

A GMDH Approach for Forecast Monthly Rainfall in Southeast of Pará

Elton Rafael Alves , Márcio Nirlando Gomes Lopes , Fabricio Silva Sales, Andson Marreiros Balieiro 
and Adônis Ferreira Raiol Leal 

Abstract—Modeling for rainfall forecasting is widely used in climate forecasting. A high volume of precipitation or scarcity brings serious problems for society. Thus, a methodology for monthly rainfall forecast is presented using Group Method of Data Handling (GMDH) and sea surface temperature (SST). The intelligent model gets the mean monthly SST, in predefined and temporally lagged areas, after a variable selection step. For model training, precipitation data from the Climate Prediction Center were used. The methodology was applied in a certain area of the municipality of Marabá, located in the southeastern region of the Pará state. The results obtained with GMDH overcame those got by using a conventional multilayer Artificial Neural Network (RNA), reaching values for regression coefficient (R^2) and root mean square error (RMSE) equal to 0.96363 and 30.8100 in the test stage, while the RNA-based model got 0.81712 and 68.2607, respectively. Results show the GMDH's effectiveness for the monthly rainfall prediction, constituting an important tool for the planning and assistance to decision makers.

Index Terms—Group Method of Data Handling, Sea Surface Temperature, Multilayer Neural Networks, Rainfall, Forecast.

I. INTRODUÇÃO

A região amazônica sofre influência de diversos sistemas meteorológicos que produzem índices pluviométricos elevados, caracterizando-a com um clima tropical e úmido [1]. A precipitação é uma das mais importantes variáveis meteorológicas, pois influencia diversos segmentos produtivos da sociedade, como agricultura, transporte e geração de energia e impacta significativamente nos ecossistemas amazônicos.

Os oceanos têm grande influência nos sistemas que causam precipitação sobre a Amazônia. A Temperatura da Superfície do Mar (TSM) é uma das variáveis climáticas capazes de alterar a dinâmica atmosférica, modulando os sistemas meteorológicos que atuam em diferentes escalas de tempo e espaço. Tais sistemas, como a Zona de Convergência Intertropical-ZCIT, as linhas de instabilidade-LI, os Vórtices Ciclônicos de Altos Níveis-VCAN, a Alta da Bolívia-AB, os Distúrbios Ondulatórios de Leste-DOL, as frentes frias e a Zona de

Convergência do Atlântico Sul-ZCAS são responsáveis pela geração de chuvas na região [2], [3].

Geralmente, os modelos de previsão de tempo e clima utilizam a TSM devido os oceanos ocuparem três quartos (71%) da superfície do globo e 81% da superfície no hemisfério Sul, além de exercerem uma influência muito grande nas características e comportamento das variáveis meteorológicas nas relações oceano-atmosfera.

No Brasil, o regime climático do sudeste do estado do Pará é bem definido, sendo composto basicamente por uma estação seca e outra chuvosa. Na estação chuvosa, os índices pluviométricos são elevados, ocasionando, eventualmente, inundações, alagamento e outras adversidades que impactam tanto no setor produtivo, quanto no cotidiano das pessoas. Na estação seca, por sua vez, os extremos climáticos podem agravar o período com pouca ou nenhuma precipitação, tornando o ambiente mais suscetível às ocorrências de queimadas, repercutindo em problemas de saúde da população, perdas econômicas e da biodiversidade.

Assim, a previsão de chuvas permite o planejamento e a tomada de decisão por parte da iniciativa privada e do poder público, especialmente de órgãos da defesa civil, de maneira a reduzir as perdas econômicas, mitigar os riscos de desastres naturais e proteger a população dos extremos climáticos, sejam estes ocasionados pelo excesso ou falta de chuvas. Em Marabá, no sudeste paraense, durante o período chuvoso, inundações de áreas urbanas e danos a propriedades ocasionados pela cheia do rio Tocantins são frequentes [4]. Neste aspecto, estudos que visam o entendimento científico das causas e consequências da variabilidade de chuva e modelos que possam prevê-la são imprescindíveis para gerar informações e produtos climatológicos úteis ao benefício da sociedade.

Os fenômenos climáticos são de origem natural e com gênese não linear, tornando difícil a sua modelagem e complexa a previsibilidade por meio de modelos físicos dinâmicos, onde a inteligência artificial tem sido empregada para endereçar fenômenos. Dessa forma, este trabalho propõe uma metodologia de modelagem inteligente para previsão mensal de chuvas na região sudeste do Pará. A proposta considera a TSM como variável preditora e emprega a técnica de aprendizagem profunda *Group Method of Data Handling* (GMDH) [5] para modelar o comportamento preditivo da precipitação pluviométrica e, assim, auxiliar os gestores na tomada de decisão. A inteligência artificial é uma abordagem amplamente adotada para solucionar problemas complexos, como a predição de chuvas. Com isso, o modelo baseado em GMDH foi comparado com uma tradicional Rede Neural Artificial de Múltiplas Camadas

Elton Rafael Alves é professor da Faculdade de Engenharia da Computação da Universidade Federal do Sul e Sudeste do Pará, Marabá, Pará, Brasil, e-mail: eltonalves@unifesspa.edu.br.

Márcio Nirlando G. Lopes é Analista em Ciência & Tecnologia do Centro Gestor e Operacional do Sistema de Proteção da Amazônia, Belém, Pará, Brasil, e-mail: marcio.lopes@sipam.gov.br

Fabricio Silva Sales é graduando em Engenharia da Computação, Marabá, Pará, Brasil, e-mail: fabricio.sales@unifesspa.edu.br

Andson Marreiros Balieiro é professor do Centro de Informática da Universidade Federal de Pernambuco, Recife, Pernambuco, Brasil, e-mail: amb4@cin.ufpe.br

Adônis Ferreira Raiol Leal é professor da Faculdade de Engenharia Elétrica e Biomédica da Universidade Federal do Pará, Belém, Pará, Brasil, e-mail: adonisleal@ufpa.br

(RNA-MLP) que tem sido comumente empregada na predição mensal de chuvas [6], [7], [8], [9]. Essa comparação buscou mostrar a eficiência da utilização do GMDH nesta proposta. Sendo assim, as principais contribuições deste artigo são: seleção das variáveis de entrada para predição e o uso de TSM, desenvolvimento de um sistema com GMDH para previsão mensal de chuvas e sua análise de desempenho comparativa utilizando uma RNA-MLP.

Este artigo encontra-se assim organizado. A Seção II destaca os trabalhos relacionados. Seção III descreve o algoritmo do GMDH. Os dados utilizados, pré-processamento e modelo proposto são descritos na Seção IV. Resultados de avaliação são apresentados na Seção V. A Seção VI conclui esse artigo.

II. TRABALHOS RELACIONADOS

Segundo Samad et al. (2020) [10], a previsão de chuvas é uma das tarefas mais desafiadoras no campo de previsão de tempo, o que é decorrente da complexidade dos processos atmosféricos que geram as chuvas. Neste aspecto, estudos para compreender a atmosfera terrestre têm sido conduzidos e o desenvolvimento de modelos de previsão de tempo (ex. chuvas) empregando métodos físicos, estatísticos e inteligentes têm sido realizado. Por exemplo, Grace et al. (2020) [11] utilizaram regressão linear múltipla (RLM) com parâmetros meteorológicos como entradas do modelo para prever chuvas, alcançando um valor de $R^2=0,469$. Já Ahmed et al. (2020) [12] desenvolveram um modelo de regressão linear múltipla para prever a taxa de precipitação com dados anuais, baseada na temperatura do ar (TA), velocidade do vento (VV) e temperatura do ponto de orvalho (TPO), obtendo um erro médio quadrático igual a 59%.

No campo de inteligência artificial, Dash et al. (2017) [13] utilizaram uma rede neural *feedforward* (RNF) e aprendizado de máquina extremo (AME) (*extreme learning machine*) para prever chuvas intensas em Kerala, na Índia, considerando diferentes estações do ano (dados mensais). O modelo com aprendizado extremo forneceu 4,7377% de *root mean squared error* (RMSE) enquanto o com rede neural *feedforward* atingiu 6,3977% de RMSE. Em Salehin et al. (2020) [14], as temperaturas do ar e do ponto de orvalho, a velocidade e a direção do vento (DV) foram adotadas como variáveis de uma LTSM (*Long short-term memory*) para prever chuvas anualmente, obtendo acurácia de 76%, que corresponde a razão entre o número de previsões corretas pelo total de dados.

Já Basha et al. (2020) [15] empregaram uma RNA-MLP e *auto-encoders* (AE) para prever a precipitação pluviométrica na Índia. O modelo obteve valores de R^2 iguais a 0,49788 e 0,64027 nas fases de teste e treinamento, respectivamente. Nastos et al. (2013) [16], por sua vez, apresentam modelos preditivos de intensidade de chuvas (mm/dia) em Atenas, na Grécia, usando RNA-MLP e dados diários de chuvas. O melhor modelo alcançou RMSE e R^2 iguais a 1,7 e 0,515, respectivamente. Já Wang et al. (2010) [17] propuseram uma rede neural de regressão generalizada (RNRG) para predição anual de chuvas (dados anuais), utilizando o histórico de chuvas e parâmetros meteorológicos, obtendo um *mean squared error* (MSE) médio igual a 1,26%. A Tabela I exibe uma

comparação da metodologia proposta neste trabalho com as metodologias apresentadas para previsão de chuvas.

Como observado, existem muitos trabalhos relevantes sobre previsão de chuvas na literatura. Entretanto, eles não realizam uma previsão mensal empregando TSM combinado com um método inteligente (GMDH) de modo a melhorar principalmente o valor de R^2 para dados de testes.

TABELA I
COMPARAÇÃO DE MÉTODOS DE PREVISÃO DE CHUVAS.

Artigo	Técnica	Desempenho	Atributo(s)
[10]	RLM	R^2	Indefinido
[11]	RLM	MSE	TA, VV, TPO
[12]	RNF e AME	RMSE	Precipitação média
[13]	LTSM	Acurácia	TV, TPO, VV e DV
[14]	RNA-MLO e AE	R^2	Indefinido
[15]	RNA-MLP	RMSE e R^2	Dias e total de chuvas
[16]	RNRG	MSE	Precipitação
Este	GMDH	RMSE e R^2	TSM

III. GROUP METHOD OF DATA HANDLING - GMDH

Este trabalho utiliza o GMDH para desenvolver uma nova correlação entre TSM e dados de precipitação pluviométrica e assim prever mensalmente chuvas no sudeste paraense. Segundo [5], o GMDH se diferencia de outros modelos de inteligência artificial pois:

- 1) Não necessita de uma pré-estrutura de rede, como é o caso das redes neurais artificiais tradicionais;
- 2) Resulta em um modelo livre de *overfitting*, pois a escolha da estrutura do modelo é realizada de acordo com a separação dos dados em treino e teste;
- 3) Possui convergência do modelo garantida após um número determinado de etapas de treinamento.

O GMDH é muito utilizado em problemas de reconhecimento de padrões com relações complexas entre os parâmetros de entrada [5], como é o caso dos dados de TSM de diferentes áreas oceânicas utilizadas neste estudo. Além disso, ele tem a vantagem de filtrar parâmetros de entrada e usar apenas os mais influentes para construir o modelo. Estas características do GMDH e sua diversidade de usos são aspectos que levaram ao emprego deste algoritmo neste trabalho.

A Fig. 1 exibe a arquitetura do GMDH, compreendendo três camadas: a de entrada, que corresponde às variáveis de entrada; a intermediária, onde as correlações com os neurônios são feitas repetidamente até que se alcance a saída desejada e a de saída, que denota à saída desejada. O algoritmo do GMDH seleciona os melhores modelos, descartando aqueles com desempenho baixo [18], [19]. Ele consiste no desenvolvimento de um polinômio de alta ordem (Kolmogorov-Gabor), conforme Eq. 1.

$$y = a_0 + \sum_{i=1}^n a_i x_i + \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_{ijk} x_i x_j x_k + \dots \quad (1)$$

$$n = 1, 2, \dots, 2^l,$$

onde y é a saída do modelo, n é o número de padrões de entrada, l corresponde ao número de camadas do algoritmo, x_i, x_j, x_k representam os parâmetros de entrada e a_0, a_i, a_j, a_k são os coeficientes do polinômio.

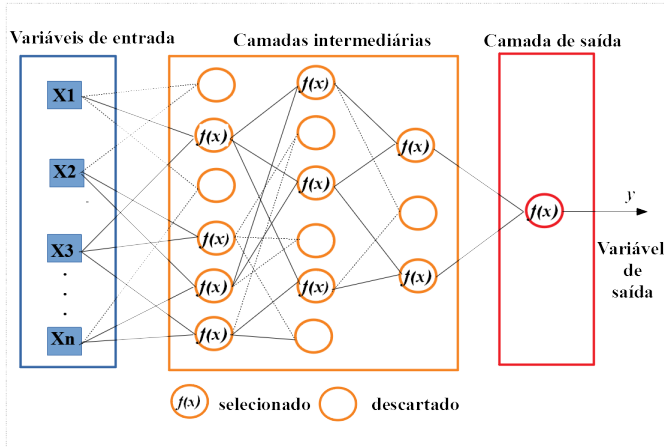


Fig. 1. Arquitetura da uma rede GMDH.

Ma et al. (2009) [20] resumem o algoritmo GMDH em seis etapas:

- 1) Seleção das variáveis de entrada: os n padrões de entrada são escolhidos para estimar a variável de saída, sendo o modelo dado por: $f(x_1, x_2, \dots, x_n)$;
- 2) Divisão aleatória dos dados em conjunto de treinamento e teste;
- 3) Construção de novas variáveis intermediárias na camada intermediária, de acordo com Eq. 2, que descreve um polinômio de grau 2 com duas entradas (x_i, x_k) e é usada para calcular a saída de cada neurônio do GMDH. Ela foi definida por Ivakhnenko [21], inspirada no polinômio de Kolmogorov-Gabor, definido na Eq.1.

$$y = a_{0i} + a_{1i}x_j + a_{2i}x_k + a_{3i}x_j^2 + a_{4i}x_k^2 + a_{5i}x_jx_k, \quad (2)$$

$$i = 1, \dots, q; j = 1, \dots, n; k = 1, \dots, n - 1; q = \frac{n(n-1)}{2}$$

- 4) Seleção de novas variáveis: ocorre a avaliação do RMSE obtido na etapa anterior. Nesta fase os melhores nós obtidos em cada camada são selecionados para passar para a próxima camada;
- 5) Computação iterativa de multicamadas: ocorre a comparação entre o melhor resultado da camada atual com a camada anterior. Caso a melhora exceda o limite definido ou atinja a próxima camada, o critério de parada do algoritmo é satisfeito, caso contrário, retorna para a etapa 3;
- 6) Obtenção da variável de saída: a variável de saída do modelo é alcançada através das variáveis intermediárias restantes na camada final.

IV. MATERIAIS E MÉTODOS

Esta seção descreve os dados utilizados, o pré-processamento dos dados e o modelo proposto.

A. Dados Utilizados

Este estudo utilizou dados mensais das seguintes variáveis:

- Temperatura da Superfície do Mar (TSM).
- Precipitação pluviométrica

1) *Dados de TSM*: este trabalho utilizou médias mensais da Temperatura da Superfície do Mar (TSM) dos oceanos Atlântico Norte (ATLA-N), Atlântico Sul (ATLA-S), Pacífico Equatorial (regiões de Niño) e Pacífico Sul (PAC-S) no período de 1982 a 2020. Os campos de TSM são dados reconstruídos na resolução espacial de $1^\circ \times 1^\circ$ de latitude e longitude, respectivamente, e foram gerados pelo *National Centers for Environmental Prediction/ National Oceanic and Atmospheric Administration* (NCEP/NOAA). Os dados de TSM para as áreas ATLA-N, ATLA-S e PAC-S foram agrupados em 96 áreas regulares de até 5° de latitude por 10° de longitude, obtendo-se a série da média mensal da TSM ponderada pelo tamanho da área. Visto que as áreas de Niño adotam regiões pré-definidas, não houve alteração quanto à delimitação das respectivas áreas, obtendo-se diretamente os valores de TSM para as áreas de Niño1+2, Niño 3, Niño 3.4 e Niño 4.

Muito embora este estudo também tenha considerado as TSM no Atlântico Tropical Norte (ATLA-N) e nas regiões de Niño, estas mostraram-se não relevantes para o modelo desenvolvido, como será detalhado nas seções seguintes, razão pela qual a Fig. 2 exibe apenas as áreas oceânicas efetivamente empregadas para a modelagem e o ponto de grade no município de Marabá, sudeste do Pará, para o qual foi realizada a previsão mensal de chuva.

2) *Dados de Precipitação*: Os dados de precipitação são provenientes do CPC/NCEP (*Climate Prediction Center/ National Centers for Environmental Prediction*) e constituem uma série diária em uma base regular global para pontos de grade com resolução espacial de $0,5^\circ$ de latitude por $0,5^\circ$ de longitude [22]. Os índices pluviométricos diários foram totalizados para obtenção de uma série de índice pluviométrico mensal, destacando-se o ponto de grade que representa o município de Marabá-PA, compreendendo o período de 1982 a 2020. A área de estudo está localizada conforme o ponto de grade mostrado na Fig. 2.

A Fig. 3 ilustra a média mensal dos valores de precipitação em mm (1982 – 2020) para área de estudo. A distribuição da precipitação mensal ao longo do ano caracteriza a região com duas estações bem marcadas: um período mais chuvoso (janeiro, fevereiro e março) e outro período mais seco (junho, julho e agosto).

B. Seleção das Variáveis de Entrada do Modelo

Para a seleção das variáveis predictoras foram realizadas defasagens (atrasos) temporais mensais nos dados de TSM. As defasagens podem ser representadas como: $x(t-1), x(t-2), \dots, x(t-12)$, onde t é o tempo em meses. Esses atrasos buscaram verificar se as janelas de tempo anteriores a um determinado período são capazes de agregar previsibilidade aos valores de precipitação desse período. Assim, buscou-se obter as áreas de TSM com melhor relação com a chuva na área de estudo. Neste processo, foram realizados atrasos na escala de 1 até 12 meses nos valores de TSM gerando um total 1.152 variáveis

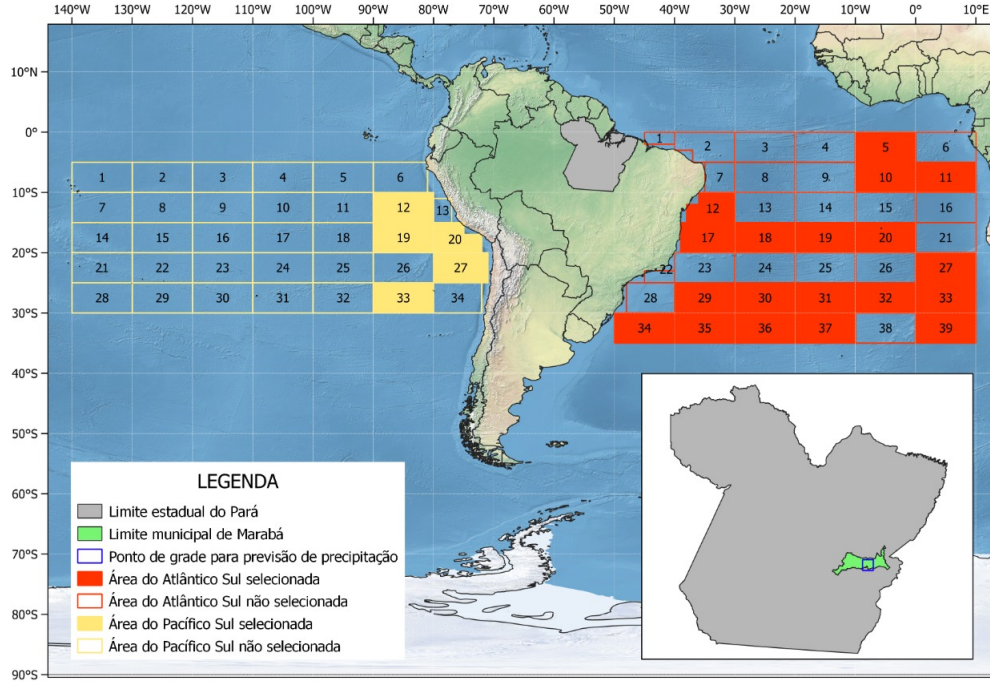


Fig. 2. Áreas oceânicas utilizadas para obter os dados de TSM de entrada do modelo e área de previsão pluviométrica mensal, em Marabá-PA.

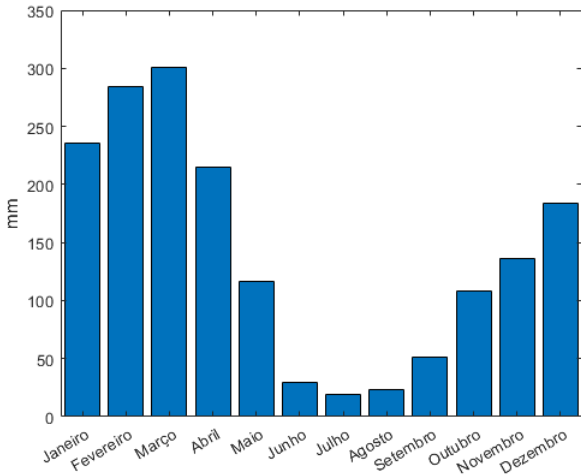


Fig. 3. Média mensal de precipitação para o período base 1982-2020.

preditoras. Dessa forma, a previsão consiste em prever chuvas nos meses posteriores a partir dos meses anteriores.

Após realizar as defasagens no tempo, todas as variáveis obtidas foram correlacionadas com a área de precipitação, utilizando o método de correlação de Spearman, dada pela Eq. 3, onde d é dado pela diferença entre o posto de x_i (dentre os valores de x) e o posto de y_i (dentre os valores de y) e n é o número de observações utilizadas (x_i, y_i). Essa correlação consistiu em dar suporte na seleção dos padrões que irão compor a entrada do modelo de predição.

$$Cor(x, y) = 1 - \frac{6 \sum_i d_i^2}{n(n^2 - 1)} \quad (3)$$

O coeficiente de correlação de Spearman gera um número na escala de -1 a 1. Quanto mais próximo dos extremos (-1 ou 1), maior é a força da correlação. Já os valores próximos de 0 implicam correlações mais fracas ou inexistentes. Esta correlação foi empregada, pois não há pressupostos de que os dados possuem um comportamento linear, sendo utilizada para relacionar dados com comportamentos não lineares.

C. Modelo Proposto

A Fig. 4 exibe o esquema genérico do modelo proposto baseado em GMDH para prever a precipitação mensal (Prec_Mensal) a partir dos dados de TSM das áreas oceânicas com defasagens no tempo.

No processo de seleção das variáveis, os padrões de entrada são representados como:

- TSM_A(n)_ATLA-N_D(n) = Dados de TSM de determinada área A(n) do oceano Atlântico Norte com defasagem temporal D (n) em meses;
- TSM_A(n)_ATLA-S_D(n) = Dados de TSM de determinada área A(n) do oceano Atlântico Sul com defasagem temporal D (n) em meses;
- TSM_A(n)_niño_D(n) = Dados de TSM de determinada área A(n) da região de niño com defasagem temporal D (n) em meses;
- TSM_A(n)_PAC-S_D(n) = Dados de TSM de determinada área A(n) do oceano Pacífico Sul com defasagem temporal D (n) em meses;

O modelo proposto foi comparado com uma Rede Neural de Múltiplas Camadas, treinada com o algoritmo *Levenberg-Marquardt Backpropagation* [23]. Segundo [6], a RNA-MLP

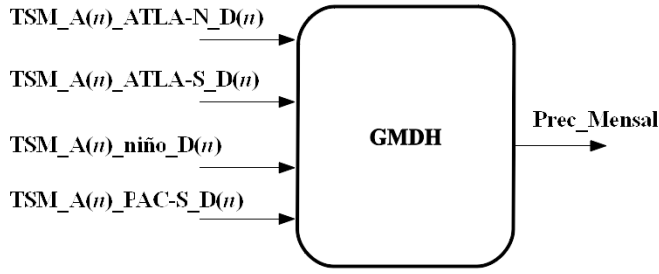


Fig. 4. Esquema de modelagem.

é uma rede neural *feedforward* que mapeia os dados de entrada em saídas apropriadas, possuindo uma ou mais camadas escondidas (neurônios escondidos) e os neurônios da camada escondida e de saída podem ter funções de ativação.

Os dados selecionados foram divididos aleatoriamente em 70% para treinamento do modelo GMDH e RNA-MLP e 30% para testes em ambos os modelos, GMDH e RNA-MLP. A fase de treinamento consistiu nos ajustes dos hiper-parâmetros e a fase de teste na avaliação do comportamento de cada modelo treinado para dados desconhecidos. Assim, os desempenhos dos modelos preditivos foram medidos em termos de *Root Mean Squared Error* (RMSE) de treinamento e teste, dado pela Eq. 4 e coeficiente de determinação R^2 , dado pela Eq. 5. O primeiro calcula a raiz quadrada da média dos erros entre os valores preditos e reais elevados ao quadrado. Já o R^2 é uma medida que mostra a proximidade dos dados a uma linha de regressão, correspondendo razão entre a variação explicada e a variação total dos dados. Os valores de R^2 ilustrados nas Fig. 6 e 7 correspondem a comparação entre a resposta do modelo e a saída real (observada) para dados de teste.

$$RMSE = \sqrt{\frac{\sum_i^n (\hat{y}_i - y_i)^2}{n}} \quad (4)$$

onde \hat{y} é o valor predito, y_i é o valor real e n é o total de amostras.

$$R^2 = \frac{\sum_i^n (\hat{y}_i - \bar{Y})^2}{\sum_i^n (y_i - \bar{Y})^2} \quad (5)$$

onde \bar{Y} é a média aritmética dos rótulos dos dados, $(\hat{y} - \bar{Y})^2$ é a variação explicada e $(y_i - \bar{Y})^2$ é a variação total.

V. RESULTADOS E DISCUSSÕES

Os modelos foram codificados no software Matlab R2019b e executados em um computador Intel Core i5-8265U CPU de 1.8GHz e 8GB de memória. O tempo de execução de cada um (compreendendo as etapas de treino e teste) foi de aproximadamente 5,23s para o GMDH e 0,55s para a RNA-MLP. Embora o GMDH tenha apresentado um tempo de execução em torno de 10 vezes o da RNA-MLP, a sua duração de poucos segundos está adequada para a previsão de chuvas, onde, em geral, adotam-se janelas de previsão de horas, dias, meses ou anuais.

A. Análises das Correlações

Considerando as 96 áreas oceânicas propostas inicialmente (Fig. 2) e as defasagens de até 12 meses aplicadas para cada área, foram obtidas 1152 variáveis predictoras possíveis, as quais foram correlacionadas com a área de precipitação de estudo. A Fig. 5 ilustra apenas uma amostra do processo total de correlações obtidas, onde se destaca o comportamento da dispersão dos valores entre as variáveis AN1, PACS6, AN17 e A. Sendo AN1, PACS6, AN17 variáveis de TSM de áreas oceânicas e A é a variável precipitação da área de estudo. Em I, VI, XI e XVI tem-se a distribuição dos valores das variáveis em histograma. Enquanto, II, III, IV, V, IX e XII expressam correlações negativas entre as variáveis, bem como os valores de correlações. Por sua vez, VII, VIII, X, XII, XIV e XV exibem correlações positivas entre as variáveis associada e os seus respectivos valores.

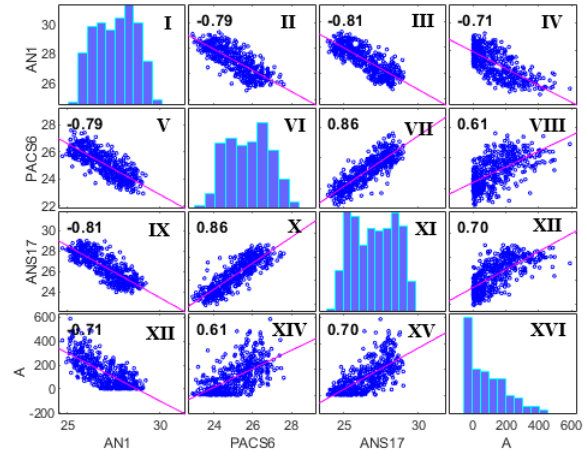


Fig. 5. Amostra de correlação obtida entre a TSM e a precipitação.

Este estudo selecionou as variáveis que obtiveram coeficiente de correlação superior a 0,65 (positiva ou negativa), descartando as demais, ou seja, as variáveis referentes às áreas de Niño e do Atlântico Norte não alcançaram o índice para compor o modelo. Desta forma, após o processo de correlação, foram obtidas 24 variáveis que compuseram a entrada do modelo, conforme segue:

TSM_12_PAC-S(5); TSM_10_ATLA-S(11); TSM_34_ATLA-S(6); TSM_20_PAC-S(6); TSM_19_PAC-S(5); TSM_20_ATLA-S(11); TSM_19_ATLA-S(11); TSM_27_ATLA-S(5); TSM_18_ATLA-S(11); TSM_11_ATLA-S(11); TSM_17_ATLA-S(11); TSM_35_ATLA-S(5); TSM_12_ATLA-S(11); TSM_33_PAC-S(5); TSM_27_PAC-S(5); TSM_33_ATLA-S(5); TSM_39_ATLA-S(5); TSM_36_ATLA-S(5); TSM_30_ATLA-S(5); TSM_29_ATLA-S(5); TSM_5_ATLA-S(11); TSM_37_ATLA-S(5); TSM_31_ATLA-S(5); TSM_32_ATLA-S(5);

B. Comparação do Modelo GMDH com a RNA MLP

Para a melhor configuração do modelo GMDH, diversos experimentos foram conduzidos, observando os resultados mais próximos do real. A comparação entre o modelo GMDH e RNA-MLP foram realizados dentro da mesma partição de treino e teste. A melhor configuração do GMDH foi a composta com quinze camadas intermediárias, quatorze neurônios na camada intermediária e com o parâmetro chamado de *Selection Pressure* igual a 0,4. Esse parâmetro corresponde ao limite para determinar o número de neurônios adequados em cada camada.

Similarmente, a definição da configuração da RNA-MLP envolveu testes com diferentes valores para os seguintes parâmetros:

- Taxa de aprendizagem (η) = [0,01, 0,045, 0,1, 0,2];
- Número de camadas escondidas (NCE) = [1, 2, 3];
- Número de neurônios na camada escondida (NNC) = [5,10,15, 20, 25, 30];
- Função de transferência da camada escondida = tansig;
- Função de transferência da camada de saída = logsig;

Combinando todos os valores desses parâmetros, foram avaliadas 144 redes. A melhor topologia obtida foi a RNA-MLP com η , NCE e NNC iguais a 0,01, 1 e 10, respectivamente.

Os resultados da avaliação de desempenho (treinamento e teste) para ambos os modelos são apresentados na Tabela II. Nela, nota-se que o modelo treinado e testado com GMDH obteve resultados superiores de RMSE e R^2 em relação ao modelo com RNA-MLP.

TABELA II
COMPARAÇÃO GMDH E RNA-MLP.

Modelo	Treinamento		Teste	
	RMSE	R^2	RMSE	R^2
GMDH	34,8739	0,95783	30,81	0,96363
RNA MLP	41,7828	0,94062	68,2607	0,81712

Nas Fig. 6 e 7 são exibidos os resultados dos modelos regressivos obtidos para os dados de testes com GMDH e RNA-MLP. Pelas figuras pode ser depreendido que o GMDH alcançou $R^2 = 0,96363$, enquanto a RNA-MLP obteve $R^2 = 0,81712$ (conforme apresentado na Tabela II). Os resultados demonstram que o modelo usando GMDH conseguiu uma forte correlação entre os dados reais (target) e a saída do modelo (output), isto é, o modelo com GMDH conseguiu prever aproximadamente 96,36% dos dados reais enquanto o modelo RNA-MLP conseguiu prever apenas 81,71%. Além disso, a partir da Fig. 7 pode ser observado que os dados reais não ficaram bem ajustados ao modelo usando RNA-MLP (linha azul), diferentemente do modelo usando GMDH (Fig. 6), onde os dados reais ficaram bem próximos.

Na Fig. 8 são exibidas as previsões obtidas através do modelo RNA-MLP (parte superior) e do modelo GMDH (parte inferior). Nota-se que o modelo GMDH conseguiu uma generalização mais eficiente em relação à RNA-MLP, realizando uma previsão mais precisa para os dados de precipitação, capturando muito bem, sobretudo, os eventos extremos (picos e vales). Diferentemente disso, o modelo RNA-MLP não obteve bom desempenho, pois não conseguiu acompanhar o

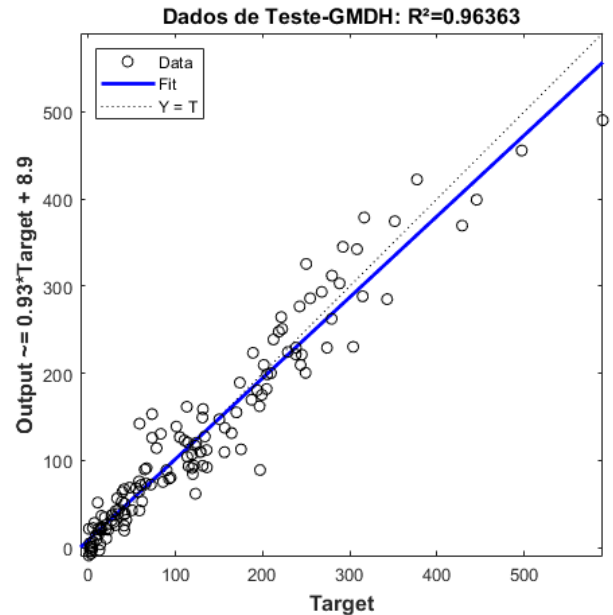


Fig. 6. Modelo regressivo de teste usando GMDH.

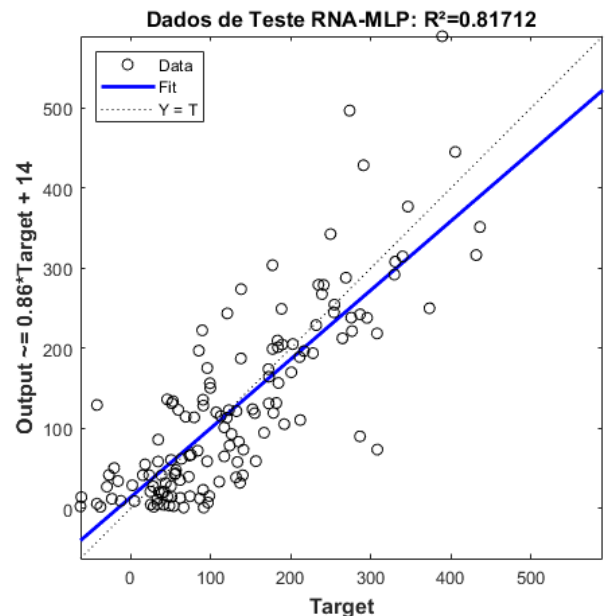


Fig. 7. Modelo regressivo de teste usando RNA-MLP.

comportamento da precipitação, mostrando grande deficiência na simulação dos extremos de chuva, sendo comprovada pelas previsões fora da faixa dos dados reais.

VI. CONCLUSÃO

O GMDH mostrou-se ser uma ferramenta robusta e eficiente para a predição do índice pluviométrico mensal em Marabá-PA, a partir da temperatura da superfície do mar. Algumas áreas oceânicas como o oceano Pacífico Sul na costa oeste da América do Sul, e diferentes áreas no oceano Atlântico

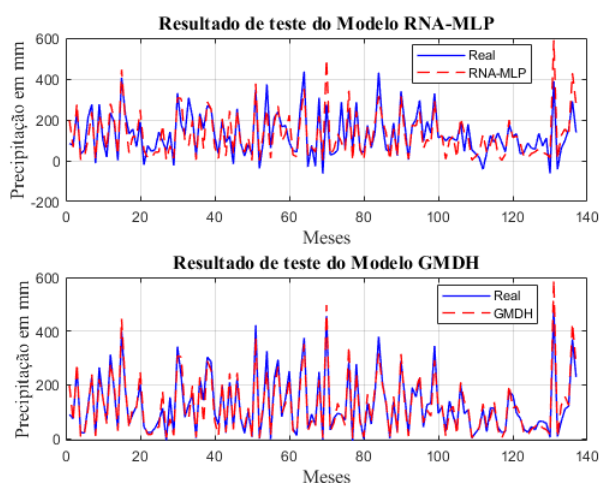


Fig. 8. Comparação entre dados previstos pelo GMDH (superior) e dados previstos pela RNA-MLP (inferior).

Tropical Sul mostraram-se mais influentes sobre a precipitação em Marabá, enquanto outras regiões como as áreas de Niño (no oceano Pacífico Equatorial) e no oceano Atlântico Tropical Norte apresentaram pouca ou nenhuma relevância na geração das chuvas em Marabá.

A modelagem da precipitação foi estudada usando comparativamente duas abordagens distintas: GMDH e RNA-MLP. A previsão da precipitação mensal usando abordagem GMDH alcançou melhor desempenho, capturando o comportamento da precipitação, especialmente os seus máximos e mínimos, representando de forma coerente os eventos climáticos extremos. A abordagem usando uma RNA-MLP convencional apresentou menor habilidade para generalização e, portanto, maior dispersão dos dados e menor eficiência na previsão.

A presente metodologia poderá ser expandida e aplicada para o desenvolvimento de modelos com abrangência em áreas maiores da Amazônia, de modo a melhor apoiar as ações de órgãos de governo e tomadores de decisão para o planejamento, a antecipação e mitigação dos impactos provenientes dos eventos adversos de precipitação. Além disso, o GMDH poderá ser comparado com outros métodos inteligentes.

REFERÊNCIAS

- [1] G. Fisch, J. Marengo, and C. A. Nobre, "Uma revisão geral sobre o clima da amazônia," *Acta amazônica*, vol. 28, pp. 101–101, 1998.
- [2] M. N. L. Lopes, "Climatologia regional da precipitação no estado do pará," *Revista Brasileira de Climatologia*, vol. 12, no. 1, 2013.
- [3] E. B. Santos, P. S. Lucio, and C. M. S. Silva, "Precipitation regionalization of the brazilian amazon," *Atmospheric Science Letters*, vol. 16, no. 3, pp. 185–192, 2015.
- [4] M. B. Dias, N. R. Galina, and C. N. Alves, "Mapeamento da suscetibilidade a inundações em áreas urbanas: estudo de caso no município de marabá, sudeste do estado do pará, região norte do brasil," *Revista Ibero-Americana de Ciências Ambientais*, vol. 12, no. 6, pp. 347–359, 2021.
- [5] M. A. Ayoub, A. Elhadi, D. Fatherlman, *et al.*, "A new correlation for accurate prediction of oil formation volume factor at the bubble point pressure using group method of data handling approach," *Journal of Petroleum Science and Engineering*, vol. 208, p. 109410, 2022.
- [6] K. Abhishek, A. Kumar, R. Ranjan, and S. Kumar, "A rainfall prediction model using artificial neural network," in *IEEE Control and System Graduate Research Colloquium*, pp. 82–87, IEEE, 2012.

- [7] M. E. Akiner, "Long-term rainfall information forecast by utilizing constrained amount of observation through artificial neural network approach," *Advances in Meteorology*, vol. 2021, 2021.
- [8] D. Shukla, V. Rajvir, and M. S. Patel, "Rainfall prediction using neural network," in *2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICSDNET)*, pp. 1–6, IEEE, 2018.
- [9] L. C. P. Velasco, R. P. Serquiña, M. S. A. A. Zamad, B. F. Juanico, and J. C. Lomoco, "Week-ahead rainfall forecasting using multilayer perceptron neural network," *Procedia Computer Science*, vol. 161, pp. 386–397, 2019.
- [10] A. Samad, V. Gautam, P. Jain, K. Sarkar, *et al.*, "An approach for rainfall prediction using long short term memory neural network," in *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, pp. 190–195, IEEE, 2020.
- [11] R. K. Grace and B. Suganya, "Machine learning based rainfall prediction," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 227–229, IEEE, 2020.
- [12] H. A. Y. Ahmed and S. W. A. Mohamed, "Rainfall prediction using multiple linear regressions model," in *2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCEEE)*, pp. 1–5, IEEE, 2020.
- [13] Y. Dash, S. Mishra, and B. Panigrahi, "Rainfall prediction of a maritime state (kerala), india using slfn and elm techniques," in *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT)*, pp. 1714–1718, IEEE, 2017.
- [14] I. Salehin, I. Talha, M. M. Hasan, *et al.*, "An artificial intelligence based rainfall prediction using lstm and neural network," in *2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*, pp. 5–8, IEEE, 2020.
- [15] C. Basha, N. Bhavana, P. Bhavya, and V. Sowmya, "Rainfall prediction using machine learning & deep learning techniques," in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp. 92–97, IEEE, 2020.
- [16] P. Nastos, K. P. Moustris, I. K. Larissi, and A. G. Paliatsos, "Rain intensity forecast using artificial neural networks in athens, greece," *Atmospheric Research*, vol. 119, pp. 153–160, 2013.
- [17] Z. liang Wang and H. hua Sheng, "Rainfall prediction using generalized regression neural network: case study zhengzhou," in *2010 International conference on computational and information sciences*, pp. 1265–1268, IEEE, 2010.
- [18] O. Dag and C. Yozgatligil, "Gmdh: An r package for short term forecasting via gmdh-type neural network algorithms," *R J.*, vol. 8, no. 1, p. 379, 2016.
- [19] M. Mrugalski, "An unscented kalman filter in designing dynamic gmdh neural networks for robust fault detection," *International Journal of Applied Mathematics and Computer Science*, vol. 23, no. 1, pp. 157–169, 2013.
- [20] C. K. S. Ma, L. Chen, and A. Wang, "Application of group method of data handling to stream-way transition," in *2009 International Joint Conference on Artificial Intelligence*, pp. 301–304, IEEE, 2009.
- [21] A. G. Ivakhnenko, "The group method of data handling—a rival of the method of stochastic approximation," *Soviet Automatic Control*, vol. 1, pp. 43–55, 1968.
- [22] P. Xie, M. Chen, S. Yang, *et al.*, "A gauge-based analysis of daily precipitation over east asia," *Journal of Hydrometeorology*, vol. 8, no. 3, pp. 607–626, 2007.
- [23] M.-H. Fun and M. T. Hagan, "Levenberg-marquardt training for modular networks," in *Proceedings of International Conference on Neural Networks (ICNN'96)*, vol. 1, pp. 468–473, IEEE, 1996.



Elton Rafael Alves is holds a degree in Computer Engineering from the Federal University of Pará and Doctor in Electrical Engineering (Energy System) from the Federal University of Pará. Is Adjunct Professor at the Federal University of the South and Southeast of Pará. His areas of interest are: computational intelligence, embedded systems and atmospheric discharges.



Márcio Nirlando Gomes Lopes Graduated in Agronomy and Meteorology, with a Master's degree in Environmental Sciences and a PhD in Electrical Engineering. He is currently a Science & Technology Analyst of the Management and Operational Center for the Amazon Protection System. His main research interests include climatology, hydrometeorology, weather radar and satellite remote sensing, lightning, computational intelligence, renewable energy and energy planning with emphasis on the Amazon region.



Fabricio Silva Sales is a computer engineering student at the Federal University of the South and Southeast of Pará.



Anderson M. Balieiro received the Ph.D. degree in Computer Science in 2015 at the Federal University of Pernambuco (UFPE), Brazil. Currently, he is a professor at the Center of Informatics (CIn) of the UFPE. He has worked in R&D projects funded by companies (e.g. Morotola Mobility and Ericsson) and governmental institution (e.g. Brazilian National Council for Scientific and Technological Development-CNPQ). He won the FET Best Paper Award in the 31th Wireless and Optical Communications Conference (WOCC). His research interests

Ultra-Reliable and Low Latency Communications, 5G/6G Networks and their key-enablers such as Cognitive Radio, Network Slicing, Network Function Virtualization, Multi-Access Edge Computing, and Software-Defined Networking as well as Machine Learning Applications.



Adônis F. R. Leal received a Master and Doctor degree in electrical engineering (Power Systems) from the Federal University of Pará, Belem, Para, Brazil in 2014 and 2018 respectively. From 2016 to 2017, he worked as a visiting researcher in the Department of Electrical and Computer Engineering at the University of Florida, Gainesville, FL, USA. Since 2018 he is an Adjunct Professor at the Federal University of Para. His main interests are development of embedded systems, lightning physics, lightning detection and location systems and

lightning occurrence in the Amazon.