

## Meteorology

# O uso de correlações máxima de TSM do dipolo do Atlântico e a precipitação no estado do Ceará

The use of maximum correlations of TSM of the Atlantic dipole and precipitation in the state of Ceará

Werbesson da Silva Freitas<sup>1</sup> 

<sup>1</sup> Universidade Estadual do Ceará, Fortaleza, CE, Brasil

## RESUMO

É sabido que a relação entre o oceano e a atmosfera é um grande precursor de uma vasta gama de eventos meteorológicos que influenciam todo o globo, incluindo o regime de chuvas em várias regiões. Entender como esses eventos explicam a precipitação é de constante interesse. Neste sentido, o presente trabalho objetivou determinar as correlações máximas entre as anomalias de Temperatura Superficial do Mar (TSM) na região do dipolo do Atlântico e a precipitação no Estado do Ceará. Para isso, foram utilizados dados observados e interpolados de precipitação diária e de anomalias de TSM oriundos do National Oceanic and Atmospheric Administration (NOAA) referentes ao período de 1982 a 2021. Usando um algoritmo de árvore entre as correlações máximas do Dipolo do Atlântico, baseados em dados diários de TSM, e a precipitação, propomos uma metodológica para prever a precipitação no Estado do Ceará.

**Palavras-chave:** Dipolo do Atlântico; Precipitação; Ceará

## ABSTRACT

It is known that the relationship between the ocean and the atmosphere is a major precursor of a wide range of meteorological events that influence the entire globe, including the rainfall regime in various regions. Understanding how these events explain precipitation is of constant interest. In this sense, the present work aimed to determine the maximum correlations between the Surface Sea Temperature (SST) anomalies in the Atlantic dipole region and precipitation in the State of Ceará. For this, we used observed and interpolated data of daily precipitation and SST anomalies from the National Oceanic and Atmospheric Administration (NOAA) for the period from 1982 to 2021. Using a tree algorithm between the maximum correlations of the Atlantic Dipole, based on in daily SST data, and precipitation we propose a methodological proposal to predict precipitation in the State of Ceará.

**Keywords:** Atlantic Dipole; Precipitation; Ceará

## 1 INTRODUÇÃO

É notório que os conhecimentos de clima e de tempo estão ligados aos setores econômicos e sociais de uma determinada região, por exemplo, ao prever eventos hidrológicos extremos, seca ou cheia, as autoridades competentes podem mitigar os impactos desses eventos, portanto, esse tipo de conhecimento serve como um ponto de apoio na tomada de decisões de Política Públicas no intuito de otimizar os recursos que serão empregados.

Devido à natureza complexa do clima, existem várias lacunas na sua representação em diferentes escalas cronológicas. Além disso, o clima não é dado como um fator constante a uma determinada área, pois o mesmo define-se por estimativas climáticas que envolvem vários fatores e elementos meteorológicos de uma determinada região. Tais fenômenos podem apresentar-se em macroescala, mesoescala, microescala ou combinação destas.

Um dos principais fatores que influenciam o clima da Terra é a interação entre a água e o ar. Ao adquirirem uma dinâmica própria na circulação atmosférica, passam a ter uma grande influência no meio (Cavalcante *et al.*, 2009). Na região Nordeste do Brasil (NEB) isso se manifesta por meio de fenômenos meteorológicos atuantes em diferentes épocas do ano e que tem forte influência no clima e no tempo nessa região. Por exemplo, a Zona de Convergência Intertropical (ZCIT), que, associada aos fenômenos de El Niño e La Niña, pode aumentar ou diminuir os totais acumulados de precipitação pluviométrica; as Frentes Frias; os Vórtices Ciclônicos em Altos Níveis (VCAN); as Linhas de Instabilidade; os Complexos Convectivos de Mesoescala (CCMs); as Brisas Marítimas e Terrestres; e as anomalias de Temperatura Superficial do Mar (TSM) no Atlântico. O correto entendimento de todos esses fenômenos é essencial para a construção de previsões climáticas (Ferreira, 2005).

As interações oceano-atmosfera possuem forte relação com os totais acumulados de precipitação no NEB; os eventos anômalos de TSM nos oceanos Pacífico e Atlântico são considerados a principal variável física na propulsão ou inibição de chuvas na região, bem como em toda faixa equatorial (Walter, 1928; Silva *et al.*, 2011). De maneira mais específica, as mudanças de TSM no Oceano Atlântico influenciam fortemente a chuva no NEB. O Dipolo do Atlântico é um índice estatístico que relaciona a mudança anômala na TSM (Servain, 1991). Vários trabalhos comprovam a relação entre o Dipolo do Atlântico e a pluviosidade no NEB, mas existem limitações quanto ao grau de previsibilidade de ocorrência de precipitação. Portanto, considera-se que novos métodos computacionais devem ser estudados com o objetivo de aprimorar essa previsibilidade (Zebiak *et al.*, 1987; Alves *et al.*, 2006).

Dessa forma, o presente trabalho objetivou determinar estatisticamente a previsibilidade da intensidade da precipitação no Estado do Ceará a partir do Dipolo do Atlântico em uma escala cronológica de meses. Isso foi feito de forma a identificar com maior precisão a influência desse fenômeno nas chuvas no Ceará utilizando o método aplicado de correlações cruzadas de Ludescher (2014). Buscou-se ainda determinar a escala de tempo de influência da TSM na pluviosidade desta região.

Ao final da introdução recomenda-se que seja realizada uma apresentação sucinta da estrutura geral do artigo de modo a permitir que o leitor compreenda como o assunto será abordado a partir de então.

A literatura presente está estruturada em cinco seções principais, a primeira delas é a introdução na qual é realçado a relação entre eventos meteorológicos e a vida humana, demonstrando a importância de ter sapiência sobre eventos climáticos, a segunda parte é denominada de materiais e métodos, na mesma há uma explanação sobre as regiões de estudo que no caso é o estado do Ceará (2,5°S - 7,5°S e 37,5°W - 41,5°W) e a região do oceano atlântico tropical (28°N - 20°S e 60°W - 15°E), além do mais é explicado como funciona o dipolo do atlântico e

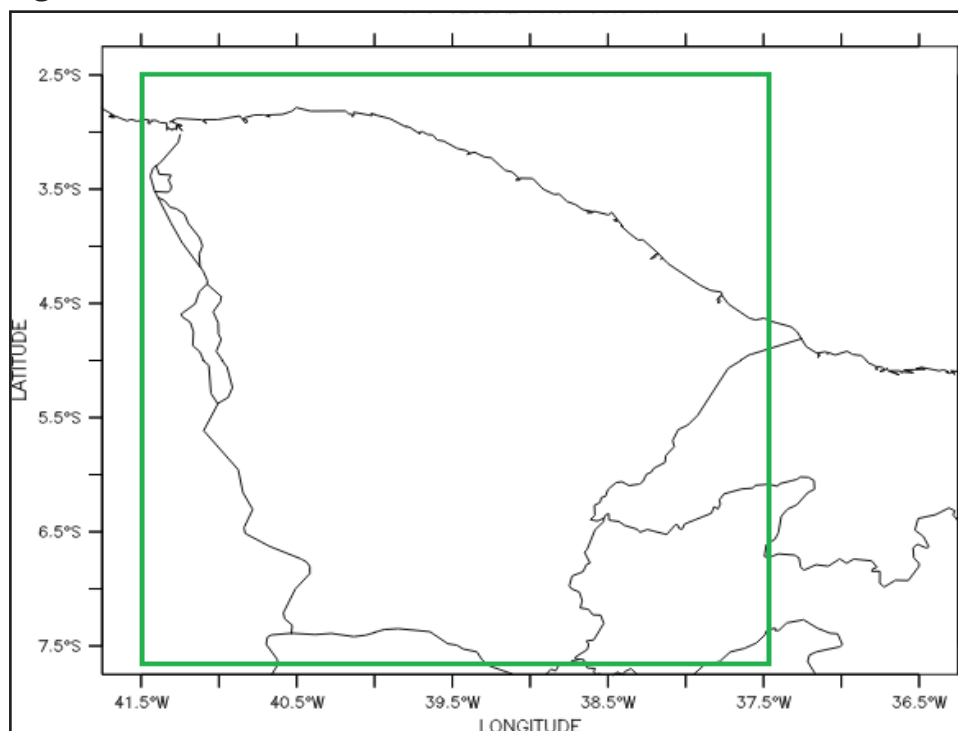
alguns parâmetros utilizados na metodologia como expoente de Hurst e séries de correlações máximas. Na terceira parte, resultados e discussões, são expostos de maneira explícita como foi realizado as correlações entre o dipolo do atlântico e o nível de precipitação no estado estado Ceará e quais resultados foram obtidos. Na quarta seccao, consideracoes finais, é realizado uma síntese das possíveis conclusões que podem ser tiradas diante do exposto na seção três, por fim, a quinta parte é composta referências utilizadas na construção da literatura.

## **2 MATERIAIS E MÉTODOS**

### **2.1 Regiões de estudo**

O Estado do Ceará tem sua área estabelecida na Região Nordeste do Brasil, fazendo fronteira com o Rio Grande do Norte e Paraíba a Leste, com o Oceano atlântico Sul ao Norte, Pernambuco ao Sul e Piauí a Oeste, utilizando as coordenadas geográficas o mesmo se encontra entre (2,5°S - 7,5°S) e (37,5°W - 41,5°W) como mostra o retângulo da Figura 1. O mesmo fica próximo a linha do equador, sua extensão territorial é de 148.826 Km<sup>2</sup>, representando assim 1,75% de todo o território nacional, toda essa extensão é subdividida em oito macrorregiões, Fortaleza, Baturité, Litoral Oeste, Sertão dos Inhamuns, Litoral Leste-Jaguaribe, Cariri centro-sul, Sertão central, Sobral-Ibiapaba, sendo integrantes dessas regiões 184 municípios, no ranking territorial de extensão dos estados, o Ceará é o quarto maior estado do NEB, dessa região cerca de 126.514 Km<sup>2</sup> é caracterizada como semiárido e inclui 150 municípios, logo 86,8% do território cearense está incluído no semiárido (IBGE, 2021; Filho, 2018).

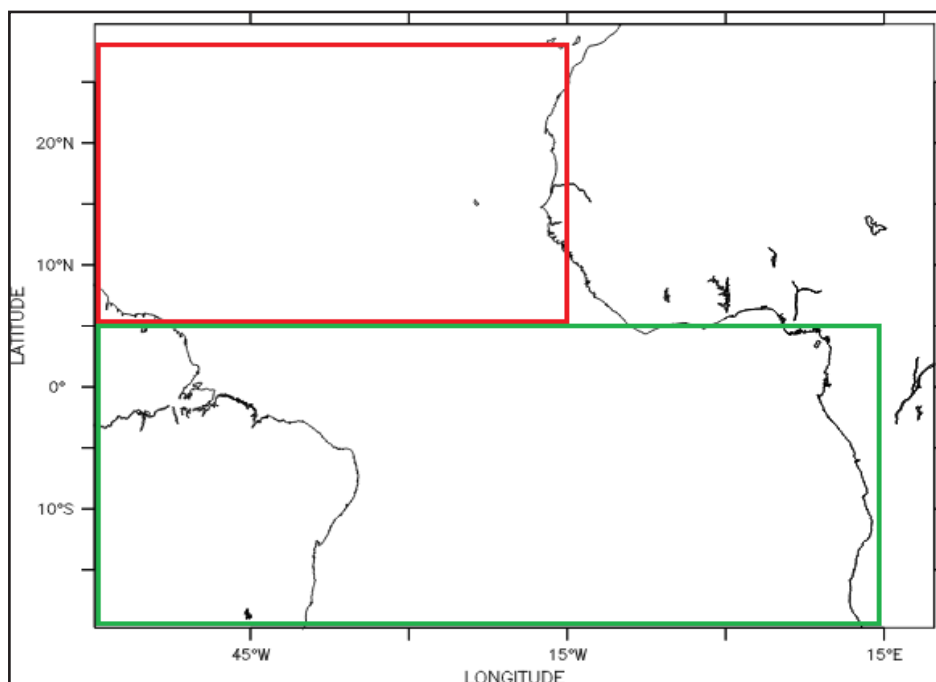
Figura 1 – Estado do Ceará



Fonte: Autores (2022)

O oceano atlântico é o segundo maior oceano do mundo, porém nesta literatura a área analisada será a do oceano atlântico tropical, esta região se encontra entre os trópicos de câncer e capricórnio sendo delimitada pela área costeira da América do Sul e pelo continente Africano, de maneira mais específica essa superfície marítima foi dividida em duas regiões conforme a metodologia aplicada no trabalho de Servain (1991), a primeira região sendo a bacias do oceano Atlântico Norte (ATN) com as seguintes coordenadas (5°N - 28°N e 60°W - 15°W), representado pelo retângulo vermelho e a bacia do oceano Atlântico Sul (ATS) (5°N - 20°S e 60°W - 15°E), representado pelo retângulo verde, como mostra a Figura 2.

Figura 2 – Área de controle do Oceano Atlântico Tropical



Fonte: Autores (2022)

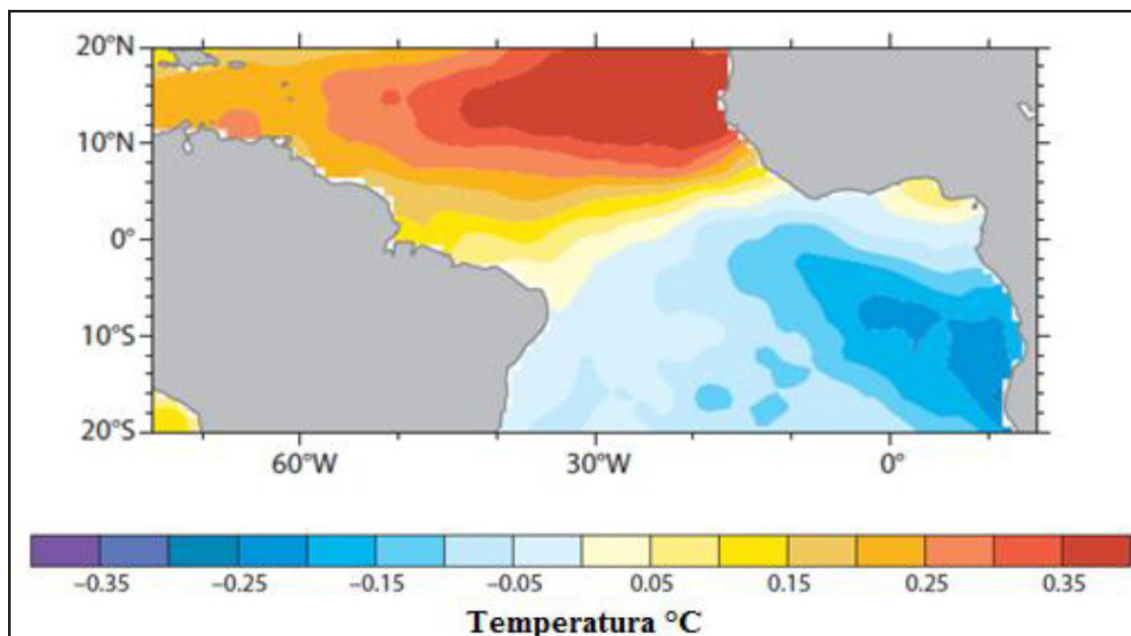
## 2.2 Dipolo do Atlântico

A Temperatura Superficial do Mar (TSM) está em constante relação com a atmosfera, os limites dessas duas camadas de fluidos geofísicos estão permanentemente trocando calor, esse fluxo de energia é governado por alguns processos e parâmetros entre esses dois ambientes, da parte oceânica podemos citar como exemplo o transporte de calor por corrente marítimas e a relação vertical de camadas oceânicas que influenciam na superfície, já com relação a atmosfera podemos citar os seguintes processos, temperatura do ar, velocidade do vento, nebulosidade e umidade do ar. Tal relação, faz com que a TSM se torne uma das principais componentes climáticas que influenciam em dinâmicas meteorológicas globais, como precipitação e formação de ciclones, podendo afetar várias regiões dos continentes, incluindo o Nordeste Brasileiro (Alves *et al.*, 2006; Silva *et al.*, 2011).

A influência da TSM do atlântico tropical sobre o regime de chuvas no Nordeste do Brasil (NEB) já está bem estabelecido na literatura, tendo como um dos estudos precursores o de Hastenrath e Heller (1977) no qual afirma que previsões sazonais

de totais acumulados de precipitação do NEB com antecedência de até seis meses poderia ser feita tendo como base a variabilidade de TSM do atlântico tropical. O modo meridional da variabilidade de TSM no Atlântico também denominado como Dipolo do Atlântico, se comporta como um gradiente térmico formado pelas bacias do oceano Atlântico Norte (ATN) (5°N - 28°N) e oceano Atlântico Sul (ATS) (5°S - 20°S) com anomalias de temperaturas médias de sinais opostos baseados na climatologia, por ter este comportamento é possível fazer uma analogia desse fenômeno com uma gangorra de temperatura, essas anomalias de TSM podem ser observadas na Figura 3.

Figura 3 – Anomalias de temperatura superficial do mar no oceano atlântico tropical



Fonte: Adaptada de Deser *et al.*, 2010

O Dipolo do Atlântico, pode ser caracterizado em duas fases, quando a temperatura superficial do mar está maior do que a sua média climatológica na bacia do ATN e as temperaturas na bacia do ATS estão abaixo da climatologia, é denominada de fase positiva do dipolo, quando a TSM está mais elevada do que a climatologia na bacia do ATS, e a TSM na bacia do ATN está abaixo da climatologia, denomina-se fase negativa do dipolo (Servain, 1991; Wagner, 1996).

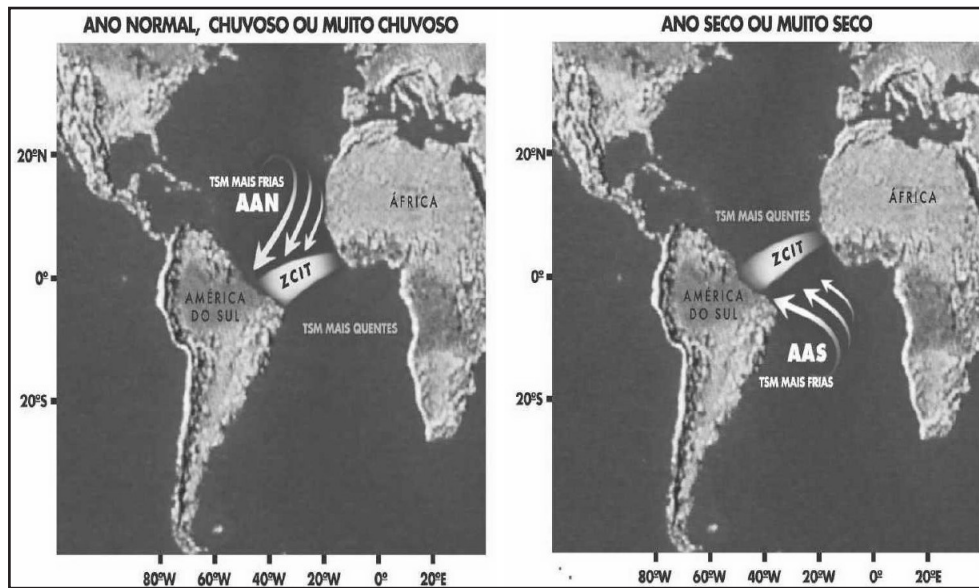
Uma das técnicas mais utilizadas para mensurar a relação de TSM entre as bacias do ATN e ATS foi proposta por Servain (1991), o mesmo denominou de índices

do Dipolo do Atlântico, tal índice é um valor numérico estatístico definido da seguinte maneira, primeiro é calculado a média mensal da TSM da bacia Norte e da bacia Sul no Atlântico, tendo como divisor dessas regiões a linha 5°N, o segundo passo é identificar as anomalias de TSM de cada bacia tendo como base a climatologia, em seguida é encontrado o desvio padrão das bacias do ATN e ATS, identificando assim as anomalias de TSM, por fim o índice é definido pela diferença entre os valores das anomalias das duas bacias. As variações de TSM das duas bacias oceânicas em questão costumam variar da climatologia em torno de 3°C, complementando que é através do índice do Dipolo do Atlântico que a seguinte classificação é realizada, quando são obtidos valores do índice abaixo de - 0,5°C denomina-se ano de dipolo negativo (DN), quando os valores encontrados estão acima de 0,5°C denomina-se ano de dipolo positivo (DP) e quando os valores dos índice oscilam entre -0,5°C e 0,5°C são caracterizados como anos neutros (Servain, 1991).

Uma das formas pelas quais o Dipolo do Atlântico influencia nas chuvas do NEB se dá pelo fato de que a Zona de Convergência Intertropical (ZCIT) possui uma relação de alto grau com a TSM na região Atlântica, quando o dipolo se encontra em sua fase positiva a ZCIT tende a se deslocar mais para o norte, como mostra a Figura 4, esse deslocamento faz com que o volume de chuva seja prejudicado no NEB, neste caso ocorrem movimentos descendentes carregando ar frio e seco que advém dos altos níveis atmosférico sobre a região setentrional, central e do sertão do NEB, que consequentemente inibe a formação de nuvens, o contrário ocorre quando o dipolo está em sua fase negativa, a ZCIT tende a se posicionar mais ao sul de sua posição climática, contribuindo assim para o aumento do índice de precipitação no NEB (Uvo *et al.*, 1988).



Figura 4 – Padrões do Dipolo na fase positiva e negativa



Fonte: Ferreira e Mello (2005)

## 2.3 Expoente de Hurst

O expoente de Hurst, é uma das maneiras de correlacionar as séries temporais, essa técnica foi desenvolvido pelo engenheiro Harold Edwin Hurts, o mesmo foi contratado para realizar a construção de uma represa de água no rio Nilo, um dos problemas iniciais se encontrava no controle da vazão do reservatório, em busca de uma solução, Hurts tentou desenvolver um modelo estatístico que se baseava na série temporal das chuvas que influenciavam no nível da água do rio, ou seja, em uma variável randômica, que segue o movimento Browniano, tal modelo desenvolvido tem uma ampla aplicação nos estudos de séries temporais que envolve os sistemas naturais, podendo identificá-las como uma série randômica ou não.

Em busca de explicar o seu método, Hurts construiu uma barragem ideal, ou seja, uma barragem que nunca secasse ou transbordasse e que tivesse uma vazão média anual constante conforme a necessidade de consumo da região, observando a flutuação do nível volumétrico de um reservatório tendo como base a média dos volumes construída no período de estudo, notou que as extensões destas flutuações, diferença entre o máximo e mínimo dos volumes, range, dependiam da extensão do

período levado em consideração, chegando a conclusão de que caso a série temporal do estudo fosse randômica a diferença das flutuações aumentariam com a raiz quadrada do tempo ( $t^{\frac{1}{2}}$ ).

A capacidade ideal (R) de um determinado reservatório, pode ser calculada para um período determinado, quando se sabe as vazões médias anuais ( $X_i$ ) que é depositada no reservatório e a vazão média de água que deve sair do reservatório ( $\bar{X}_n$ ), para isto deve ser construída uma série de totais acumulados (Y), composta pela soma das vazões  $X_i$  em relação às vazões  $\bar{X}_n$ , como mostra a equação:

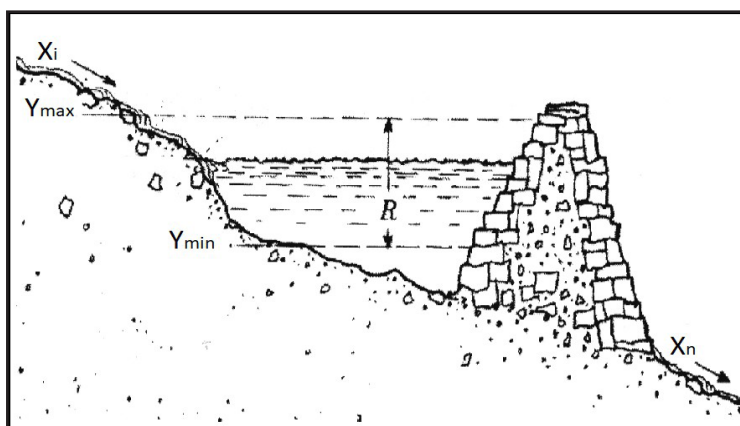
$$Y_j = \sum_{i=1}^j (X_i - \bar{X}_n) \quad j = 1, \dots, n. \quad (1)$$

A capacidade ideal do reservatório vai ser constituída pela diferença entre os maiores e os menores totais acumulados, considerando que os maiores acumulados, atingem o nível máximo da barragem de maneira que não transborde e o valor mínimo atinja um volume aceitável para manter a vazão de consumo ( $\bar{X}_n$ ), como pode ser demonstrado na equação:

$$R_n = \max(Y_1, \dots, Y_n) - \min(Y_1, \dots, Y_n). \quad (2)$$

A figura 5 demonstra um esquema ilustrativo da estrutura da barragem utilizada nos estudos de Hurts.

Figura 5 – Representação da Barragem de Hurts



Fonte: Adaptada de Feder (1989)

O trabalho de Hurts (1951) é abordado novamente pelos autores Mandelbrot e Wallis (1968), como uma metodologia incompleta na identificação de dependência estatística de longo prazo em séries temporais, tais autores formalizam o desenvolvimento da metodologia do trabalho de Hurts (1951), e os chamam de análise R/S ou de reescalonamento, mostrando que tal relação segue uma lei de potência:

$$\left(\frac{R}{S}\right)_n \sim cn^h \quad (3)$$

Lembrando que o valor de R é obtido pela equação 2, S se caracteriza como o desvio padrão e n é o número de elementos da série levada em consideração. O expoente de hurst é obtido ao aplicar o logaritmo em ambos os lados da equação 3, obtendo assim a seguinte equação:

$$\log\left(\frac{R}{S}\right)_n = \log c + H \log n \quad (4)$$

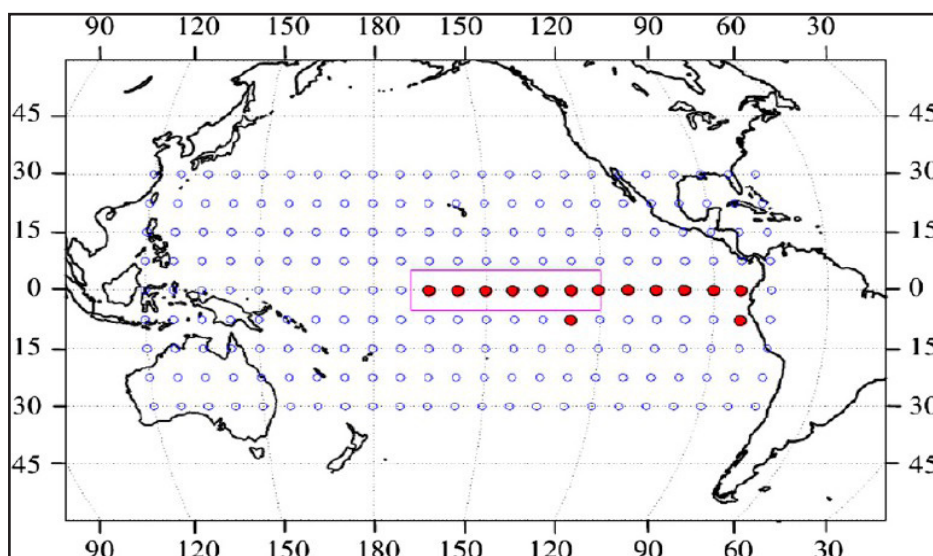
Os valores do Expoente de Hurst, H, é interpretado da seguinte maneira, se os valores de H estiverem entre 0 e 0,5 a série temporal em questão não é persistente, em outras palavras, se tivermos uma tendência positiva no passado a probabilidade dela inverter e se transformar em uma negativa ou vice-versa é maior, já quando  $H = 0,5$  o evento é tido como puramente aleatório, por fim quando os valores de H variam de 0,5 a 1 a série temporal é persistente, ou seja, se tivermos uma tendência positiva no passado existe uma maior probabilidade de que ela se conserve, o mesmo ocorre quando a tendência for negativa.

## 2.4 Série de correlações máximas S

Ludescher *et al* (2014) apresentou uma técnica para determinar uma série de correções máximas entre pontos de uma região de interesse, naquele estudo a região era a do Nino 3.4, e uma tentativa de previsão para o próximo El Nino comparando com o índice de Nino 3.4. Neste estudo, a região de interesse está sob uma área da região

do oceano passifico (Figura 6) subdividida em uma rede, que constituída de 14 pontos de grade na bacia do El Niño, representado pelas bolinhas vermelhas e 193 pontos de grade fora deste domínio, representadas pelas bolinhas azuis, notoriamente o retângulo em vermelho representa a região na qual é feita o cálculo do índice NINO3.4, cada nó representa a TSM daquela região.

Figura 6 – Região de controle aplicado na metodologia de Ludescher (2014)



Fonte: Autores (2022)

Para montar a série temporal das correlações máximas ( $S$ ) entre os nós vermelhos e azuis desses valores é necessário em cada par de ponto da grade em uma determinada janela de tempo calcular a força de correlação máxima ( $C_{i,j}^{max}$ ), valor médio ( $C_{i,j}^{med}$ ) e o desvio padrão de todas as correlação que ele encontrou durante um ano, o cálculo é realizado pela seguinte equação:

$$S(t)_{i,j} = \frac{C_{i,j}^{max} - C_{i,j}^{med}}{\sigma(C_{i,j}^t)} \quad (5)$$

De maneira que a função de correlação máxima cruzada retrocedendo no tempo em um período de dias que chamaremos de tal ( $\tau$ ), tem como pontos de grade as anomalias de TSM do oceano ATN, representados pela variável  $i$ , bem como a TSM do oceano ATS representado pela variável  $j$ . Como essa correlação se baseia nas observações e relações de duas séries espaciais distintas  $i$  e  $j$ , que se encontra em um

mesmo período, é necessário fixar uma das variáveis enquanto a outra retrocede no tempo, no caso, quando fixamos o  $i$  e retardamos o  $j$ , o cálculo é representado pela equação 6, já quando fixamos o  $j$  e retrocedemos o  $i$ , a conta é realizada conforme a equação 7.

$$C_{ij}^{(t)}(-\tau) = \frac{\langle T'_i(t)T'_j(t-\tau) \rangle - \langle T'_i(t) \rangle \langle T'_j(t-\tau) \rangle}{\sqrt{\langle (T'_i(t) - \langle T'_i(t) \rangle)^2 \rangle} \cdot \sqrt{\langle (T'_j(t-\tau) - \langle T'_j(t-\tau) \rangle)^2 \rangle}} \quad (6)$$

$$C_{ij}^{(t)}(\tau) = \frac{\langle T'_i(t-\tau)T'_j(t) \rangle - \langle T'_i(t-\tau) \rangle \langle T'_j(t) \rangle}{\sqrt{\langle (T'_i(t-\tau) - \langle T'_i(t-\tau) \rangle)^2 \rangle} \cdot \sqrt{\langle (T'_j(t) - \langle T'_j(t) \rangle)^2 \rangle}} \quad (7)$$

Para montar a série temporal das correlações máximas  $S(t)$  é necessário determinar a anomalia de temperatura diária utilizando a equação 8, em todos os nós da rede montada sobre o oceano atlântico nas área representadas pela Figura 2, durante o período de teste (1981-2001).

$$T'_k(t) = T_k(t) - \langle T_k \rangle_{1986-2001} \quad (8)$$

onde  $T'_k(t)$  representa o índice de anomalia de temperatura superficial do mar diária, o  $T$  indica a temperatura real, o  $t$  representa o tempo, a letra  $K$  representa a região do nó e o  $\langle T_k \rangle$  é a média climatológica da temperatura superficial do mar. É importante realçar que a média das correlações cruzadas estão para um período de 365 dias, observem na equação 9.

$$\langle f(t) \rangle = \frac{1}{365} \sum_{m=0}^{365} f(t-m) \quad (9)$$

## 2.5 Machine Learning

A aprendizagem é uma característica intrínseca da espécie humana, porém com o avanço da ciência computacional e sua atuação em ramos científicos, econômicos e sociais, essa característica se tornou uma necessidade em máquinas, tal adaptação é denominada de Machine learning (ML). O ML é um dos ramos que compõem a área

de inteligência artificial, podendo ser definida como a técnica de treinamento de algoritmos em prol da identificação de padrões e tomadas de decisões realizadas por máquinas, ou seja, é uma forma de ensinar os computadores a decodificar informações por meio de análise de dados e algoritmos dedutivos (Shinde, 2018).

O aprendizado supervisionado aplicado no ML, é um método que tenta prever ou classificar uma variável dependente por meio de um conjunto de dados de variáveis independentes, no qual o algoritmo do modelo é submetido a uma fase de treinamento para poder ajustar e relacionar os atributos descritivos (x) a uma determinada classe (y), realçando que os dados de treinamento são pré classificados em rótulos, após esse período de treinamento é iniciado a fase de teste, nesta ocasião o modelo é exposto a uma gama novas de dados (x) e espera-se que o algoritmo possa fornecer os valores/classes estimados (y), realçando que, conforme a natureza da saída dos dados estimados (y), o aprendizado supervisionado se divide em duas categorias, uma utiliza o modelo de regressão e o outro de classificação (Hao, 2019).

No caso do aprendizado não supervisionado de ML, não existem exemplos rotulado de dados aplicados aos modelos, os próprios algoritmos buscam identificar estruturas semelhantes que estão contidas em uma base de dados e então os agrupa em classes com características semelhantes, de maneira geral, esses modelos são utilizados na busca de novos padrões e tendências de um conjunto de dados, podendo assim, aumentar a compreensão de um determinado evento. De maneira geral o ML é uma ferramenta poderosa que pode ser utilizada em vários viés de pesquisas e um deles é no mapeamento e em estimativas futuras de sistemas não lineares como a TSM e na precipitação, tal aplicação pode ser vistas no trabalho desenvolvido por Xiao (2019) que utilizou o ML especificamente a técnica de AdaBoost e um modelo de rede neural para prever eventos de de TSM no mar da China oriental.

O machine learning e seus algoritmos de classificação possuem uma ampla aplicação na ciência, tal afirmação pode ser vista no trabalho de Shiraishi *et al.* (2014) que analisou cinco algoritmos de classificação Naïve Bayes, AdaBoost, multi-layer perceptron, random

forest e support vector machine, aplicados em mapa de classificação de florestas tropicais, tendo como base dados de sensoriamento remoto óptico, chegando a conclusão de que essas técnicas são altamente eficientes. No trabalho desenvolvido por Leduc *et al.* (2017) o mesmo utiliza ML para prever possíveis terremotos por meio de sinal acústico emitidos por uma falha de laboratório, o algoritmo de classificação neste caso em questão tem por objetivo identificar sinais ocultos que precedem terremotos.

### 2.5.1 Árvore de Decisão

Os algoritmos de classificação são técnicas utilizadas para organizar e classificar dados por meio de um algoritmo de maneira mais autônoma, para isso o mesmo utiliza um conjunto de regras e dados prévios de classificação, baseados nessas informações o programa realiza de maneira mais autônoma a classificação de novos elementos. A Árvore de Decisão (AD) é um método estatístico de classificação e previsão que atua no campo de mineração de dados. Destaca-se que este campo busca extrair informações de uma vasta quantidade de dados, tais extrações podem encontrar novos ou elucidar padrões existentes, conseqüentemente, geram regras que predizem as correlações dos dados e aumentam a veracidade das informações extraídas (Camilo, 2009).

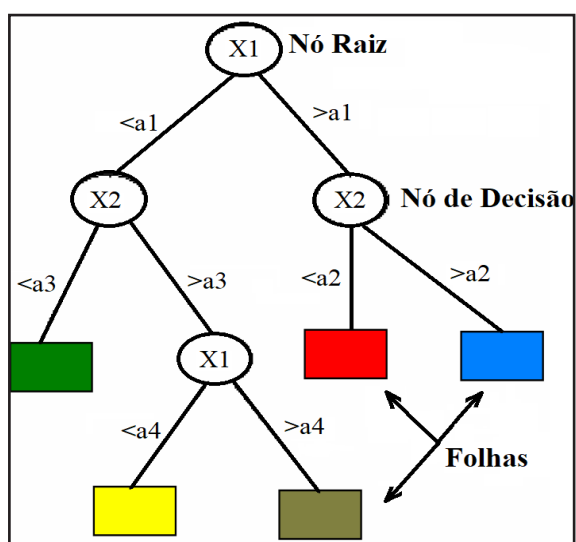
A técnica de AD utiliza um algoritmo do qual é baseado em aprendizagem supervisionada, utilizado para classificação e regressão de dados não parametrizados. A mesma em sua construção, utiliza uma estratégia conhecida como dividir-para-conquistar, um determinado problema complexo é decomposto em subproblemas mais simples, tal atividade se repete para os subproblemas e assim sucessivamente, cada sub-espço é associado a uma classificação, essa organização estrutural é o que traz a capacidade de discriminação desse método (Monard, 2003).

Uma exemplificação de árvore de decisão pode ser vista na Figura 7, na qual cada elipse representa um nó de decisão, o mesmo contém um teste para aferir os dados de um conjunto a um atributo, após a realização do teste, esses dados passam por uma aresta, que filtra as informações e as separam em uma nova categoria, definidas conforme o objetivo



do trabalho, essas aresta podem estar ligadas a um novo nó de decisão ou a uma folha, representada por um retângulo, na qual indica uma classe, lembrando que a construção da árvore de decisão se inicia pelo nó raiz que contém todos os dados.

Figura 7 – Árvore de Decisão



Fonte: Adaptada de Gama (2002)

A aplicação e a sistematização da técnica da árvore de decisão pode ser vista em detalhes no trabalho desenvolvido por Braimah *et al.*, (1984) e Quinlan (1996), tais trabalhos descrevem minuciosamente a construção desses algoritmos, já a aplicação dessa técnica em ambientes industriais podem ser agraciados na pesquisa realizada por Michel *et al.*, (1996). A utilidade de algoritmos de decisão é tão abrangente que pode ser notada em várias áreas, por exemplo no trabalho de Priyam (2013), que utiliza dados de alunos para prever seu desempenho.

### 3 RESULTADO E DISCUSSÕES

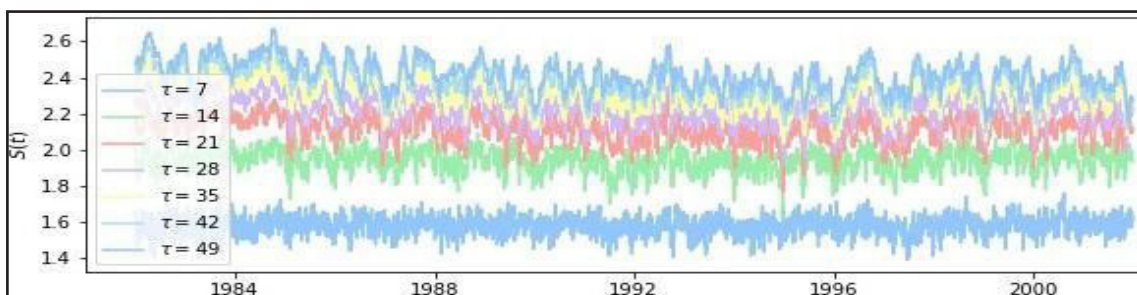
Nessa seção são apresentadas as características da série de correlações máximas de  $S(t)$  e a precipitação  $p(t)$  no Estado do Ceará para alguns cenários de tempo de relaxamento  $\tau$ . A seção 3.1 apresenta uma descrição dos dados e algumas estatística e a seção 3.2 apresenta um dos resultados obtidos considerando o treinamento do algoritmo de Machine Learning entre  $S(t)$ .



### 3.1 Descrição e Análise dos Dados

Nesta seção são apresentados a série de correlações máximas  $S(t)$  calculadas com a série temporal das anomalias de TSM entre as bacias do oceano ATN e ATS, as áreas dessas regiões estão definidas conforme a metodologia de Servain (1991), o primeiro gráfico montado faz alusão ao período de teste que vai de 1981 a 2002 com valores de tau ( $\tau$ ) variando de 7 a 49 dias, conforme mostra a figura 8.

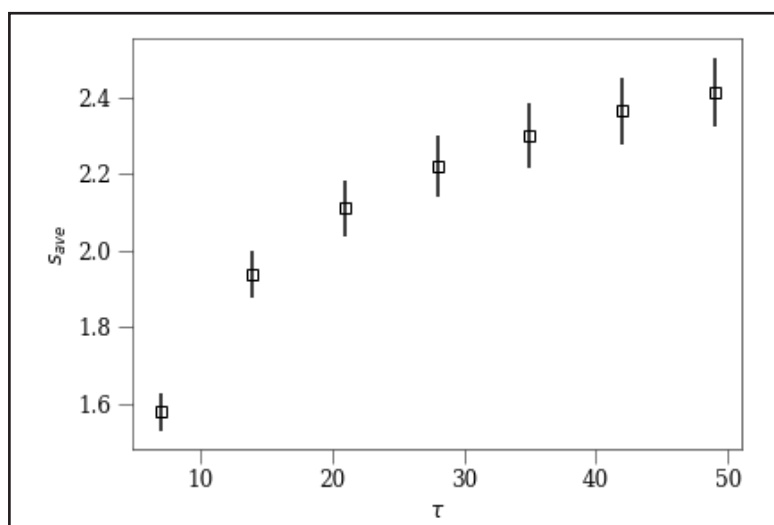
Figura 8 – Valores de correlação  $S(t)$  com o tau variando de 7 a 49 dias (1981-2002)



Fonte: Autores (2022)

Com o intuito de demonstrar melhor o comportamento dos valores das séries de correlação máxima com a variação crescente dos valores de  $\tau$ , um gráfico dos desvios padrões de  $S(t)$  é apresentado na figura 9, observando o mesmo é possível inferir que o aumento do valor de  $\tau$  traz como consequência um aumento tanto do range no desvio padrão  $S(t)$  quanto em seus valores de média, porém é necessário realçar que com o aumento do  $\tau$  acima de trinta dias a comparação progressiva dos valores do range vão diminuindo.

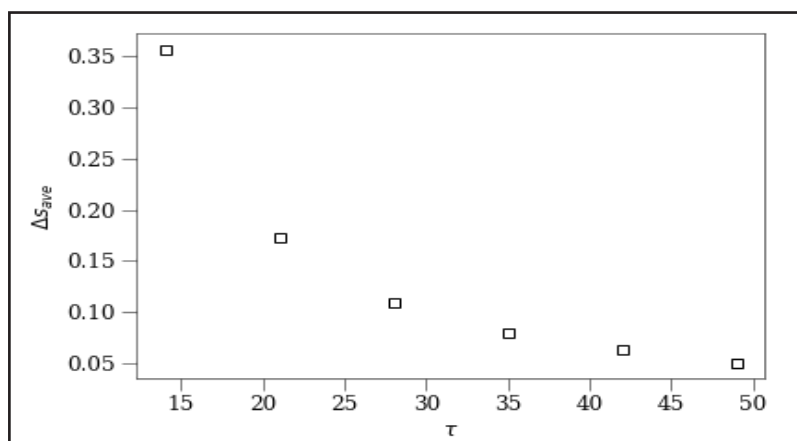
Figura 9 – Desvio padrão ( $\sigma$ ) dos valores das séries de correlações máxima com relação aos valores de tau ( $\tau$ )



Fonte: Autores (2022)

A mudança dos valores médios de  $S(t)$  quando comparadas com valores posteriores, há uma tendência de diminuir de acordo com o aumento do  $\tau$ , esse comportamento se mantém de maneira progressiva como pode ser visto na figura 10, logo a série  $S(t)$  para valores de  $\tau$  muito grande podem ser desnecessários por uma possível similaridade, já que o seu aumento segue uma curva exponencial decrescente, este fato, faz com possa haver um benefício computacional no momento de processar os dados de correlação.

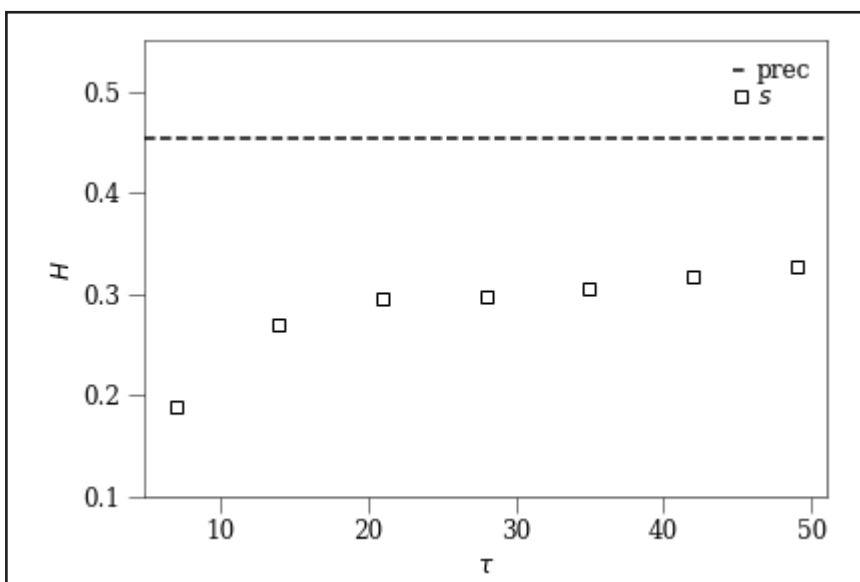
Figura 10 – Diferença progressiva entre os valores consecutivos das médias da  $S(t)$ . A diferença  $\Delta S$  é obtida usando um método de diferença recuada



Fonte:

A série  $S(t)$  armazena as correções máximas obtidas entre as áreas de cálculo do dipolo. Há um conjunto variado de métricas que podem ser usadas para medir a similaridade entre  $S(t)$  e a precipitação ( $prec$ ), neste trabalho, usamos o coeficiente de Hurst para comparar o nível de similaridade entre os dados. Na figura 11 é possível notar que a distribuição temporal entre as série de correlações máximas e a precipitação se comportam de maneira diferente. Um indicador importante sobre o Hurst é que uma série cujo expoente seja de 0.5, essa série temporal é caracterizada por um movimento browniano.

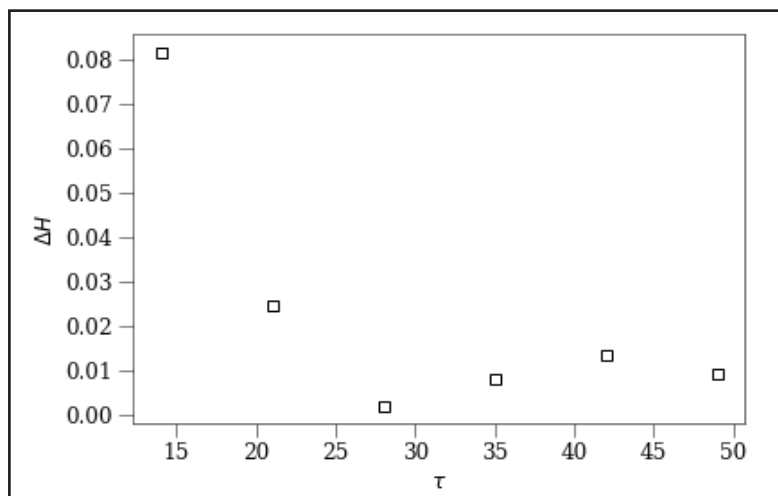
Figura 11 – Expoente de Hurst para as séries temporais de precipitação ( $prec$ ) e a série de correlações máximas  $S$



Fonte: Autores (2022)

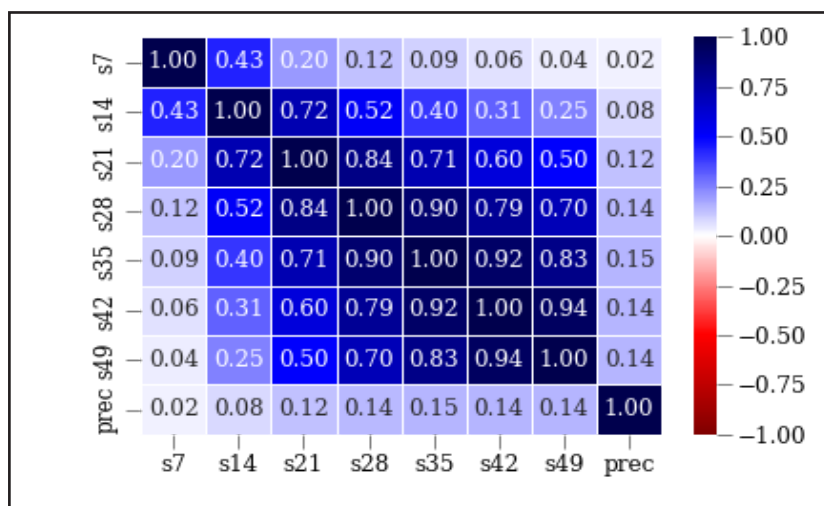
Uma informação importante é que a série de precipitação tem um valor aproximado de 0,42 e próximo a 0.5, que indica uma evidência robusta na dificuldade de previsão numérica. A figura 12 apresenta que o Hurst tende a estabilizar na medida que o valor de tau aumenta.

Figura 12 – Diferenças recuada dos coeficientes de Hurst



Fonte: Autores (2022)

A figura 13 apresenta a matriz de correlação de Pearson entre as séries de correlação máxima  $S(t)$  em função do  $(\tau)$  com a precipitação ( $prec$ ). As correlações são positivas e não há nenhum sinal de autocorrelação, porém possuem um índice de correlação muito baixo mesmo para  $\tau = 49$  dias. Diante do exposto, através de técnicas simples estatísticas, as correlações obtidas tornam difícil a previsibilidade da precipitação em função simples dos valores de  $S$ . No entanto, o uso de algoritmos de *Machine Learning* pode contribuir para estabelecer uma relação entre as séries temporais.

Figura 13 – Matriz de correlação de Pearson entre os valores de  $S$  e a precipitação para o período de 1981 a 2001

Fonte: Autores (2022)

### 3.1 Tentativa de Treinamento usando ML

Nessa seção apresentamos uma amostra das tentativas de usar os valores de  $S$  como variáveis preditoras para prever a precipitação no Ceará. Inicialmente, foram realizadas tentativas de treinar um período longo de tempo de 1981 a 2001 como período de treinamento e de 2002 até 2020 como período de testes. Para esse conjunto de primeiras tentativas não foi possível obter resultados bons, por esse motivo os mesmos não estão expostos na literatura presente.

Os testes seguiram o uso de algoritmos mais simples, por exemplo, o de árvore de decisão. A escolha do algoritmo é baseada em uma alternativa mais sofisticada, que foi aplicada no trabalho desenvolvido por Ludescher *et al.* (2014).

Depois de muitas tentativas, uma delas se mostrou pertinente e que parece abrir caminhos alternativos na compreensão do uso de algoritmos mais simples de ML para compreender as relações entre séries temporais. Sugerimos que de alguma maneira, a taxa de variação da precipitação pode ser explicada através da equação,

$$dp_{ML} = \alpha dS_{\tau=28} + \beta dS_{\tau=35} + \gamma \Delta S_1 + \epsilon \Delta S_2 \quad (10)$$

De maneira que  $dS$  é a taxa diária da série de correlação máxima para  $\tau = 28$  dias e  $\tau = 35$  dias, a diferença entre os valores das correlações máximas utilizando 28 e 21 dias de atraso  $\Delta S$  está representado na equação 11 e a diferença entre os valores de 21 e 14 dias da mesma correlação é representada pela equação 12

$$\Delta S_1 = dS_{\tau=28} - dS_{\tau=21} \quad (11)$$

$$\Delta S_2 = dS_{\tau=21} - dS_{\tau=14} \quad (12)$$

Os coeficientes  $\alpha, \beta, \gamma$  e  $\epsilon$  foram o conjunto de treinamento  $\Sigma$  calculado pelo algoritmo de ML. A taxa de precipitação observada é a diária e foi obtida através da seguinte equação,

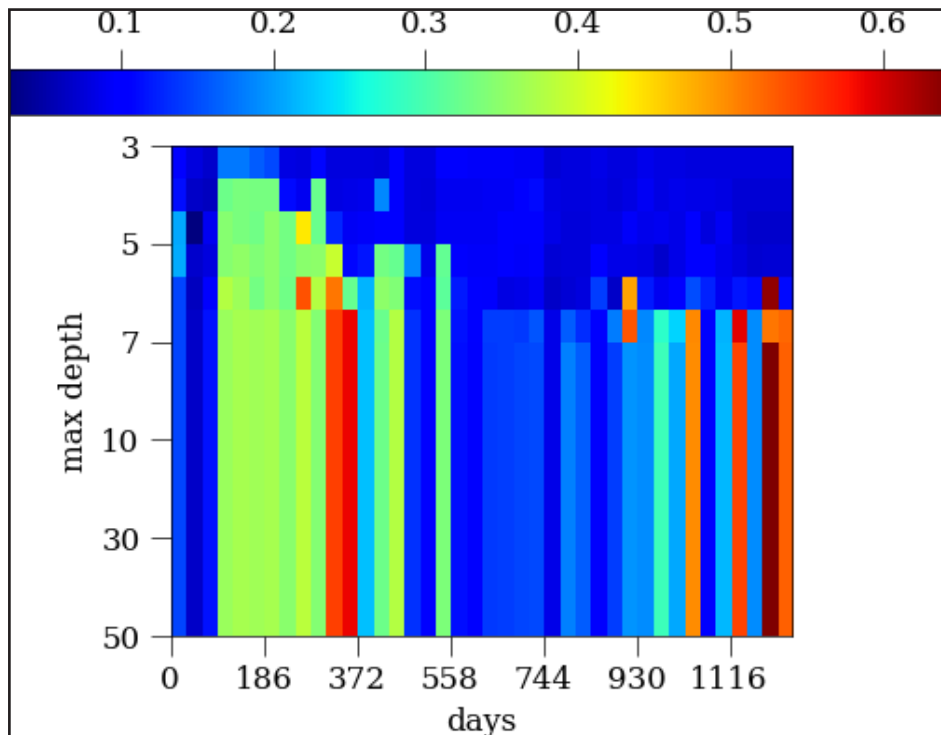
$$dp_{obs} = (prec(t) - prec(t - 1))/(1 \text{ dia}) \quad (13)$$

Em busca de facilitar a compreensão do método, os procedimentos de treinamento são descritos as seguir:

1. Uma data de referência é escolhida.
2. Um conjunto de treinamentos entre as taxas de precipitação teórica  $dp_{ML}$  e observada  $dp_{obs}$  de 1 até 1200 dias antes da data de referência são realizados e configuram o período de treinamento.
3. Um subconjunto de treinamento que possuem erro quadrático médio pequeno  $mse < 1$  e coeficiente de correlação positivo  $R^2 > 0$  são permitidos para serem usados no período de teste. Portanto, esse conjunto de dados configura o conjunto de testes.
4. O subconjunto de treinamento é usado para prever 7, 10 e 14 dias após a data de referência.

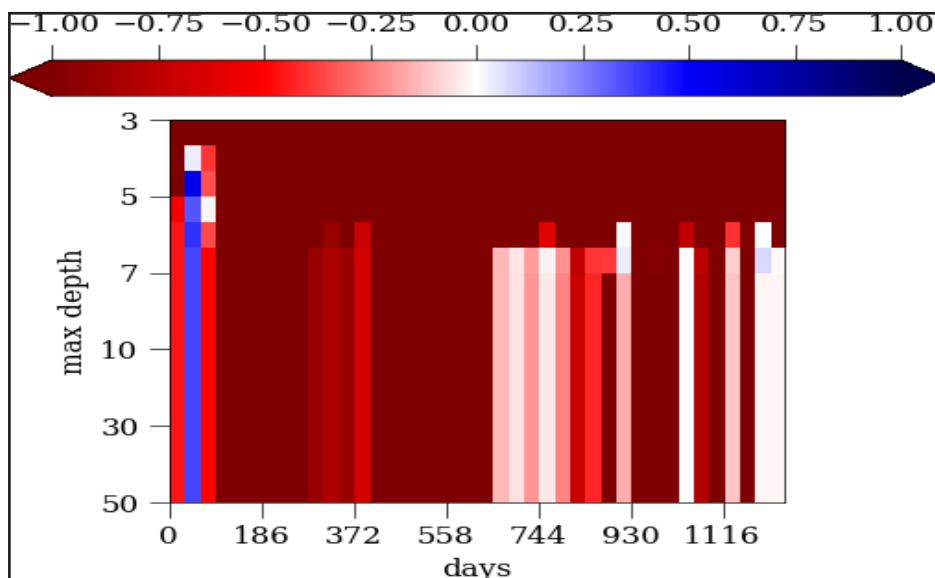
As figuras 14 e 15 apresentam uma matriz de correlação obtida da etapa 2 dos procedimentos. A paisagem de correlações é heterogênea e pouco pode ser explicado observando apenas elas. O objetivo desse conjunto é filtrar o subconjunto de constantes que possuem erro quadrático médio e erro quadrático médio  $mse < 1$  e  $R^2 > 0$ . Neste exemplo do gráfico, esse subconjunto de valores fica limitado a poucas amostras. Realçando que nas figuras 14 e 15, o *maxdepth* é a quantidade de ramos permitidos no treinamento. Days é o número de dias que antecedem a data de referência.

Figura 14 – Matriz de correlação de Pearson entre os valores de  $dp_{ML}$  e da taxa de precipitação  $dp_{obs}$  cuja data de referência é dia 15/12/2008



Fonte: Autores (2022)

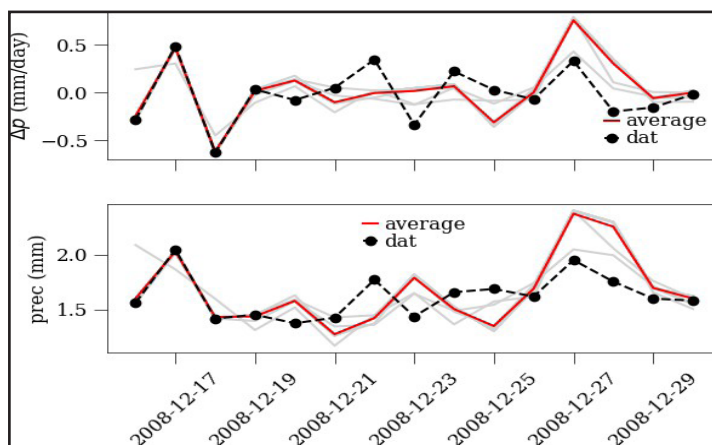
Figura 15 – Erro quadrático médio  $R^2$  entre os valores de precipitação  $dp_{ML}$  e da taxa de precipitação  $dp_{obs}$  cuja data de referência é dia 15/12/2008



Fonte: Autores (2022)

Após a utilização dos filtros  $mse$  e  $R^2$  citados anteriormente, um novo gráfico é gerado, figura 16, para demonstrar a série temporal da taxa de precipitação e da precipitação para o período de teste (15 dias) com data de referência no dia 15/12/2008, de maneira que, a linha pontilhada é o  $dp_{obs}$ , as linhas de cor cinza é o resultado de cada subconjunto permitido  $dp_{ML}$  e a linha vermelha representa o valor médio de todos os valores  $dp_{ML}$ .

Figura 16 – Série temporal (a) da taxa de precipitação e (b) da precipitação para o período de teste (15 dias) com data de referência no dia 15/12/2008



Fonte: Autores (2022)

A figura 16 apresenta a série temporal da taxa de precipitação no período de teste usando o algoritmo de árvores para prever o comportamento da taxa de precipitação para o estado do Ceará. O período de testes é curto para obter um coeficiente de correlação robusto, no entanto, em uma observação é possível verificar que o treinamento com ML pode contribuir com a previsão de uma variável meteorológica de interesse que é a precipitação.

## CONCLUSÕES

O algoritmo de árvore foi usado para realizar testes de treinamento entre as correlações máximas de anomalia da temperatura da superfície do mar entre dois setores que são calculados no dipolo do atlântico. Diferente do procedimento comum



usado para esse tipo de problemas, o ML não se comportou bem para prever séries temporais com grande volume de amostras temporais, no entanto, teve melhor comportamento de treinamento usando um período de dados relativamente curto, até 1200 dias, aproximadamente 3 anos. Por meio desta literatura é possível notar o potencial da aplicação do ML e que é conveniente realizar mais testes considerando outras métricas de filtragem de dados, por exemplo, usando correlação de pearson ou spearman. Também é possível notar a necessidade de novos testes para verificar o limite de dias para uma previsão confiável e de maneira mais específica verificar se a paisagem de correlações entre *maxdepth* e dias exibe algum padrão ao longo dos anos e na ocorrência de eventos climáticos pertinentes.

## REFERÊNCIAS

- CAVALCANTI, I. F. A.; FERREIRA, M.; ASSUNÇÃO, F. S. D.; GERTRUDES, A.; SILVA, J. **Tempo e Clima no Brasil**. São Paulo: Oficina de Textos, p. 182, 2009.
- FERREIRA, A. G.; MELLO, N. G. 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, 2005.
- WALTER, G. T. Ceará(Brazil) famines and the general air movement. **Beitrage zur Physik der freien Atmosphäre**, v. 14, p. 88-93, 1928.
- SERVAIN, J. Simple climatic indices for the tropical Atlantic Ocean and some applications. **Journal of Geophysical Research: Oceans**, v. 96, n. C8, p. 15137-15146, 1991.
- SILVA, A. P. N.; MOURA, G. B. A.; GIONGO, P. R.; MEDEIROS, S. R. R. Correlação entre astemperaturas do mar e a quantidade de precipitação na estação chuvosa no nordeste do estado de Pernambuco. **Revista Brasileira de Meteorologia**, v. 26, n.1, p. 149-156, 2011.
- ZEBIK, S. E.;CANE, M. A. A model El Nino- Southern Oscillation. **Monthly Weather Review**, v. 15, p. 2262-2278, 1987.
- ALVES, J. M. B.; SOUZA, R. O.; CAMPOS, J. N. B. Previsão de anomalias de temperatura da superfície do mar (TSM) no Atlântico tropical, com equação da difusão de temperatura. **Revista Climanálise**, v. 1, n. 3, p. 6-19, 2006.
- LUDESCHER, J.; GOZOLCHIANI, A.; BOGACHEV, M. I.; BUNDE, A.; HAVLIN, S.; SCHELLNHUBER, H. J. Very early warning of next El Niño.**PNAS**, n. 6, v.111, p. 2064-2066, 2014.

IBGE. **Divisão Territorial Brasileira**, 2020. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/23701-divisao-territorial-brasileira.html?=&t=downloads>. Acesso em: 29 ago. 2021.

FILHO, F. A. Estudo Setorial Especial - Recursos Hídricos. **Ceará 2050**. Fortaleza, 2018. FUNCEME. Área de Meteorologia, 2013. Disponível em: <http://www.funceme.br/?p=843> Acesso em: 21 mar. 2021.

ALVES, J. M. B.; SOUZA, R. O.; CAMPOS, J. N. B. Previsão de anomalias de temperatura da superfície do mar (TSM) no Atlântico tropical, com equação da difusão de temperatura. **Revista Climanalise**, v. 1, n. 3, p. 6-19, 2006.

SILVA, A. P. N.; MOURA, G. B. A.; GIONGO, P. R.; MEDEIROS, S. R. R. Correlação entre as temperaturas do mar e a quantidade de precipitação na estação chuvosa no nordeste do estado de Pernambuco. **Revista Brasileira de Meteorologia**, v. 26, n.1, p. 149-156, 2011.

HASTENRATH, S.; HELLER, L. Dynamics of climatic hazards in Northeast Brazil. **Quarterly Journal Royal Meteorological Society**. v. 103, p. 77-92, 1977.

WAGNER, R. G. Mechanisms controlling variability of the interhemispheric sea surface temperature gradient in the tropical Atlantic. **Journal Climate**, v. 9, p. 2010-2019, 1996.

UVO, C. R. B.; NOBRE, C. A.; CITRAN, J. Análise da posição da Zona de Convergência Intertropical (ZCIT) no Atlântico Equatorial e sua relação com a precipitação no nordeste do Brasil. In: Congresso brasileiro de meteorologia, 1988. Rio de Janeiro. **Anais[...]** Rio de Janeiro: SBMET, 1988. p. 323-327.

FERREIRA, A. G.; MELLO, N. G. 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, 2005.

DESER, C.; ALEXANDER, M. A.; XIE, S. P.; PHILLIPS, A. S. Sea Surface Temperature Variability: Patterns and Mechanisms. **Annual Review Of Marine Science**, v. 2, p.115-143, 2010.

CAMILO, C. O.; SILVA, J. C. D. Mineração de dados: conceitos, tarefas, métodos e ferramentas. **Relatório Técnico**. Goiânia: Universidade Federal de Goiás, 2009.

MAY, S.; PRADO, R. T. A. Estudo de Viabilidade de Aproveitamento de Água de Chuva MONARD, M. C.; BARANAUSKAS, J. A. Indução de Regras de Árvores de Decisão. Sistema inteligente. Ed. Manole Ltda. p.115-139. 2003.

GAMA, J. **Árvore de decisões**, 2002. Disponível em: [http://www.dcc.fc.up.pt/~ines/aulas/MIM/arvores\\_de\\_decisao.pdf](http://www.dcc.fc.up.pt/~ines/aulas/MIM/arvores_de_decisao.pdf). Acesso em: 20 out. 2021.

GAN, M. A. **Um estudo observacional sobre as baixas frias da alta em estudo observacional sobre as baixas frias da alta troposfera, nas latitudes subtropicais do Atlântico sul e leste do Brasil**. 1982. Dissertação (Mestrado em Meteorologia) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1982.

PRIYAM, A.; GUPTA, R.; RATHEE, A.; SRIVASTAVA S. Comparative Analysis of Decision Tree Classification Algorithms. **International Journal of Current Engineering and Technology**, Vol.3, No.2, 2013.

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. **Classification and Regression Trees**. Wadsworth, Belmont, CA. 1984.

QUINLAN, J. R. **Decision trees and instance-based classifiers**. Handbook of Computer Science and Engineering. A. B. Boca Raton, FL. 1996.

MICHIE, D., SPIEGELHALTER, D.J., AND TAYLOR, C. C. **Machine Learning, Neural and Statistical Classification**. Ellis Horwood, Hertfordshire, UK. 1994.

SHINDER, P.P.; SHAH, S. **A review of machine learning and deep learning applications**. Fourth International Conference on Computing Communication Control and Automation (ICCUBE). 2018.

## Contribuição de Autoria

### 1 – Werbeson da Silva Freitas

Mestre em Ciências Físicas Aplicadas, Especialista em Ciências da Natureza, suas Tecnologias e o mundo do trabalho

<https://orcid.org/0000-0002-0544-9443> • [prof.werbeson.freitas@gmail.com](mailto:prof.werbeson.freitas@gmail.com)

**Contribuição:**

## Como citar este artigo

Freitas, W. S. O uso de correlações máximas de TSM do dipolo do Atlântico e a precipitação no estado do Ceará. **Ciência e Natura**, Santa Maria, v. 45, e71512, 2023. DOI 10.5902/2179460X71512. Recovered in: <https://doi.org/10.5902/2179460X71512>