressão Lasso</i> .....	43
4.3	Comparação dos modelos .....	45
5	CONCLUSÃO .....	47
	REFERÊNCIAS .....	48



## 1 INTRODUÇÃO

O Nordeste Brasileiro (NEB) é a segunda região mais populosa do Brasil, com uma população estimada de aproximadamente 57 milhões de habitantes, que corresponde a aproximadamente a 27,0% da população total do país (IBGE, 2022). O clima predominante nessa região é o semiárido, que é caracterizado pela presença significativa de variações temporais e espaciais da precipitação, e elevadas temperaturas anuais (NÓBREGA et al., 2014).

A região apresenta um volume anual de precipitação de 400 mm a 800 mm, o que é considerado elevado para uma região semiárida. No entanto, a grande concentração de chuva entre dois a quatro meses, a presença de um solo com baixa infiltração, alta taxa de evaporação, dentre outros fatores, fazem com que essa região seja marcada por um grande déficit hídrico (NÓBREGA; SANTIAGO; SOARES, 2016).

Essa deficiência hídrica que acontece no NEB pode ser ainda agravada pelo processo de mudanças climáticas que acontecem no mundo. Essas mudanças climáticas tem potencial de causar diversos impactos significativos nos setores naturais, sociais e econômicos (SANTOS et al., 2009).

O Oceano Pacífico e Atlântico Tropical são responsáveis por influenciar os eventos de precipitação no NEB. Dentre esses mecanismos que governam o regime de chuva nessa região, tem-se o El Niño-Oscilação Sul (ENOS), Temperatura da superfície do mar (TSM) no oceano Atlântico, Ventos alísios, Pressão ao Nível do Mar (PNM), Zona de Convergência Intertropical (ZCIT), Frentes Frias, Vórtices Ciclônicos de Altos Níveis (VCAN), Linha de Instabilidade (LI) e Brisas Marítima e Terrestre (FERREIRA; MELLO, 2005; UVO; BERNDTSSON, 1996).

O Estado do Ceará, que está localizado no NEB, apresenta uma precipitação anual que varia entre 500 e 800 mm/ano (CEARÁ, 2018). Segundo Hiera; Júnior; Zanella (2016), os principais eventos que influenciam a variação de chuva nesse estado são a Zona de Convergência Intertropical e o El Niño Oscilação Sul.

A região hidrográfica dos Sertões de Crateús, que está localizada na macrorregião dos sertões de Crateús, apresenta uma população estimada de aproximadamente 350 mil habitantes. A cidade de Crateús é a mais populosa da região, sendo estimada em 75 mil habitantes, correspondendo a 21,3 % de toda a população (IPECE, 2017).

Segundo Rosa (2019), essa região apresenta rios importantes, entretanto, devido à baixa qualidade e a disponibilidade de água apenas em um curto período do ano, torna-se necessário a utilização de reservatórios. Porém, os reservatórios da região mostram dados

históricos preocupantes. Segundo Ceará (2015), o sistema de abastecimento no ano de 2009 correspondia a 99,63% da capacidade de armazenagem, enquanto no ano de 2015 esse valor chegou a 0,61%.

Dessa forma, o uso racional da água disponível nos sistemas de abastecimento e o planejamento adequado dos recursos hídricos por parte do poder público são medidas essenciais para mitigar os problemas hídricos dessa região.

Araújo et. al (2019) afirmam que devido as regiões semiáridas apresentarem vulnerabilidade às flutuações climáticas e sofrerem com impactos nas reservas de água, a utilização de modelos de previsão de precipitação são essenciais para operação eficiente e eficaz dos recursos hídricos e a alocação da água.

Nesse contexto, esse trabalho está delimitado a avaliar a previsibilidade da precipitação sazonal da região hidrográfica nos sertões de Crateús. Os métodos empregados para realizar os modelos de previsão sazonal foram a Regressão Linear Múltipla Tradicional, Regressão Ridge e a Lasso relacionando os dados de precipitação com a Temperatura Superficial do Mar (TSM).

## **1.1 Objetivos**

### ***1.1.1 Objetivo geral***

Avaliar a previsibilidade da precipitação sazonal através de índices de TSM para a Região Hidrográfica dos Sertões de Crateús utilizando métodos estatísticos de Regressão Linear Múltipla.

### ***1.1.2 Objetivos específicos***

- Gerar modelos de previsão de precipitação a partir da associação de índices TSM.
- Verificar a acurácia dos diferentes modelos de precipitação;
- Definir o modelo preditor de melhor desempenho para os diferentes períodos preditivos adotados.
- Comparar os modelos com relação ao tipo de regressão empregada.

## 2 REFERENCIAL TEÓRICO

Esta seção apresenta o embasamento teórico necessário para a compreensão deste presente trabalho, sendo abordados: os principais sistemas meteorológicos que influenciam a precipitação sobre o NEB e o Ceará; os índices TSM mais relevantes para a área de estudo; o Algoritmo Iterative Input Selection, que foi utilizado para a seleção de variáveis mais influentes; o Método Random Forest; Regressão OLS; Regressão Rigde e Lasso.

### 2.1 Caracterização climática

O Nordeste Brasileiro é influenciado por diversos sistemas atmosféricos, que são responsáveis pelas variações intra e inter anual de precipitação. Dentre os principais fenômenos que influenciam diretamente e indiretamente a precipitação sobre o Nordeste Brasileiro, destaca-se os Vórtices Ciclônicos em Ar Superior, Distúrbios Ondulatórios de Leste, Zona de Convergência Intertropical (ZCIT), Zona de Convergência do Atlântico Sul e os Sistemas Frontais. (SILVA *et al.*, 2012).

Porém, devido à grande extensão territorial do NEB, várias regiões apresentam sistemas atmosféricos específicos que tem maior influência sobre a precipitação. Mollion e Bernado (2002), discretizam o Nordeste Brasileiro em sub-regiões que são influenciados pelos mesmos sistemas atmosféricos. Os autores realizam uma análise do NEB a partir de três sub-regiões, sendo elas: Norte do Nordeste (NNE), Sul do Nordeste (SNE) e Leste do Nordeste (LNE).

Para a região Sul do Nordeste (SNE), que cobre grande parte da Bahia, norte de Minas Gerais, e partes do sul do Maranhão e Piauí e extremo sudoeste de Pernambuco, os principais mecanismos são: a estacionaridade dos sistemas frontais, sistemas pré-frontais, convecção local e brisas de mar e terra no litoral. Já para o Leste do Nordeste (ENE), que inicia no Rio Grande do Norte até o Sul da Bahia, Molion (2002) expõe os principais fatores causadores de precipitação a circulação de brisa, frentes frias, perturbações ondulatórias nos ventos Alísios, brisas marinha e terrestre e convecção local.

Por fim, os autores definem os fatores relevantes para o Norte do Nordeste (NNE), que é composto pelo estado do Ceará e partes do Rio Grande do Norte, Piauí, Maranhão e oeste da Paraíba e de Pernambuco, sendo eles a Zona de Convergência Intertropical (ZCIT), a convergência de umidade, a convecção local e, com menor influência, frentes frias.

Mas especificamente, o principal sistema atmosférico que influencia o estado do Ceará é a ZCIT, sendo o maior responsável pela geração de chuva nessa região. A Zona de Convergência Intertropical (ZCIT) pode ser definida como sendo uma formação de nuvens que ficam localizadas próximo ao equador e que são formadas pelo transporte, por parte dos ventos alísios, de massas de ar quentes e úmidas para altos níveis.

## 2.2 Temperatura de Superfície do Mar

Ferreira *et al.* (2016) afirmam que os índices TSM “representam um importante forçante dos modelos numéricos acoplados (oceano/atmosfera) que servem de ferramentas para auxiliar a previsão climática em diversas regiões do mundo.”

### 2.2.1 Dipolo do atlântico

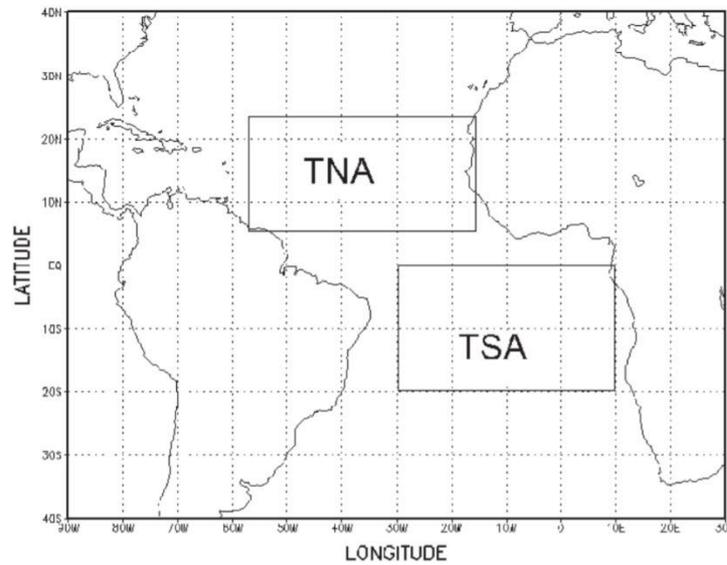
O Dipolo do atlântico é caracterizado como sendo a mudança anômala da temperatura da superfície do mar no Oceano Atlântico Tropical, existindo uma fase positiva e negativa. (NÓBREGA; SANTIAGO; SOARES, 2016).

A fase positiva do Dipolo acontece quando as águas do Atlântico Tropical Norte (TNA – *Tropical Northern Atlantic*) estão mais quentes e as águas do Atlântico Equatorial e Tropical Sul (TSA – *Tropical Southern Atlantic*) estão mais frias, gerando assim movimentos descendentes que transportam ar frio e seco dos altos níveis da atmosfera sobre a região setentrional, central e sertão do Nordeste, inibindo a formação de chuva. Em contrapartida, a fase negativa do Dipolo ocorre quando as águas do Atlântico Tropical Norte estão mais frias e as águas do Atlântico Tropical Sul estão mais quentes, acarretando no aumento ascendente nessas regiões, intensificando a formação de nuvens e aumentando os índices pluviométricos.

O Dipolo do Atlântico é responsável por influenciar a posição e a intensidade da Zona de Convergência Intertropical (ZCIT), que é o principal sistema que influencia no regime de chuvas da região Nordeste. No caso de Dipolo positivo (temperatura ao norte mais alta do que ao sul), a ZCIT é deslocada para o norte, o que gera uma redução da precipitação sobre o Nordeste (CARMO; LIMA, 2020).

Como principal forma para descrever esse fenômeno, tem-se os índices TNA e o TSA, que nada mais são do que a média da anomalia mensal para a área 5°N-23,5°N, 15°W-57,5°W e 0°-20°S, 10°E-30°W, respectivamente. A Figura 1 representa as áreas que são utilizadas para calcular os índices.

Figura 1 – Áreas referentes aos índices TNA e TSA



Fonte: Menezes *et al.*, Adaptado (2008)

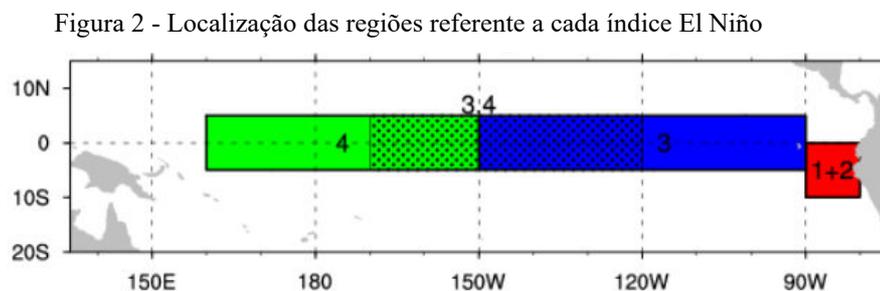
### 2.2.2 *El Niño – Oscilação Sul*

O El Niño e a Oscilação do Sul (ENOS) são um fenômeno global do oceano e da atmosfera que geram anomalias climáticas que duram vários meses, sendo exemplo de eventos relacionados a esse fenômeno: a precipitação elevada no Peru, Equador, Ilhas do Pacífico central e leste e secas na Indonésia, Austrália e norte do Nordeste do Brasil. O El Niño é o aquecimento no Pacífico Tropical da costa do Peru/Equador até o oeste do Pacífico. Já a Oscilação do Sul, que é uma resposta aérea ao El Niño, é definido como a variação anômala da pressão atmosférica tropical e sendo associado a mudanças na circulação geral da atmosfera (ARAGÃO, 1998).

Esse fenômeno apresenta duas fases extremas, uma fase negativa (El Niño) e positiva (La Niña). Enquanto o El Niño está associado a elevação da TSM, o que representa a fase quente, a La Niña está associado ao resfriamento da TSM, o que representa a fase fria (BUENO *et al.*, 2020).

Marengo *et al.* (1993) notaram que o El Niño interfere na posição do ZCIT, de forma que na presença desse evento a ZCIT encontra-se mais ao norte de sua posição normal, fazendo com que os ventos alísios de NE sejam mais fracos, diminuindo a umidade que penetra no NEB.

Existem vários índices usados para monitorar as alterações da TSM no Oceano Pacífico e o Fenômenos de El Niño e La Niña, dentre eles tem-se: Niño 1 + 2 (0-10S, 90W-80W), Niño 3 (5N-5S,150W-90W), Niño 3.4 (5N-5S,170W-120W) e Niño 4 (5N-5S,160E-150W). A Figura 2 mostra a posição referente a cada índice (NCAR, 2022).



Fonte: NCAR, 2022 (ADPTADO)

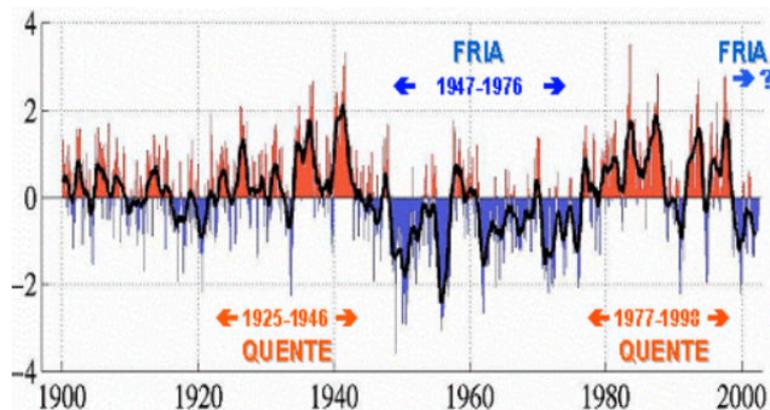
Além desses índices, outro índice que representam o ENOS é o Niño Oceânico (ONI – *Oceanic Niño Index*), que utiliza a mesma região do índice Niño 3.4, sendo calculado usando a média contínua de 3 meses. (NOAA, 2022; NCAR, 2022).

### 2.2.3 Oscilação Decadal do Pacífico

A Oscilação Decadal do Pacífico (PDO – *Pacific Decadal Oscillation*) são variações de longo prazo nas temperaturas da superfície do Oceano Pacífico, que persistem por 20 a 30 anos, sendo distinto do ENOS, que apresenta uma variação de 6 a 18 meses. A ODP é subdividida em duas fases, uma fase fria e outra quente. A fase fria é caracterizada por anomalias negativas de TSM no Pacífico Tropical, juntamente com anomalias positivas no Pacífico Extratropical. Já a fase quente é definida pela presença de anomalias de TSM positiva no Pacífico Tropical e negativas no Pacífico Extratropical. (MOLION, 2005)

A Fase quente do PDO faz com que exista uma tendência de ocorrência de El Niño e com maior intensidade, que por consequência cria uma menor ocorrência de La Niña e com uma menor intensidade. Já para a fase fria, existe um tendência de maiores ocorrência de La Niña e com intensidade maior, e menor ocorrência de El Niño, que tendem a serem curtos e rápidos (ANDREOLI; KAYANO, 2004).

Figura 3 – Série temporal da ODP



Fonte: Molion (2005)

Molion (2005), indica que a PDO pode ter entrado novamente em sua fase fria em 1999, onde pode permanecer até cerca de 2025, criando assim uma tendência de aumento (redução) da frequência de eventos de La Niña (El Niño). Assim, nota-se uma situação favorável para a precipitação no NEB nesse período, uma vez que se tem uma tendência de aumento de La Niña e um enfraquecimento do El Niño, devido a fase fria do PDO.

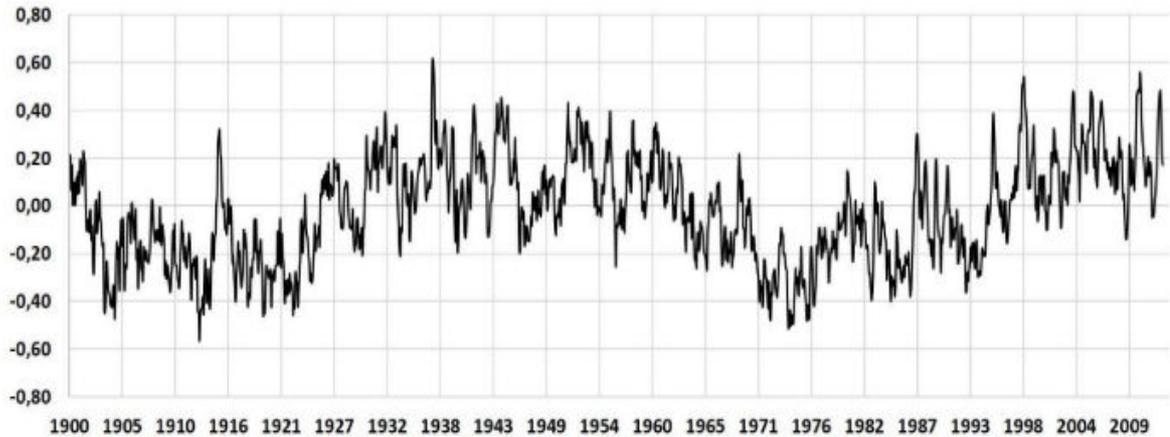
#### 2.2.4 Oscilação Multidecadal do Atlântico

A oscilação Multidecadal do Atlântico (AMO) é definida como sendo uma série contínua de longa duração Multidecadal de temperatura do mar do Oceano Atlântico Norte, apresentando duas fases, uma fria e outra quente, assim como na PDO, porém com uma duração de 20-40 anos (SILVA, 2012).

As fases frias e quentes da AMO são responsáveis por deslocar a ZCIT na escala Multidecadal. Dessa forma, a precipitação nas regiões do Atlântico que apresenta relação com à ZCIT sofrem variações nessa escala. Para o NEB, a fase fria da AMO está relacionada a condições úmidas, caso contrário a fase quente, que está associada a condições mais secas (KNIGHT; FOLLAND; SCAIFE, 2006).

A Figura 4 mostra a série temporal da AMO ao longo dos anos. Santos; Siqueira; Melo (2016) notou que a fase quente da série apresentou 36 anos de duração, sendo definido de 1927 à 1964, enquanto que a fase fria é definido de 1964 a 1995, com uma duração de 32 anos.

Figura 4 – Serie temporal da AMO



Fonte: Knight *et al.* (2005)

### 2.2.5 Oscilação do Atlântico Norte

A oscilação do Atlântico Norte (North Atlantic Oscillation – NAO) é definida por um dipolo com orientação aproximada Norte/Sul, sendo que um de seus centros está localizado sobre a Islândia, enquanto o outro centro se encontra nas latitudes do Atlântico Norte e centrado cerca dos 35°N, na zona dos Açores (IPMA, 2022).

A fase quente do dipolo é caracterizada pelo fortalecimento dos ventos de oeste em latitudes médias, enquanto que na fase fria existe um enfraquecimento dos ventos de oeste (HURRELL, 1995).

MARENGO E VALVERDE (2007) mostra que o NAO, junto com outros fatores, é responsável pelas variações interdecadal nas vazões de rios tanto na Amazônia, como no Nordeste Brasileiro (NEB).

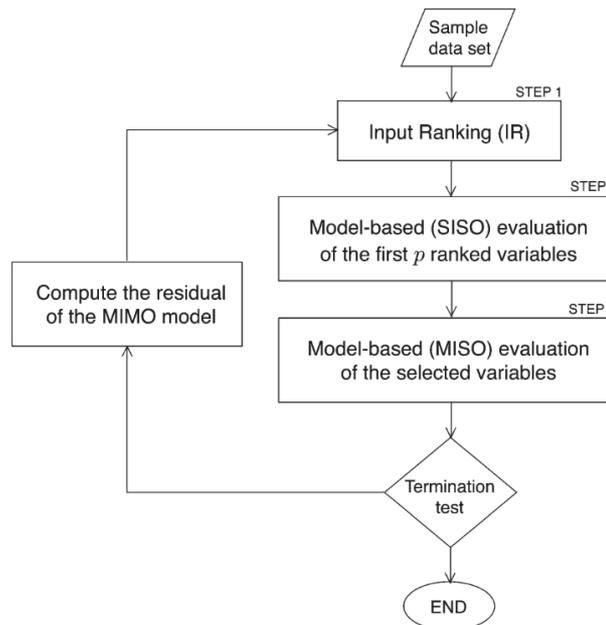
## 2.3 Interactive Input Selection

Um das etapas essenciais na geração de modelos de previsão é a escolha adequada das variáveis explicativas que compõem os modelos. Na seleção dessas variáveis dentro de um conjunto de variáveis possivelmente explicativas para um dado modelo, é possível encontrar três dificuldades básicas: a relação desconhecida entre a variável explicativa e a variável preditora pode ser complexa; o número de variáveis possivelmente explicativas pode ser

elevado; a correlação entre as variáveis explicativas pode gerar redundância ao modelo (MAIER *et al.*, 2010).

Esse método é demonstrado no trabalho de Galelli e Castelletti (2013), o fluxograma desse algoritmo é mostrado na Figura 5.

Figura 5 – Fluxograma do Algoritmo IIS



Fonte: Galelli e Castelletti (2013)

Como pode-se notar na Figura 5, o algoritmo é composto por três etapas, sendo elas: o Input Ranking (IR), avaliação das variáveis por intermédio de um modelo de entrada-única-saída-única (SISO - *Single-Input-Single-Output*) e multi-entrada-única-saída (MISO - *multiinput single-output*).

Uma vez inseridos um conjunto de dados de amostra no IIS, a primeira etapa consiste em realizar a classificação das  $n$  entradas candidatas com relação a uma medida estatística não linear de significância. Em tese, a primeira variável deve ser a mais importante para a previsão da variável de saída, porém, devido a possível existência de redundância entre as variáveis potencialmente explicativas, isso pode não acontecer. Assim, para evitar esse tipo de erro, na etapa seguinte as  $p$  primeiras variáveis são avaliadas individualmente (GALELLI; CASTELLETTI, 2013).

Na próxima etapa as primeiras  $p$  variáveis selecionadas são analisadas individualmente por meio de  $p$  modelos SISO. O desempenho de cada um dos  $p$  modelos SISO

é registrado e o melhor dentre eles é classificado para a próxima etapa, sendo inserido no conjunto  $X_y^i$ .

Por fim, a partir dos conjuntos de variáveis selecionadas momentaneamente  $X_y^i$  é gerado um modelo MISO. Em seguida, as etapas anteriores são repetidas usando agora os resíduos ( $y - m(X_y^i)$ ) como variável de saída. O algoritmo continua a realizar interações até que a melhor variável retornada pelo Input Ranking (IR) já esteja no conjunto  $X_y^i$  ou a diferença entre a métrica de performance do modelo nas iterações  $i$  e  $i-1$  seja menor ou igual a um valor  $\epsilon$ .

## 2.4 Random Forest

O Random Forest é considerado um eficaz algoritmo de aprendizado de máquina para classificação de variáveis, sendo que o mesmo consiste em uma coleção de classificadores  $\{h(\mathbf{x}, \theta_k), k = 1, \dots\}$  onde os  $\theta_k$  são vetores aleatórios independentes e identicamente distribuídos (BREIMAN, 2022).

O RF é constituído de um conjunto de Árvores de Decisão (AD), que é um método de classificação estruturado em forma de árvore, sendo formado por três componentes: nó interno, ramo e nós folha. O nó interno condiz a um teste de um atributo, o ramo a um resultado do teste e os nós folhas representam classes ou distribuições de classes (CASTRO; FERRARI, 2019; RIQUETE; RIBEIRO; ZÁRATE, 2018).

À medida que as AD são geradas cria-se aleatoriedade no modelo, aumentando a diversidade e tornando o modelo mais robusto (LIU; MOTODA, 2007; STAŃCZYK; ZIELOSKO; JAIN, 2018). Como exemplo da aplicação do Random Forest, tem-se o trabalho desenvolvido por Evangelista (2019), que utilizou o algoritmo para realizar a seleção de variáveis e assim gerar os modelos de previsão de precipitação para a Região Hidrográfica de Crateús.

## 2.5 Regressão Linear Múltipla

Este tópico tem como objetivo apresentar os tipos de regressão linear múltipla que serão utilizados para gerar os modelos de previsão, sendo essas a regressão Tradicional, Lasso e Ridge.

### 2.5.1 Regressão Tradicional - OLS

Um modelo de regressão linear é utilizado quando se tem a necessidade de analisar, qualitativamente e quantitativamente, relações entre variáveis. Denomina-se variável dependente,  $y$ , aquela cujo comportamento será explicado pela variável  $x$ , chamada de variável explicativa ou independente. Quando a variável dependente é explicada somente por uma variável independente, tem-se um modelo de regressão linear simples, entretanto, quando a variável dependente apresenta mais de uma variável explicativa, tem-se um modelo de regressão linear múltipla (CHEIN, 2019).

A partir de um modelo de regressão linear é possível realizar inferências estatísticas. Uma inferência estatística consiste em gerar estimativas para a variável dependente a partir da verificação das correlações existentes entre as variáveis (CHEIN, 2019).

Em um caso onde existam um conjunto de  $n$  pares de valores de duas variáveis,  $X_i$ ,  $Y_i$  (Onde  $i = 1, 2, \dots, n$ ), pode-se estabelecer uma regressão linear simples, onde o modelo estatístico é definido pela equação 2.1 (HOFFMANN, 2016).

$$Y_i = \alpha + \beta X_i + u_i \quad (2.1)$$

Esse modelo pode ser discretizado em dois termos, um termo que representa o valor estimado de  $Y_i$  pela equação da reta que melhor se ajusta aos dados, que apresenta coeficiente linear  $\alpha$  e coeficiente angular  $\beta$ , e um termo que expressa o erro existente entre o valor real da variável  $Y_i$  e do valor estimado pela regressão, representado por  $u_i$ .

Diferente da regressão linear simples, onde existe somente uma variável explicativa, na maioria dos problemas, é necessário utilizar mais de uma variável no modelo de regressão. Nesse caso, torna-se necessário a utilização de um modelo de regressão linear múltipla (WALPOLE *et al.*, 2009).

O modelo de uma regressão linear múltipla, com  $k$  variáveis independentes pode ser definido pela equação 2.2 (HOFFMANN, 2016).

$$Y_i = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} \dots + \beta_k X_{kj} + u_j, \quad j = 1, \dots, n \quad (2.2)$$

Hoffmann (2016) expressa o modelo de regressão linear múltipla de maneira matricial, que é definido pela equação 2.3.

$$Y_i = X\beta + u \quad (2.3)$$

Em que:

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} ; \boldsymbol{\beta} = \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} ; \mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} ; \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

A equação 2.3 é composta por dois termos, sendo que o primeiro termo é um vetor (1 x n) que contém os valores estimados de  $Y_i$ , e o segundo termo é um vetor (1 x n) que contém os erros existentes entre o valor real da variável  $Y_i$  e do valor estimado pela regressão.

Em problemas de regressão linear múltipla, busca-se encontrar os melhores coeficientes  $\boldsymbol{\beta}$ , de tal forma a reduzir os erros. Para obter o vetor  $\boldsymbol{\beta}$  que melhor se ajuste aos dados da regressão, é realizado a minimização da soma dos quadrados dos erros, essa minimização é definida como sendo o estimador do modelo.

$$RSS = \mathbf{u}^T \mathbf{u} = \sum_{i=1}^n u_i^2 \quad (2.4)$$

Em Hoffmann (2016), é possível ver a dedução da minimização da equação 2.4, que tem como resposta a equação 2.5, onde  $\mathbf{b}$  é o vetor dos parâmetros estimados.

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (2.5)$$

### 2.5.2 Regressão Lasso e Ridge

Algumas análises de regressão linear múltipla tradicional podem apresentar algumas dificuldades, principalmente quando existem problemas de multicolinearidade e/ou alta dimensionalidade dos conjuntos de dados (CASAGRANDE, 2016).

Uma maneira de resolver o problema de multicolinearidade é utilizar a regressão Ridge e Lasso, que basicamente acrescenta um pequeno vício aos estimadores dos coeficientes de regressão encontrado pelo método tradicional, afastando o sistema da singularidade. Esses métodos, fazem com que os erros quadráticos médios sejam menores, à medida que viciam o estimador e diminuem consideravelmente a variância dos estimadores (CASAGRANDE, 2016).

Na regressão linear múltipla tradicional, o erro pode ser expresso pelo eq. 2.6, que é basicamente a equação 2.2, deixando o erro em evidência.

$$u_i = \sum_{k=1}^n (Y_k - \alpha - \sum_{k=1}^n (\beta_j X_{kj})) \quad (2.6)$$

Logo, o erro quadrado, que é equação que deve ser minimizada para encontrar as melhores estimativas para os coeficientes, é definido pela equação 2.7.

$$u_i^2 = \left( \sum_{k=1}^n (Y_k - \alpha - \sum_{j=1}^n (\beta_j X_{kj})) \right)^2 \quad (2.7)$$

Assim, de maneira análoga ao que é feito na regressão tradicional, para encontrar os melhores coeficientes da regressão da Lasso e Ridge é necessário minimizar a equação 2.7, porém, com um acréscimo de um fator que é definido no próprio modelo, que é chamado de fator de penalização.

Para a Regressão Lasso e Ridge, as equações do erro quadrado que será minimizado para encontrar os coeficientes são expressas pelas equações 2.8 e eq. 2.9, respectivamente. Salienta-se, que o parâmetro lambda, definido nos dois modelos, é encontrado por intermédio do método da validação cruzada.

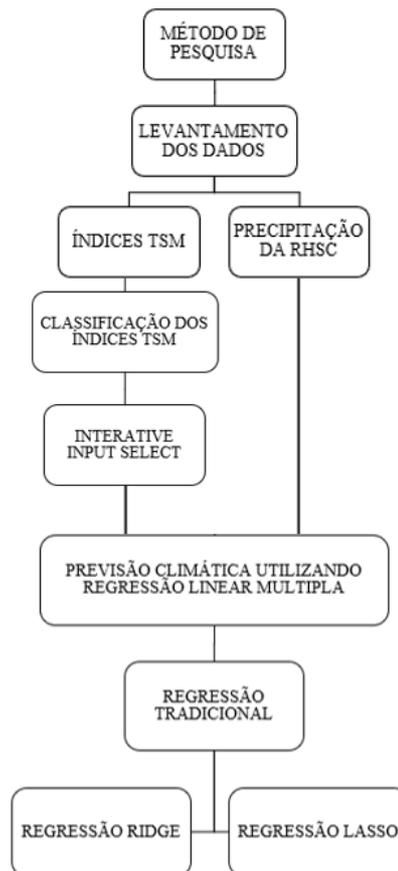
$$u_i^2 = \sum_{k=1}^n \left( Y_k - \alpha - \sum_{j=1}^d (\beta_j X_{k,j}) \right)^2 + \lambda \sum_{j=1}^n |\beta_j|, \text{ onde } \lambda > 0 \quad (2.8)$$

$$u_i^2 = \sum_{k=1}^n \left( Y_k - \alpha - \sum_{j=1}^d (\beta_j X_{k,j}) \right)^2 + \lambda \sum_{j=1}^n |\beta_j^2|, \text{ onde } \lambda > 0 \quad (2.9)$$

### 3 METODOLOGIA

Nesta seção serão apresentadas a caracterização da área de estudo e o método de pesquisa empregado, sendo detalhados os procedimentos necessários para alcançar os objetivos estabelecidos. A Figura 06 mostra um fluxograma das etapas executadas.

Figura 6 – Fluxograma do método de pesquisa.



Fonte: Autor (2022)

A metodologia de pesquisa classifica-se com relação a natureza como sendo quantitativa de caráter preditivo. De acordo com Richardson (2011), uma abordagem quantitativa é caracterizada pela quantificação das informações, desde a coleta até o tratamento por meio de técnicas estatísticas.

Essa pesquisa tem caráter preditivo, uma vez que por meio de uma lógica indutiva dos dados, os valores de precipitação futura podem ser estimados por meio de Regressões Linear múltiplas que relacionam os valores históricos de precipitação na região Hidrográfica dos Sertões de Crateús com os valores de anomalia TSM na mesma região.

Os valores históricos de precipitação e de anomalias de TSM que serão utilizados nas análises podem ser obtidos nos sites da Agência Nacional de Águas e no National Oceanic and Atmospheric Administration (NOAA), respectivamente. Como tem-se que esses dados estão previamente disponíveis nesses sites, a coleta de dados é classificada como sendo de natureza documental.

Com relação a estratégia metodológica empregada, essa pesquisa se enquadra em um estudo de caso. Como afirma Yin (2010), o estudo de caso é uma investigação empírica de um fenômeno contemporâneo dentro de seu contexto real, possuindo uma estratégia de pesquisa abrangente.

### 3.1 Delimitação da área de estudo

A Bacia hidrográfica dos Sertões de Crateús é uma unidade de planejamento para a gestão dos recursos hídricos do Ceará. Compreendendo uma área de 10.821 km<sup>2</sup>, a RHSC tem uma população total de 240.714 habitantes e é composta por nove municípios, sendo eles: Ararendá, Crateús, Independência, Ipaporanga., Novo Oriente, Poranga, Ipueiras, Quiterianópolis e Tamboril (COGERH, 2020). A população de cada um dos nove municípios dessa região é apresentada na tabela 1.

Além da RHSC, o Ceará apresenta 11 regiões hidrográficas. A localização da área de estudo, bem como a sua posição em relação as demais regiões são apresentadas na Figura 7.

Figura 7 – Regiões hidrográficas do Ceará



Fonte: SRH (2013)

Tabela 1– População de cada município que compõe a RHSC

<b>Município</b>	<b>População (Censo 2010)</b>
Crateús	72.812
Independência	25.573
Novo Oriente	27.453
Quiterianópolis	19.921
Tamboril	25.451
Ipaporanga	11.343
Ararendá	10.491
Nova Russas	30.965
Monsenhor Tabosa	16.705

Fonte: SRH (2013).

### 3.2 Levantamento dos dados

Os dados de precipitação utilizados para gerar os modelos de previsão foram obtidos a partir da média das precipitações de estações pluviométricas que estão dentro da Região Hidrográfica dos Sertões de Crateús. Para calcular essa média das precipitações, utilizou-se um código produzido na linguagem R que realiza o Thiessen com bases em todas as estações disponíveis e que são relevantes para a região de estudo. A tabela 2 mostra todas as estações pluviométricas utilizadas para realizar o Thiessen.

Tabela 2 – Estações Pluviométricas utilizadas para geração do Thiessen

<b>Estação</b>	<b>Nome do Posto</b>	<b>Estação</b>	<b>Nome do Posto</b>	<b>Estação</b>	<b>Nome do Posto</b>
640037	Baixio	540031	Jaburu	440020	Sucesso
640041	Algodões	540027	Vila Santo Antônio	440055	Ipaporanga
540052	Quiterianópolis	540019	Independência	440007	Tamboril
540002	Coutinho	540016	Independência	440080	Açude Carão
540042	Cruz	540035	Recife	440060	Canindezinho
540041	Monteiro	540111	Fazenda Gracinha	440044	Ararendá
540030	Iapi	540028	Tucuns	440017	Nova Russas
540033	Várzea Grande	540034	Ematuba	440045	Santo Antônio
540115	Muquem	540020	Aeroporto		
540113	Fazenda Realeza	540000	Crateús		
540007	Iapi	540026	Irapuã		
540005	Tranqueiras	540025	Morro Alegre		
540038	Cachoeira do Fogo	540039	Barreiros		
540003	Novo Oriente	540044	Oliveira		
540054	Caminho da Cachoeira	540043	Logradouro		
540040	Emaus	540107	Poti		
540036	Jandrangoeira	540023	Ibiapaba		
540108	Desejo	540029	Santa Terezinha		
540032	Monte Sinai	440003	Curatis		
540024	Monte Nebo	440068	Curatis		

Fonte: Autor 2022.

É importante salientar que uma vez que a precipitação gerada tem início em 1911 e termino em 2020, mas existem índices climáticos com dados compreendidos de 1950 à 2020, tornou-se necessário utilizar os valores de precipitação somente para esse período. Os índices climáticos utilizados nesse presente trabalho são mostrados na tabela 03.

Tabela 3 – Índices TSM utilizados

<b>Índices</b>	<b>Sigla</b>
<i>Tropical Atlantic Dipole</i>	<b>TAD</b>
<i>Tropical Northern Atlantic</i>	<b>TNA</b>
<i>Tropical Southern Atlantic</i>	<b>TSA</b>
<i>Niño 1 + 2</i>	<b>NINO1.2</b>
<i>Niño 3</i>	<b>NINO3</b>
<i>Niño 3.4</i>	<b>NINO3.4</b>
<i>Niño 4</i>	<b>NINO4</b>
<i>North Atlantic Oscillation</i>	<b>NAO</b>
<i>Pacific Decadal Oscillation</i>	<b>PDO</b>
<i>Atlantic Multidecadal Oscillation</i>	<b>AMO</b>
<i>Oceanic Niño Index</i>	<b>ONI</b>

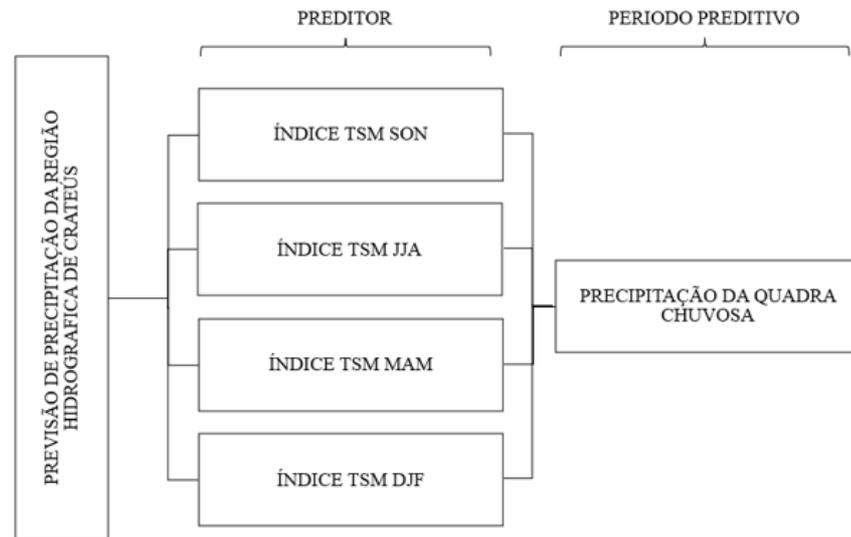
Fonte: Autor 2022.

### 3.3 Estratégia metodológica

Para estimar os modelos, utilizou-se como período preditivo a quadra chuvosa, que é composta pelos meses de fevereiro, março, abril e maio (FMAM). Esse período foi escolhido por apresentar as maiores precipitações ao longo do ano. Já como preditores foram definidos quatro trimestres de dados TSM de meses imediatamente anteriores, sendo esses os trimestres: setembro, outubro e novembro (SON); junho, julho e agosto (JJA); março, abril e maio (MAM); dezembro, janeiro, fevereiro (DJF).

Os dados de precipitação da quadra chuvosa, que é formada pela soma das precipitações dos meses que a compõem, foram obtidos a partir do Thiessen da região hidrográfica. Além disso, os dados dos índices TSM utilizados para cada trimestre foi definido como sendo a média dos valores dos índices nos meses que constitui o trimestre. A Figura 8 mostra esquematicamente toda a estratégia metodológica empregada.

Figura 8– Estratégia metodológica utilizada para determinação da precipitação na RH de Crateús



Fonte: Autor (2022).

### 3.4 Seleção dos preditores e geração dos modelos de previsão.

Para a seleção dos preditores dos modelos, utilizou-se o algoritmo IIS, sendo implementado esse algoritmo no programa R-Studio. Salienta-se que apenas a regressão tradicional foi inserida como modelo SISO e MISO, para assim gerar o ranqueamento das variáveis. Enquanto que no IR foi empregado o algoritmo RF para realizar a seleção das entradas candidatas do modelo.

O algoritmo foi definido para funcionar até que a melhor variável retornada pelo IR já tivesse sido escolhida ou a diferença da métrica de performance entre uma interação e sua anterior fosse menor que um valor épsilon 0,1%, sendo que a distância métrica utilizada para interromper o programa foi o R quadrado. Ainda se destaca que o processo de seleção de variáveis ocorre para cada um dos quatro trimestres definidos na estratégia metodológica.

Uma vez que as variáveis foram selecionadas, gerou-se 11 modelos de previsão da quadra chuvosa para cada período preditor, utilizando a regressão tradicional. Sendo que o modelo 01 é composto pela primeira variável selecionada no Ranking, o modelo 02 é formado pelas duas primeiras variáveis selecionadas, e assim sucessivamente até o modelo 11 que é composto por todas as variáveis.

Ademais, mesmo que o IIS realize a seleção das variáveis de forma a considerar a redundância entre as mesmas, com o fim de minimizar esse parâmetro, realizou-se 11 novos modelos de previsão de precipitação utilizando a regressão Ridge e Lasso. Essas regressões são caracterizadas por justamente considerar as multicolinearidade na geração dos modelos e realizar a diminuição dos coeficientes no intuito de resolver esse tipo de problema.

A acurácia dos modelos foi analisada a partir de dois parâmetros, sendo eles o coeficiente de correlação Pearson e o valor de Nash-Sutcliffe, parâmetros esse utilizados no trabalho de Evangelista (2019).

O coeficiente de correlação Pearson foi calculado a partir dos valores de precipitação real e dos encontrados pelos métodos de previsão, sendo que o cálculo desse parâmetro é definido pela equação 3.1.

$$R = \frac{1}{N-1} \sum_{i=1}^n \left( \frac{P_i - P_{méd}}{S_p} \right) \left( \frac{O_i - O_{méd}}{S_o} \right) \quad (3.1)$$

Onde  $P_i$  e  $P_{méd}$  são referentes aos valores previstos de precipitação;  $O_i$  e  $O_{méd}$  correspondem aos valores observados de precipitação;  $S_p$  e  $S_o$  são os valores do desvio padrão dos valores de precipitação previstos e observados, especificamente.

Esse coeficiente é responsável por mensurar a direção e o grau de relação linear entre duas variáveis quantitativas, variando no intervalo de -1 a 1. Definido pela letra  $r$ , valores iguais a -1,00 e 1,00, representam uma correlação perfeita negativa e positiva, respectivamente. Em outras palavras, isso indica que o escore de uma variável pode ser determinado exatamente ao saber o escore da outra. Já para valores de  $r$  próximo a zero, isso representa a inexistência de relação entre as variáveis (PARANHOS *et al.*, 2009).

Assim, para correlações próximas a 1,00 nos modelos gerados, isso indica que a precipitação real e a prevista apresentam valores próximos. Cohen (1988), indica alguns valores para classificar as correlações, sendo que os valores de 0,10 à 0,29 são consideradas pequenos, 0,30 a 0,49 são considerados valores médios, e de 0,50 à 1,00 são considerados valores elevados.

O Nash-Sutcliffe é um coeficiente capaz de mostrar a eficiência de desempenho de um modelo de previsão. Os valores de Nash variam de  $-\infty$  a 1,00, sendo que quanto mais próximo de 1,00 melhor é o modelo para representar previsões assertivas. A Equação 3.2 mostra a formulação para o cálculo desse coeficiente.

$$Nash = 1 - \frac{(P_i - O_i)^2}{(O_i - O_{méd})^2} \quad (3.2)$$

Valores de Nash iguais a zero, indicam que a previsão realizada apresenta a mesma eficiência da média dos dados observados. Já para valores negativos e positivos, indicam uma menor ou maior eficiência do modelo com relação à média, respectivamente (MORIASI *et al.*, 2007).

O coeficiente de correlação Pearson foi obtido diretamente pelo software R-Studio durante a geração dos modelos. Em contrapartida, o Nash teve que ser implementado em uma planilha Excel.

## 4 RESULTADOS E DISCUSSÕES

Esta seção é destinada a apresentação e discussão dos resultados obtidos a partir das etapas mostradas na seção de metodologia. Mais especificamente, a seleção das variáveis a partir do IIS e a geração de modelos de previsão de precipitação para cada um dos métodos de regressão linear múltipla.

### 4.1 Seleção das variáveis

Foram gerados quatro ranqueamentos das variáveis predictoras, referindo-se a cada trimestre de dados de índices TSM (SON, JJA, MAM e DJF). Enfatiza-se novamente que a seleção ocorreu utilizando apenas o IIS implementado com a Regressão OLS. A tabela 4 mostra os resultados obtidos.

Tabela 4 - Classificação dos índices TSM

Ordem de classificação	Índices			
	SON	JJA	MAM	DJF
1º	NINO3	NINO3	NINO3	AMO
2º	NINO3.4	NAO	ONI	NINO4
3º	NINO4	NINO4	NINO1.2	PDO
4º	PDO	AMO	NINO4	NINO3.4
5º	NAO	ONI	NINO3.4	TSA
6º	ONI	NINO1.2	TNA	NINO3
7º	TNA	TSA	NAO	TAD
8º	AMO	TNA	TSA	ONI
9º	NINO1.2	NINO3.4	AMO	NINO1.2
10º	TAD	TAD	PDO	NAO
11º	TSA	PDO	TAD	TNA

Fonte: Autor (2022)

Em geral os índices relacionados a ocorrência de El Niño e La Niña tiveram um bom desempenho. Nota-se que o Índice NINO 3 é um dos índices mais importantes, uma vez que para três preditores (SON, JJA e MAM) esse índice foi considerado o mais importante, e para o preditor DJF, o mesmo ainda foi classificado como o sexto mais importante. Além disso, outro índice que foi classificado com um dos mais importantes foi o NINO4, que para os preditores SON e JJA ficou na terceira posição, e para os preditores MAM e DJF foi classificado como sendo o quarto e segundo mais importante, respectivamente. Ademais, os índices NINO 3.4 apresentaram resultados variados, ocupando posições superiores, intermediárias e inferiores

para diferentes preditores, enquanto que o NINO1.2 foi classificado na maioria das vezes em posições mais baixas.

O índice que apresentou uma das piores classificações foi o TAD, que para os preditores SON e JJA foi classificado em penúltimo, e para os preditores MAM e DJF foi classificado com sendo o último e o sétimo, respectivamente. Já os demais índices relacionados ao dipolo do atlântico, TSA e TNA, apresentaram colocações diversas para os diferentes preditores a partir da quinta posição.

O PDO por sua vez se apresentou como sendo um dos índices mais importantes para em alguns casos, como para os preditores SON e DJF, enquanto que para os demais ostentou as últimas posições. Outrossim, os índices AMO, NAO e ONI apresentam classificações diversas para os diferentes preditores, desde posições elevadas a mais inferiores.

## 4.2 Modelos de previsão

### 4.2.1 Regressão Linear Múltipla tradicional

Os Resultados dos 11 Modelos de regressão linear múltipla para cada um dos quatro preditores são mostrados nas tabelas 5, 6, 7 e 8. Os resultados são referentes aos valores de Nash e de coeficiente de correlação Person, tanto para a etapa de Treinamento como também para a de validação.

Tabela 5 – Resultados dos modelos de previsão utilizando regressão tradicional e como período preditor o trimestre SON.

OLS: SON	Treinamento		Validação	
	R	Nash	R	Nash
Modelo 01	40,03%	16,02%	16,00%	-6,02%
Modelo 02	40,04%	16,03%	16,30%	-5,65%
Modelo 03	41,97%	17,61%	7,25%	-12,73%
Modelo 04	42,28%	17,88%	6,45%	-12,87%
Modelo 05	42,30%	17,89%	7,57%	-12,33%
Modelo 06	42,30%	17,90%	7,47%	-12,34%
Modelo 07	50,06%	25,06%	28,29%	-0,55%
Modelo 08	50,12%	25,12%	29,49%	0,70%
Modelo 09	51,31%	26,33%	29,07%	-0,80%
Modelo 10	69,58%	48,42%	46,66%	11,55%
Modelo 11	67,26%	45,24%	49,39%	19,41%

Fonte: Autor 2022

Tabela 6 - Resultados dos modelos de previsão utilizando regressão tradicional e como preditor o trimestre JJA.

OLS: JJA	Treinamento		Validação	
	R	Nash	R	Nash
Modelo 01	34,67%	12,02%	20,38%	-0,05%
Modelo 02	38,37%	14,72%	26,67%	-0,12%
Modelo 03	38,80%	15,05%	27,98%	1,06%
Modelo 04	41,64%	17,34%	27,21%	1,31%
Modelo 05	41,67%	17,37%	28,07%	1,99%
Modelo 06	42,40%	17,98%	32,92%	5,02%
Modelo 07	56,59%	32,03%	65,17%	34,52%
Modelo 08	57,37%	32,91%	66,69%	36,59%
Modelo 09	57,41%	32,96%	67,15%	37,21%
Modelo 10	56,35%	31,58%	62,93%	32,41%
Modelo 11	56,39%	31,80%	62,99%	31,83%

Fonte: Autor 2022

Tabela 7 - Resultados dos modelos de previsão utilizando regressão tradicional e como preditor o trimestre MAM.

OLS: MAM	Treinamento		Validação	
	R	Nash	R	Nash
Modelo 01	21,42%	4,59%	2,08%	-7,16%
Modelo 02	21,78%	4,75%	6,74%	-5,31%
Modelo 03	28,65%	8,21%	8,68%	-6,21%
Modelo 04	37,26%	13,88%	2,78%	-23,03%
Modelo 05	37,59%	14,13%	0,28%	-20,88%
Modelo 06	38,02%	14,45%	0,41%	-23,34%
Modelo 07	40,75%	16,60%	8,87%	-39,44%
Modelo 08	53,40%	28,51%	16,62%	-15,94%
Modelo 09	53,58%	28,71%	19,55%	-13,80%
Modelo 10	60,56%	36,68%	7,22%	-65,83%
Modelo 11	55,44%	30,74%	4,66%	-22,01%

Fonte: Autor 2022

Tabela 8 - Resultados dos modelos de previsão utilizando regressão tradicional e como preditor o trimestre DJF.

OLS: DJF	Treinamento		Validação	
	R	Nash	R	Nash
Modelo 01	3,22%	0,10%	29,02%	-6,39%
Modelo 02	12,22%	1,49%	10,98%	-8,50%

Tabela 8 - Resultados dos modelos de previsão utilizando regressão tradicional e como preditor o trimestre DJF.

OLS: DJF	Treinamento		Validação	
	R	Nash	R	Nash
Modelo 03	20,17%	4,07%	18,21%	-13,35%
Modelo 04	23,71%	5,62%	5,37%	-7,99%
Modelo 05	27,10%	7,35%	24,94%	4,80%
Modelo 06	29,51%	8,71%	29,46%	6,63%
Modelo 07	32,82%	10,77%	30,14%	6,39%
Modelo 08	32,87%	10,80%	32,13%	8,08%
Modelo 09	33,88%	11,48%	32,23%	8,18%
Modelo 10	33,92%	11,50%	32,09%	8,25%
Modelo 11	32,54%	10,59%	33,60%	10,01%

Fonte: Autor 2022

A partir da tabela 5 é possível notar que na etapa de Treinamento os valores de correlação para o preditor SON aumentaram à medida que crescia o número de variáveis nos modelos. Iniciando com modelo 01, que é composto apenas pelo índice NINO 3 e apresentou uma correlação de 40,03%, até o modelo 10, que apresentou uma correlação de 69,58%, os valores foram sempre crescentes, apresentando apenas um arrefecimento no modelo 11. Outrossim, os valores de Nash para a mesma etapa também apresentaram o mesmo comportamento, crescendo de 16,02%, para o modelo 01, até o valor de 48,42%, para o modelo 10, mostrando também um decréscimo no modelo 11, que teve um valor de 45,24%.

Já para a etapa de validação, os valores de correlação e Nash foram significativamente menores do que a etapa de Treinamento. Mais especificamente, apenas os modelos 08, 10 e 11 tiveram Nash positivo. Os maiores valores de correlação e Nash foram obtidos no modelo 11, com 49,39% e 19,41%, respectivamente.

Como pode-se notar na tabela 6, os modelos gerados utilizando o preditor JJA tiveram correlações e Nash na etapa de Treinamento próximos dos encontrados com os modelos que tinham como preditor o trimestre SON. Os valores de correlações variaram de 34,67% à 57,41 e de Nash de 12,02% à 32,96%, correspondentes aos modelos 01 e 09, respectivamente. Entretanto, a correlação e o Nash da validação foram maiores, sendo que o maior resultado encontrado foi o do modelo 09, com 67,15% e 37,21%, respectivamente.

Na tabela 7, nota-se que o Nash foi negativo para todos os modelos na validação, sendo que o melhor resultado foi -5,31% referente ao modelo 03, e pior -65,83% para o modelo

10. Já na etapa de Treinamento, no modelo 10 foi encontrado uma correlação de 60,56% e Nash de 36,68%, sendo o melhor modelo para essa etapa.

Ademais, como mostra a tabela 8, na Treinamento a maior correlação foi do modelo 09 que teve um valor de 33,88%, entretanto o maior valor de Nash foi o do modelo 10. Já na validação, a maior correlação foi do modelo 11, com 33,88%, sendo que esse modelo também apresentou maior Nash, com valor um valor de 10,01%.

#### 4.2.2 Regressão Ridge

Para os 11 modelos gerados para cada preditor a partir da Regressão Ridge, utilizando a mesma seleção de variáveis dos modelos da Regressão Tradicional também foram calculados os valores de Nash e do coeficiente de Regressão Pearson para as etapas de Treinamento e validação. Os resultados podem ser conferidos nas tabelas 9, 10, 11 e 12.

Tabela 9 - Resultados dos modelos de previsão utilizando regressão Ridge e como preditor o trimestre SON

Ridge: SON	Treinamento		Validação	
	R	Nash	R	Nash
Modelo 01	40,03%	16,02%	16,00%	-6,02%
Modelo 02	39,06%	14,68%	13,56%	-4,38%
Modelo 03	38,23%	13,99%	14,69%	-3,22%
Modelo 04	37,95%	12,09%	13,57%	-1,52%
Modelo 05	37,91%	11,53%	13,08%	-1,34%
Modelo 06	37,38%	11,41%	14,53%	-0,72%
Modelo 07	44,25%	16,52%	16,88%	-0,23%
Modelo 08	45,58%	17,09%	18,74%	0,82%
Modelo 09	45,60%	17,34%	17,70%	0,23%
Modelo 10	63,76%	38,47%	42,91%	17,32%
Modelo 11	64,50%	38,73%	45,08%	19,76%

Fonte: Autor 2022

Tabela 10 - Resultados dos modelos de previsão utilizando regressão Ridge e como preditor o trimestre JJA.

Ridge: JJA	Treinamento		Validação	
	R	Nash	R	Nash
Modelo 01	34,67%	12,02%	20,38%	-0,05%
Modelo 02	36,74%	14,64%	26,85%	1,89%

Tabela 10 - Resultados dos modelos de previsão utilizando regressão Ridge e como preditor o trimestre JJA.

Ridge: JJA	Treinamento		Validação	
	R	Nash	R	Nash
Modelo 03	36,94%	11,83%	24,56%	1,47%
Modelo 04	39,84%	12,31%	24,65%	1,17%
Modelo 05	39,08%	11,97%	24,51%	1,42%
Modelo 06	39,28%	11,84%	23,59%	1,19%
Modelo 07	49,15%	20,00%	47,81%	14,69%
Modelo 08	47,77%	18,03%	43,67%	11,38%
Modelo 09	46,85%	16,99%	40,40%	9,96%
Modelo 10	48,30%	18,40%	45,87%	13,03%
Modelo 11	48,82%	18,12%	45,82%	12,26%

Fonte: Autor 2022

Tabela 11 - Resultados dos modelos de previsão utilizando regressão Ridge e como preditor o trimestre MAM.

Ridge: MAM	Treinamento		Validação	
	R	Nash	R	Nash
Modelo 01	21,42%	4,59%	2,08%	-7,16%
Modelo 02	19,92%	2,20%	3,65%	-5,80%
Modelo 03	25,11%	4,19%	2,32%	-5,20%
Modelo 04	24,56%	4,00%	0,90%	-5,87%
Modelo 05	25,45%	4,48%	2,91%	-6,58%
Modelo 06	24,90%	3,42%	3,47%	-6,08%
Modelo 07	25,09%	2,94%	5,03%	-6,11%
Modelo 08	32,36%	5,66%	9,09%	-3,72%
Modelo 09	31,44%	4,52%	8,74%	-3,98%
Modelo 10	35,60%	5,64%	5,31%	-4,73%
Modelo 11	34,68%	4,78%	9,21%	-4,14%

Fonte: Autor 2022

Tabela 12 - Resultados dos modelos de previsão utilizando regressão Ridge e como preditor o trimestre DJF

Ridge: DJF	Treinamento		Validação	
	R	Nash	R	Nash
Modelo 01	3,22%	0,10%	29,02%	-6,39%
Modelo 02	12,22%	1,49%	10,98%	-8,22%
Modelo 03	20,16%	4,06%	18,03%	-12,85%
Modelo 04	23,51%	5,47%	9,30%	-8,68%
Modelo 05	26,90%	7,19%	22,36%	3,57%
Modelo 06	27,27%	7,35%	22,76%	3,76%
Modelo 07	30,82%	9,41%	26,86%	5,70%
Modelo 08	31,75%	9,9754%	28,44%	9,50%
Modelo 09	31,76%	9,9816%	30,83%	8,10%

Tabela 13 - Resultados dos modelos de previsão utilizando regressão Ridge e como preditor o trimestre DJF

Ridge: DJF	Treinamento		Validação	
	R	Nash	R	Nash
Modelo 10	32,16%	10,26%	33,08%	9,67%
Modelo 11	32,18%	10,30%	32,82%	9,49%

Fonte: Autor 2022

De acordo com a tabela 09, a etapa de treinamento para o preditor SON apresentou correlações que variavam de 37,38%, para o modelo 06, até 64,50%, para o modelo 11, sendo esses modelos o de menor e maior Nash, com valores de 11,41% e 38,73%, respectivamente. Sendo que o modelo 11 também foi o que teve maior correlação e Nash para a etapa de validação. Já a pior correlação e Nash ficou para os modelos 05 e 01, que apresentaram um valor de 13,08% de correlação e -6,02% de Nash, respectivamente.

Com relação ao preditor JJA, a maior correlação e Nash para etapa de Treinamento foi encontrado no modelo 07, apresentando uma correlação de 49,15% e 20,00%, como pode-se notar pela tabela 10. Além disso, nota-se que o mesmo modelo obteve os maiores valores também para a etapa de validação, com uma correlação de 47,81% e Nash de 14,69%.

A tabela 11 mostra que existiu variações modestas entre os resultados dos modelos que tinham como preditor o trimestre MAM. No treinamento, a menor correlação encontrada foi no modelo 02 com 19,92% e a maior foi obtido no modelo 11 com 32,18%, mostrando assim uma variação de apenas 12,26%. Já na validação os valores de correlação variaram de 0,90% até 9,21%, que correspondem aos modelos 04 e 11, respectivamente. Além disso, o Nash apresentou valores baixos tanto na etapa de treinamento como na validação. Enquanto que o maior Nash de treinamento teve como maior valor 5,66%, referente ao modelo 08, todos os valores de Nash de validação apresentaram valores negativos.

Ademais, para a etapa de treinamento os modelos para o preditor DJF apresentam valores de correlações e Nash crescente, sendo o modelo 11 o que apresentou resultados mais elevados, com uma correlação de 32,18% e Nash de 10,30%. Já na validação, as correlações variaram de 9,30% à 33,08%, referente aos modelos 04 e 10, respectivamente. Com relação ao Nash. Os primeiros 04 modelos tiveram valores negativos, enquanto que os demais apresentaram valores que partiam de 3,57% até 9,67%, correspondentes aos modelos 05 e 10, respectivamente.

### 4.2.3 Regressão Lasso

Por fim, para os 11 modelos gerados para cada preditor a partir da Regressão Lasso, utilizando a mesma seleção de variáveis dos modelos da Regressão Tradicional também foram calculados os valores de Nash e do coeficiente de Regressão Pearson para as etapas de Treinamento e validação. Os resultados podem ser conferidos nas tabelas 13, 14, 15 e 16.

Tabela 14 - Resultados dos modelos de previsão utilizando regressão Lasso e como preditor o trimestre SON

Lasso: SON	Treinamento		Validação	
	R	Nash	R	Nash
Modelo 01	40,03%	16,02%	16,00%	-6,02%
Modelo 02	39,82%	15,40%	16,00%	-2,73%
Modelo 03	39,82%	15,40%	16,00%	-2,73%
Modelo 04	39,82%	14,70%	16,00%	-1,40%
Modelo 05	39,82%	14,70%	16,00%	-1,40%
Modelo 06	39,82%	14,70%	16,00%	-1,40%
Modelo 07	47,24%	20,33%	16,22%	-1,90%
Modelo 08	47,23%	19,94%	16,24%	-1,53%
Modelo 09	47,23%	19,94%	16,24%	-1,53%
Modelo 10	64,49%	39,92%	44,50%	18,58%
Modelo 11	64,42%	39,15%	44,70%	19,17%

Fonte: Autor 2022

Tabela 15 - Resultados dos modelos de previsão utilizando regressão Lasso e como preditor o trimestre JJA.

Lasso: JJA	Treinamento		Validação	
	R	Nash	R	Nash
Modelo 01	34,67%	12,02%	20,38%	-0,05%
Modelo 02	38,27%	14,68%	26,51%	0,47%
Modelo 03	38,24%	14,47%	26,22%	1,20%
Modelo 04	40,86%	15,93%	24,91%	1,91%
Modelo 05	40,76%	15,46%	24,60%	2,01%
Modelo 06	40,24%	13,78%	23,44%	1,87%
Modelo 07	53,89%	28,40%	58,04%	26,76%
Modelo 08	53,86%	28,02%	58,04%	26,40%
Modelo 09	53,84%	27,77%	58,04%	26,15%
Modelo 10	53,85%	28,09%	58,02%	26,46%
Modelo 11	46,54%	17,10%	53,09%	14,96%

Fonte: Autor 2022

Tabela 16 - Resultados dos modelos de previsão utilizando regressão Lasso e como preditor o trimestre MAM.

Lasso: MAM	Treinamento		Validação	
	R	Nash	R	Nash
Modelo 01	21,42%	4,59%	2,08%	-7,16%
Modelo 02	21,35%	4,50%	2,08%	-6,38%
Modelo 03	28,20%	7,36%	8,67%	-4,79%
Modelo 04	37,99%	13,88%	2,70%	-22,71%
Modelo 05	29,95%	8,35%	2,71%	-7,73%
Modelo 06	29,81%	7,61%	3,91%	-6,09%
Modelo 07	29,62%	7,06%	4,82%	-5,36%
Modelo 08	37,59%	12,29%	29,57%	4,45%
Modelo 09	37,03%	10,05%	32,00%	3,21%
Modelo 10	61,20%	36,67%	6,91%	-64,21%
Modelo 11	61,20%	36,68%	6,96%	-64,50%

Fonte: Autor 2022

Tabela 17 - Resultados dos modelos de previsão utilizando regressão Lasso e como preditor o trimestre DJF.

Lasso: DJF	Treinamento		Validação	
	R	Nash	R	Nash
Modelo 01	3,22%	0,10%	29,02%	-6,39%
Modelo 02	12,18%	1,46%	9,07%	-7,12%
Modelo 03	20,13%	4,03%	17,05%	-11,52%
Modelo 04	23,24%	5,29%	9,39%	-8,32%
Modelo 05	26,71%	7,02%	21,76%	3,32%
Modelo 06	26,71%	7,02%	21,76%	3,33%
Modelo 07	30,03%	8,91%	27,02%	5,91%
Modelo 08	30,74%	9,33%	29,42%	7,30%
Modelo 09	30,74%	9,33%	29,41%	7,30%
Modelo 10	31,35%	9,70%	32,99%	9,62%
Modelo 11	31,54%	9,82%	31,88%	8,87%

Fonte: Autor 2022

A partir da Tabela 13, é possível verificar que para o preditor SON o Nash mostrou resultados negativos para os primeiros nove modelos na etapa de validação, já para os modelos 10 e 11, apresentou valores de 18,58% e 19,17%, respectivamente. Já os valores de correlação para a mesma etapa variaram de 16,00% à 44,70, sendo o maior referente ao modelo 11. Na etapa de treinamento, o modelo 11 também foi o que apresentou maiores resultados, com uma correlação de 64,42% e Nash 39,15%. É importante salientar que alguns modelos tiveram valores iguais quando analisando apenas duas casas decimais, porém esses valores são distintos, fato esse que acontece com os demais modelos.

Na tabela 14 é possível notar que para o preditor JJA os maiores valores de correlação e Nash encontrados foram para o modelo 07, com 53,89% e 28,40%, respectivamente. Esse modelo também foi o que apresentou os maiores valores para a etapa de validação, com uma correlação de 58,04% e Nash de 26,76%.

Já com relação a predito MAM, verifica-se que o Nash de validação apresentou valores negativo para a maioria dos casos, exceto para os modelos 08 e 09, que tiveram resultados de 4,45% e 3,21%, respectivamente. Como maior valor de correlação para essa etapa tem-se o modelo 09, com um valor de 32,00%. Ademais, a maior correlação foi de 61,20% e Nash 36,68%, ambos referentes ao modelo 11.

Por fim, o preditor DJF apresentou resultados crescentes, sendo o modelo 11 o que mostrou os maiores valores, a correlação encontrada foi de 31,54% e o Nash de 9,82%. Já o modelo 10 o que teve valores de correlação e Nash mais elevados para a etapa de validação, com 32,99% e 9,62%.

### 4.3 Comparação dos modelos

Os modelos que tiveram melhores resultados nem sempre permaneceram os mesmos para todos os tipos de regressão. As tabelas 17 e 18 mostra os três primeiros modelos com maior correlação e Nash para cada tipo de preditor e regressão, respectivamente.

Tabela 18 – Comparação dos resultados obtidos de correlação para cada tipo de regressão

	Preditor	Treinamento						Validação					
		1°		2°		3°		1°		2°		3°	
		R	M	R	M	R	M	R	M	R	M	R	M
OLS	SON	69,58%	10	67,26%	11	51,31%	9	49,39%	11	46,66%	10	29,49%	8
	JJA	57,41%	9	57,37%	8	56,59%	7	67,15%	9	66,69%	8	65,17%	7
	MAM	60,56%	10	55,44%	11	53,58%	9	19,55%	9	16,62%	8	8,87%	7
	DJF	33,92%	10	33,88%	9	32,87%	8	33,60%	11	32,23%	9	32,13%	8
Ridge	SON	64,50%	11	63,76%	10	45,60%	9	45,08%	11	42,91%	10	18,74%	8
	JJA	49,15%	7	48,82%	11	48,30%	10	47,81%	7	45,87%	10	45,82%	11
	MAM	35,60%	10	34,68%	11	32,36%	8	9,21%	11	9,09%	8	8,74%	9
	DJF	32,18%	11	32,16%	10	31,76%	9	33,08%	10	32,82%	11	30,83%	9
Lasso	SON	64,49%	10	64,42%	11	47,24%	7	44,70%	11	44,50%	10	16,24%	8 e 10
	JJA	53,89%	7	53,86%	8	53,85%	10	58,04%	7, 8 e 9	58,02%	10	53,09%	11
	MAM	61,20%	10 e 11	37,59%	8	37,03%	9	32,00%	9	29,57%	8	6,96%	11
	DJF	31,54%	11	31,35%	10	30,74%	8 e 9	32,99%	10	31,88%	11	29,42%	8

Fonte: Autor 2022

Tabela 19 - Comparação dos resultados obtidos de Nash para cada tipo de regressão.

	Preditor	Treinamento						Validação					
		1°		2°		3°		1°		2°		3°	
		Nash	M	Nash	M	Nash	M	Nash	M	Nash	M	Nash	M
OLS	SON	48,42%	10	45,24%	11	26,33%	9	19,41%	11	11,55%	10	0,70%	8
	JJA	32,96%	9	32,91%	8	32,03%	7	37,21%	9	36,59%	8	34,52%	7
	MAM	36,68%	10	30,74%	11	28,71%	9	-5,31%	2	-6,21%	3	-7,16%	1
	DJF	11,50%	10	11,48%	9	10,80%	8	10,01%	11	8,25%	10	8,18%	9
Ridge	SON	38,73%	11	38,47%	10	17,34%	9	19,76%	11	17,32%	10	0,82%	8
	JJA	20,00%	7	18,40%	10	18,12%	11	14,69%	7	13,03%	10	12,26%	11
	MAM	5,66%	8	5,64%	10	4,78%	11	-3,72%	8	-3,98%	9	-4,14%	11
	DJF	10,30%	11	10,26%	10	9,98%	9	9,67%	10	9,50%	8	9,49%	11
Lasso	SON	39,92%	10	39,15%	11	20,33%	7	19,17%	11	18,58%	10	-1,40%	4, 5 e 6
	JJA	28,40%	7	<b>28,09%</b>	<b>10</b>	28,02%	8	26,76%	7	26,46%	10	26,40%	8
	MAM	36,68%	11	36,67%	10	13,88%	4	4,45%	8	3,21%	9	-4,79%	3
	DJF	9,82%	11	9,70%	10	9,33%	8 e 9	9,62%	10	8,87%	11	7,30%	8 e 9

Fonte: Autor 2022

Analisando os resultados obtidos para cada preditor, nota-se que os modelos gerados a partir da regressão Ridge e Lasso apresentaram valores menores de correlação e Nash para todos os modelos da etapa de treinamento, quando comparados com a regressão tradicional, exceto para os modelos 04, 09 e 10 para a regressão Lasso e preditor MAM.

Em contrapartida, alguns modelos tiveram maiores valores de Nash de validação quando os modelos utilizavam regressão Ridge ou Lasso. Nota-se, por exemplo, que o preditor SON a regressão Ridge obteve resultados maiores em todos os modelos quando comparado com a regressão Lasso, e que para o preditor MAM, para todos os modelos, exceto o modelo 02 e 11, os valores de Nash foram maiores aos encontrados na regressão tradicional.

## 5 CONCLUSÃO

Este trabalho teve como intuito gerar modelos de previsão de precipitação para a quadra chuvosa da Região Hidrográfica de Crateús e verificar a eficácia desses modelos. Foram utilizados três tipos de regressões e quatro preditores TSM, sendo esse os trimestres SON, JJA, MAM e DJF, gerando assim um total de 33 modelos. Com relação a seleção dos índices TSM, o mesmo ocorreu mediante a aplicação do Algoritmo Iterative Input Selection.

Esse Algoritmo indicou os índices TSM relacionados ao fenômeno El Niño e La Niña como as variáveis mais importantes para os preditores, em especial o índice NINO3 e o NINO4. Já os índices relacionados ao dipolo do Atlântico foram classificados em posições intermediárias e inferiores, sendo que o índice TAD é que apresentou pior classificação. Os demais índices, por sua vez tiveram posições variadas dependendo do período preditor.

Diante dos resultados obtidos nos modelos gerados, notou-se que foram satisfatórios. Com relação as correlações, as mesmas apresentaram valores elevados, tanto para a etapa de treinamento e validação. Diversos modelos apresentaram correlações compreendidas entre 0,30 a 0,49, enquanto que outros alcançaram valores maiores que 0,5, caracterizando correlações intermediárias e elevadas, respectivamente, segundo Cohen (1988). Além disso, a maioria dos modelos apresentaram valores de Nash positivo, o que mostra que as previsões geradas tiveram um resultado melhor do que a média dos valores observados.

Com relação aos desempenhos das regressões, a OLS foi a que apresentou resultados mais elevados para a etapa de treinamento para a maioria dos modelos. Enquanto os valores de correlação e Nash para as demais regressões mostrou-se com melhor desempenho para alguns casos na etapa de validação.

Para estudos futuros recomenda-se:

- a) Utilizar diferentes tipos de regressões na geração de modelos de previsão de precipitação para a RHSC.
- b) Verificar a eficiência dos modelos com relação a diferentes períodos preditivos e preditores.
- c) Analisar a influência de outros índices TSM na precipitação da região.

## REFERÊNCIAS

ANDREOLI, R. V.; KAYANO, M. T. Multi-scale variability of the sea surface temperature in the Tropical Atlantic. **Journal of Geophysical Research C: Oceans**, v. 109, n. 5, p. 1–12, 2004.

ARAGÃO, JOR de. O Impacto do ENSO e do Dipolo do Atlântico no Nordeste do Brasil. **Bull. Inst. fr. études andines**, v. 27, n. 3, p. 839-844, 1998.

BREIMAN, L. **Random Forests** Kluwer Academic Publishers, 2022. .

BUENO, J. C. M. *et al.* **Fases do Enos e características da precipitação pluvial e temperatura na região de Guarapuava, Sul do Brasil**, 2020.

CARMO, M. V. N. S. do; LIMA, C. H. R. Caracterização Espaço-Temporal das Secas no Nordeste a partir da Análise do índice SPI. **Revista Brasileira de Meteorologia**, v. 35, n. 2, p. 233–242, 2020.

CASAGRANDE, M. H. **Comparação de métodos de estimação para problemas com colinearidade e/ou alta dimensionalidade**. 2016. Universidade Federal de São Carlos, 2016.

CASTRO, L. N. de; FERRARI, D. G. **Introdução à Mineração de Dados**. v. 1

CEARÁ. **PLANO ESTADUAL DE CONVIVÊNCIA COM A SECA**, 2018. .

CEARÁ, G. E. Plano estadual de convivência com a seca ações emergenciais e estruturantes. Fortaleza, 2015. ARAÚJO, C. B. C. *et al.* Previsão de vazões: Utilização de RNA. In: ADAPTA: Gestão adaptiva do risco climático de seca/ organizado por Francisco de Assis de Souza Filho *et al.* Fortaleza: Expressão Gráfica e Editora, 2019.

CHEIN, F. **Introdução aos modelos de regressão linear: um passo inicial para compreensão da econometria como uma ferramenta de avaliação de políticas públicas**.

COGERH. **Sertões de Crateús**. Disponível em: <<https://portal.cogerh.com.br/serto-es-de-crateus/>>.

DOS SANTOS, Nayara Arroxelas; SIQUEIRA, Anderlan Henrique Batista; DE MELO, Maria Luciene Dias. PRECIPITAÇÃO SOBRE A AMÉRICA DO SUL E SUAS RELAÇÕES COM A OSCILAÇÃO MULTIDECADAL DO ATLÂNTICO E A OSCILAÇÃO DO ATLÂNTICO NORTE (RAINFALL OVER SOUTH AND ITS RELATIONS WITH AMERICA SWING MULTIDECADAL ATLANTIC AND SWING THE NORTH ATLANTIC). **Revista Brasileira de Climatologia**, v. 18, 2016.

FERREIRA, A. G.; MELLO, N. G. da S. Principais Sistemas Atmosféricos Atuantes Sobre a Região Nordeste Do Brasil E a Influência Dos Oceanos Pacífico E Atlântico No Clima Da Região. **Revista Brasileira de Climatologia**, v. 1, n. 1, p. 15–28, 2005.

FERREIRA, F. F. *et al.* Comportamento das águas superficiais nos oceanos Pacífico e Atlântico durante o período de 1982 a 2014. **Revista Brasileira de Meteorologia**, v. 31, n. 3, p. 366–373, 2016.

GALELLI, S.; CASTELLETTI, A. Tree-based iterative input variable Selectionion for hydrological modeling. **Water Resources Research**, v. 49, n. 7, p. 4295–4310, 2013.

HIERA, Mitchel Druz; JUNIOR, Antonio Ferreira Lima; ZANELLA, Maria Elisa. Análise da tendência da precipitação no Estado do Ceará no período 1974 a 2016. **Revista Brasileira de Climatologia**, v. 24, 2019.

HOFFMANN, Rodolfo; VIEIRA, Sônia. Análise de regressão: uma introdução à econometria. **São Paulo**, 1998.

HURRELL, J. W. Decadal trends in the North Atlantic oscillation: Regional temperatures and precipitation. **Science, New Series**, v. 269, n. 5224, p. 676–679, 1995.

IBGE. **Estimativa da população residente no Brasil e unidades da federação com data de referência em 1º de julho de 2022**. Disponível em:

<[https://ftp.ibge.gov.br/Estimativas\\_de\\_Populacao/Estimativas\\_2022/estimativa\\_dou\\_2022.pdf](https://ftp.ibge.gov.br/Estimativas_de_Populacao/Estimativas_2022/estimativa_dou_2022.pdf)>.

INSTITUTO DE PESQUISA E ESTRATÉGIA ECONÔMICA DO CEARÁ - IPECE. **Perfil das Regiões de Planejamento**. Fortaleza: IPECE 2017. .

IPMA. **Área educativa - NAO [North Atlantic Oscillation - Oscilação do Atlântico Norte]**. Disponível em: <<https://www.ipma.pt/pt/oclima/normais.clima/>>.

IPMA, Instituto Português do Mar e da Atmosfera. Área educativa (2014) NAO (North Atlantic Oscillation - Oscilação do Atlântico Norte). Acesso em: 22 de março de 2014. Disponível em: <http://www.ipma.pt/pt/educativa/tempo.clima/index.jsp?>

KNIGHT, J. R.; FOLLAND, C. K.; SCAIFE, A. A. Climate impacts of the Atlantic multidecadal oscillation. **Geophysical Research Letters**, v. 33, n. 17, p. 2–5, 2006.

KAYANO, Mary Toshie; CAPISTRANO, Vinicius Buscioli. **How the Atlantic multidecadal oscillation (AMO) modifies the ENSO influence on the South American rainfall**. **International Journal of Climatology**, v. 34, n. 1, p. 162-178, 2014.

LACHENBRUCH, Peter A. Statistical power analysis for the behavioral sciences. **Journal of the American Statistical Association**, v. 84, n. 408, p. 1096-1097, 1989.

LIU, Huan; MOTODA, Hiroshi. In: *Computational methods of feature Selectionion*. Chapman and Hall/CRC, 2007. p. 19-34.

MAIER, H. R. *et al.* Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. **Environmental Modelling and Software**, v. 25, n. 8, p. 891–909, 2010.

MARENGO, J. a; VALVERDE, M. C. Caracterização do clima no Século XX e Cenário de Mudanças de clima para o Brasil no Século XXI usando os modelos do IPCC-AR4. **Revista Multiciência**, n. 8, p. 5–28, 2007.

MARENGO, J., DRUYAN, L., HASTENRATH, S. Observational and modelling studies of Amazonia interannual climate variability. *Climatic Change*, v. 23, n.3, p.267-286, 1993.

MENEZES, Hudson Ellen Alencar et al. A relação entre a temperatura da superfície dos oceanos tropicais e a duração dos veranicos no estado da Paraíba. **Revista Brasileira de Meteorologia**, v. 23, n. 2, p. 152-161, 2008.

MOLION, L. C. B. Aquecimento Global, El Niño, Manchas Solares, Vulcões e Oscilação Decadal do Pacífico. **Revista climanálise**, 2005.

MOLION, Luiz Carlos Baldicero; BERNARDO, S. de O. Uma revisão da dinâmica das chuvas no nordeste brasileiro. **Revista Brasileira de Meteorologia**, v. 17, n. 1, p. 1-10, 2002.

MORIASI, Daniel N. *et al.* Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. **Transactions of the ASABE**, v. 50, n. 3, p. 885-900, 2007.

NCAR. **Dados Climáticos Figuras chave Cite esta página**. Disponível em: <<https://climatedataguide.ucar.edu/climate-data/nino-sst-indices-nino-12-3-34-4-oni-and-tni>>.

NOAA. Índice ENSO multivariado versão 2 (MEI.v2). v. 2, p. 1–5, 2022. Disponível em: <<https://psl.noaa.gov/enso/mei/>>.

NÓBREGA, J. N. DA *et al.* RELAÇÕES COM A TSM DOS OCEANOS TROPICAIS Universidade Federal de Campina Grande ( UFCG ), Unidade Acadêmica de Ciências Atmosféricas ,. **Revista Brasileira de Meteorologia**, p. 197–208, 2014.

SILVA NÓBREGA, Ranyére; CHAGAS FELIPE SANTIAGO, Gabriela Ayane. Tendências do controle climático oceânico sob a variabilidade temporal da precipitação no Nordeste do Brasil. **Revista de Geografia Norte Grande**, n. 63, p. 9-26, 2016.

PARANHOS, R. *et al.* Desvendando os Mistérios do Coeficiente de Correlação de Pearson: o Retorno. **Leviathan (São Paulo)**, n. 8, p. 66, 2009.

RICHADSON, Roberto Jarry. **Pesquisa Social: Métodos e Técnicas**. 3. ed. São Paulo, SP: Atlas, 2011.

RIQUETI, G. A.; RIBEIRO, C. E.; ZÁRATE, L. E. Classificando perfis de longevidade de bases de dados longitudinais usando floresta aleatória. *In: Symposium on Knowledge Discovery; mining and Learning – KMMILE*, 2018.

ROSA, C. S. **Operação De Reservatórios Do Sistema De Abastecimento Dos Municípios Da Região Hidrológica Dos Sertões De Crateús**. 2019. 2019.

SANTOS, C. A. C. dos *et al.* Tendências dos índices de precipitação no estado do Ceará. **Revista Brasileira de Meteorologia**, v. 24, n. 1, p. 39–47, 2009.

SILVA, D. F. Influência Interdecadal (ODP e OMA) nas Cotas do Rio São Francisco. **Revista Brasileira de Geografia Física**, v. 06, p. 1275–1291, 2012.

SILVA, Vicente de Paulo Rodrigues da; PEREIRA, Emerson Ricardo Rodrigues; ALMEIDA, Rafaela Silveira Rodrigues. Estudo da variabilidade anual e intra-anual da precipitação na região nordeste do Brasil. **Revista Brasileira de Meteorologia**, v. 27, n. 2, p. 163-172, 2012.

STAŃCZYK, Urszula; ZIELOSKO, Beata; JAIN, Lakhmi C. (Ed.). *Advances in feature Selectionion for data and pattern recognition*. Springer International Publishing, 2018.

UVO, C. B.; BERNDTSSON, R. Regionalization and spatial properties of Ceará State rainfall in northeast Brazil. **Journal of Geophysical Research Atmospheres**, v. 101, n. D2, p. 4221–4233, 1996.

WALPOLE, R. E. *et al.* Probabilidade e estatística para engenharia e ciências. p. 319–321, 2009.

YIN, R. K. **Estudo de caso: planejamento e métodos**. 4. Ed. Porto Alegre: Bookman, 2010.