



IV-500 – AVALIAÇÃO DA PREVISÃO DE PRECIPITAÇÃO SAZONAL NA REGIÃO HIDROGRÁFICA DOS SERTÕES DE CRATEÚS UTILIZANDO REGRESSÃO LINEAR MÚLTIPLA

Samuel Martins Pereira

Engenheiro Civil pela Universidade Federal do Ceará, Campus Crateús (2022).

Alan Michell Barros Alexandre ⁽¹⁾

Engenheiro Civil pela Universidade de Fortaleza (2002), mestre e doutor em Engenharia Civil (Recursos Hídricos) pela Universidade Federal do Ceará (2005 e 2012). Professor Adjunto da Universidade Federal do Ceará, Campus de Crateús.

José Kerlly Soares de Araújo

Engenheiro Civil pela Universidade Federal do Ceará, Campus Crateús (2024). Mestrando em Engenharia Civil (Recursos Hídricos), POSDEHA/UFC.

Endereço ⁽¹⁾: Avenida Professora Machadinho Lima, S/N, Príncipe Imperial, Crateús-CE, CEP 63708-825. Fone: +55 (88) 3691-9700 - e-mail: alanmichell@crateus.ufc.br.

RESUMO

O Nordeste Brasileiro (NEB) semiárido é caracterizado por baixos índices pluviométricos anuais, uma concentração de precipitação em apenas alguns meses do ano. Os cenários de mudanças climáticas apresentados até agora mostram um aumento na temperatura considerável, tendo como consequência a diminuição da disponibilidade hídrica e aumento dos eventos de chuvas intensa, situação está desfavorável para transposição deste recurso hídrico no tempo. A realização de previsão de precipitação tem se mostrado uma alternativa valiosa no suporte a decisão da alocação negociada efetuada anualmente. Assim, o presente estudo tem como objetivo gerar modelos de previsão de precipitação sazonal para a Região Hidrográfica dos Sertões de Crateús (RHSC), utilizando a regressão linear tradicional (OLS – *ordinary least squares*), e as regressões do tipo Ridge e Lasso. O período de precipitação a ser previsto consiste na quadra mais chuvosa na RHSC concentrada entre os meses de fevereiro a maio (FMAM), como elementos preditores dos modelos foram utilizados os índices da anomalia da Temperatura da Superfície do Mar (TSM), uma vez que a precipitação do NEB como um todo é fortemente influenciada pela oceanos Atlântico e Pacífico. O algoritmo *Iterative Input Selection* (IIS) foi utilizado para realizar o ranqueamento desses índices em ordem de importância, e em seguida foram gerados os diferentes modelos propostos, tendo a acurácia medida pelos coeficientes de correlação de Pearson e de eficiência de Nash. De um modo geral os índices ligados ao El Niño Oscilação Sul (ENSO) se mostraram mais representativos. Os modelos apresentaram uma baixa capacidade preditiva, em termos de desempenho a regressão tradicional mostrou-se um pouco melhor que as demais, os melhores resultados para todos os métodos de regressão avaliados foi o de junho a agosto (JJA), na maioria dos modelos observasse uma melhora quando se aumenta o número de variáveis explicativas.

PALAVRAS-CHAVE: Temperatura da superfície do Mar (TSM), Regressão Lasso e Regressão Ridge.

INTRODUÇÃO

O Nordeste Brasileiro (NEB) predomina o semiárido, marcado por notáveis variações temporais e espaciais na precipitação, além de elevadas temperaturas anuais (Nóbrega *et al.*, 2014). Essa condição climática desafia a gestão dos recursos hídricos na região, pois além da precipitação total anual média varia entre 400 e 800 mm, dependendo da localização, a concentração significativa de chuvas em apenas três ou quatro meses, aliada a um solo com baixa capacidade de infiltração e altas taxas de evaporação que resulta em um considerável déficit hídrico durante a maior parte do ano (Nóbrega; Santiago; Soares, 2016).



Além disso, a deficiência hídrica no NEB tende a ser agravada de acordo com os cenários de mudanças climáticas apresentados até agora. Estes mostram um aumento significativo na temperatura, tendo como consequência a diminuição da disponibilidade hídrica pela diminuição do total anual de precipitação e o aumento dos eventos de chuvas intensas, situação essa desfavorável para transposição do recurso hídrico no tempo, resultando em um potencial causador de impactos significativos nos setores naturais, sociais e econômicos da região (Santos *et al.*, 2009).

O estado do Ceará tem a Zona de Convergência Intertropical (ZCIT) como principal sistema causador de chuvas, notadamente entre os meses de fevereiro a maio, representando os quatro meses mais úmidos no estado, chamado de quadra chuvosa. O fenômeno é intensificado em fevereiro e no período do verão para o inverno, ocasionando precipitações e reduzindo a influência dos vórtices ciclônicos, que também podem influenciar negativamente as chuvas, inibindo a translação do ZCIT do hemisfério norte para o sul. (Coutinho; Sokomoto, 2017).

A posição da ZCIT, seja mais ao norte, permanecendo sobre o oceano, ou mais ao sul, entrando no continente e trazendo chuva para mesmo, é fortemente influenciada pela temperatura da superfície do mar (TSM) no oceano Pacífico equatorial representada pelo El Niño-Oscilação Sul (ENOS) e pela TSM do oceano Atlântico tropical representado pelo Dipolo do Atlântico (TAD), função das TSMs do Atlântico norte e sul. (Ferreira; Mello, 2005).

Por tanto, dada a vulnerabilidade do estado do Ceará às flutuações climáticas e seus impactos na disponibilidade hídrica em reservatórios superficiais, Araújo *et al.* (2019) destacam a importância da utilização de modelos de previsão de precipitação como uma ferramenta valiosa no planejamento do uso desse recurso e gerando subsídio a tomada de decisão no processo de alocação negociada como realizado atualmente no estado.

Assim, compreender a intensidade e frequência das anomalias de TSM é fundamental, pois estas podem desempenhar um papel como variável determinante nos modelos numéricos acoplados (oceano/atmosfera), os quais são empregados como ferramentas para melhorar a previsão climática (Ferreira *et al.*, 2016), ou em modelos matemáticos empíricos que usam a variabilidade da TSM para explicar o regime pluviométrico local.

Nesse contexto, o presente estudo tem como objetivo principal, avaliar a previsibilidade da precipitação na quadra chuvosa da RHSC a partir de índices de TSM utilizando métodos de regressão linear tradicional (OLS – *ordinary least squares*) e regressões do tipo Ridge e Lasso. Como objetivos específicos, tem-se:

- Classificar por ordem de influência as variáveis preditivas através do algoritmo *Iterative Input Selection*;
- Gerar modelos de previsão sazonal de precipitação a partir da associação de índices TSM por ordem de influência;
- Avaliar o desempenho dos diferentes modelos de previsão sazonal de precipitação a partir das métricas de avaliação;
- Comparar os modelos com relação ao tipo de regressão empregada.

MATERIAIS E MÉTODOS

Área de estudo

A Região Hidrográfica dos Sertões de Crateús (RHSC) é uma das 12 regiões hidrográficas de planejamento para aplicação da Política Estadual de Recursos Hídricos. Ela está localizada na porção centro-oeste do Ceará, delimitada pelas RHs Serra da Ibiapaba e Acaraú ao norte, RH Banabuí a leste, RH Alto Jaguaribe ao sul, e o estado do Piauí a oeste (Figura 1). A RHSC possui uma área de drenagem de 10.821 km² e população de aproximadamente 240 mil habitantes, divididos em nove municípios, destes o mais populoso é Crateús com cerca de 75 mil habitantes. O abastecimento humano e demais usos da água na RHSC é realizada a partir de 10 reservatórios superficiais estratégicos, capacidade de acumulação de 446,69 hm³. A recarga desses reservatórios é sempre de difícil previsão, devido à alta variabilidade das precipitações interanuais.



Figura 1: Regiões hidrográficas do Estado do Ceará. Fonte: SRH-CE (2013).

Estratégia metodológica e levantamento dos dados.

Inicialmente, 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 médios da Temperatura da Superfície do Mar (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), conforme ilustrado na Figura 2.

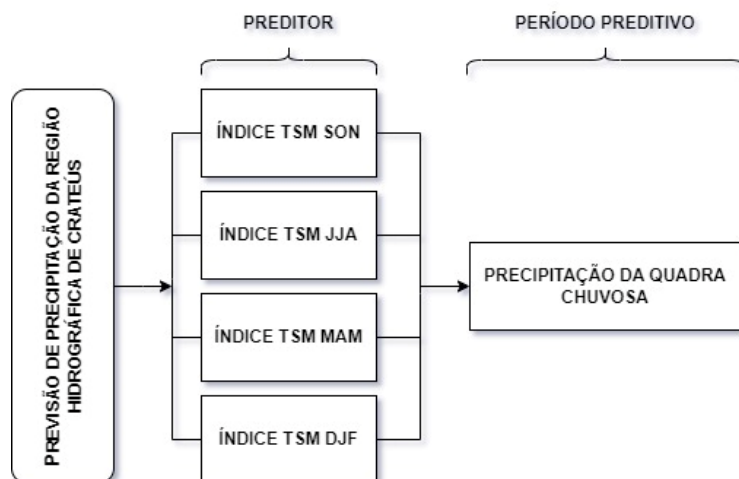


Figura 2: Estratégia metodológica utilizada para determinação da precipitação.

Os índices de anomalias de TSM utilizados foram definidos com base no estudo de Evangelista *et al.* (2024), com enfoque no estado Ceará, a partir dos índices disponíveis no sítio do *Earth System Research Laboratory* (ESRL) do *National Oceanic and Atmospheric Administration* (NOAA). Os índices utilizados podem ser observados na Tabela 1.

Tabela 1: Índices TSM utilizados como variáveis preditivas

SIGLA	ÍNDICES	
	Inglês	Português
TAD	<i>Tropical Atlantic Dipole</i>	Dipolo do Atlântico Tropical
TNA	<i>Tropical Northern Atlantic</i>	Atlântico Tropical Norte
TSA	<i>Tropical Southern Atlantic</i>	Atlântico Tropical Sul
NINO1.2	<i>Niño 1 + 2</i>	Niño 1 + 2
NINO3	<i>Niño 3</i>	Niño 3
NINO3.4	<i>Niño 3.4</i>	Niño 3.4
NINO4	<i>Niño 4</i>	Niño 4
NAO	<i>North Atlantic Oscillation</i>	Oscilação do Atlântico Norte
PDO	<i>Pacific Decadal Oscillation</i>	Oscilação Decadal do Pacífico
AMO	<i>Atlantic Multidecadal Oscillation</i>	Oscilação Multidecadal do Atlântico
ONI	<i>Oceanic Niño Index</i>	Niño Oceânico

Seleção dos preditores e geração dos modelos de previsão

Para realizar a classificação das variáveis por ordem de influência, foi empregado o algoritmo *Iterative Input Selection* (IIS) composto por três etapas, 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 – *Multi-Input-Single-Output*), conforme ilustrado na Figura 2.

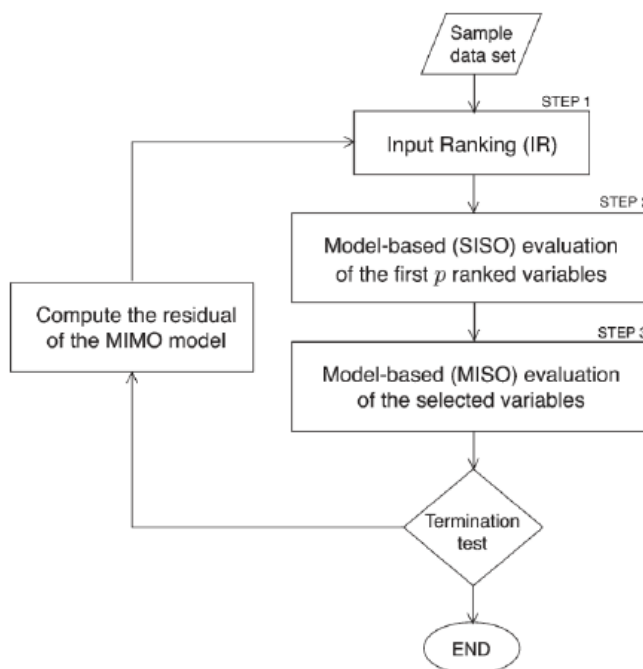


Figura 3: Fluxograma do Algoritmo *Iterative Input Selection*. Fonte: Galelli e Castelletti (2013).

Inicialmente, as variáveis foram submetidas ao *Input Ranking* (IR) usando o algoritmo *Random Forest* (RF) para avaliar sua importância inicial. 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).

No processo, destaca-se que, enquanto apenas a regressão linear tradicional (OLS – *ordinary least squares*) foi utilizada como modelo SISO e MISO para o ranqueamento das variáveis, no IR foi adotado o algoritmo RF para a seleção das entradas candidatas do modelo.

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 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 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 ϵ .

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 ϵ 0,1%, sendo que a distância métrica utilizada para interromper o programa foi o coeficiente de determinação (R^2). 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 (FMAM) para cada período preditor (SON, JJA, MAM, DJF), 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. O modelo de uma regressão linear múltipla, com k variáveis independentes pode ser definido pela Equação 1 (Hoffmann, 2016).

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

Em problemas de regressão linear múltipla, busca-se encontrar os melhores coeficientes β , de tal forma a reduzir os erros. Para obter o vetor β que melhor se ajuste aos dados da regressão, é realizada a minimização da soma dos quadrados dos erros (SQE), essa minimização é definida como sendo o estimador do modelo. Dessa forma, a Equação 2 deve ser minimizada para encontrar as melhores estimativas para os coeficientes.

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

Entretanto, 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). Dessa forma, para mitigar a multicolinearidade (correlação entre variáveis independentes) e minimizar a redundância, foram criados 11 novos modelos de previsão de precipitação utilizando as técnicas de regressão Ridge e Lasso.

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, 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 3 e 4, respectivamente. Salienta-se, que o parâmetro λ , definido nos dois modelos, é encontrado por intermédio do método da validação cruzada.

$$u_i^2 = \left(\sum_{k=1}^n \left(Y_k - \alpha - \sum_{j=1}^d (\beta_j X_{kj}) \right) \right)^2 + \lambda \sum_{j=1}^n |\beta_j|, \text{ em que } \lambda > 0 \quad \text{equação (3)}$$

$$u_i^2 = \left(\sum_{k=1}^n \left(Y_k - \alpha - \sum_{j=1}^d (\beta_j X_{kj}) \right) \right)^2 + \lambda \sum_{j=1}^n |\beta_j^2|, \text{ em que } \lambda > 0 \quad \text{equação (4)}$$



A acurácia dos modelos de previsão foi analisada a partir de dois parâmetros, sendo eles os coeficientes de correlação Pearson (R) e eficiência proposto Nash; Sutcliffe (1970) (Nash), parâmetros esse utilizados no trabalho de Evangelista *et al.* (2024). Para a correlação de Pearson, Dancey e Reidy (2006) classificam os valores da seguinte forma: correlações fracas para valores entre 0,10 e 0,30, correlações moderadas para valores entre 0,40 e 0,60, e correlações fortes para valores entre 0,70 e 1,00. Conforme Moriasi *et al.* (2007), valores do coeficiente de Nash superiores a 0,75 são considerados muito bons; valores entre 0,65 e 0,75 são bons; valores entre 0,50 e 0,65 são satisfatórios; e valores iguais ou inferiores a 0,50 são considerados insatisfatórios.

RESULTADOS

Ranqueamento das variáveis preditoras

Quatro ranqueamentos das variáveis preditoras foram gerados, abrangendo cada trimestre dos dados de índices temperatura da superfície do mar (TSM) (SON, JJA, MAM e DJF). É importante ressaltar que a seleção foi realizada exclusivamente utilizando o algoritmo *Iterative Input Selection* (IIS) implementado com *Random Forest* (RF) e regressão linear múltipla tradicional (RLM). Os resultados obtidos são apresentados na Tabela 2.

Tabela 2: Classificação dos índices TSM para os diferentes períodos preditores

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

Analisando a Tabela 2, observa-se que os índices relacionados ao El Niño-Oscilação Sul (ENOS) mostraram um desempenho consistente, destacando-se o índice NINO3 como um dos mais influentes. Em três trimestres (SON, JJA e MAM), ele foi considerado o mais relevante, e no trimestre DJF, classificado como o sexto mais importante. Outro índice significativo foi o NINO4, posicionando-se em segundo para DJF, terceiro lugar para SON e JJA, e quarto para MAM. Os resultados do NINO3.4 variaram, entre segundo (SON), quarto (DJF) e nono (JJA), enquanto o NINO1.2 geralmente teve classificações mais baixas com um terceiro, um sexto e dois nono lugares.

Ainda na Tabela 2, o índice TAD teve uma das piores classificações, ficando em penúltimo para os preditores SON e JJA, e em último e sétimo para MAM e DJF, respectivamente. Os índices TSA e TNA, relacionados ao Dipolo do Atlântico, tiveram colocações variadas, a partir da quinta posição. O índice 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 para os demais obteve as últimas posições. Outrossim, os índices AMO, NAO e ONI apresentam classificações diversas para os diferentes preditores, desde posições elevadas a mais inferiores.

Modelos de previsão

As Figuras 4, 5, 6 e 7, apresentam os resultados dos 11 Modelos para cada tipo de regressão utilizada em relação aos quatro preditores, os resultados são referentes aos valores de eficiência Nash e correlação Pearson (R), referente a etapa de validação dos modelos.

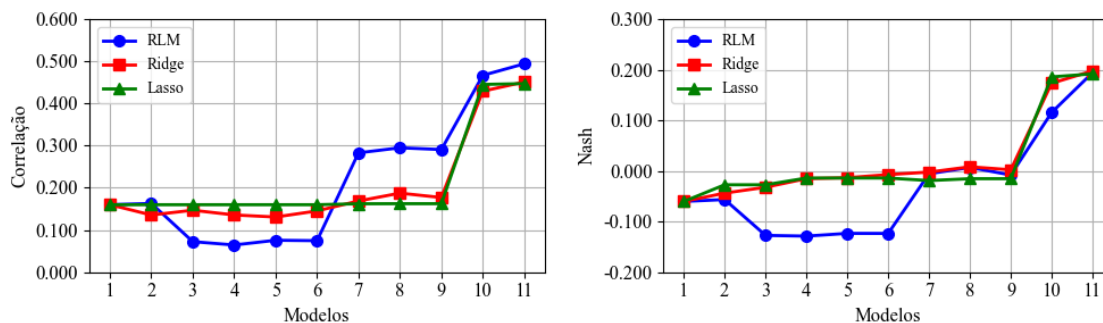


Figura 4: Correlação de Pearson e eficiência Nash para o período predictor SON.

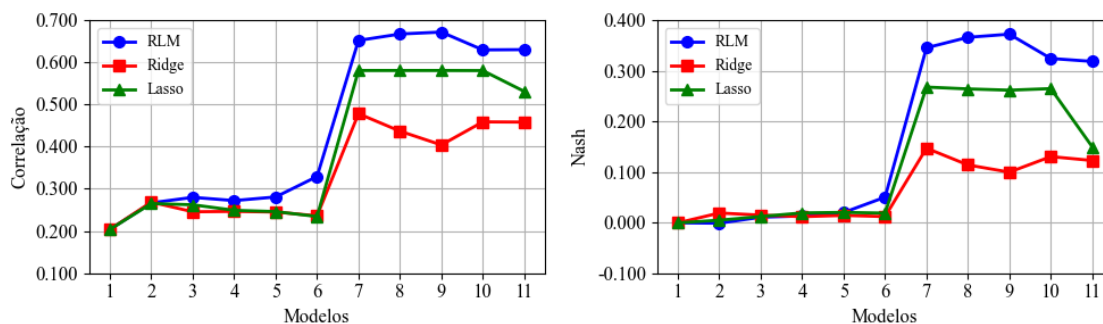


Figura 5: Correlação de Pearson e eficiência Nash para o período predictor JJA.

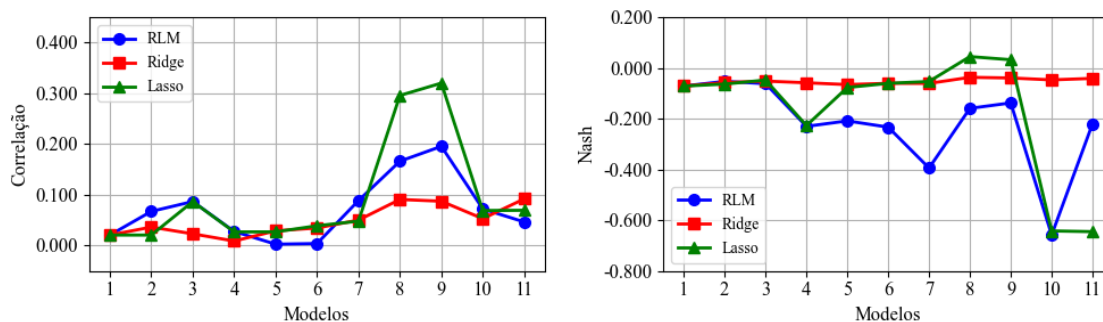


Figura 6: Correlação de Pearson e eficiência Nash para o período predictor MAM.

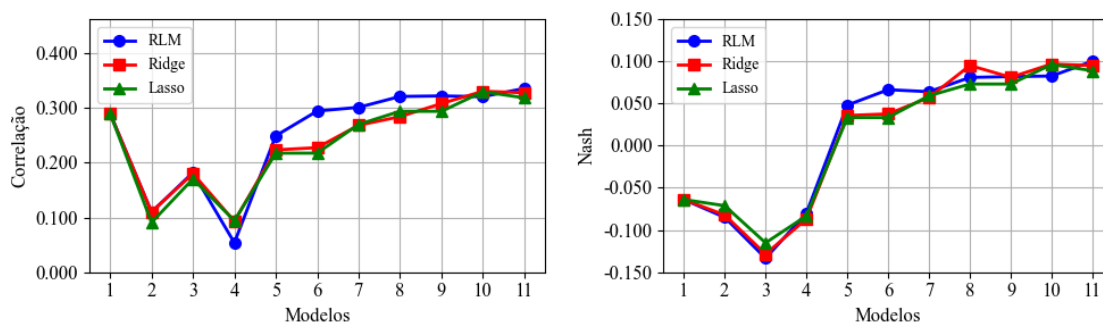


Figura 7: Correlação de Pearson e eficiência Nash para o período predictor DJF.



A análise das Figuras 4 a 7, que comparam os três tipos de regressão para os 11 modelos, evidencia que os períodos preditores que se destacam são SON e JJA. Notadamente, para o período JJA, a Regressão Linear Múltipla tradicional (RLM) se sobressai, apresentando correlações de Pearson superiores a 0,6 e um índice de Nash acima de 0,3. Isso indica correlações moderadas, conforme Dancey e Reidy (2006), entretanto, o coeficiente de Nash é classificado como insatisfatório, de acordo com Moriasi *et al.* (2007).

Além disso, os modelos que apresentaram os melhores resultados variaram conforme o tipo de regressão e a quantidade de variáveis associadas para cada período predictor. Durante o período predictor SON, os modelos que obtiveram os melhores desempenhos foram aqueles com mais de 9 variáveis associadas, especificamente os modelos M10 e M11. Isso sugere que, para este período, a inclusão de um número maior de variáveis preditivas permite capturar melhor as nuances e tendências dos dados, resultando em valores maiores das métricas de avaliação. No entanto, é necessário avaliar a dimensão dessa variação, pois se um modelo com poucas variáveis apresentar comportamento somente um pouco pior, significa que o acréscimo de variáveis nos modelos não proporciona grandes ganhos em termos de desempenho, por tanto a opção de uso fica por conta do usuário do modelo.

Por outro lado, no período predictor JJA, os modelos começaram a mostrar um desempenho representativo a partir da associação de 7 variáveis preditivas. Isso indica que, para esses meses, um número menor de variáveis preditivas é suficiente para fornecer resultados significativos. Essa diferença no número de variáveis necessárias para otimizar o desempenho dos modelos entre os dois períodos pode ser atribuída a variações sazonais e ao comportamento específico dos dados em cada intervalo de tempo. Dessa maneira, esses resultados ressaltam a importância de ajustar o número de variáveis preditivas de acordo com o período analisado para maximizar a eficácia dos modelos de regressão.

As Tabelas 3 e 4 apresentam os melhores resultados para cada predictor, observa-se que os modelos gerados com regressão Ridge e Lasso mostraram correlações de Pearson (R) e valores de eficiência de Nash menores durante o treinamento, em comparação com a RLM. Esta tendência foi consistente, com exceção dos modelos 04, 09 e 10 da regressão Lasso para o predictor MAM. Por outro lado, alguns modelos demonstraram maiores valores de Nash durante a validação quando utilizaram regressão Ridge ou Lasso. Especificamente, para o período do predictor SON, a regressão Ridge superou a Lasso em todos os modelos, enquanto para o predictor MAM, com exceção dos modelos 02 e 11, os valores de Nash foram consistentemente maiores com regressão Ridge ou Lasso em comparação com a RLM.

Tabela 3: Comparação dos resultados obtidos de correlação (R) para cada tipo de regressão.

Método	Predictor	Treinamento						Validação					
		1°		2°		3°		1°		2°		3°	
		R	M	R	M	R	M	R	M	R	M	R	M
RLM	SON	0,696	10	0,673	11	0,513	9	0,494	11	0,467	10	0,295	8
	JJA	0,574	9	0,574	8	0,566	7	0,672	9	0,667	8	0,652	7
	MAM	0,606	10	0,554	11	0,536	9	0,196	9	0,166	8	0,089	7
	DJF	0,339	10	0,339	9	0,329	8	0,336	11	0,322	9	0,321	8
Ridge	SON	0,645	11	0,638	10	0,456	9	0,451	11	0,429	10	0,187	8
	JJA	0,492	7	0,488	11	0,483	10	0,478	7	0,459	10	0,458	11
	MAM	0,356	10	0,347	11	0,324	8	0,092	11	0,091	8	0,087	9
	DJF	0,322	11	0,322	10	0,318	9	0,331	10	0,328	11	0,308	9
Lasso	SON	0,645	10	0,644	11	0,472	7	0,447	11	0,445	10	0,162	8 e 10
	JJA	0,539	7	0,539	8	0,539	10	0,580	7 a 9	0,580	10	0,531	11
	MAM	0,612	10 e 11	0,376	8	0,370	9	0,320	9	0,296	8	0,070	11
	DJF	0,315	11	0,314	10	0,307	8 e 9	0,330	10	0,319	11	0,294	8

Tabela 4: Comparação dos resultados obtidos de Nash para cada tipo de regressão.

Método	Preditor	Treinamento						Validação					
		1º		2º		3º		1º		2º		3º	
		Nash	M	Nash	M	Nash	M	Nash	M	Nash	M	Nash	M
RLM	SON	0,484	10	0,452	11	0,263	9	0,194	11	0,116	10	0,007	8
	JJA	0,330	9	0,329	8	0,320	7	0,372	9	0,366	8	0,345	7
	MAM	0,367	10	0,307	11	0,287	9	-0,053	2	-0,062	3	-0,072	1
	DJF	0,115	10	0,115	9	0,108	8	0,100	11	0,083	10	0,082	9
Ridge	SON	0,387	11	0,385	10	0,173	9	0,198	11	0,173	10	0,008	8
	JJA	0,200	7	0,184	10	0,181	11	0,147	7	0,130	10	0,123	11
	MAM	0,057	8	0,056	10	0,048	11	-0,037	8	-0,040	9	-0,041	11
	DJF	0,103	11	0,103	10	0,100	9	0,097	10	0,095	8	0,095	11
Lasso	SON	0,399	10	0,392	11	0,203	7	0,192	11	0,186	10	-0,014	4 a 6
	JJA	0,284	7	0,281	10	0,280	8	0,268	7	0,265	10	0,264	8
	MAM	0,367	11	0,367	10	0,139	4	0,045	8	0,032	9	-0,048	3
	DJF	0,098	11	0,097	10	0,093	8 e 9	0,096	10	0,089	11	0,073	8 e 9

Destaca-se que a diferença nos resultados entre a RLM e as técnicas de regressão Ridge e Lasso pode ser atribuída ao efeito de penalização dos últimos. Segundo Hastie *et al.* (2009), enquanto a regressão Ridge penaliza a soma dos quadrados dos coeficientes e a regressão Lasso penaliza a soma dos valores absolutos dos coeficientes, essas penalizações podem afetar a precisão da estimativa dos coeficientes, especialmente em conjuntos de dados pequenos, dado que esse fator de penalização pode não ser bem calibrado. Em contraste, a RLM não impõe restrições aos coeficientes, permitindo uma interpretação direta das relações entre as variáveis preditoras e a variável de resposta.

CONCLUSÃO

Este estudo teve como intuito a classificação dos índices de TSM por ordem de influência em relação a precipitação observada, geração e verificação da eficácia de modelos de previsão para a quadra chuvosa (FMAM) da Região Hidrográfica dos Sertões de Crateús (RHSC). Foram utilizados três tipos de regressões, a regressão linear múltipla tradicional (RLM) e as regressões Ridge e Lasso, para quatro períodos preditores trimestrais da temperatura da superfície do mar TSM (SON, JJA, MAM e DJF), gerando assim um total de 11 modelos associados aos índices TSM para cada tipo de regressão.

A seleção dos índices de TSM foi realizada por meio do Algoritmo *Iterative Input Selection* (IIS), que destacou os índices relacionados aos fenômenos El Niño-Oscilação Sul (ENOS), com ênfase nos índices NINO3 e NINO4, como os mais relevantes. Os índices associados ao oceano Atlântico foram classificados em posições intermediárias e inferiores, sendo o índice dipolo do Atlântico (TAD) o menos significativo. Os demais índices mostraram variações de importância dependendo do período de previsão.

Na avaliação do desempenho dos modelos, observou-se que as correlações de Pearson (R) obtiveram valores significativos tanto no estágio de treinamento quanto de validação. Alguns modelos apresentaram correlações na faixa de 0,30 a 0,49, enquanto outros ultrapassaram o valor de 0,5, indicando correlações moderadas, conforme definido por Dancey e Reidy (2006). Além disso, a maioria dos modelos demonstrou valores positivos no índice de Nash, indicando que as previsões foram mais precisas do que a média dos valores observados, conforme descrito por Moriasi *et al.* (2007). De um modo geral a RLM apresentou resultados mais acurados que as do tipo Lasso e Ridge.



REFERÊNCIAS BIBLIOGRÁFICAS

1. 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.
2. CASAGRANDE, M. H. Comparação de métodos de estimação para problemas com colinearidade e/ou alta dimensionalidade ($p > n$). 2016. Dissertação (Mestrado em Estatística) – Universidade Federal de São Carlos, São Carlos, 2016. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/7954>.
3. COUTINHO, M. D. L.; SAKAMOTO, M. S. Análise sinótica da intensa precipitação observada no litoral do Ceará em 7 de abril de 2016. In: Geosp – Espaço e Tempo, v. 21, n. 3, p. 858-871, 2017.
4. DANCEY, C. P.; REIDY, J. ESTATÍSTICA SEM MATEMÁTICA PARA PSICOLOGIA: Usando SPSS para Windows. 3ª. ed. Porto Alegre - RS: Artmed, 2006. 608 p. ISBN 978-85-363-0688-9.
5. EVANGELISTA, F. I. S. *et al.* Análise da Variabilidade da Temperatura da Superfície do Mar e o Regime Pluviométrico nos Sertões de Crateús. In: 21º Simpósio Luso-Brasileiro de Engenharia Sanitária e Ambiental, Recife. Anais [...]. Recife: ABES, 2024. p. 1-10.
6. 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.
7. 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, p. 366-373, 2016.
8. GALELLI, S.; CASTELLETTI, A. Tree-based iterative input variable Selection for hydrological modeling. Water Resources Research, v. 49, n. 7, p. 4295–4310, 2013.
9. HASTIE, T. *et al.* The elements of statistical learning: data mining, inference, and prediction. New York: Springer, 2009.
10. HOFFMANN, Rodolfo; VIEIRA, Sônia. Análise de regressão: uma introdução à econometria. São Paulo, 2016.
11. MORIASI, D. N. *et al.* Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. American Society of Agricultural and Biological Engineers, v. 50, n. 3, p. 885-900, 2007.
12. NASH, J. E.; SUTCLIFFE, J. V. River flow forecasting through conceptual models. Part 1: A discussion of principles, J. Hydrol., 10, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6), 1970.
13. NÓBREGA, J. N. da. *et al.* Eventos extremos de precipitação nas mesorregiões da Paraíba e suas relações com a TSM dos oceanos tropicais. Revista Brasileira de Meteorologia, v. 29, p. 197-208, 2014.
14. NÓBREGA, R. S.; SANTIAGO, G. A. C. F.; SOARES, D. B. Tendências do controle climático oceânico sob a variabilidade temporal da precipitação no Nordeste do Brasil. Revista Brasileira de Climatologia, [S.l.], v. 18, out. 2016. ISSN 2237-8642.
15. SANTOS, C. A. C. dos, *et al.* Tendências dos índices de precipitação no Estado do Ceará. Revista Brasileira de Meteorologia, v. 24, p. 39-47, 2009.
16. SECRETARIA DE RECURSOS HÍDRICOS. Comitê de bacias hidrográficas. Fortaleza: SRH, 2013.