

PREVISÃO DE SÉRIES DE VAZÕES COM REDES NEURAIS DE ESTADOS DE ECO

Hugo Valadares Siqueira^a, Levy Boccato^b, Romis Attux^b, Christiano Lyra Filho^a

^aDepartamento de Engenharia de Sistemas (DENSIS), Faculdade de Engenharia Elétrica e de Computação (FEEC),
Universidade Estadual de Campinas – UNICAMP
hugo@densis.fee.unicamp.br, chrlyra@densis.fee.unicamp.br

^bDepartamento de Engenharia da Computação e Automação Industrial (DCA), Faculdade de Engenharia Elétrica e de
Computação – FEEC, Universidade Estadual de Campinas – UNICAMP
lboccato@dca.fee.unicamp.br, attux@dca.fee.unicamp.br

Resumo – As redes neurais com estados de eco são alternativas promissoras em aplicações como previsão de séries temporais pelo fato de serem redes recorrentes dotadas de um método de treinamento simples e eficiente computacionalmente, baseado essencialmente numa metodologia de regressão linear. O caráter recorrente advém da existência de um reservatório de dinâmicas da rede – o qual é alimentado pelo histórico de amostras de entrada –, enquanto a camada de saída corresponde a um combinador linear ajustável. Recentemente, Boccato et al. [1] propuseram uma nova arquitetura para essa classe de redes, na qual se introduz o uso de um processo de compressão baseado em análise de componentes principais (PCA) aliado ao emprego de uma camada de saída estruturada como um filtro de Volterra. Vale frisar que essa proposta não compromete a simplicidade do processo de treinamento, uma vez que o filtro de Volterra é linear nos parâmetros [6]. Este trabalho faz um estudo comparativo, no contexto da previsão da série de vazões médias mensais do posto de FURNAS em dois períodos distintos, das arquiteturas de redes neurais de estado de eco mais solidamente estabelecidas - as propostas por Jaeger [6] e por Ozturk et al. [8] - e aquela decorrente da proposta de Boccato et al [1]. Os resultados mostram que as redes neurais de estados de eco são uma opção muito efetiva no âmbito de séries de vazões e, em particular, revelam que a nova arquitetura é capaz de trazer ganhos de desempenho relevantes.

Palavras Chaves – Redes neurais de estados de eco, previsão de séries de vazões, filtros de Volterra, PCA

1 Introdução

Redes neurais artificiais são muito utilizadas na previsão de séries temporais tanto na forma de arquiteturas *feedforward* quanto de arquiteturas recorrentes, as quais se caracterizam pela existência de laços de realimentação [5]. Uma vez que, no contexto do problema de previsão, é crucial explorar a relação temporal entre as amostras da série, a existência desses laços pode suscitar um elemento de memória capaz de trazer ganhos significativos de desempenho. Entretanto, o treinamento de redes recorrentes clássicas (RNNs, do inglês *recurrent neural networks*) é bastante complexo, envolvendo aspectos como estabilidade, dificuldade de estimação das derivadas da função custo e risco de convergência local.

Uma possibilidade de estabelecer uma solução de compromisso para o treinamento das redes recorrentes foi proposta por Jaeger [7], que estabelece uma arquitetura caracterizada pela existência de um reservatório de dinâmicas com parâmetros fixos – o qual funciona como uma espécie de camada intermediária recorrente – e uma camada de saída linear que é ajustada de maneira supervisionada. A essa proposta deu-se o nome de rede neural com estados de eco (ESNs, do inglês *echo state networks*). Estabelecido o novo paradigma, que terminou por contribuir para o estabelecimento do ramo mais amplo de computação por reservatórios [7], foram surgindo novas estratégias de projeto, dentre as quais destacamos a de Ozturk et al. [9], que explorou o princípio de Kautz de projeto de sistemas lineares para obter, na saída do reservatório, um repertório de dinâmicas mais diversificado, no sentido da entropia média.

Mais recentemente, Boccato et al. [1] propuseram uma nova abordagem para projeto da camada de saída de uma ESN, com o intuito de explorar o seu potencial não-linear no mapeamento entrada-saída. Esta proposta mantém a simplicidade no treinamento da rede, uma vez que a camada de saída é formada por um filtro de Volterra, que é linear com respeito aos parâmetros livres, mas tem a vantagem de vincular a etapa de treinamento supervisionado a um estágio não-linear da rede. Uma vez que a estrutura de Volterra pode levar a um aumento significativo no número de parâmetros a serem ajustados de acordo com a dimensão do reservatório de dinâmicas, os autores lançaram mão de um estágio de compressão baseado na estratégia de Análise de Componentes Principais (PCA). Isto permite a redução do número de estados de eco efetivamente utilizados, por meio da retirada de redundância, sem, idealmente, perda significativa de informação. Vale ressaltar que a proposta é aplicável a reservatórios projetados segundo qualquer estratégia.

As características das ESNs as qualificam como estratégias potencialmente atraentes para lidar com o problema de previsão de séries de vazões médias mensais, problema este que é de grande importância para países como o Brasil, que geram energia elétrica, em sua maior parte, por meio de usinas hidrelétricas. Isso foi percebido por Sacchi et al. [10], que, de maneira pioneira, fizeram uso de ESNs no âmbito do referido problema. Este trabalho faz uma comparação entre arquiteturas de ESN's,

trazendo duas inovações principais: 1) o estudo do desempenho da arquitetura clássica com a metodologia de projeto de reservatório de Ozturk et al. [9] e 2) o uso da nova arquitetura de Boccato et al., que apresentou bons resultados em outras tarefas de processamento de sinais [1]. Os cenários de estudos serão construídos a partir das séries de vazões médias mensais do posto de FURNAS, em dois períodos distintos.

O trabalho está organizado da seguinte forma: a Seção 2 apresenta as redes neurais de estados de eco segundo as abordagens clássicas de Jaeger e Ozturk et al.; a Seção 3 aborda as novas perspectivas apresentadas por Boccato et al., com uso de filtros de Volterra e PCA; os resultados experimentais são mostrados na Seção 4. A seção 5 apresenta as principais conclusões e discute perspectivas de desdobramentos do trabalho.

2 Redes Neurais de Estado de Eco

As redes neurais recorrentes constituem uma importante classe de ferramentas dentro da neurocomputação devido ao grande potencial que possuem para lidar com problemas de natureza dinâmica e temporal, uma vez que são estruturas dotadas de laços de realimentação, os quais contribuem para a formação de uma memória interna capaz de, em muitos casos, oferecer valioso auxílio no tratamento de séries temporais. Entretanto, as dificuldades associadas ao seu treinamento segundo abordagens clássicas, tais como a ameaça de instabilidade num processo de adaptação *online*, a complexidade do processo de estimação de derivadas e o risco de convergência local [4], motivaram a proposta, por parte de Jaeger [7], das chamadas redes neurais com estados de eco, descritas a seguir.

Uma ESN genérica é apresentada na figura 1, a qual possui M entradas, N neurônios na camada intermediária e L saídas:

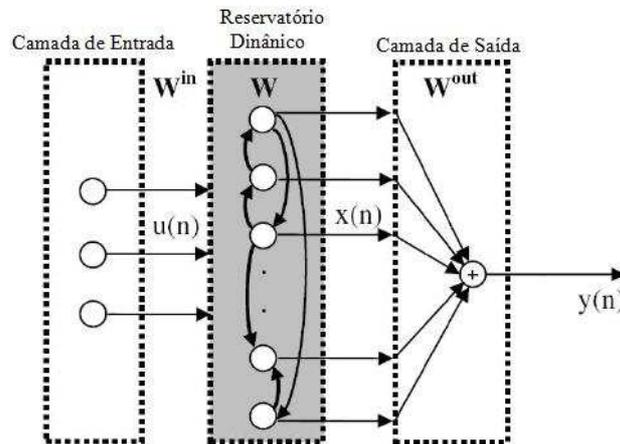


Figura 1– Rede de Estado de Eco

Definem-se as seguintes variáveis:

- i- $\mathbf{u}(n)$ – vetor de entradas da rede, de dimensão M ;
- ii- $\mathbf{x}(n)$ – vetor de estados de eco, de dimensão N ;
- iii- $\mathbf{y}(n)$ – saída da rede;
- iv- $\mathbf{d}(n)$ – saída desejada;
- v- L – número de neurônios na camada de saída;
- vi- \mathbf{W}^{in} – matriz de pesos da camada de entrada, de dimensão $N \times M$;
- vii- \mathbf{W} – matriz de recorrências da camada intermediária de dimensão $N \times N$;
- viii- \mathbf{W}^{out} – matriz dos pesos da camada de saída de dimensão $L \times N$.

O vetor de entrada da rede $\mathbf{u}(n) = [\mathbf{u}(n), \mathbf{u}(n-1), \dots]^T$, corresponde à ativação das entradas, transmitida para os neurônios por uma combinação linear. A matriz \mathbf{W}^{in} contém os coeficientes dessa combinação. A camada intermediária é denominada reservatório de dinâmicas e é composta por neurônios não-lineares totalmente interconectados por meio de laços de realimentação. O vetor $\mathbf{x}(n)$, de estados de eco, é composto das variáveis de estado dessa rede, sendo dinamicamente atualizado como a seguir:

$$\mathbf{x}(n+1) = f(\mathbf{W}^{in} \mathbf{u}(n+1) + \mathbf{W}\mathbf{x}(n)) \quad (1)$$

onde $f(\cdot) = (f_1(\cdot), f_2(\cdot), \dots, f_N(\cdot))$ reúne as funções de ativação de todos os neurônios do reservatório. A saída da rede, ou seja, o vetor $\mathbf{y}(n) = (y_1(n), y_2(n), \dots, y_N(n))^T$, é atualizado de acordo com a seguinte expressão:

$$\mathbf{y}(n+1) = f^{out}(\mathbf{W}^{out} \mathbf{x}(n+1)) \quad (2)$$

onde $f^{out}(\cdot) = (f_1^{out}(\cdot), f_2^{out}(\cdot), \dots, f_L^{out}(\cdot))$ reúne as funções de ativação dos neurônios da camada de saída.

Jaeger observou que nessas redes, sob certas circunstâncias, os estados $x(n)$ tornam-se assintoticamente independentes da condição inicial. Ou seja, a partir de duas condições iniciais distintas $x_1(0)$ e $x_2(0)$, sob a égide do mesmo sinal de entrada, há convergência para valores próximos. Dessa forma, apenas o efeito do histórico de entradas é considerado em longo prazo pelo reservatório dinâmico, e, assim, a rede possui estados de eco. Além disso, Jaeger demonstrou que a existência de estados de eco guarda estreita relação com propriedades da matriz de pesos do reservatório. Mais especificamente, no caso de uma rede sem realimentações da saída para a entrada, com não-linearidades do tipo tangente hiperbólica, e admitindo entrada nula, garante-se a existência de estados de eco se o maior valor singular da matriz \mathbf{W} estiver dentro do círculo de raio unitário.

Outra característica desejada para o reservatório é que ele seja capaz de gerar um repertório dinâmico tão rico quanto possível a fim de que haja uma boa aproximação do sinal desejado, o que significa que o ajuste dos pesos da matriz \mathbf{W} também deve considerar este aspecto. Na abordagem de Jaeger, estes pesos são definidos segundo a geração de valores aleatórios de acordo com distribuições definidas *a priori*, enquanto na abordagem de Ozturk et al., seus valores surgem de metodologias de posicionamento dos autovalores de \mathbf{W} inspiradas em formulações de sistemas dinâmicos. Em ambos os casos, o processo de treinamento da rede como um todo obedece à seguinte seqüência [7]:

- 1- crie uma matriz \mathbf{W} com raio espectral menor ou igual a 1;
- 2- defina arbitrariamente os pesos da matriz de entrada \mathbf{W}^{in} , mantendo as regras descritas em [7] que garantem a validade da propriedade de estados de eco;
- 3- ajuste o combinador linear da camada de saída usando a metodologia de mínimos quadrados ou algoritmos iterativos, como o LMS e o RLS [4].

Embora seja possível, neste trabalho não serão abordadas arquiteturas com realimentação da resposta de saída da rede.

A próxima seção discute a extensão das abordagens de Jaeger e Ozturk et al. proposta por Boccatto et al. [1].

3 Redes Neurais de Estado de Eco com Filtro de Volterra e PCA

As ESNs têm como principal característica o compromisso entre desempenho e simplicidade de atualização de pesos. Em outras palavras, trata-se de um tipo de rede que busca, em alguma medida, aliar o poder de processamento de uma estrutura recorrente à simplicidade de treinamento de uma estrutura *feedforward*. Um fator decisivo para que esse compromisso seja efetivo é o projeto do reservatório, mas há outras possibilidades, entre elas a introduzida por Boccatto et al. [1]: a inserção de elementos não-lineares na camada de saída, sem que esta perca a propriedade de ser linear com respeito aos parâmetros livres, o que permite, de forma direta, o uso de metodologias clássicas de mínimos quadrados para o seu treinamento.

A opção para gerar o sinal de saída da rede, nesse caso, foi pela utilização de um filtro de Volterra [5], caracterizado por uma combinação linear de termos polinomiais:

$$\begin{aligned} y(n) = & h_0 + \sum_{m_1=0}^{N-1} h_1(m_1) x_{m_1}(n) + \sum_{m_1=0}^{N-1} \sum_{m_2=0}^{N-1} h_2(m_1, m_2) x_{m_1}(n) x_{m_2}(n) \\ & + \sum_{m_1=0}^{N-1} \sum_{m_2=0}^{N-1} \sum_{m_3=0}^{N-1} h_3(m_1, m_2, m_3) x_{m_1}(n) x_{m_2}(n) x_{m_3}(n) + \dots \end{aligned} \quad (3)$$

na qual $x_m(n)$ representa o k -ésimo estado de eco no instante n , e $\mathbf{y}(n)$ a saída da rede. Observe que os coeficientes $h_i(m)$ estão linearmente relacionados com a saída, como desejado.

Apesar disto, surge um novo problema que precisa ser contornado: com o aumento do número de estados de eco, a

quantidade de coeficientes $h_i(m)$ tende a crescer rapidamente, o que pode numa aplicação prática dificultar o uso desta ferramenta. Por conta disso, faz-se uso de uma técnica de compressão bastante conhecida, a análise de componentes principais [5], com o intuito de reduzir a quantidade de sinais efetivamente transmitidos à camada de saída, o que, tendo em vista a redundância entre os estados de eco apontada, por exemplo, em [8], não deve comprometer significativamente o potencial de desempenho da rede.

Seja \mathbf{x} , de dimensão $N \times T_s$, a matriz que contém os valores dos estados de eco da rede considerando T_s padrões de entrada. Admitindo que a média do vetor de estados de eco para cada amostra de entrada tenha sido ajustada para zero, a matriz de covariância dos estados pode ser estimada por:

$$C = \frac{\mathbf{x}\mathbf{x}^T}{T_s} \quad (4)$$

Suponha agora que seja realizado o cálculo dos autovalores e autovetores dessa matriz, e que estes últimos estejam agrupados numa matriz \mathbf{V} ($N \times N$). Assim, os N_{PC} primeiros componentes principais são dados por:

$$S_i = \mathbf{V}_i^T \mathbf{x} \quad (5)$$

com $i=1, \dots, N_{PC}$.

A aplicação de PCA sobre os estados de eco, em conjunto com o uso do filtro de Volterra na camada de saída, é, portanto, capaz de dar à rede um efetivo potencial de processamento supervisionado não-linear sem perda da característica fundamental de que o processo de treinamento da ESN deve ser simples, já que os coeficientes da camada de saída podem ser ajustados segundo uma abordagem de mínimos quadrados. No contexto do problema de equalização de canais, Boccato et al. mostraram ganhos de desempenho em relação a ESNs clássicas que nos motivaram a estender a investigação ao problema de previsão de séries de vazões, que discutiremos a seguir.

4 Previsão de Séries de Vazões

4.1 Séries de Vazões e pré-processamento

Séries de vazões médias mensais são caracterizadas por serem não-estacionárias e possuem componentes sazonais, que refletem os períodos de chuvas e de seca nas proximidades dos rios brasileiros. Essas componentes podem afetar o desempenho de um preditor, seja ele linear ou não [3]. Em todo caso, a técnica estatística conhecida por padronização permite que essas componentes sejam retiradas e reinseridas ao fim do processo de previsão. A equação (6) descreve tal procedimento [3]:

$$z(n) = \frac{x(n) - \mu(n)}{\sigma(n)} \quad (6)$$

sendo que as observações $x(n)$ que formam a série original são transformadas em uma nova série padronizada $z(n)$, que possui média zero e desvio padrão unitário. As variáveis $\mu(n)$ e $\sigma(n)$ representam a média de cada mês e o respectivo desvio padrão, e são calculadas da seguinte forma:

$$\mu(n) = \frac{1}{N} \sum_{t=1}^N x(n) \quad (7)$$

$$\sigma(n) = \sqrt{\frac{1}{N} \sum_{t=1}^N (x(n) - \mu(n))^2} \quad (8)$$

Dessa forma, utilizaremos as ESNs para realizar a previsão da série $z(n)$ nos conjuntos de treinamento e teste. Ao final, a padronização é revertida, e a componente sazonal reinserida para fins de análise de desempenho.

4.2 Resultados Computacionais

Os estudos de caso apresentados nesta seção foram realizados para dois períodos da série histórica de vazões médias mensais da usina hidrelétrica de FURNAS, localizada no Rio Grande. Os dados estão disponíveis no sitio da ELETROBRÁS (www.eletrobras.gov.br). Foi utilizado o histórico de 1931 a 1990. Os conjuntos de testes selecionados foram de 1972 a 1976 - mediano e de média 882,63 m³/s - e de 1952 a 1956 - seco e com média 656,41 m³/s. Tais períodos compreendem cinco anos e 60 amostras, sendo bastante usuais neste tipo de aplicação [10]. Cenários distintos como estes favorecem a análise de

desempenho também de uma perspectiva de robustez devido à variedade de suas características. Os conjuntos de treinamento serão as amostras na janela de dados disponível, excluindo os cinco anos do conjunto de testes.

O primeiro passo na previsão foi aplicar o procedimento de padronização descrito na seção 4.1 a cada caso, e, em seguida, optou-se pela retirada da média de cada conjunto separadamente antes da entrada de dados no preditor. Em seguida, as quatro configurações de redes neurais foram treinadas com duas entradas *forward*, correspondentes a dois atrasos, sendo essa escolha embasada por testes preliminares visando um compromisso entre desempenho satisfatório e parcimônia. O horizonte de previsão foi sempre de um passo à frente. Outro fato relevante é que, também através de testes preliminares, percebeu-se que os melhores resultados eram obtidos utilizando apenas os termos de 1ª e 3ª ordens do filtro de Volterra, não tendo sido relevantes os termos de segunda ordem, os quais, por esta razão, não serão doravante considerados. O ajuste do reservatório seguiu as distribuições de probabilidade descritas em [6] para o caso da estratégia de Jaeger, e, para o caso da estratégia de Ozturk et al., usou-se um raio espectral igual a 0,8 [8]. No caso da rede proposta por Boccato et al., foram utilizadas apenas duas componentes principais. Além disso, o número de neurônios no reservatório foi, mais uma vez, definido por meio de ensaios preliminares para todas as redes.

As Tabelas 1 e 2 apresentam os desempenhos das ESNs em termos dos valores de erro quadrático médio (EQM) – média de 20 simulações independentes – para os conjuntos de treinamento e teste.

ESN	Número de neurônios	Trein real (x 10 ⁴)	Trein padron.	Teste real (x 10 ⁴)	Teste padron.
Jaeger	15	10.895	0.5730	6.9702	0.4341
Ozturk	25	10.958	0.5886	7.4581	0.4534
Boccato+Jaeger	30	10.998	0.5704	6.5583	0.3690
Boccato +Ozturk	80	10.904	0.5662	5.9899	0.3522

Tabela 1 – Erro quadrático médio e melhores desempenhos de previsão para série FURNAS 1952/1956

ESN	Número de neurônios	Trein real (x 10 ⁴)	Trein padron.	Teste real (x 10 ⁴)	Teste padron.
Jaeger	20	11.187	0.6068	7.2666	0.2957
Ozturk	12	10.960	0.5781	7.2728	0.2858
Boccato+Jaeger	30	11.106	0.5786	5.4881	0.2521
Boccato +Ozturk	70	11.003	0.5726	5.7796	0.2554

Tabela 2 – Erro quadrático médio e melhores desempenhos de previsão para série FURNAS 1972/1976

Os resultados sumarizados nas Tabelas 1 e 2 permitem algumas observações interessantes. Em primeiro lugar, é possível observar que, para o conjunto de treinamento no domínio padronizado, a rede proposta por Boccato et al. obteve um desempenho ligeiramente superior, o que revela que o uso de uma camada de saída não-linear foi capaz de agregar um potencial de exploração maior do sinal proveniente do reservatório. Entretanto, no domínio real da série, o quadro se inverte, o que se deve ao fato de, na abordagem escolhida para a padronização da série, meses com diferentes desvios padrões serem tratados de maneira similar.

Com respeito ao projeto do reservatório de dinâmicas, podemos observar que a abordagem proposta por Ozturk et al., ainda que baseada em uma estratégia interessante de maximização da entropia dos estados de eco, não necessariamente fornece melhores desempenhos. Por outro lado, tanto no domínio padronizado, quanto no domínio real, podemos verificar que a arquitetura proposta por Boccato et al. conduz a um desempenho significativamente superior considerando ambos períodos de teste (1952/1956 e 1972/1976). Isto significa que, embora esta rede não tenha atingido o melhor desempenho frente ao conjunto de treinamento, ela foi capaz de absorver de maneira equilibrada as características da série temporal, extraindo informações relevantes do período de treinamento, sem, contudo, comprometer sua capacidade de generalização, o que, por fim, possibilitou uma aproximação com maior precisão da série nos dois períodos distintos de teste.

É importante mencionar que a rede proposta por Boccato et al. alcançou desempenhos melhores para um número mais

elevado de neurônios, o que sugere uma necessidade de mais estados de eco para que o processo de compressão mantenha um nível mais alto de informações acerca da entrada. Ressaltamos também que a opção por um número bastante reduzido de componentes principais (duas) foi feita de maneira a impor à rede uma situação de parcimônia.

Estes resultados, portanto, evidenciam os benefícios obtidos com a introdução de uma camada de saída mais flexível, como a proposta por Boccato et al., além de enfatizarem que a abordagem através de ESN's para o problema de previsão de vazão é bastante promissora. A fim de ilustrar o desempenho de cada rede, as Figuras 2 a 5 apresentam os gráficos das séries de vazões reais e padronizadas, juntamente com os melhores resultados obtidos para cada método de previsão contido nas tabelas 1 e 2.

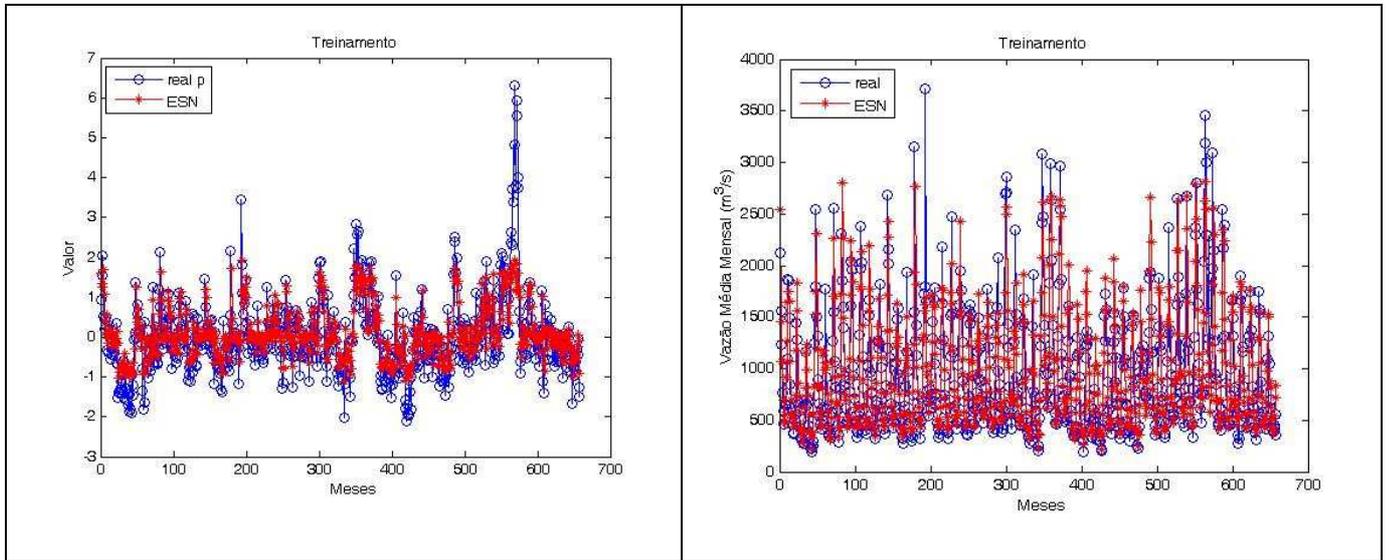


Figura 2 – Resultados das melhores previsões para série FURNAS 1952/1956 nos espaços padronizado e real - conjunto de treinamento

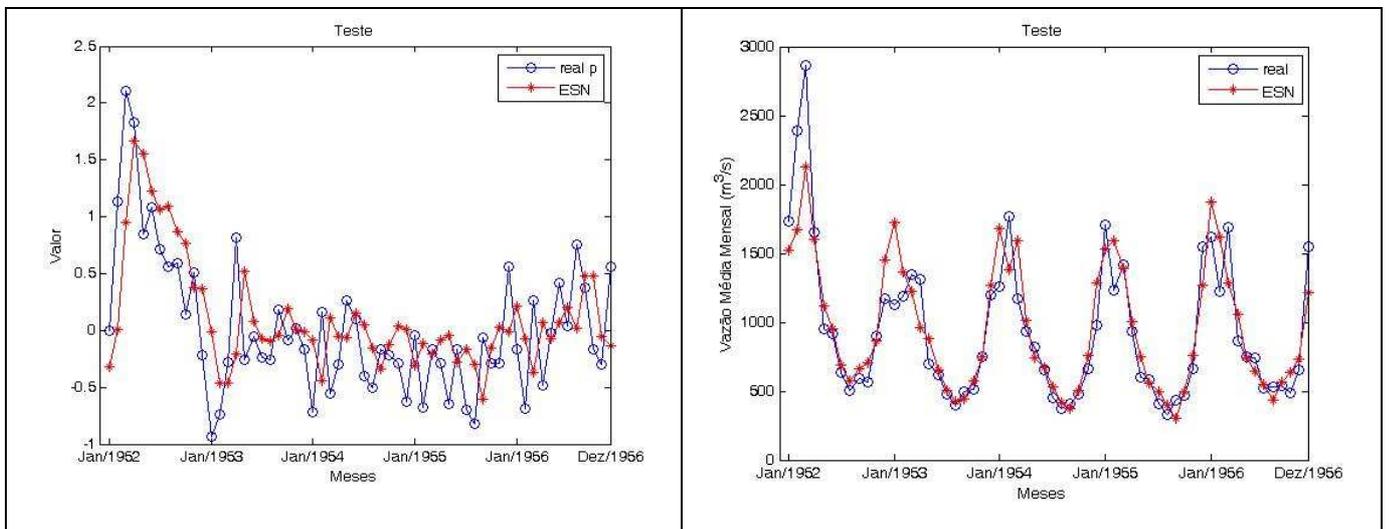


Figura 3 – Resultados das melhores previsões para série FURNAS 1952/1956 nos espaços padronizado e real - conjunto de teste

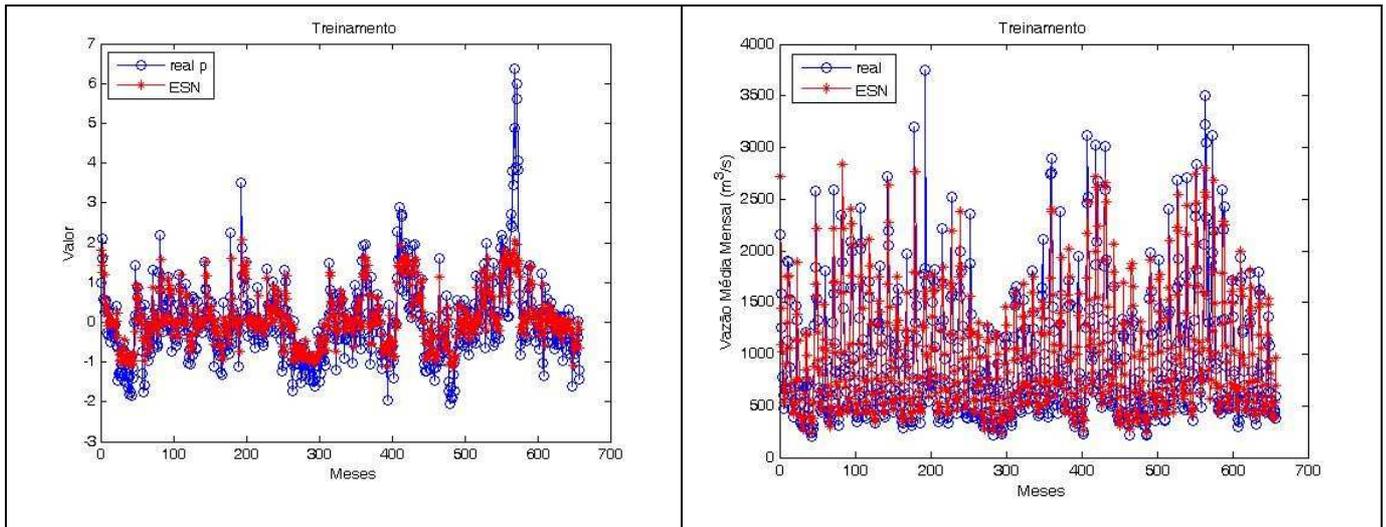


Figura 4 – Resultados das melhores previsões para série FURNAS 1972/1976 nos espaços padronizado e real - conjunto de treinamento

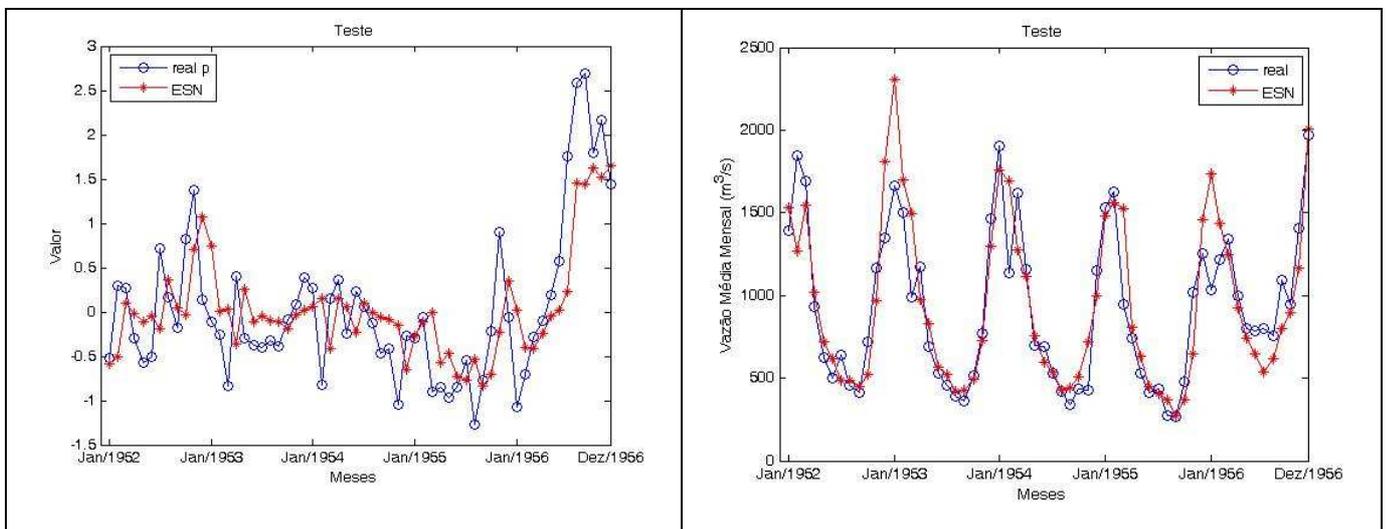


Figura 4 – Resultados das melhores previsões para série FURNAS 1972/1976 nos espaços padronizado e real - conjunto de teste

5 Conclusões

Este trabalho apresentou um estudo comparativo do desempenho de diferentes arquiteturas de neurais com estados de eco junto ao problema de previsão de séries de vazões: as redes de estados de eco propostas por Jaeger [7] e Ozturk et al.[9], além da proposta mais recente de Boccatto et al [1], caracterizada pela introdução de uma camada de saída não-linear e pelo emprego de uma técnica de compressão via análise de componentes principais. A principal motivação para o uso destas redes é a capacidade que possuem de preservar características desejáveis da existência de laços de realimentação no reservatório, mas com um processo de treinamento simples, o qual se reduz a um problema de regressão linear para o ajuste dos pesos de saída.

No âmbito do problema de previsão de dois períodos da série de vazões médias mensais da usina hidrelétrica de FURNAS – com horizonte de previsão de um passo à frente – foi possível verificar que a proposta de Boccatto et al alcançou um desempenho superior em relação ao das demais abordagens na etapa de teste, tanto no domínio padronizado quanto no real. Os resultados obtidos também indicaram que ESNs e, particularmente, a proposta caracterizada por uma camada de saída não-linear, são alternativas que merecem ser consideradas em problemas de previsão, e, mais particularmente, em problemas vinculados a séries hidrológicas.

Investigações adicionais sobre as ideias apresentadas no trabalho deverão estudar a possibilidade de utilizar um maior número de passos à frente nos processos de previsões. Deverão também considerar um conjunto maior de séries de vazões, envolvendo bacias hidrográficas com características distintas. Outro aspecto a ser investigado é o desenvolvimento de

padronizações que mantenham a proporcionalidade entre as séries padronizadas e as respostas em domínio real, já que há casos em que a resposta da rede é degradada no processo de reversão da padronização.

Agradecimentos

Este trabalho contou com o suporte das agências CAPES, CNPq e FAPESP.

Referências

- [1] Boccato, L., Lopes A., Attux, R., Von Zuben F. J., An Echo State Network Architecture Based on Volterra Filtering and PCA with Application to the Channel Equalization Problem, aceito para publicação em **IEEE Proceedings of International Joint Conference on Neural Networks, San Jose - CA, USA** (2011).
- [2] Box, G., Jenkins, G. e Reinsel, G. C. - Time Series Analysis, Forecasting and Control, **3rd ed, Holden Day, Oakland, California, EUA**, (1994).
- [3] Ballini, R., Análise e Previsão de Vazões Utilizando Séries Temporais, Redes Neurais e Redes Neurais Nebulosas, **Tese de Doutorado, FEEC-Unicamp, Brasil**, (2000).
- [4] Haykin, S., Adaptive Filter Theory, **Prentice Hall**, (1997).
- [5] Haykin, S., Neural Networks: A Comprehensive Foundation, 2nd Edition, **Prentice-Hall** (1999).
- [6] Hyvärinen A., Karhunen J. e Oja E., Independent Component Analysis, **New York: John Wiley & Sons**, (2001).
- [7] Jaeger, H., The echo state approach to analyzing and training recurrent neural networks, **Bremem: German National Research Center for Information Technology, Tech. Rep. GMD Report 148**, (2001).
- [8] Jaeger, H., Short term memory in echo state networks, **Bremem: German National Research Center for Information Technology, Tech. Rep. 152**, (2002).
- [9] Ozturk, M. C., Xu, D. e Príncipe J. C., Analysis and design of echo state networks, **Neural Computation**, vol. 19, pp. 111–138, (2007).
- [10] Sacchi, R., Ozturk, M. C., Príncipe, J. C., Carneiro, A. A. F. M., da Silva, I. N, Water Inflow Forecasting Using the Echo State Network: a Brazilian Case Study, **IEEE Proceedings of International Joint Conference on Neural Network, Orlando FL, USA**, (2007).
- [11] Siqueira, H. V., Previsão de Séries de Vazões com Redes Neurais Artificiais e Modelos Lineares Ajustados por Algoritmos Bio-Inspirados – **Dissertação de Mestrado, FEEC-UNICAMP, Brasil**, (2009).