

PREVISIBILIDADE DE EVENTOS HIDROLÓGICOS NO SEMIÁRIDO BRASILEIRO – PARTE II

*Francisco Jácome Sarmiento*¹

RESUMO – Na segunda parte do artigo investiga-se a viabilidade do uso prático dos resultados de previsões de eventos hidrológicos na bacia de drenagem do reservatório Orós, localizado no estado do Ceará (Brasil), lançando-se mão de redes neurais artificiais. As variáveis climáticas consideradas são o El Niño e o Dipolo do Atlântico. As variáveis dependentes a serem previstas com seis meses de antecedência são a precipitação média anual, convertida no Standardized Precipitation Index – SPI, sobre a bacia focada e o deflúvio médio anual na seção fluvial mencionada. Os resultados encontrados apresentam-se superiores comparativamente a modelos bem mais complexos, entretanto, também com limitações para uso prático na gestão dos recursos hídricos.

ABSTRACT– The second part of the article investigates the feasibility of the practical use of the results of forecasting hydrological events in the water basin of the Orós reservoir, located in the state of Ceará (Brazil), using artificial neural networks. The climatic variables considered are the El Niño and the Atlantic Dipole. The dependent variables to be predicted six months in advance are the mean annual precipitation, converted into the Standardized Precipitation Index – SPI, over the basin in question, and the mean annual runoff from the river section mentioned. The results found are superior compared to much more complex models, however, also with limitations for practical use in the management of water resources.

Palavras-Chave –SPI, Deflúvio médio anual, Redes Neurais Artificiais (RNA)

REDES NEURAIAS ARTIFICIAIS (RNA)

Costuma-se agrupar as RNA de acordo com sua arquitetura e sua dinâmica neural. A arquitetura de uma rede neural é representada pela sua estrutura, formada pelos neurônios artificiais (também referidos como nós ou unidades) distribuídos em camadas e pelas ligações que os integram. A dinâmica neural é caracterizada por propriedade através da quais a rede aprende, recupera, associa, classifica e compara novas informações com o conhecimento assimilado.

No que diz respeito à arquitetura, as RNA sempre apresentam (i) uma camada inicial por onde ingressa o conjunto de informações utilizados nas fases de treinamento e teste da rede e; (ii) uma camada de saída, por onde são apresentados os atributos resultantes da realização da tarefa demandada à rede. Adicionalmente, as RNA podem ou não apresentar uma ou mais camadas neurais intermediárias, designadas camadas ocultas ou escondidas, bem como outros tipos de nós, referidos como tendência (bias). Esse conjunto de neurônios dispostos em camadas organizadas da esquerda para a direita, desde a camada de entrada dos dados, até a última camada, por onde saem as informações de interesse, é interconectado por ligações que estabelecem um fluxo de informação da esquerda para direita (horizontal, nunca vertical). A primeira camada (camada de entrada ou retina), deverá ter tantos neurônios quantos sejam os atributos considerados para caracterizar os objetos que compõem o conjunto de dados. Então, se temos m atributos, teremos um igual número de neurônios na camada de entrada da rede.

O número de camadas escondidas, bem como o número de neurônios por cada uma dessas camadas irá depender do problema considerado. Por exemplo, em problemas envolvendo Engenharia de Recursos Hídricos, uma única camada de neurônios oculta ou no máximo duas pode dar conta das

1) Afiliação: Professor Titular, Universidade Federal da Paraíba, Campus 1, Centro de Tecnologia, João Pessoa – PB, jacomesarmiento@hotmail.com

tarefas mais comuns, sendo o número de neurônios em cada uma dessas camadas determinado por tentativa e erro, ou seja, testando-se diferentes arquiteturas para a rede.

É evidente que a performance da RNA para o atingimento dos objetivos de seu uso é influenciada pela adoção da arquitetura da rede. Embora a definição da arquitetura de uma RNA seja feita por tentativa e erro, em relação ao número de neurônios numa camada oculta n_0 , pode ser útil para iniciar o processo estimar seu valor em função do número de neurônios na camada de entrada (m) ou em função desse e do número de vetores do conjunto de treino (n_t):

$$n_0 \leq 2m + 1 \quad (1)$$

$$n_0 \leq \frac{n_t}{m + 1} \quad (2)$$

Há uma relação entre o número de neurônios em uma camada escondida e a taxa de aprendizado. Quando a taxa de aprendizado adotada é pequena, pode haver a necessidade de se aumentar o número de neurônios na(s) camada(s) oculta(s). A camada de saída poderá ser constituída por um ou mais neurônios, a depender do número de atributos que caracterizam o objeto resultante. Ou seja, haverá tantos neurônios de saída quantas forem as variáveis a serem determinadas, visto os termos “atributo” e “variável” se equivalerem na terminologia empregada na área da Inteligência artificial (IA).

Concernente à dinâmica neural, esta pode ser utilizada como elemento classificador das RNA quanto à direção do fluxo de informações e processamento em três modalidades: (i) RNA progressivas, nas quais a informação flui da primeira à última camada, da esquerda para a direita, com todas as unidades de cada camada conectadas com todas as unidades da camada subsequente e assim por diante, até o fluxo alcançar a camada de saída; (ii) RNA recorrentes, redes em que a informação flui através dos nós em ambas as direções, desde a camada de entrada a camada de saída e vice versa; (iii) RNA com conexões laterais, são redes que além das conexões descritas para as demais (conexões intercamadas) também dispõem de conexões entre os neurônios de uma mesma camada. Principalmente na engenharia de recursos hídricos, assim como em grande parte das aplicações práticas, as redes progressivas são as mais utilizadas e serão elas o nosso instrumento de modelagem na aplicação a ser apresentada.

Função de ativação

O tipo de função de ativação determina a resposta de um neurônio do sinal que o mesmo recebe através das conexões que lhes chegam. As funções de ativação geralmente empregadas podem ser lineares ou não lineares. Utilizaremos a função sigmoide que é contínua e possui derivadas contínuas. Sua equação é:

$$g(u) = \frac{1}{1 + e^{-u}} \quad (3)$$

Os valores produzidos pela sigmoide variam entre zero e um. Sua derivada é simples, o que facilita seu uso em processos de treinamento. Essa é uma das funções de ativação mais empregadas nas redes arquitetadas para resolver problemas de engenharia. Mas não é a única a apresentar as características de ser contínua e derivável.

Algoritmo de treinamento da RNA

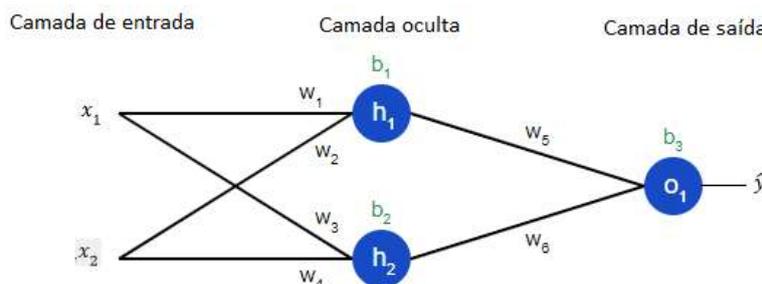
Em particular nas aplicações de engenharia, as funções de ativação mais utilizadas exigem que os dados de entrada, ou seja, os atributos ou componentes de cada vetor disponível, sejam padronizados, uma vez que essas funções de ativação trabalham com dados de entrada que variam, por exemplo, entre 0 e 1 no caso da função de ativação sigmoide e de -1 a 1 em se tratando da função de ativação tangente hiperbólica. Convém lembrar que a função sigmoide nunca assume seus valores extremos ($]0; 1[$). Por essa razão, deve-se usar um método mais adequado de padronização, como,

por exemplo, uma equação que compacte os dados em intervalo escolhido, como o faz a expressão abaixo:

$$z_i = 0,1 + \frac{0,8(x_i - x_{min})}{x_{max} - x_{min}} \quad (4)$$

Para o desenvolvimento do equacionamento básico que permite o treinamento da RNA iremos considerar de início a rede mostrada na Figura 1, formada por uma camada receptora dos dados, no caso, com dois atributos, uma camada escondida ou intermediária, contando com dois neurônios e uma camada de saída com um neurônio gerador da resposta.

Figura 1: Exemplo de rede neural com uma camada de entrada, uma camada escondida e uma camada de saída.



Matematicamente, uma rede neural pode ser entendida como uma composição de funções que promove a transformação da informação fornecida pela camada de entrada ao longo do fluxo estabelecido pelas ligações entre as camadas de neurônios que a compõem. Na notação adotada para a rede mostrada na Figura 1, a letra h informa tratar-se de um neurônio da única camada oculta da rede exemplo, enquanto o índice subscrito é o número do neurônio. A letra o designa neurônio da camada de saída (output), no caso, há apenas um neurônio de saída, que realiza a última transformação, fornecendo o valor estimado ou predito designado por \hat{y} . A transformação promovida quando a informação ingressada em x_1 e x_2 é processada pelo neurônio h_1 será ponderada pelos pesos das ligações que a ele chegam, no caso, w_1 e w_2 , acrescida da tendência (b_1) referente ao neurônio em questão.

$$z = w_1x_1 + w_2x_2 + b_1 \quad (5)$$

A resposta fornecida pelo primeiro neurônio h_1 (Figura 1) é transformada pela função de ativação da camada. Na explanação do equacionamento básico do algoritmo *back-propagation* utilizaremos em ambas as camadas de neurônios de processamento a função sigmoideal $g(z)$. Note que, como a camada escondida possui dois neurônios, duas serão as saídas fornecidas para o único neurônio da camada de saída. Assim, as saídas produzidas pelos dois neurônios integrantes da camada escondida serão:

$$h_1 = g(w_1x_1 + w_2x_2 + b_1) \quad (6)$$

$$h_2 = g(w_3x_1 + w_4x_2 + b_2) \quad (7)$$

Analogamente, na camada de saída produz-se a estimativa final combinando-se de maneira ponderada pelos pesos w_5 e w_6 os resultados advindos da camada imediatamente anterior, ou seja:

$$\hat{y} = g(w_5h_1 + w_6h_2 + b_3) \quad (8)$$

Substituindo-se os valores de h_1 e h_2 tem-se a composição de funções que expressa a rede neural exemplificada:

$$\hat{y} = g(w_5 g(w_1x_1 + w_2x_2 + b_1) + w_6 g(w_3x_1 + w_4x_2 + b_2) + b_3) \quad (9)$$

A ideia da dedução acima pode ser generalizada para redes que possuam um número maior de camadas, bem como um maior número de neurônios por camadas. Evidente que para um número maior de camadas e/ou de neurônios a composição de funções ganha proporções muito grandes.

Expressar a rede neural do exemplo em consideração propicia um melhor entendimento do funcionamento do algoritmo *back-propagation*, uma vez que sua base consiste no uso da Regra da Cadeia para determinar os vetores gradiente das variáveis envolvidas na atualização dos pesos. A atualização é feita em função do erro (E), cujos valores calculados são retropropagados, distribuindo-os adequadamente entre os pesos. No processo de treino, essa função traduz a diferença entre os valores calculados pela rede neural e o valor esperado (y) ou desejável (observação). A função que empregaremos para medir a magnitude dessas diferenças, contando com n objetos para a fase de treinamento, será a que traduz o erro quadrático médio, dada por:

$$E = \frac{1}{n} \sum_{j=1}^n (\hat{y}_j - y_j)^2 \quad (10)$$

Como todo algoritmo de otimização, o algoritmo *back-propagation* busca encontrar o ponto ótimo de uma função que, geralmente, expressa diferenças entre os valores calculados e observados. A busca pelo ponto de mínimo da função erro é orientada pelo gradiente de cada variável envolvida. Na função E acima, como o valor estimado \hat{y} do valor esperado y depende de $w_1, w_2, \dots, w_6, b_1, b_2, b_3$, a função erro é uma função multivariada:

$$E(w_1, w_2, w_3, w_4, w_5, w_6, b_1, b_2, b_3) \quad (11)$$

Em resumo, com a aplicação da Regra da Cadeia, a taxa de variação da função erro pode ser expressa como o produto de três parcelas:

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_1} \frac{\partial h_1}{\partial w_1} \quad (12)$$

De maneira análoga, podem ser obtidas as taxas de variação da função erro em relação às demais variáveis envolvidas. Importante ressaltar que a aplicação da regra da cadeia para obter a expressão acima onde aparece o peso w_1 , percorreu a rede neural desde a sua última camada até se chegar à ligação representada pelo peso em questão, ou seja, percorreu-se a rede no sentido contrário ao fluxo da informação, por essa razão o método de distribuição do erro a partir do cálculo dos gradientes dos pesos e tendências envolvidos é denominado *back-propagation* (Rumelhard e McClelland, 1986).

O gradiente da função erro tem como componentes suas taxas de variação em relação a cada uma das variáveis. Uma vez calculados os valores para cada objeto apresentado para o treinamento da rede, o erro quadrático médio calculado pela equação (10) deve ser retropropagado, distribuindo-o entre os pesos e as tendências (bias), considerando sua respectiva influência no valor desse erro, influência essa expressa pela derivada da função erro em relação ao parâmetro a ser atualizado. O algoritmo de otimização da função erro é geralmente referido como gradiente descendente. Para o caso do peso w_1 a expressão de atribuição, usando notação computacional (\leftarrow), seria:

$$w_1 \leftarrow w_1 - \alpha \frac{\partial E}{\partial w_1} \quad (13)$$

Onde α é a taxa de aprendizagem fixada para a rede. Ao multiplicar a taxa de variação da função erro em relação ao peso w_1 , estabelece-se a rapidez com que a rede é treinada. Note que, se essa taxa de variação é positiva, o novo valor de w_1 será menor, caso contrário, o novo valor de w_1 irá aumentar. Em ambos os casos o valor da função erro irá diminuir.

Inicialização da rede

Precede ao início do treinamento da rede a atribuição de valores aleatórios tanto para os pesos sinápticos, como para os limiares (*bias*). Como regra empírica para a amplitude dos valores aleatórios

pode-se utilizar o intervalo definido pelos seguintes quocientes $\frac{-3}{n}$ e $\frac{3}{n}$, onde n é o número de neurônios na camada de entrada. Valendo-se da linguagem de programação *Visual Basic*, como foi o caso no presente artigo, para o qual foi desenvolvido um aplicativo nessa linguagem de programação, pode-se utilizar conjuntamente as funções *Rnd* e *Randomize*. A função *Randomize* evita que a mesma sequência de números pseudoaleatórios seja repetida a cada chamada da função *Rnd* (Sarmiento, 2020).

Outra opção pode ser inicializar os pesos com valores constantes, recomendando-se apenas que os valores adotados sejam diferentes, para diferentes ligações que conectam as camadas da rede. Valores nulos ou unitários não são recomendados na inicialização dos pesos, pois no primeiro caso, anula-se o fluxo de informação entre os neurônios interligados. No segundo caso, a informação passa inalterada na direção do fluxo. Vale aqui lembrar que a informação que transita entre camadas é o somatório de um produto. Quanto ao limiar de ativação, geralmente se atribui um valor igual a -1. Não é problemático se atribuir qualquer outro valor de pequena monta. Na verdade, sequer o uso de um limiar é obrigatório.

Taxa de aprendizado

Qualquer que seja o algoritmo utilizado no treinamento das RNA, o que se busca sempre é maximizar a capacidade da rede atingir o objetivo estabelecido. Isso é alcançado mediante a correção sistemática dos pesos sinápticos. As chamadas regras de aprendizado que se prestam a isso têm em comum um parâmetro denominado taxa de aprendizado (α) que multiplica a parcela a ser acrescentada ou subtraída na correção do valor de determinado peso sináptico.

É importante a escolha de um valor adequado para α , pois esse parâmetro influi diretamente na correção dos pesos sinápticos e, conseqüentemente, no aprendizado da RNA. Esse parâmetro assume valores entre zero e um. Atribuir um valor próximo à unidade significa que a correção do peso sináptico irá ser feita acrescentando-se ou subtraindo-se a quase totalidade da correção calculada pela regra de aprendizado empregada. Uma escolha assim pode provocar oscilação no aprendizado levando a rede à extremos, sem que visite as zonas desejáveis do espaço de busca pelos pesos sinápticos ótimos. Por outro lado, uma taxa de aprendizado pequena impede o aprendizado da rede, além de exigir muitas iterações.

Métricas de avaliação

As métricas mais usuais voltadas à avaliação quantitativa da performance das RNA em reproduzir as saídas observadas (S_i^{Obs}) são os coeficientes de correlação e de determinação entre os valores observados e seus correspondentes calculados (S_i^{Cal}), o erro absoluto médio e o erro quadrático médio ou sua raiz quadrada. Além das duas primeiras métricas mencionadas utilizaremos também o coeficiente de eficiência de Nash-Suttcliffe (Nash e Suttcliffe, 1970), que varia de $-\infty$ a 1, denotando esse limite superior a perfeição do modelo na descrição da relação entre preditores e variável predita. Um valor nulo para essa métrica significa igualdade de precisão entre o resultado oferecido pelo modelo e a média das observações da variável dependente ($\overline{S^{Obs}}$). Já valores negativos implicam na inadequação do modelo, sendo a adoção da média da variável predita um estimador mais adequado para os valores futuros.

$$NASH = 1 - \frac{\sum_{i=1}^n (S_i^{Obs} - S_i^{Cal})^2}{\sum_{i=1}^n (S_i^{Obs} - \overline{S^{Obs}})^2}$$

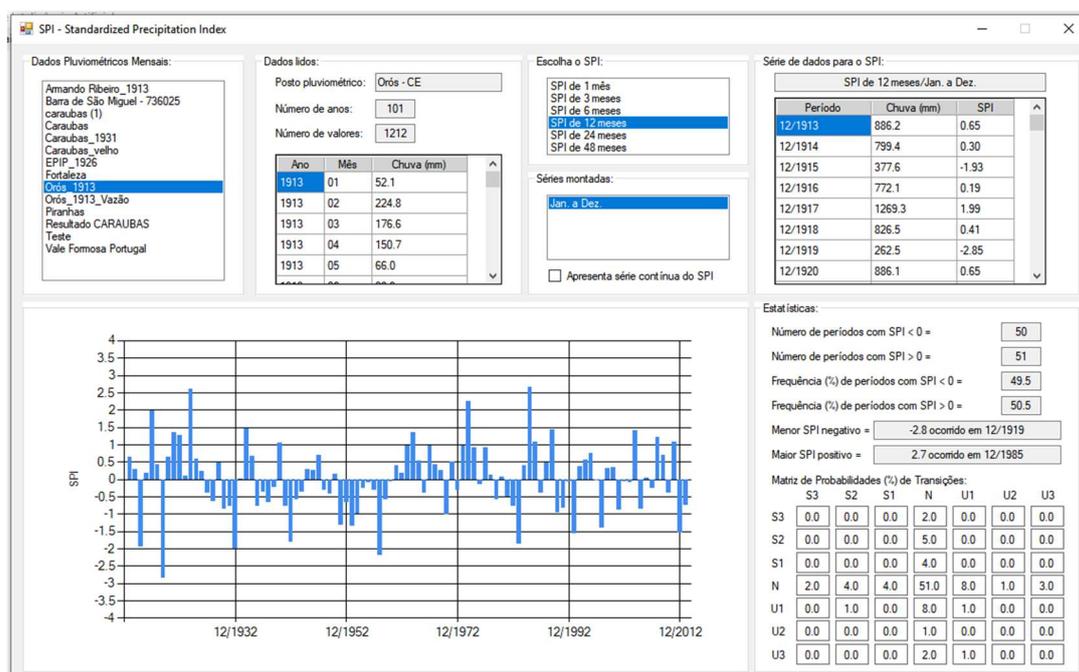
APLICAÇÃO E RESULTADOS

São inúmeras as aplicações das RNA na solução de problemas relacionados com a engenharia. Por exemplo, na década de 1990, Sarmiento (1996) estudou a relação de chuva-vazão para uma bacia localizada na região semiárida do Nordeste brasileiro, onde a modelagem desse fenômeno de

transformação se torna bastante complexa, entre outras razões, pelo caráter intermitente dos rios, onde o escoamento perdura apenas durante os períodos influenciados pelos meses mais chuvosos. O mesmo artigo investigou ainda a viabilidade das redes de perceptrons para a previsão das vazões médias diárias na bacia do rio Leine, localizaria na região central da Alemanha. Previsões de 1 e 2 dias foram realizadas para o posto Herrenhausen (área de drenagem > 3.500 km²) a partir da vazão atual no próprio posto e das vazões registradas nos postos de Popenburg (área de drenagem > 2.500 km²) e Greene (área de drenagem > 2.300 km²) localizados a montante.

Para a série de precipitações médias anuais na bacia de drenagem do açude Orós (Ceará-Brasil) foram calculados os correspondentes valores do SPI, valendo-se de software desenvolvido para esses fins, conforme mostra a Figura 2, abaixo. A série de dados tem início em janeiro de 1913 e se estende até dezembro de 2013 (101 anos). Para a modelagem incluindo variáveis climáticas, como a série disponível se inicia em 1950 e a previsão usa tais variáveis para obter os valores das variáveis hidrológicas no ano subsequente, conta-se com 63 vetores de dados. O conjunto de treinamento contou com 51 vetores e o de teste com 12 vetores, sendo esses os 12 últimos da série disponível, quando a rede foi treinada sem reamostragem aleatória. Observe na Figura 2 que a frequência de valores do SPI menores e maiores do que zero praticamente se igualam ao longo da série de dados. Um padrão perceptível no gráfico dos valores mostrados na mesma Figura 2 é o fato da ocorrência agrupada de anos secos, bem como de anos úmidos. Na parte inferior direita pode-se ver a matriz de transição envolvendo os sete estados tabelados para o SPI (N = Normal, U1 = moderadamente úmido, U2 = severamente úmido, U3 = extremamente úmido, S1 = moderadamente seco, U2 = severamente seco e U3 = extremamente seco). Praticamente metade das transições (51%) ocorrem entre anos com SPI classificado como normal.

Figura 2: SPI 12: série de precipitações médias anuais sobre a bacia controlada pelo reservatório Orós (Ceará-Brasil).



Se aplicarmos a mesma transformação matemática representada pelo SPI à série de deflúvios médios (mm) resultantes da série correspondente de precipitações utilizada na Figura 2 acima, obteremos valores de SPI dentro de uma escala de variação menor do que a visualizada na Figura 2, mas preservando o mesmo padrão de ocorrência agrupada de anos secos e anos úmidos.

Foram testadas diversas arquiteturas para a RNA. Seguindo-se a premissa de dispor de um modelo com o menor número de parâmetros possível, mas sem prejuízos para a qualidade dos resultados, optou-se por iniciar como uma rede bastante simples, com três neurônios na camada de

entrada, uma camada intermediária contando com seis neurônios e um neurônio na camada de saída, correspondendo à variável modelada.

A grande diversidade de resultados obtíveis provém principalmente das variações da arquitetura da RNA, da inicialização dos parâmetros (mesmo quando mantida a mesma arquitetura), do número de iteração e da taxa de aprendizagem, além, claro, das variáveis a serem consideradas como preditoras. As descrições do SPI da precipitação média anual, bem como dos deflúvios médios anuais afluentes para algumas variantes analisadas são mostradas nas Figuras 3 a 7, como reprodução das abas da tela de interface do aplicativo desenvolvido para o estudo. Aqui foi empregada uma rede com duas camadas escondidas: a primeira contando com 10 neurônios e a segunda com 14 neurônios. A taxa de aprendizagem foi de 5% com apenas 5.000 iterações. Os preditores do referido índice foram em número de três: Niño 3, TNA e TSA. Note a dispersão dos resultados exibidos no conjunto de treino, o que leva a um $NASH = 0,324$ (Figuras 3 e 4), inaceitável para quaisquer fins práticos. Um exemplo para o qual os vetores disponíveis foram reorganizados aleatoriamente (validação cruzada aleatória) tem os resultados obtidos mostrados na Figura 5.

CONCLUSÕES

Conforme visto, há uma infinidade de variantes baseadas em RNA que podem ser testadas na modelagem das variáveis aqui consideradas. A exploração dessa diversidade levou às seguintes conclusões, sempre consideradas à luz da utilidade prática dos resultados das previsões:

- 1) Em função da arquitetura, da taxa de aprendizagem, do número de iterações e da inicialização dos parâmetros, a RNA exibe sua condição de aproximador universal propiciando amplo espectro de ajustamentos para os conjuntos de treino e de teste, desde a grande dispersão na modelagem do primeiro (*underfitting*), até a condição de sobreajuste (*overfitting*, exemplo na Figura 8), o que implica na necessidade de uso de um balanço entre viés e variância para propiciar poder de generalização à rede quando processada com o conjunto de teste.
- 2) Caso não possam ser utilizadas técnicas de reamostragem ou validação cruzada não se tem nenhuma estimativa acerca do risco (no sentido empregado no âmbito da *Machine Learning*) de se utilizar o modelo para previsões futuras. A incerteza se torna ainda maior quando, no conjunto de treino, não constam eventos extremos raros, com ocorrência no conjunto de teste, sendo ainda mais difícil, se não impossível a previsão quando os preditores climáticos no ano em questão não apontam com tanta ênfase na direção esperada.
- 3) A disponibilidade de um modelo estocástico de geração de vazões (ou deflúvios) representa um meio para viabilizar a avaliação de risco de se admitir determinada arquitetura e respectivos preditores.
- 4) Em se tratando de modelagem de deflúvios médios ou vazões afluentes a um dado reservatório do semiárido setentrional, onde os coeficientes de variação das séries temporais são bastante elevados, a adoção de intervalos de confiança pode implicar em faixas de variação tão largas para a variável predita que inviabiliza o uso desses valores para fins práticos.
- 5) As variáveis preditoras aqui consideradas não têm poder explicativo suficiente para proporcionar previsões aceitáveis do ponto de vista prático da gestão dos recursos hídricos;
- 6) Sofisticar o modelo de previsão não nos parece uma forma relevante para sobrepujar o baixo poder dos preditores aqui considerados.
- 7) A transformação matemática propiciada pelo SPI, sem sombra de dúvidas conduz a previsões menos dispersas, entretanto os resultados obtidos permanecem inaceitáveis do ponto de vista prático.
- 8) Ao se optar pela utilização de pacotes computacionais disponíveis no mercado para fins de modelagem com RNA deve-se conhecer a fundo os mecanismos algorítmicos empregados

pelos mesmos, sob pena de contar com uma modelagem inteiramente a mercê dos autores do pacote, ou seja, sem que o modelista tenha qualquer poder de interferência nos processos envolvidos, limitando sua atuação à simples aceitação dos resultados. Na presente aplicação optou-se pelo desenvolvimento de um aplicativo específico para o desenvolvimento da tarefa.

Figura 3: Comparação entre valores observados e calculados para os conjuntos de treinamento e teste (SPI da precipitação média anual na bacia do Orós – Ceará - Brasil) – 5000 iterações – sem reamostragem aleatória- tempo de execução de cerca de 45 segundos.

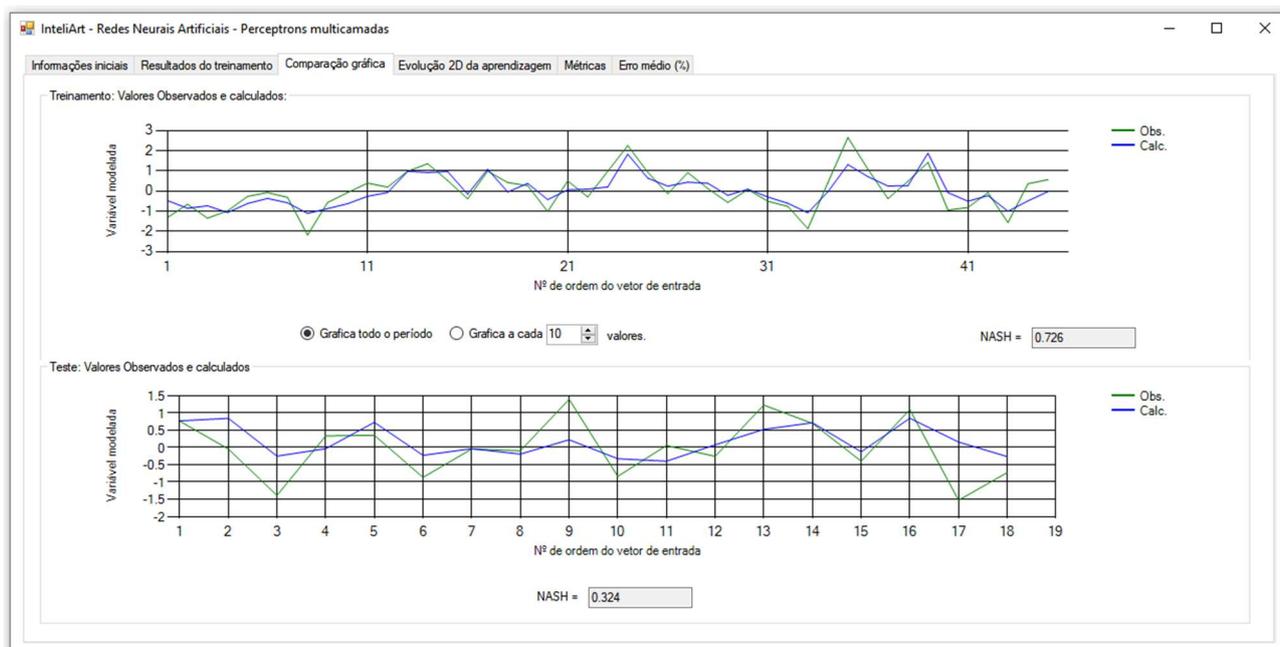


Figura 4: Vetores observados e calculados para os conjuntos de treinamento e teste (SPI da precipitação média anual na bacia do Orós – Ceará - Brasil) – 5000 iterações – sem reamostragem aleatória- tempo de execução de cerca de 45 segundos.

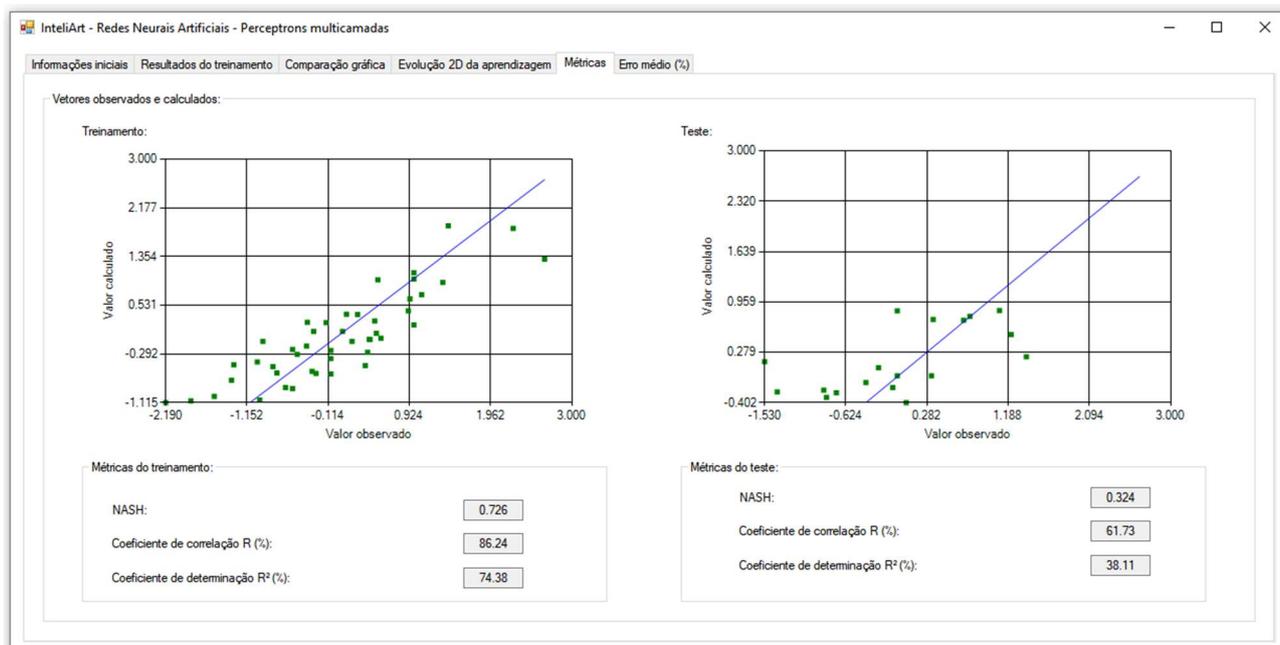


Figura 5: Vetores observados e calculados para os conjuntos de treinamento e teste (SPI dos deflúvios médios anuais) – 500 iterações – reamostragem aleatória- tempo de execução de cerca de 5 segundos.

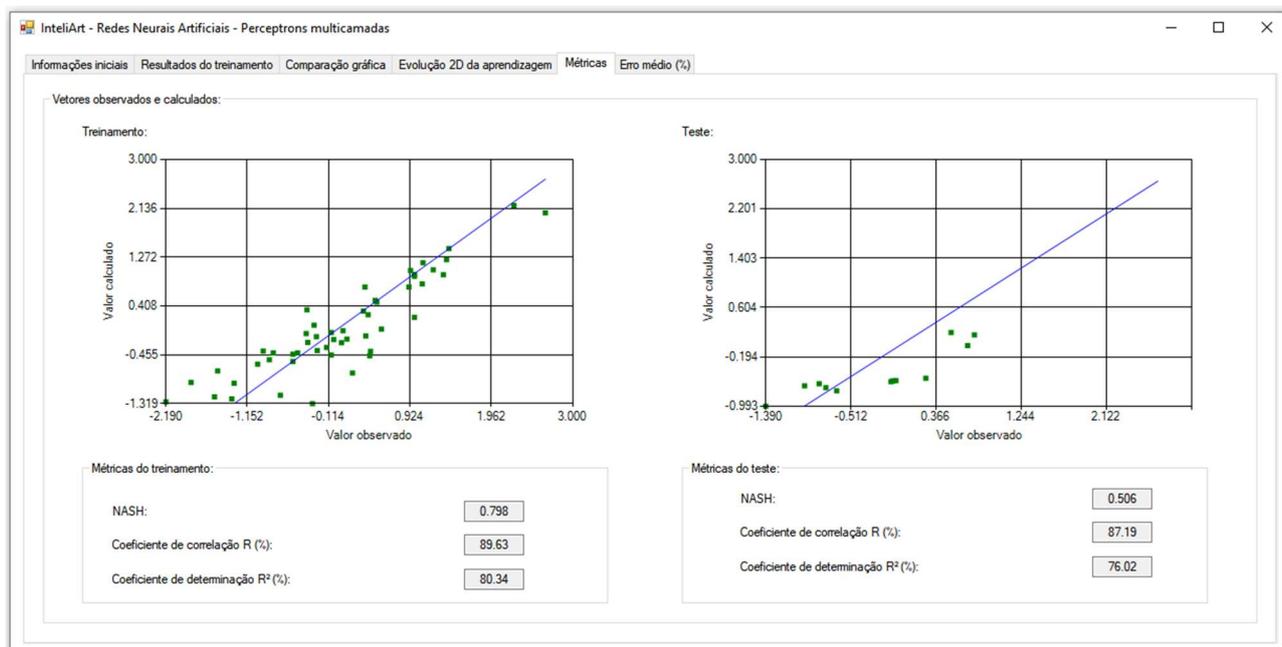


Figura 6: Comparação sequencial entre valores observados e calculados para os conjuntos de treinamento e teste (SPI do deflúvio médio anual na bacia do Orós – Ceará - Brasil) – 5000 iterações – sem reamostragem aleatória- tempo de execução de cerca de 45 segundos.

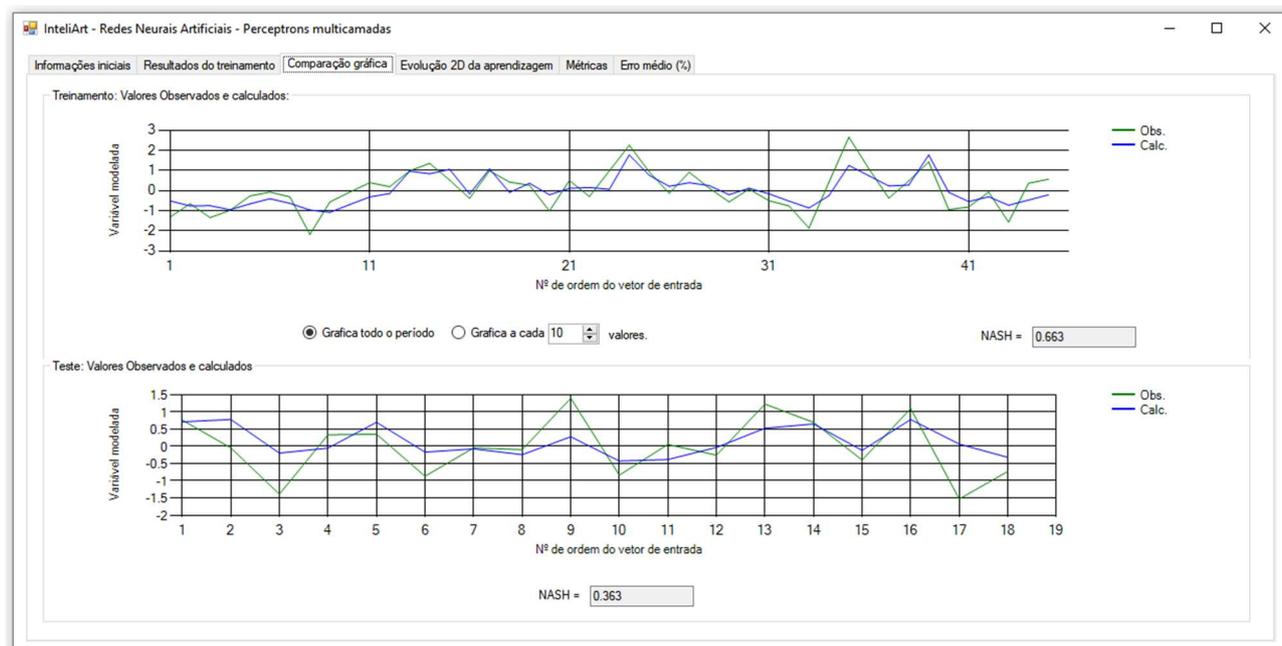


Figura 7: Vetores observados e calculados para os conjuntos de treinamento e teste (SPI do deflúvio médio anual na bacia do Orós – Ceará - Brasil) – 5000 iterações – tempo de execução de cerca de 45 segundos.

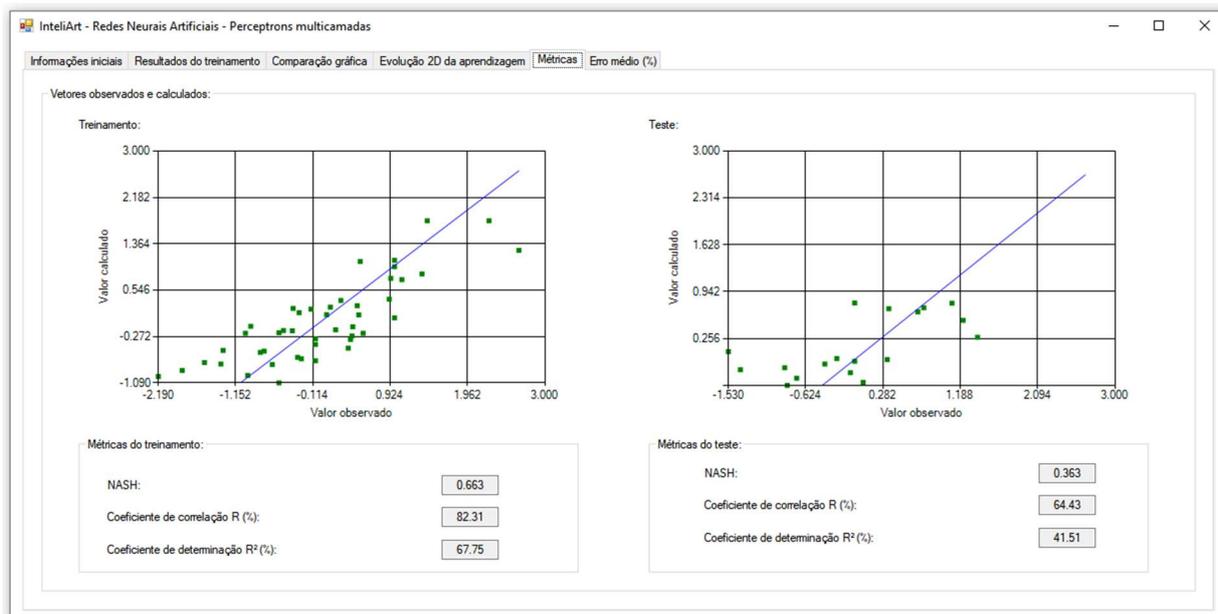
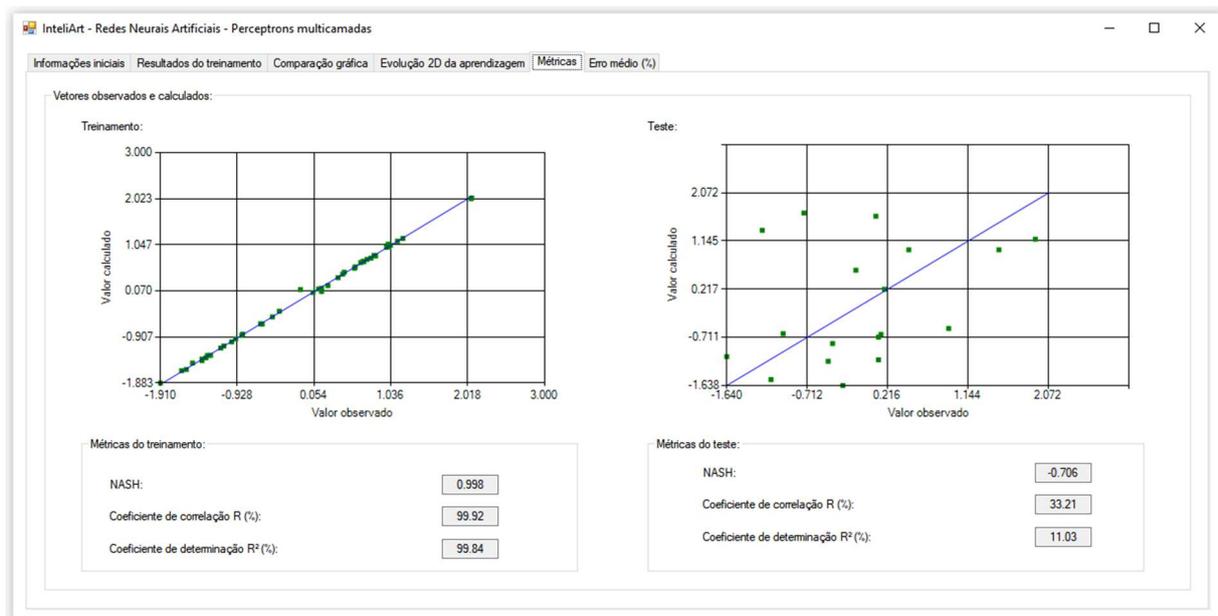


Figura 8: Exemplo de *overfitting*: SPI do deflúvio médio anual na bacia do Orós – Ceará - Brasil



REFERÊNCIAS

- RUMELHARD, D. F.; MCCLELLAND, J. L. “Parallel Distributed Processing Exploration in the Microstructure of Cognition”, Vol 1 Foundations, MIT Press, 1986.
- SARMENTO, F.J. “Modelagem de séries hidrológicas através de redes de neurônios”. RBRH – Revista Brasileira de Recursos Hídricos, Porto Alegre, v.1, n.2, p.19-31, jul/dez 1996.
- SARMENTO, F. J. (2020). *Engenharia Civil e Ambiental – Uma abordagem computacional – Volume 1 - Programação com Visual Basic*. Edição Kindle – Amazon, ISBN: 978-85-924138-2-8, 347 p.
- NASH J.E.; SUTCLIFFE J.V. River flow forecasting through conceptual models Part I - a discussion of principles. *Journal of Hydrology*, v. 10, 3 ed., p. 282-290, 1970.