UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO

HUGO VALADARES SIQUEIRA

Máquinas Desorganizadas para Previsão de Séries de Vazões

CAMPINAS
2013

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO

Hugo Valadares Siqueira

Máquinas Desorganizadas para Previsão de Séries de Vazões

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas para obtenção do título de Doutor em Engenharia Elétrica, na área de Energia Elétrica.

Orientador: Prof. Dr. Christiano Lyra Filho

Co-orientador: Prof. Dr. Romis Ribeiro de Faissol Attux

Este exemplar corresponde à versão final da tese defendida pelo aluno Hugo Valadares Siqueira, e orientada pelo Prof. Dr. Christiano Lyra Filho

_____

CAMPINAS

2013

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Elizangela Aparecida dos Santos Souza - CRB 8/8098

Si75m

Siqueira, Hugo Valadares, 1983-
Máquinas desorganizadas para previsão de séries de vazões / Hugo Valadares Siqueira. – Campinas, SP : [s.n.], 2013.

Orientador: Christiano Lyra Filho.
Coorientador: Romis Ribeiro de Faissol Attux.
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Previsão de vazões. 2. Pronósticos (Modelos Box Jenkins). 3. Redes neurais (Computação). 4. Seleção de variáveis. I. Lyra Filho, Christiano,1951-. II. Attux, Romis Ribeiro de Faissol,1978-. III. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Unorganized machines to seasonal streamflow series forecasting
Palavras-chave em inglês:
Streamflow forecasting
Predictions (Box-Jenkins Models)
Neural networks (Computer)
Variable selection
Área de concentração: Energia Elétrica
Titulação: Doutor em Engenharia Elétrica
Banca examinadora:
Christiano Lyra Filho [Orientador]
Reinaldo Castro Souza
André Carlos Ponce de Leon Ferreira de Carvalho
Fernando José Von Zuben
João Marcos Travassos Romano
Data de defesa: 28-11-2013
Programa de Pós-Graduação: Engenharia Elétrica

COMISSÃO JULGADORA - TESE DE DOUTORADO

Candidato: Hugo Valadares Siqueira

Data da Defesa: 28 de novembro de 2013

Título da Tese: "Máquinas Desorganizadas para Previsão de Séries de Vazões"

Prof. Dr. Christiano Lyra Filho (Presidente): _____

Prof. Dr. Reinaldo Castro Souza: _____

Prof. Dr. André Carlos Ponce de Leon Ferreira de Carvalho: _____

Prof. Dr. Fernando José Von Zuben: _____

Prof. Dr. João Marcos Travassos Romano: _____

Aos quatro pilares da minha vida:

Mãe, pai, irmão e amore.

Agradecimentos

Agradeço a Deus;

aos quatro;

aos mestres;

a família (nova e antiga);

aos avaliadores;

aos amigos;

aos colegas de trabalho e pesquisa;

a escola;

aos ilhenses;

a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

RESUMO

Este trabalho explora a possibilidade de aplicação de arquiteturas de redes neurais artificiais - redes neurais de estado de eco (ESN) e máquinas de aprendizado extremo (ELM) - aqui denominadas coletivamente por máquinas desorganizadas (MDs), para a previsão de séries de vazões. A previsão de vazões é uma das etapas fundamentais no planejamento da operação dos sistemas de energia elétrica com predominância hidráulica, como é o caso brasileiro.

Os modelos mais comumente utilizados para previsão de vazões pelo Setor Elétrico Brasileiro (SEB) são baseados na metodologia Box & Jenkins, lineares, sobretudo modelos periódicos auto-regressivos (PAR). Todavia, técnicas mais abrangentes, que alcancem melhores desempenhos, vêm sendo investigadas. Destacam-se as redes neurais artificiais, sobretudo arquiteturas do tipo *perceptron* de múltiplas camadas (MLP), muito conhecidas por serem aproximadores universais com elevada capacidade de aprendizado e mapeamento não-linear, características desejáveis para solução do problema em questão.

Por outro lado, as máquinas desorganizadas têm apresentado resultados promissores na previsão de séries temporais. Estes modelos têm um processo de treinamento simples, baseado em encontrar os coeficientes de um combinador linear; em particular, não precisam fazer ajuste dos pesos de sua camada intermediária, ao contrário das redes MLP. Por isso, este trabalho investigou as MDs do tipo ESN e ELM, versões recorrente e não-recorrente, respectivamente, para previsão de vazões médias mensais.

Serão avaliadas também três técnicas para retirada da componente sazonal característica destas séries – médias móveis, padronização e diferenças sazonal – além da exploração de técnicas de seleção de variáveis do tipo filtro e *wrapper*, no intuito de melhorar performance dos modelos preditores.

Na maioria dos casos estudados, os resultados obtidos pelas MDs na previsão das séries associadas a importantes usinas hidrelétricas brasileiras - Furnas, Emborcação e Sobradinho - em cenários com horizontes variados, mostraram-se de melhor qualidade do que os obtidos pelo modelo PAR e as redes neurais MLPs.

Palavras-chave: Previsão de séries de vazões médias mensais, seleção de variáveis, modelos Box & Jenkins, redes neurais, máquinas desorganizadas, máquinas de aprendizado extremo, redes de estado de eco.

ABSTRACT

This work explores the possibility of application of neural network architectures – echo state networks (ESN) and extreme learning machines (ELM) – collectively referred as unorganized machines (UMs), to seasonal streamflow series forecasting. Streamflow forecasting is one of the key steps in the planning of operation of power systems with hydraulic predominance, as in the Brazilian case.

The models most commonly used to streamflow prediction by the Brazilian Electric Sector are based on the Box & Jenkins methodology, with linear and especially periodic autoregressive models. However, more extensive techniques that achieve better performances have been investigated to this task. We highlight artificial neural networks, especially architectures such as multilayer perceptron (MLP), known to be universal approximators with high learning ability skills ability to perform nonlinear mapping, desirable characteristics for the solution of this problem.

On the other hand, unorganized machines have shown promising results in time series forecasting. These models have a simple training process, based on finding the coefficients of a linear combiner; they do not require adjustments in the weights of the hidden layer, which are necessary with MLP architecture. Therefore, this study investigated the UMs such as ESN and ELM, recurrent and nonrecurrent versions, respectively, to seasonal streamflow series forecasting.

Three techniques to remove the seasonal component of streamflow series will also be evaluated - moving averages, standardization and seasonal differences. In addition, In order to improve the performance of predictive models techniques for variable selection, such as filters and wrappers, will also be explored.

In the most cases, the computational results obtained by the UMs in streamflow series forecasting associated to important Brazilian hydroelectric plants - Furnas, Emborcação and Sobradinho - with scenarios including several horizons, presented better performance when compared to forecasting obtained with PAR models and MLPs.

Keywords: Monthly seasonal streamflow series forecasting, variable selection, Box & Jenkins models, neural networks, unorganized machines, extreme learning machines, echo state networks.

Sumário

Lista de Abreviaturas

AIC – Critério de Informação de Akaike

AR – Modelo auto-regressivo

BIC – Critério de Informação Bayesiano

CEPEL – Centro de Pesquisas de Energia Elétrica

ELM - Máquinas de Aprendizado Extremo (*Extreme Learning Machine*)

EPE – Empresa Brasileira de Pesquisas Energéticas

ESN - Rede Neural com Estados de Eco (*Echo State Network*)

FACP – Função de Autocorrelação Parcial

JAE-ESN - ESN com reservatório de Jaeger

JAE-ESN-ELM - ESN com reservatório de Jaeger e uma ELM como camada de saída

JAE-PV-ESN - ESN com reservatório de Jaeger e um filtro de Volterra como camada de saída

LMS – algoritmo *least mean square*

MAE –Erro Médio Absoluto (*Mean Absolute Error*)

MD - Máquinas Desorganizadas

MI – Critério de Informação Mútua

MLP - Perceptron de Múltiplas Camadas (*Multilayer Perceptron*)

MLT – Média de Longo Termo

MSE - Erro Quadrático Médio (*Mean Squared Error*)

ONS – Operador Nacional do Sistema Elétrico

OZT-ESN - ESN com reservatório de Ozturk et al.

OZT-ESN-ELM - ESN com reservatório de Ozturk et al. e uma ELM como camada de saída

OZT-PV-ESN - ESN com reservatório de Ozturk et al.e um filtro de Volterra como camada de saída

PAR – Modelo Periódico Auto-Regressivo

PARMA – Modelo Periódico Auto-Regressivo e Médias Móveis

PCA - Análise de Componentes Principais (*Principal Component Analysis*)

PDF - Função Densidade de Probabilidade (*Probability Density Function*)

RBF - Função de Base Radial (*Radial Basis Function*)

RC - Computação com Reservatórios (*Reservoir Computing*)

RNA – Rede Neural Artificial (*Artificial Neural Network*)

RNN - Rede Neural Recorrente (*Recurrent Neural Network*)

SEB – Sistema Elétrico Brasileiro

SCG – Método do Gradiente Conjugado Escalonado (*Scaled Conjugate Gradient*)

SCGM - Método do Gradiente Conjugado Escalonado Modificado (*Modified Scaled Conjugate Gradient*)

SVM - Máquina de Vetores Suporte (*Support Vector Machine*)

Lista de Figuras

Lista de Tabelas

Capítulo 1. Introdução

As vazões afluentes das usinas hidrelétricas são informações fundamentais para o planejamento energético brasileiro, pois aproximadamente 77% de o todo potencial energético disponível no Brasil provém deste tipo de unidade geradora (EPE 2013). Vazões afluentes formam uma série histórica que é comumente utilizada como entrada de modelos de simulação.

A água utilizada para fins de geração de energia é insumo renovável e não poluente características que, aliadas à grande quantidade de rios com possibilidade de serem aproveitados, fizeram com que o parque gerador brasileiro fosse formado predominantemente por usinas de base hidráulica. Embora hoje já estejam disponíveis outras fontes de energia renováveis, como eólica e energia solar, a participação dessas na matriz elétrica brasileira ainda é pequena (EPE 2013). Por disso, a alternativa imediata às hidrelétricas são as termelétricas, alimentadas por combustíveis fósseis, os quais são caros, não-renováveis e poluentes. A utilização de usinas térmicas deve ser reduzida, através do melhor uso possível de recursos hídricos, o que requer informações precisas sobre vazões afluentes a usinas hidrelétricas (Souza, Marcato, et al. 2010, Ballini 2000).

O objetivo geral deste trabalho é investigar as possibilidades de aperfeiçoar a previsão de vazões médias mensais de usinas hidrelétricas brasileiras, com a utilização de arquiteturas de redes neurais do tipo máquinas desorganizadas – Redes de Estado de Eco e Máquinas de Aprendizado Extremo. Um estudo comparativo do desempenho destas redes será elaborado, em relação a técnicas já estabelecidas para solução de problemas desse tipo, como os modelos periódicos auto-regressivos (PAR) e redes tipo *Perceptron de Múltiplas Camadas* (MLP).

1.2 Previsão de vazões

O processo de previsão de séries de vazões médias mensais é realizado através de um conjunto de etapas articuladas, sumarizadas na Figura 1.1 e discutidas a seguir.

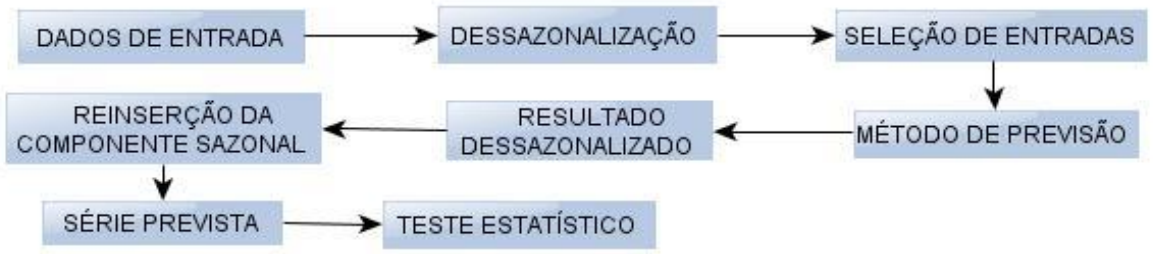


Figura 1.1 - Metodologia completa de previsão de vazões

Os dados de entrada, são vazões médias mensais observadas em um ponto de interesse, normalmente associados a usinas hidrelétricas do país. No Brasil, o Operador Nacional do Sistema Elétrico (ONS) publica essas informações em sua página da web para postos de observação associados as principais usinas hidrelétricas. Neste trabalho, serão usados os dados observados entre janeiro de 1931 e dezembro de 2010, o que totaliza 960 amostras mensais para cada posto, correspondente a esses 80 anos.

Uma característica destas séries é a presença de uma componente sazonal , consequência da variação de densidade pluviométrica ao longo do ano. A forma mais usual para se trabalhar com essas séries é fazer uma transformação nos dados de entrada, de forma que a componente sazonal seja retirada, deixando a série aproximadamente estacionária em relação à média e a variância (Souza, Marcato, et al. 2010).

Em seguida, processos de seleção das melhores entradas são aplicados às séries para que os melhores atrasos sejam escolhidos, com uma perspectiva de melhorar a performance dos modelos de previsão. Tais modelos, podem ser lineares, por exemplo modelos Box & Jenkins, ou não-lineares, como arquiteturas de redes neurais.

Após essas etapas preliminares de processamento, tendo também as entradas iniciais, o modelo preditor irá responder com um resultado no domínio dessazonalizado de forma que, para a análise da qualidade da previsão no espaço real, a componente sazonal é reinserida para cálculo das medidas que indicarão a performance de cada modelo.

Por fim, é comum a aplicação de um teste estatístico com intuito de verificar se as respostas dos modelos são significativamente diferentes. Neste trabalho a opção foi pelo teste de Friedman (Friedman 1937, Hollander e Wolfe 1999, Ballini 2000).

Serão utilizadas séries de vazões naturais mensais de usinas localizadas em regiões diferentes e com comportamento hidrológico distintos. Os conjuntos de testes serão formados por observações de 10 anos amostrados em 120 meses, em períodos que apresentaram diversidade em relação ao regime de chuvas. Além disso, os horizontes de previsão utilizados serão de 1, 3, 6 e 12 passos à frente.

No setor elétrico são utilizadas as vazões naturais afluentes aos reservatórios, que correspondem àquelas que ocorreriam em uma seção do rio, caso não houvesse a operação de reservatórios a montante, nem a vazão evaporada pelos lagos artificiais, nem mesmo retiradas de água consumidas com abastecimento e irrigação a montante. Em síntese, a vazão provida pela própria natureza.

1.3 Trabalhos Relacionados

Nesta seção serão discutidos trabalhos que possuem relação com as abordagens propostas para a previsão de vazões, investigadas nesta tese para resolver este problema como discussões sobre modelos preditores, formas de retirada da componente sazonal e técnicas de seleção de entradas. Os trabalhos apresentados estão organizados nos seguintes tópicos: dessazonalização, seleção de variáveis, previsão múltiplos passos à frente, modelos lineares de previsão, redes neurais artificiais MLP e *neuro-fuzzy*, máquinas de aprendizado extremo, redes neurais de estado de eco e Máquinas Desorganizadas .

- Dessazonalização

A geração de energia em uma usina hidrelétrica depende das vazões dos rios onde estão localizadas. Estas formam séries temporais que possuem comportamento periódico com relação à média, à variância e função de autocorrelação. (Andrade, et al. 2012).

As séries de vazões possuem componente sazonal que precisa ser retirada antes da aplicação de metodologias de previsão. Tal procedimento torna a série aproximadamente estacionária, condição necessária para aplicação de modelos lineares da metodologia Box & Jenkins (Box, Jenkins e Reinsel 2008). Em Nelson et al. (1999) e Zhang & Qi (2005) os proponentes mostraram que trabalhar com as séries no nível dessazonalizado melhora o desempenho em previsões com redes neurais. No trabalho de Siqueira (2009), este comportamento foi observado para previsões utilizando redes MLP.

- Seleção de variáveis

Um aspecto importante é avaliar a relação das variáveis mais relevantes para realização de previsões. Guyon & Elisseeff (2003) classificaram os métodos de seleção como filtros, *wrappers* (envoltório) e *embeddeds* (embutido).

Os filtros definem as entradas por meio de critérios de correlação linear ou não-linear que dependem apenas das relações de dependência estatística dos dados. O modelo de seleção por meio da função de autocorrelação parcial (FACP) (Box et al., 2008), linear, é uma alternativa comumente utilizada. O número máximo de entradas ou atrasos usualmente adotado é seis, uma vez que modelos de ordens superiores aumentam a possibilidade de coeficientes auto-regressivos negativos (Maceira e Damázio 2004). Este fato foi corroborado em Andrade et al. (2012), trabalho no qual os autores mostraram que a capacidade preditiva dos modelos PAR limitava-se a seis passos à frente. Para meses com alta variabilidade, o limitante superior do erro era alcançado para três passos apenas.

Souza & Oliveira (2009) propuseram o uso de técnicas de computação intensiva de reamostragem, ou *bootstrap*, que foram mais parcimoniosos que a FACP na identificação de modelos PAR. Stedinger (2001) sugeriu que atrasos de um sistema hidrológico não consecutivos selecionados pelo método de FACP não têm sentido físico, propondo a supressão destas entradas. Segundo este trabalho, algumas séries históricas como as hidrológicas sazonais possuem uma estrutura de autocorrelação relativa tanto ao tempo entre as observações quanto ao período observado.

Os critérios de informação mútua (MI) foram inicialmente desenvolvidos como alternativas as FACPs e sua concepção tem raiz na medição da quantidade de informação transmitida em um canal de comunicação (Cover e Thomas 1991). Esta métrica estatística capta relações não-lineares entre as variáveis (Wang e Lochovsky 2004). Luna et al. (2006) utilizaram o critério de informação mútua parcial (Sharma 2000) para previsão de séries temporais com redes neurais nebulosas, com o intuito de eliminar ao máximo possíveis redundâncias ao maximizar o critério MI (Bonnlander e Weigend 1994).

Os métodos de envoltório ou *wrappers* (Kohavi e John 1997) são caracterizados por um mecanismo de iteração entre o processo de seleção e o preditor os quais irão avaliar (dar uma nota de desempenho) para cada subconjunto de variáveis formado para solução do

problema. Estes métodos definem as entradas *a posteriori*, após os modelos serem ajustados e testados. Em Villanueva et al. (2006) os *wrappers* são utilizados como seleção de variáveis para previsão de séries temporais de diversas naturezas.

O critério de informação bayesiano - BIC (Schwarz 1978) e critério de informação de Akaike - AIC (Akaike 1978) são utilizados como função de avaliação de um *wrapper,* medindo a qualidade do conjunto das variáveis de entrada com base no princípio da parcimônia (Haber e Unbehauen 1990) e penalizando o excesso de entradas para um determinado modelo. Em McLeod (1994) foi proposta uma modificação no critério BIC para identificação da ordem de modelos PAR.

Em Luna (2007), o critério BIC foi utilizado no contexto de previsão de séries de temporais, enquanto Magalhães (2004) aplica o mesmo critério e o estimador de máxima verossimilhança (EMV) para identificação e estimação dos parâmetros de modelos PARMA de séries de vazões mensais. Santurio & Gomes (2011) fizeram um estudo comparativo no qual os critérios BIC e AIC são utilizados para definição da ordem de modelos auto-regressivos, enquanto usam os EMV e estimadores bayesianos exatos para a otimização dos coeficientes do modelo de previsão.

Métodos tipo embutido ou *embedded* são aqueles nos quais o treinamento também define quais as entradas serão selecionadas para um modelo, processo este realizado de maneira concomitante (Villanueva 2006). Esta técnica é muito característica (embora não exclusiva) de classificadores do tipo Máquinas de Vetores Suporte, SVM (Guyon e Elisseeff 2003).

- Previsão Múltiplos Passos à Frente

Em Atiya et al. (1999) foi feito um estudo comparativo entre técnicas de pré-processamento e previsão multi-passos, com os métodos direto e recursivo (Sorjamaa, et al. 2007), chegando a melhores resultados com o método direto.

No entanto, num trabalho específico para previsão de vazões, Marinho (2005) propôs a utilização de redes MLP para a previsão múltiplos passos a frente, com a comparação entre o método agregado e o direto, obtendo melhores resultados gerais com o primeiro para a série da usina hidrelétrica de Furnas.

- Modelos lineares de previsão

O modelo periódico auto-regressivo (PAR), da família Box & Jenkins, PAR, é empregado na modelagem das séries de vazões hidrológicas e/ou de energias naturais afluentes utilizadas no planejamento da operação energética no Brasil pelo Sistema Newave (CEPEL 2001a). A geração de cenários para o planejamento de médio prazo abrange um horizonte de até cinco anos à frente e é feita a partir do histórico de Energias Naturais Afluentes - ENAs dos subsistemas, utilizando o Modelo de Geração de Séries Sintéticas de Energias e Vazões – GEVAZP (CEPEL 2001a). Este é um modelo estocástico multivariado de geração de séries sintéticas de ENAs, que se baseia em modelos PAR e utiliza informações de até 6 (seis) meses anteriores.

Os modelos PARMA são munidos de realimentação e dispõem de mais informações para formação da resposta do preditor, sendo o uma generalização do modelo PA, que inclui realimentação. No entanto, segundo Rasmussen et al. (1996), estender os modelos PAR para os modelos auto-regressivos e médias móveis periódicos (PARMA) não é tarefa trivial, alegando ainda que os bons resultados dos modelos PAR acabam por não justificar essa alternativa para previsão de séries hidrológicas.

Por outro lado, Siqueira (2009) e Siqueira et al. (2010) mostraram que modelos auto-regressivos e médias móveis (ARMA) podem alcançar melhores resultados em comparação a modelos auto-regressivos (AR) na previsão de vazões com todo o histórico, para casos em que um único modelo é considerado para toda a série.

- Redes neurais artificiais MLP e *neuro-fuzzy*

Nas últimas décadas as redes neurais artificiais tornaram-se modelos conhecidos e amplamente aceitos em diversos tipos de aplicações e áreas do conhecimento. Serão discutidos alguns trabalhos históricos relevantes para a área no Capítulo 3.

Em um amplo estudo utilizando trabalhos da época sobre a aplicação de redes neurais à previsão de séries temporais, Hippert et al. (2001) mostraram que ainda havia ceticismo com relação aos verdadeiros ganhos de desempenho sob essa abordagem, embora diversos resultados em estudos já apontassem os benefícios da aplicação de RNAs à previsão de séries de curto prazo. Trabalhos como o de Tang et al. (1991) contribuíam para

esta impressão, já que em uma série específica as previsões de curto prazo foram melhores com a aplicação da metodologia Box & Jenkins, enquanto as previsões de longo prazo foram favoráveis às redes MLP. O mesmo vale para o trabalho de Sharda & Patil (1992), no qual RNAs tiveram desempenho equivalente a modelos lineares.

Nos dias atuais esta "desconfiança" já foi superada. Alguns estudos realizados nos anos 1990 contribuíram para essa maior aceitação, como Srinivasan et al. (1994), Kohzadi et al. (1996) e Tang & Fishwick (1993).

Lachtermacher & Fuller (1995) aplicaram uma metodologia híbrida para análise de séries temporais, com os modelos Box & Jenkins explorando as relações estatísticas dos dados de entrada de uma rede, como meio de diminuir o número de parâmetros a serem treinados numa rede neural. Esta metodologia obteve melhores resultados do que os modelos puros Box & Jenkins.

Francelin et al. (1996) utilizaram redes MLP para previsões de longo prazo e fizeram um estudo comparativo com modelos lineares Box & Jenkins, sendo que as MLPs alcançaram melhores resultados no tocante aos erros percentuais.

O trabalho de Mason et al. (1996) comparou as redes MLP com redes RBF (Redes de função base radial) na previsão de vazões, chegando a resultados computacionais semelhantes. Todavia, as redes RBF tiveram menor tempo de processamento.

Zealand et al. (1999) fizeram testes para previsão de vazões de curto prazo em bacias do Canadá com redes neurais artificiais chegando a bons resultados e apontando facilidades da proposta, a qual não depende de uma modelagem prévia da estrutura destas bacia para ser utilizada. A dificuldade, segundo os autores, é determinar a rede ótima para a tarefa de previsão, assim como a definição do número de padrões de entrada.

No trabalho de Kentel (2009) o autor investigou o impacto de diversos padrões de treinamento, número de iterações e valores iniciais de pesos na performance de redes neurais, concluindo que a inclusão de parâmetros, como dados e precipitação, contribui para melhoria dos resultados, a depender de qual entrada nova é atribuída a rede. Além disso, prever eventos extremos, como cheias e secas, degradava a resposta do preditor.

Chiang et al. (2004) utilizaram modelos de MLP treinados com o algoritmo *backpropagation* e gradiente conjugado, comparando com uma rede dinâmica recorrente. Os resultados mostraram que os modelos estáticos apresentaram menores erros de previsão, mas os modelos recorrentes conseguiram capturar os picos de vazões. Em testes com dados faltantes, a memória da rede recorrente foi fator que tornou os resultados significativamente melhores.

Os trabalhos de Ballini (2000), Magalhães (2004) e Luna (2007) apresentaram alternativas para melhoria na previsão de vazões com modelos não-lineares, dentre os quais redes neurais e sobretudo redes *neuro-fuzzy*. Estas últimas possuem natureza adaptativa, com a inserção de um sistema de inferência nebulosa quando o desempenho não é satisfatório. Os resultados comparativos com vários modelos de previsão, lineares ou não, apresentaram erros menores com este tipo de rede.

- Máquinas de Aprendizado Extremo (ELM)

Huang et al. (2004) propuseram as Máquinas de Aprendizado Extremo (*Extreme Learning Machines* - ELM): redes *feedforward* semelhante às MLP's, mas com a diferença de que a camada intermediária não é treinada, o que diminui bastante o custo computacional envolvido na tarefa de previsão. Neste trabalho pioneiro está a demonstração de que tais máquinas são aproximadores universais.

Investigações têm sido realizadas no sentido de melhorar o desempenho das ELMs como em Miche et al. (2010), no qual são adicionados passos extras para a classificação e regressão de dados. Os resultados foram comparados com os obtidos por redes MLP e *support vector machines* (SVM), mostrando que as ELMs são mais eficientes computacionalmente e, no caso em estudo, apresenta resultados próximos aos das SVMs.

No artigo de Deng et al. (2009), os autores mostram que *outliers* presentes nos dados de entrada podem trazer perdas de performance para a rede. Por isso, propuseram uma forma de regularização que melhorou a capacidade de generalização sem mudança na velocidade de convergência.

- Redes Neurais de Estado de Eco (ESN)

O trabalho de Jaeger (2001) propôs as redes neurais de estado de eco (do inglês *Echo State Networks,* ESN), modelo de rede neural recursiva que, assim como as ELMs, não realiza ajuste nos pesos dos neurônios de sua camada intermediária – aqui chamada de reservatório de dinâmicas - o que torna seu treinamento rápido e simples, baseado na solução de um problema de mínimos quadrados. Além disso, a proposta é eficiente em termos de erro final de previsão. No entanto, critérios prévios devem ser respeitados na formação do reservatório, para que seja válida a chamada *propriedade de estados de eco*, a qual garante a formação de memória da rede.

O trabalho de Jaeger inicia uma nova área de investigação chamada de *computação de reservatório*. Extensões desta rede, como outras ideias para criação do reservatório podem ser vistas em Ozturk et al. (2007) e Boccato et al. (2013). Em Verplancke et al. (2010), Wyffels et al. (2008) e Sheng et al. (2012) outras propostas de melhoria na performance e aplicações de ESNs são apresentadas.

No trabalho de Showkati et al. (2010) e de Deihimi & Showkati (2012), as ESNs são aplicadas na previsão de carga elétrica de curto prazo, considerando informações de cargas horárias e temperatura. Em Lin et al. (2009) os autores utilizaram as ESNs para previsão de preços de ações a curto prazo, com um estudo comparativo com MLPs, redes RBF e redes de Elman, chegando a melhores resultados com as redes ESNs.

As ESNs vêm sendo aplicadas com êxito em tarefas que vão desde equalização de canais (Boccato, Lopes, et al. 2011a) a previsão de séries caóticas (Jaeger 2001, Ozturk, Xu e Principe 2007), e até aplicações em estruturas de linguagem gramatical (Tong, et al. 2007).

No trabalho de Butcher et al. (2010), foi sugerida uma arquitetura híbrida na qual uma ELM opera como camada de saída de uma ESN, com o intuito de ampliar as não-linearidades dos estados de eco e aumentar a capacidade de mapeamento da rede. Novamente, não acontece o treinamento de camadas ocultas, o que mantém a simplicidade da proposta.

Em Boccato et al. (2011a), os autores apresentaram uma nova arquitetura de ESN na qual um filtro de Volterra (Mathews e Sicuranza 2001) serve como camada de saída da rede. Este filtro é não-linear, mas linear com relação aos parâmetros a serem determinados, o que também mantém a simplicidade característica das ESNs no processo de treinamento. Por conta do risco de um esforço computacional excessivo, a técnica de Análise de Componentes Principais (PCA) (Hyvärinen, Karhunen e Oja 2001) foi utilizada para redução do número de estados de eco efetivamente transmitidos à camada de saída, sem grandes perdas de informação.

- Máquinas Desorganizadas

O trabalho de Boccato et al. (2011b) evocou os trabalhos sobre aprendizado de máquina de Alan Turing (1968) para agrupar ELMs e ESNs sob o termo *máquinas desorganizadas*.

Sacchi et al. (2007) e Sacchi (2009) foram os pioneiros na aplicação de máquinas desorganizadas em séries de vazões mensais, com bons resultados na previsão frente a modelos conhecidos.

As investigações relatadas em Siqueira (2009), Siqueira et al. (2011, 2012a, 2012b, 2012c) aprofundaram a aplicação de diversas arquiteturas de máquinas desorganizadas na previsão de vazões afluentes de diversas usinas hidrelétricas brasileiras. Os resultados computacionais mostraram a viabilidade da proposta, apresentando ganhos de desempenho em termos de erro de previsão com diversos horizontes. Além disso, têm um custo computacional reduzido, mesmo em relação a modelos lineares.

1.4 Organização Geral do Documento

Este trabalho está divido em cinco capítulos. O Capítulo 2 faz uma apresentação dos principais conceitos associados a previsão de séries de vazões médias mensais. Serão discutidos métodos de análise de séries temporais e de dessazonalização de séries de vazões. As metodologias utilizadas para este fim serão padronização, médias móveis e diferenças sazonais. São também apresentados os modelos lineares de previsão de séries temporais: modelos Box & Jenkins, auto-regressivo (AR) e periódico auto-regressivo (PAR).

O Capítulo 3 apresenta as arquiteturas de redes neurais artificiais tipo MLP e as máquinas desorganizadas: máquinas de aprendizado extremo (ELMs) e rede de estado de eco (ESNs). No caso desta última, apresentaremos novas propostas de camadas de saída que podem melhorar sua performance. Um breve relato histórico e alguns modelos de neurônios artificiais são também discutidos.

O Capítulo 4 aborda alguns modelos de seleção de variáveis do tipo filtro (função de autocorrelação parcial e informação mútua), e o modelo *wrapper,* com função de avaliação baseada no mínimo erro quadrático médio e nos critérios BIC e AIC.

O Capítulo 5 apresenta estudos de casos baseados em séries históricas das usinas de Furnas, Sobradinho e Emborcação. Após a verificação dos valores dos erros de previsão, será aplicado o teste de Friedman para averiguarmos se os resultados são significativamente diferentes.

Conclusões e possíveis desdobramentos da investigação desenvolvida na tese são apresentadas no Capítulo 6.

Capítulo 2. Análise de Séries Temporais e Modelos Auto-Regressivos

A análise destas séries tem diversos objetivos, dentre os quais pode-se destacar (Morettin e Toloi 2006):

a) Modelar um fenômeno natural subjacente;
b) Investigar o mecanismo gerador da série;
c) Descrever seu comportamento ou fazer previsões de valores futuros;
d) Obter conclusões sobre o seu comportamento estatístico;
e) Avaliar se um determinado modelo é um adequado regressor dos dados disponíveis.

Em geral, a principal tarefa ao lidar-se com séries temporais é definir um modelo de representação do processo envolvido. Para isso, é necessário passar por diversas etapas, que incluem a observação dos dados, o pré-processamento e tratamento das amostras do sinal formado, a seleção das entradas mais significativas e a determinação dos coeficientes deste modelo. Por fim, é necessária uma análise de adequação baseada em alguma métrica de erro.

Modelos lineares de previsão são largamente utilizados por possuírem características desejáveis como facilidade de implementação e relativa simplicidade no tratamento matemático, além de bons e confiáveis resultados empíricos. A abordagem de Box e Jenkins (Box, Jenkins e Reinsel 2008) para séries estacionárias tem grande destaque, e, por isso, vamos discuti-la mais detalhadamente neste capítulo. O foco estará nos modelos auto-regressivos (AR) e periódicos auto-regressivos (PAR), uma vez que são estes os modelos adotados pelo Setor Elétrico Brasileiro (Souza, Marcato, et al. 2010).

A seguir, serão discutidos conceitos fundamentais para o estudo de séries temporais, como estacionariedade, covariância e correlação. Além disso, serão feitas considerações sobre as séries de vazões médias mensais, alvo das investigações deste trabalho, e sobre métodos para a retirada da componente sazonal destes dados.

2.1 Conceitos Básicos de Séries Temporais

Segundo Morettin & Toloi (2006), uma série temporal é um conjunto de observações ordenadas no tempo. Em Ehlers (2007), o autor ressalta que a característica

mais marcante deste tipo de dado é que observações consecutivas são, em geral, dependentes, o que é fundamental para a definição de um modelo representativo.

Pode-se definir formalmente uma série temporal da seguinte maneira (Box, Jenkins e Reinsel 2008):

Definição 2.1: Série temporal é um conjunto de observações $x_t, t \in T \subset \Re,$ de uma variável aleatória, geradas sequencialmente no tempo, sendo $\Re$ o conjunto dos números reais

Como dito, o estudo e a análise de séries temporais têm por meta entender seu comportamento estatístico e encontrar um modelo que se adeque ao padrão gerador. Dessa forma, busca-se um modelo suficientemente parcimonioso para a realização de regressões, previsões ou geração de séries sintéticas com a mesma probabilidade de ocorrência do fenômeno observado.

Os dados encontram-se no domínio do tempo e, por conta disso, procura-se a magnitude e as relações dos eventos que acontecem em cada instante. Tais fenômenos podem ser descritos por leis probabilísticas, sendo possível associar o estudo de séries temporais à teoria de processos estocásticos. Elementos dessa teoria serão discutidos na próxima seção.

2.2 Processos Estocásticos

De acordo com Morettin & Toloi (2006), a definição formal de um processo estocástico é:

Definição 2.2 – Seja Ω um conjunto arbitrário. Um processo estocástico é uma família $X = \{x_t, t \in \Omega\}$, tal que para cada $t \in \Omega$, x_t é uma variável aleatória (v. a.). O conjunto Ω pode ser contínuo ou discreto.

Segundo Box et al. (2008), há um mecanismo intrínseco subjacente a cada processo estocástico, sendo que uma série temporal é uma de suas possíveis realizações (ou trajetórias). Se assumirmos que $\Omega \in Z_+$, sendo este o conjunto dos números inteiros positivos, as observações que caracterizam uma série temporal podem ser descritas por uma v. a. $\{x_t, t \in \Omega\}$, com função de probabilidade conjunta $p(x_1, x_2, ..., x_N)$.

Um processo estocástico pode ser classificado de acordo com o comportamento estatístico associado ao seu desenvolvimento. Ele será estacionário se suas propriedades estatísticas não se alteram com o tempo. Classicamente, consideram-se duas formas de estacionariedade: estrita (forte) e ampla (ou fraca). A primeira classe é definida a seguir (Ehlers 2007)

Definição 2.3 – Um processo estocástico é *estritamente estacionário* (ou fortemente estacionário) se a distribuição de probabilidade conjunta das observações permanece a mesma sob translações no tempo, ou seja,

$$p(x_1, x_2, ..., x_N) = p(x_{1+k}, x_{2+k}, ..., x_{N+k})$$

para qualquer k e $N \geq 1$.

Isto quer dizer que a escolha da origem do eixo t não afeta as características estatísticas do processo. Em outras palavras, para um processo discreto ser estritamente estacionário, a distribuição conjunta de qualquer grupo de observações não deve ser afetada por translações no tempo, para qualquer k inteiro.

Algumas implicações se vinculam à ideia de estacionariedade estrita. Como a distribuição de x_t será a mesma para todo t, os dois primeiros momentos serão finitos, e média (ou valor esperado) $\mu_t = E[x_t]$ e variância $\sigma_t^2 = E[x_t - \mu_t]^2$ são constantes, sendo $E[.]$ o operador esperança matemática. Assim, as correlações entre termos adjacentes dependem apenas de seu espaçamento temporal, ou seja, as relações temporais entre as amostras são estáveis e podem ser generalizadas.

Todavia, em diversas situações práticas, é muito difícil a determinação prévia de todas as distribuições conjuntas de x_t, o que permitiria lidar com toda a estrutura de dependência estatística entre as amostras. Por isso, é usual utilizar uma forma de classificação menos restritiva de estacionariedade (Magalhães 2004, Ehlers 2007). Este conceito é definido a seguir:

Definição 2.4 – Um processo estocástico é *fracamente estacionário* (ou estacionário no sentido amplo, ou de segunda ordem) se e somente se sua média e variância são constantes e sua covariância depende apenas do intervalo k, ou seja:

- A média é constante: $E[x_t] = \mu_t = \mu$, para todo t;
- A variância é constante: $Var[x_t] = E[(x_t - \mu)^2] = \sigma_t^2 = \sigma^2$, para todo t;
- A covariância $Cov[x_t, x_{t+k}] = E[(x_t - \mu)(x_{t+k} - \mu)] = \gamma_k$ é uma função exclusiva de *k*.

Como é possível observar, nesse caso, nenhuma suposição é feita com respeito a comportamentos estatísticos de ordens superiores à segunda. Além disso, a média e a variância precisam ser finitas. Quando se lida com modelos lineares, tipicamente, se tem em mente essa definição menos restrita.

2.3 Ferramentas de Análise de Séries Temporais

Um processo estocástico estacionário pode ser analisado por meio de algumas ferramentas estatísticas derivadas de suas distribuições de probabilidade, as quais auxiliam na sua identificação e modelagem. Abordaremos métricas como a média, variância e autocorrelação, as quais são essenciais na caracterização de séries temporais estacionárias.

2.3.1 Média e Variância

Adotada a hipótese de estacionariedade de um processo estocástico, vale a implicação de que a distribuição de probabilidade $p(x_t)$ é a mesma para todo *t*: podemos então reescrevê-la como $p(x)$. Assim, a métrica que define o nível em torno do qual a série oscila é a média (ou valor esperado), e a amplitude desta oscilação é definida pela variância. Ambas são constantes e definidas pelas Equações (2.1) e (2.2), respectivamente (Box, Jenkins e Reinsel 2008):

$$\mu = E[x_t] = \int_{-\infty}^{+\infty} x p(x) dx \tag{2.1}$$

$$\sigma^2 = E[x_t - \mu] = \int_{-\infty}^{+\infty} (x - \mu)^2 \, p(x) dx \tag{2.2}$$

Como se pode observar, as integrais são definidas para uma distribuição de probabilidade constante e séries contínuas. Para o caso de séries discretas a média e a variância amostrais podem ser obtidas da seguinte forma

$$\hat{\mu} = \frac{1}{N} \sum_{t=1}^{N} x_t \tag{2.3}$$

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{t=1}^{N}(x_t - \hat{\mu})^2 \tag{2.4}$$

sendo *N* o número total de amostras disponíveis.

2.3.2 Autocovariância e Autocorrelação

Duas outras ferramentas importantes para identificação das propriedades de uma série temporal são as funções de autocovariância e a autocorrelação amostrais. Dada a hipótese de estacionariedade, tem-se que a distribuição conjunta $p(x_t, x_{t+k})$ é a mesma para todo *t*, $t+k$, separados por um intervalo constante *k.* Em particular, define-se a covariância entre x_t e x_{t+k}, que será chamada autocovariância de passo *k,* como

$$\gamma_k = E[(x_t - \mu)(x_{t+k} - \mu)] \tag{2.5}$$

sendo μ dada pela Equação (2.1). De forma similar, a autocorrelação é dada por

$$\rho_k = \frac{E[(x_t - \mu)(x_{t+k} - \mu)]}{\sqrt{E[(x_t - \mu)^2]E[(x_{t+k} - \mu)^2]}} = \frac{\gamma_k}{\sigma^2} \tag{2.6a}$$

na qual é importante notar que, como $\sigma^2 = \gamma_0$, tem-se:

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{2.6b}$$

Como, para um processo fracamente estacionário, a variância é igual para *t* e $t+k$, há a implicação que $\rho_0 = 1$ (Box, Jenkins e Reinsel 2008).

Em uma série temporal estacionária e discreta, os coeficientes de autocorrelação e autocovariância serão os valores destas funções para cada *k.* Estes podem ser estimados, sendo uma das alternativas o método dos momentos (Box, Jenkins e Reinsel 2008), no qual o termo γ_k é calculado como uma variável c_k do tipo:

$$c_k = \frac{1}{N-k}\sum_{t=1}^{N-k}(x_t - \hat{\mu})(x_{t+k} - \hat{\mu}),\ k=0, 1,\ldots, K \tag{2.7}$$

onde $\hat{\mu}$ é a média estimada pela Equação (2.3).

De forma similar, utilizando-se das expressões (2.6b) e (2.7), o coeficiente de autocorrelação amostral r_k é:

$$r_k = \frac{c_k}{c_0} \tag{2.8}$$

Verifica-se que r_k é calculado para cada $k = 0, 1,\ldots, K,$ e este valor será dependente do número de observações utilizadas para estimar c_k. Segundo Box et al. (2008), deve-se utilizar K menor que $N/4$ quando $N \cong 100$.

Um gráfico com os k primeiros coeficientes de autocorrelação é chamado correlograma, e pode ser útil na identificação de características de uma série temporal (Ehlers 2007). Este gráfico determina as conhecidas funções de autocovariância $\{\gamma_k\}$ e autocorrelação $\{\rho_k\}$ de um processo estocástico. Estas funções são adimensionais, independentes da escala de medida e obedecem a algumas propriedades:

a) As funções $\{\gamma_k\}$ e $\{\rho_k\}$ são simétricas, já que as covariâncias entre $[x_t, x_{t+k}]$ e $[x_{t-k}, x_t]$ são iguais;
b) Os valores dos coeficientes de correlação respeitem o intervalo $-1 \le \rho_k \le 1$;
c) A estrutura da autocovariância é única para cada processo estocástico ou série temporal, mas processos distintos podem apresentar a mesma função de autocorrelação. Isto pode dificultar o processo de identificação via correlograma;
d) Um processo estacionário normal é completamente caracterizado por sua média e função de autocorrelação.

Após a discussão de características e definições sobre processos estocásticos e séries temporais em geral, passaremos às séries de vazões mensais.

2.4 Séries de Vazões

Como discutido no Capítulo 1, as séries de vazões de usinas hidrelétricas são parte essencial no planejamento da operação do Sistema Elétrico Brasileiro. Neste trabalho, serão consideradas as vazões médias mensais, com valores observados no tempo e denotados por $[x_1, x_2, ..., x_N]$, sendo x_t a observação no instante $t = 1,2,\ldots, N.$ A Figura 2.1 mostra a série

histórica de Furnas, uma das mais frequentemente utilizadas no caso brasileiro em estudos desta natureza (Ballini 2000, Magalhães 2004, Sacchi 2009). A série possui valores compreendidos entre janeiro de 1931 e dezembro de 2010, totalizando 960 amostras. A média e o desvio padrão estimados são $\hat{\mu} = 926{,}6177$ m³/s e $\hat{\sigma} = 613{,}1671$ m³/s.

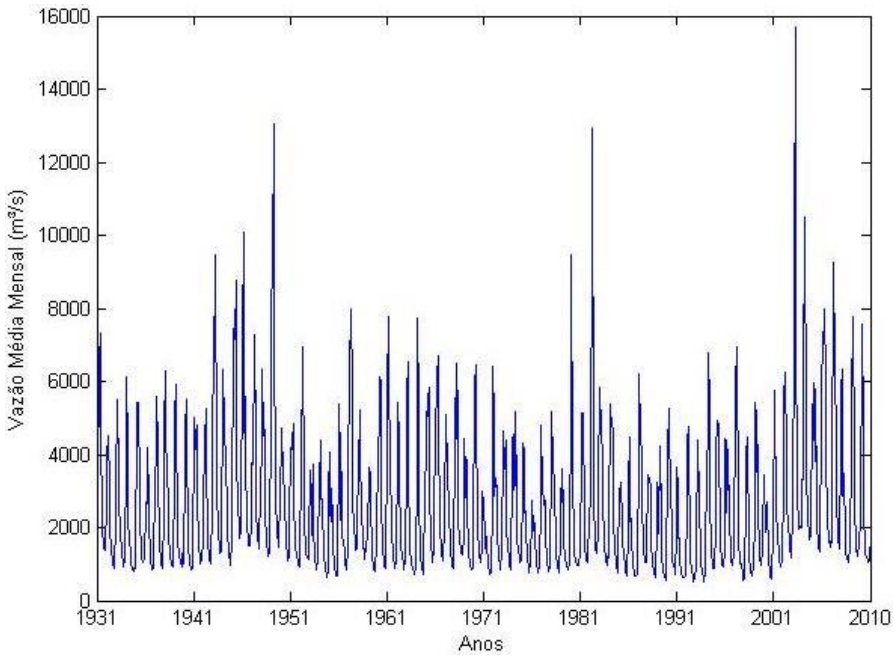


Figura 2.1 - Série de Vazões Médias Mensais da Usina de Furnas

Séries temporais usualmente são analisadas por meio da sua decomposição em quatro componentes (Morettin e Toloi 2006, Bouzada 2012):

i) tendência (τ_t) - movimento temporal de longo prazo que caracteriza a mudança no nível médio da série;

ii) componente sazonal (s_t) - movimento temporal periódico que ocorre em períodos infra anuais, que, para os fenômenos naturais, são decorrentes de características meteorológicas;

iii) componente cíclica (ω_t) - movimentos temporais oscilatórios recorrentes, mas sem periodicidade específica eu regularidade que permita serem deterministicamente previsíveis;

iv) componente aleatória (ou ruído) (a_t) - movimentos temporais aleatórios de natureza imprevisível.

Como é possível observar na Figura 2.1, há uma oscilação periódica característica que mostra uma variação sazonal de período aproximado de 12 meses. Empiricamente, consideram-se como sazonais fenômenos que ocorram regularmente em períodos determinados, como de ano para ano (Morettin e Toloi 1987).

As séries de vazões são não estacionárias, mas a maneira mais usual de trata-las não considera componentes de tendência e cíclica (Ballini 2000). Deste modo, a vazão x_t pode ser representada por uma componente sazonal (s_t) e outra estacionária $(\tilde{z}_t)$:

$$x_t = s_t + \tilde{z}_t + a_t \tag{2.9}$$

na qual a_t é a componente aleatória de média zero, variância constante e chamada de ruído branco. Na prática, a componente aleatória será o erro inerente à previsão, já que por melhor que sejam os preditores, não se consegue acertar perfeitamente todos os pontos da série.

Por conta da sazonalidade anual dos dados, é útil tratá-los também com uma discretização mensal, observando as relações e padrões entre os dados de cada mês separadamente. Denotaremos por $x_{i,m}$ as vazões sazonais, sendo *i* o índice do ano para *n* anos do histórico disponível e *m* o mês de referência: *m*=1 equivalente a janeiro e *m*=12 a dezembro. A Tabela 2.1 e a Figura 2.2 mostram a estimativa da média e da variância para a todos os meses da série de Furnas. Estas medidas são calculadas pelas Equações 2.10 e 2.11:

$$\hat{\mu}_m = \frac{1}{n}\sum_{i=1}^{n} x_{i,m} \tag{2.10}$$

$$\hat{\sigma}_m^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_{i,m} - \hat{\mu}_m)^2 \tag{2.11}$$

nas quais $x_{i,m}$ é a vazão observada no ano *i*=1, 2,..., *n* e mês *m*=1, 2,..., 12.

Tabela 2.1 – Médias e Desvios Padrões mensais para a Série da Usina de Furnas

Mês	Média - $\hat{\mu}_m$	Desvio padrão - $\hat{\sigma}_m^2$
Janeiro	1755,075	688,969
Fevereiro	1665,587	624,769
Março	1474,087	583,299
Abril	1013,362	348,446
Maio	741,037	227,413
Junho	613,787	241,340
Julho	506,112	151,377
Agosto	417,537	120,179
Setembro	438,087	224,886
Outubro	514,962	220,589
Novembro	730,366	302,715
Dezembro	1249,475	454,605

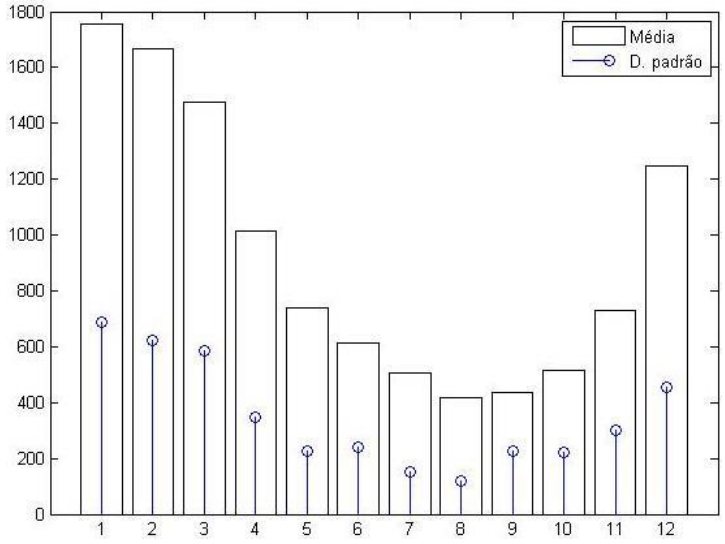


Figura 2.2–Médias e Desvios Padrões mensais para a Série da Usina de Furnas

Como é possível observar, as maiores vazões ocorrem geralmente no mês de janeiro, e as menores em agosto, fato que é compatível com o regime de chuvas no país, que varia de acordo com as estações do ano. O conjunto das 12 médias mensais ($\hat{\mu}_m$) também é conhecido como média de longo termo (MLT).

Para aplicação dos modelos lineares, é necessário que a componente sazonal descrita seja retirada da série original por meio de transformações estatísticas, processo este chamado de dessazonalização (Siqueira, Boccato, et al. 2012c). Discutiremos três formas de conduzir esse processo a seguir.

- Padronização

A padronização é a metodologia que visa eliminar a componente sazonal subtraindo-se de cada dado de vazão a média e dividindo-se pelo desvio padrão referentes ao mês ao qual este dado pertence, como mostra a Equação (2.12):

$$z_{i,m}^{PA} = \frac{x_{i,m} - \hat{\mu}_m}{\hat{\sigma}_m} \tag{2.12}$$

na qual os índices e variáveis são aqueles definidos em (2.10) e (2.11). A nova série padronizada $z_{i,m}^{PA}$ possui média aproximadamente igual a zero e variância unitária. A Figura 2.3 mostra a série de Furnas padronizada

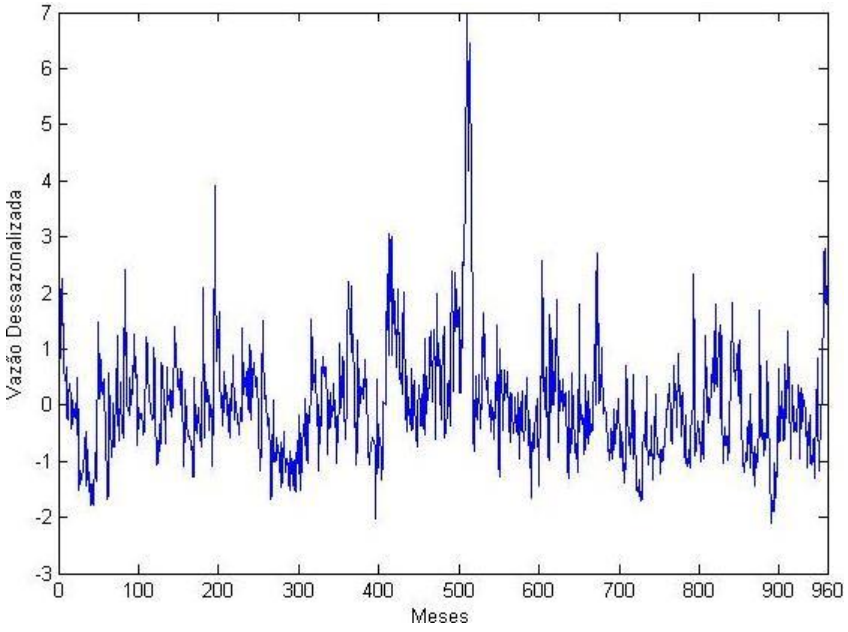


Figura 2.3 - Série da usina de Furnas dessazonalizada via padronização

- Médias Móveis

O procedimento de médias móveis procura estimar a componente sazonal s_t de uma série temporal de modo a retirá-la do processo de previsão, havendo uma subtração com respeito à série original. Este método é adequado quando a componente sazonal é estocástica, ou seja, varia com o tempo, mas é frequentemente aplicado para um padrão sazonal constante (Morettin e Toloi 1987).

Inicialmente, deve-se utilizar as médias de cada mês, conforme a Equação (2.10). Como a soma dos *m* valores de $\hat{\mu}_m$ não é zero, tomamos como estimativa as constantes sazonais:

$$\hat{s}_m^{MM} = \hat{\mu}_m - \overline{\mu} \tag{2.13}$$

na qual $\overline{\mu}$ é a media dos 12 valores de $\hat{\mu}_m$:

$$\overline{\mu} = \frac{1}{12}\sum_{m=1}^{12} \hat{\mu}_m \tag{2.14}$$

Assim, a série dessazonalizada será:

$$z_t^{MM} = x_t - \hat{s}_m^{MM} \tag{2.15}$$

de tal forma que a componente subtraída corresponda ao mês ao qual a observação pertença. A Figura 2.4 mostra a série dessazonalizada por este método.

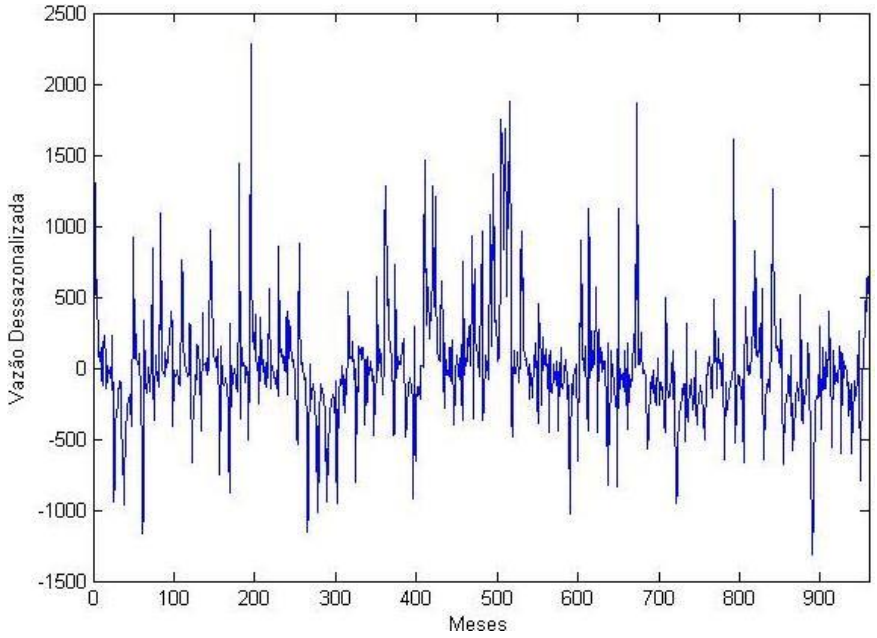


Figura 2.4 - Série da usina de Furnas dessazonalizada via médias móveis

- Diferença Sazonal

O método de diferenças sazonais é adequado para dessazonalização de séries com sazonalidade determinística (Morettin 1986). Aqui, parte-se do princípio de que a

componente sazonal se repete cada período, ou seja, $s_t^{DS} = s_{t+hL}^{DS}$, sendo h um número inteiro e L o período ou o tamanho da sazonalidade (no caso anual, L=12).

O método consiste em fazer diferenças de ordem igual ao período da série. Logo, a série dessazonalizada será:

$$z_t^{DS} = x_t - x_{t-L} \tag{2.16}$$

A Figura 2.5 apresenta a série dessazonalizada por este método:

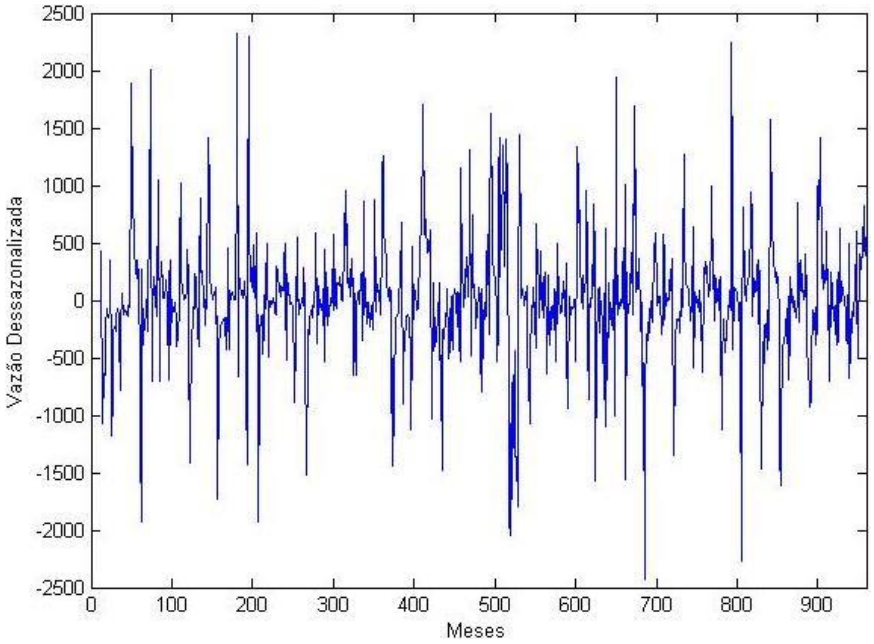


Figura 2.5 - Série da usina de Furnas dessazonalizada via diferenças sazonais

2.4.1 Discussão sobre os processos de dessazonalização

As metodologias de dessazonalização são muito importantes na etapa de pré-processamento dos dados de vazões. Entretanto, conforme ilustram os histogramas da Figura 2.6, ocorre certa assimetria na distribuição dos dados resultantes. Esta observação é importante, pois alguns modelos para inferência de parâmetros, como estimadores de máxima verossimilhança, perdem capacidade de aproximação já que pressupõem simetria na distribuição para sua aplicação (Ballini 2000).

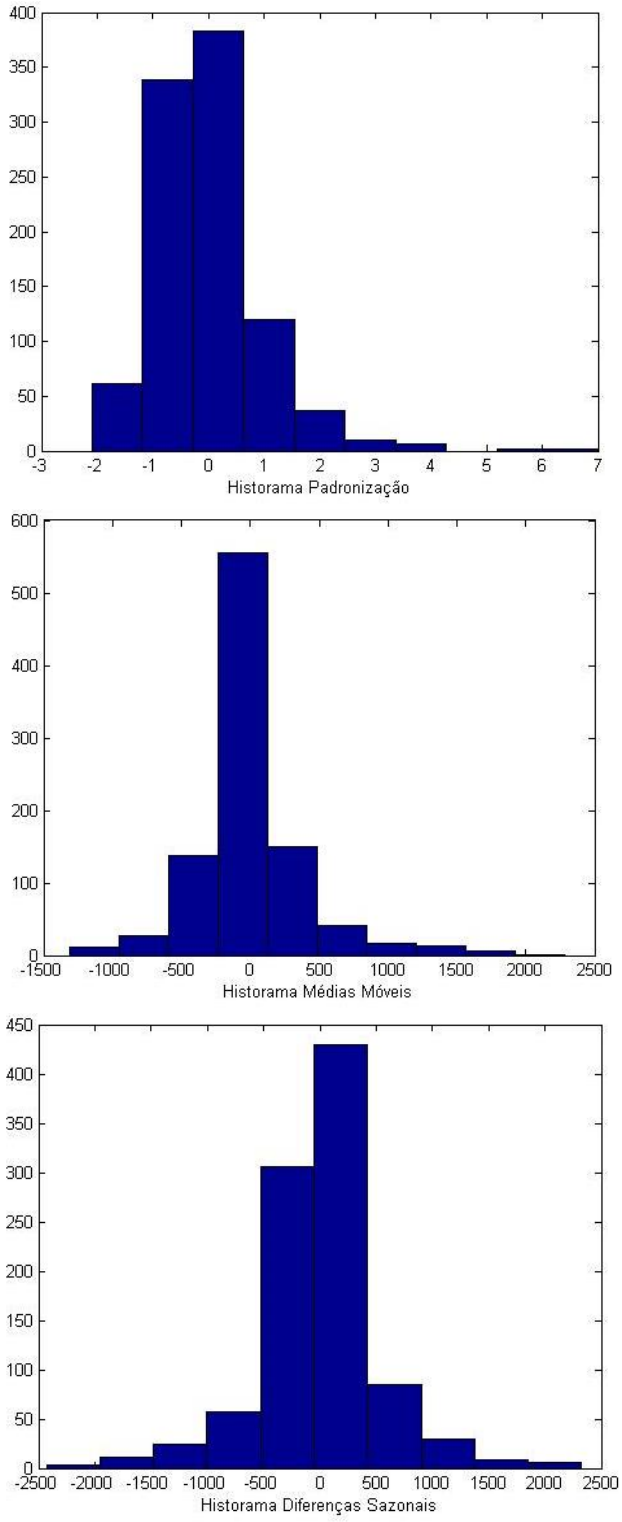


Figura 2.6 – Histograma das séries de vazões dessazonalizadas – (a) padronização, (b) médias móveis, (c) diferenças sazonais

Como forma de definir um modelo de dessazonalização adequado às séries de vazões mensais, procuraram-se evidências acerca da adequação dos três métodos aos ensaios deste trabalho. No Capítulo 5, serão apresentados resultados computacionais da previsão de vazões com modelos lineares e não-lineares que indicaram o método de padronização como o mais adequado ao problema, corroborando a escolha desta metodologia pelo Setor Elétrico Brasileiro.

2.5 Análise da Função de Autocorrelação Dessazonalizada

Após a descrição da função de autocorrelação e dos métodos de dessazonalização, investigaremos em mais detalhe séries de vazões, tomando como base a série da usina de Furnas. Para isso, foram calculados os 36 primeiros valores de r_k, conforme as Equações (2.7) e (2.8) da série dessazonalizada. O histograma da Figura 2.7 mostra que há um decaimento dos valores de r_k à medida que *k* aumenta, e uma variação que obedece a um comportamento senoidal. Na verdade, a função formada tem características análogas às de uma combinação de senóides e exponenciais amortecidas, o que é indicativo de que o processo gerador se associa a um modelo auto-regressivo (Box, Jenkins e Reinsel 2008).

Tabela 2.2 – Valores da Autocorrelação Série de Furnas

k	r_k	k	r_k	k	r_k
1	0.7165	13	0.1745	25	0.0048
2	0.6192	14	0.1787	26	-0.0201
3	0.5436	15	0.1801	27	-0.0325
4	0.4674	16	0.1809	28	-0.0438
5	0.4297	17	0.1576	29	-0.0391
6	0.3601	18	0.1763	30	-0.0339
7	0.2962	19	0.1488	31	-0.0205
8	0.2732	20	0.1186	32	0.0014
9	0.2498	21	0.1020	33	-0.0151
10	0.2204	22	0.0383	34	0.0052
11	0.2100	23	0.0474	35	0.0232
12	0.1745	24	0.0130	36	0.0208

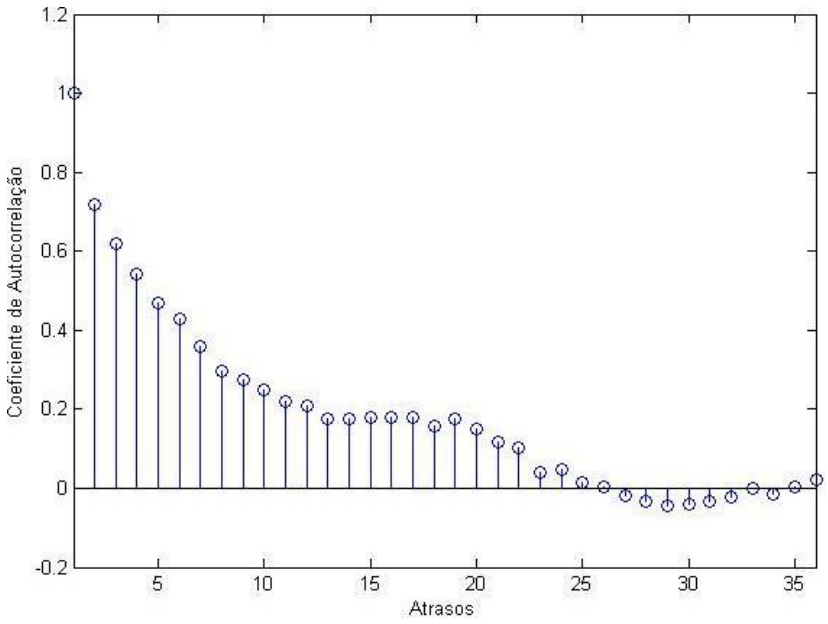


Figura 2.7 – Função de Autocorrelação Série de Furnas

2.6 Modelos Lineares de Previsão

Como já foi mencionado, os modelos lineares auto-regressivos (AR) serão utilizados neste trabalho. Uma parte importante da metodologia Box e Jenkins, da qual faz parte o modelo AR, foi desenvolvida para descrever séries temporais estacionárias. A construção de modelos é realizada por meio de um processo iterativo, que condiciona a estrutura do modelo aos próprios dados disponíveis. Para isso, é necessário passar pelos seguintes estágios:

i) **Especificação**: considera-se uma classe de modelos para análise;
ii) **Identificação**: de posse das funções de autocorrelação, autocorrelação parcial e outros indicadores, um modelo é proposto para o problema;
iii) **Estimação**: etapa de cálculo dos parâmetros livres do modelo sugerido no passo anterior;
iv) **Verificação**: através de alguma métrica de erro ou análise de resíduos, observa-se o grau de adequação do modelo.

No caso de o modelo selecionado não ser suficientemente adequado à representação dos dados, o processo é repetido a partir do passo *ii*. Este passo, inclusive, é fundamental para todo procedimento, já que metodologias aplicadas ao mesmo conjunto de dados

podem identificar modelos diversos (Morettin e Toloi 2006). Vale salientar que um critério da parcimônia com relação ao número de parâmetros deve ser buscado constantemente.

Os modelos lineares estacionários de previsão são amplamente utilizados por sua relativa facilidade de implementação e pelos bons resultados empíricos associados. A definição destes modelos por Box e Jenkins evoca diversos conceitos relativos à temática de filtragem, como mostra a Figura 2.8 (Box, Jenkins e Reinsel 2008).

Figura 2.8 – Representação de uma série temporal como saída de um filtro linear

Formalmente, este modelo pode ser representado por:

$$x_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + ... = \mu + a_t + \sum_{k=1}^{\infty} \psi_k a_{t-j} \tag{2.17}$$

sendo μ a média da série.

Tem-se, então, o seguinte operador linear, que transforma uma soma ponderada dos elementos a_t em x_t, incorporando o termo $k = 0$ para o elemento fora do somatório:

$$\mathbf{\Psi} = \sum_{k=0}^{\infty} \psi_k \tag{2.18}$$

Os modelos podem ser considerados como casos particulares de um filtro linear de *função de transferência* $\mathbf{\Psi}$, sendo x_t a saída. Entende-se aqui que função de transferência refere-se à soma dos coeficientes (pesos) de um operador linear que mapeia um determinado sinal em uma saída desejada, como a Equação (2.17) explicita. Para o caso em questão, ruído branco gaussiano a_t (também chamado de choque aleatório) é transformado em uma série temporal x_t, que possui correlação entre observações sucessivas. Considera-se que o sinal a_t tenha, por hipótese, distribuição normal, média zero e variância constante.

Se a soma dos pesos $\sum_{k=0}^{\infty}\psi_k$ for finita ($k < \infty$), ou infinita e convergente, diz-se que o filtro é estável (somável) e $\tilde{x}_t$ é estacionária, e μ será a média do processo. Caso contrário, a série é não estacionária (Morettin e Toloi 2006).

É possível também representar determinado sinal x_t como uma ponderação de valores passados adicionada de um ruído, caso os coeficientes da função de transferência sejam absolutamente somáveis. Esta forma de representação nos é conveniente, e é definida pela Equação (2.19):

$$x_t = \mu + \pi_1 x_{t-1} + \pi_2 x_{t-2} + \ldots + a_t = \mu + \sum_{j=1}^{\infty} \pi_j x_{t-j} + a_t \tag{2.19}$$

de forma que, isolando o termo a_t

$$a_t = \mu + x_t - \sum_{j=1}^{\infty} \pi_j x_{t-j} \tag{2.19a}$$

De onde pode-se agrupar os coeficientes da seguinte forma

$$\mathbf{\Pi} = \sum_{j=0}^{\infty} \pi_j \tag{2.20}$$

Se simplificarmos a notação de (2.17) e (2.19a) e adotarmos um certo abuso de notação, eliminamos a média e colocarmos os termos agrupados da seguinte forma:

$$x_t = \mathbf{\Psi} a_t \qquad \text{e} \qquad \mathbf{\Pi} x_t = a_t$$

De maneira que igualando as expressões, chega-se a

$$\mathbf{\Pi} = \mathbf{\Psi}^{-1} \tag{2.21}$$

a qual mostra uma relação direta entre os pesos das duas formas de representação da série x_t, além de explicitar que um conjunto pode ser gerado pelo conhecimento do outro.

2.7 Modelos Auto-regressivos e Equações de Yule-Walker

Considere uma série temporal estacionária x_t. Um processo auto-regressivo de ordem p - AR(p) – é definido como a combinação linear dos p atrasos relativos à observação x_t, ou seja, $x_{t-1}, x_{t-2}, ..., x_{t-p}$, com a adição de um ruído branco gaussiano a_t.

Este processo tem relação direta com a Equação (2.19), sendo que os coeficientes π_j são iguais a zero para $k>p$. Dessa forma, o modelo pode ser escrito como:

$$\tilde{x}_t = \phi_1 \tilde{x}_{t-1} + \phi_2 \tilde{x}_{t-2} + ... + \phi_p \tilde{x}_{t-p} + a_t \tag{2.22}$$

de forma que $\tilde{x}_t = x_t - \mu$ e sendo substituídos os termos π_j por ϕ_p, como convencionado em Box et al. (2008). De acordo com os mesmos autores, a condição que garante a estacionaridade é que todos os coeficientes tenham valores absolutos menores do que a unidade, ou seja, $|\phi_p| < 1$. Esta condição garante que uma mudança finita incremental na entrada acarretará em uma mudança finita incremental na saída (Box, Jenkins e Reinsel 2008).

Na implementação do modelo AR(p), o termo a_t será equivalente ao erro inerente ao processo de regressão, que, posteriormente, será associado ao erro de previsão. O cálculo dos coeficientes ϕ_p ótimos em termos do erro quadrático médio (i. e. a obtenção da solução de Wiener) (Haykin 1997) pode ser feito analiticamente, o que é um aspecto computacionalmente vantajoso. Por isso, este modelo tornou-se um dos mais utilizados em tarefas de previsão. Esta solução é originária de uma importante relação de recorrência que emerge da sua função de autocorrelação. Para encontrá-la, tomemos a Equação (2.22) e multipliquemos ambos os lados por x_{t-j}. Em seguida apliquemos o operador esperança matemática, gerando a Equação (2.23):

$$E(x_{t-j}x_t) = E(\phi_1 x_{t-j}x_{t-1}) + E(\phi_2 x_{t-j}x_{t-2}) + ... + E(\phi_p x_{t-j}x_{t-p}) + E(a_t x_{t-j}) \tag{2.23}$$

Como $E(a_t x_{t-j}) = 0$ para $j>0$ e a média do processo também é zero, tem-se exatamente valores de covariância entre os termos, como visto no Capítulo 2, que resultam em:

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + ... + \phi_p \gamma_{j-p}, \text{ com } j>0 \tag{2.24}$$

É válida a relação entre covariância e correlação: portanto, se dividirmos (2.24) por γ_0, chegamos a Equação (2.25)

$$\rho_j = \phi_1\rho_{j-1} + \phi_2\rho_{j-2} + ... + \phi_p\rho_{j-p}\text{, com } j>0 \tag{2.25}$$

Se fizermos *j*=1,2,...,*p,* obtém-se um conjunto de equações lineares para $\phi_1, \phi_2, ..., \phi_p$ em função de $\rho_1, \rho_2, ..., \rho_p$:

$$\begin{aligned}
\rho_1 &= \phi_1\rho_0 + \phi_2\rho_1 + ... + \phi_p\rho_{p-1} \\
\rho_2 &= \phi_1\rho_1 + \phi_2\rho_0 + ... + \phi_p\rho_{p-2} \\
&.... \\
\rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + ... + \phi_p\rho_0
\end{aligned} \tag{2.26a}$$

Estas são as importantes *equações de Yule-Walker* para estimação dos coeficientes de um modelo AR(*p*) (Box, Jenkins e Reinsel 2008). É possível substituir os termos ρ_j por suas estimativas r_j definidas em (2.8). Reescrevendo o sistema na forma matricial, e lembrando que $\rho_0 = 1$:

$$\mathbf{P}_p = \begin{bmatrix} 1 & \rho_1 & ... & \rho_{p-1} \\ \rho_1 & 1 & ... & \rho_{p-2} \\ ... & ... & & ... \\ \rho_{p-1} & \rho_{p-2} & ... & 1 \end{bmatrix} \qquad \boldsymbol{\rho}_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ ... \\ \rho_p \end{bmatrix} \boldsymbol{\Phi}_p = \begin{bmatrix} \phi_1 \\ \phi_2 \\ ... \\ \phi_p \end{bmatrix} \tag{2.26b}$$

A solução de (2.26), então, determina os parâmetros $\boldsymbol{\Phi}_p$, que podem ser calculados como:

$$\boldsymbol{\Phi}_p = \mathbf{P}_p^{-1}\boldsymbol{\rho}_p \tag{2.27}$$

Assim, um modelo AR pode ser otimizado tendo por base os parâmetros de autocorrelação, que formam um sistema de equações com solução analítica. A Figura 2.9 foi traçada em formato de curvas de nível por meio da função de erro quadrático médio amostral $J(\phi) = \sum_{t=1}^{N}(x_t - \hat{x}_t)^2 = \sum_{t=1}^{N} e_t^2$ de um modelo AR(2) para a série dessazonalizada de

Furnas entre os anos de 1931 e 2000. Variamos os valores dos coeficientes no intervalo $[-1;+1]$.

Em seguida, utilizamos as equações de Yule-Walker - Equação (2.27) - para estimar os coeficientes ótimos do modelo e os valores encontrados foram: $\phi_1 = 0{,}5704$ $\phi_2 = 0{,}2104$. Se observarmos a Figura 2.9, é possível notar que esta resposta é compatível com o mínimo global da função $J(\phi)$.

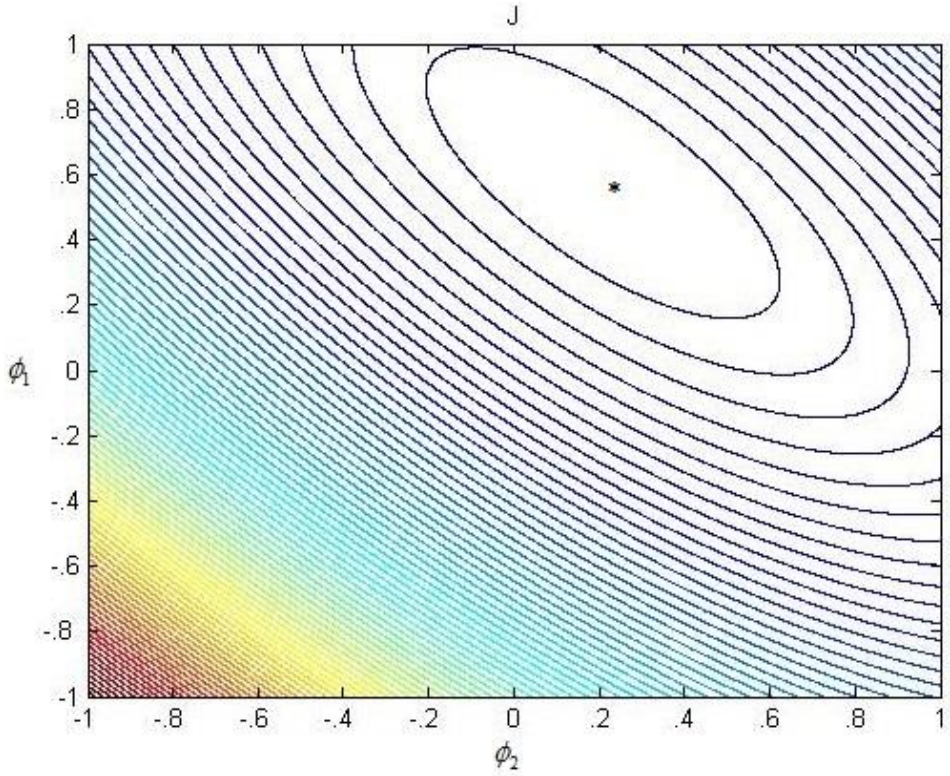


Figura 2.9 – Função de Erro quadrático médio de um modelo AR(2) em curvas de nível

Por fim, também é possível estimar a variância do resíduo a_t:

$$\sigma_a^2 = c_0(1 - r_1\phi_1 - r_2\phi_2 - ... - r_p\phi_p) \tag{2.28}$$

onde c_0 é a variância estimada do processo σ_a^2.

2.7.1 Modelos Periódicos Auto-regressivos

Os modelos periódicos auto-regressivos de ordem p_m - PAR(p_m) - são extensões do modelo AR, e são utilizados em séries temporais que apresentam variações na sua estrutura (Vecchia 1985).

Segundo Hippel & McLeod (1994), algumas séries históricas como as hidrológicas com comportamento sazonal apresentam uma estrutura de autocorrelação não apenas ligada ao atraso de tempo entre as observações, mas também ao período observado. O trabalho de Maceira (1989) mostrou que séries hidrológicas com periodicidade de até o ano, como as vazões mensais, possuem comportamento periódico em relação a suas propriedades estatísticas como a média, a variância e a estrutura de correlação.

A condição de estacionariedade definida na Seção 2.2 é extensível para o caso periódico se a distribuição de probabilidade conjunta associada às k observações respeitar a condição $p(x_{1,m}, x_{2,m},..., x_{N,m}) = p(x_{1+h,m}, x_{2+h,m},..., x_{N+h,m})$, sendo N o número de anos observados, h um fator qualquer e m=1,2,...,12 o índice de cada mês. A padronização, processo de retirada da componente sazonal, pode ser aplicada de forma direta.

Um processo estacionário PAR de ordem p_m pode ser escrito formalmente como:

$$\tilde{x}_t = \phi_1^m \tilde{x}_{t-1} + \phi_2^m \tilde{x}_{t-2} + ...\phi_{p_m}^m \tilde{x}_{t-p_m} + a_t \tag{2.29}$$

onde o índice superior m denota o mês ao qual a observação pertence.

Como em um processo auto-regressivo, os coeficientes de um modelo PAR(p_m) também podem ser calculados via equações de Yule-Walker, mas a forma da Equação (2.27) precisa ser adaptada para o caso mensal. Basicamente, é necessário um ajuste de índices para o cálculo das estimativas r_j^m dos coeficientes de correlação ρ_j^m, que são diferentes para cada mês. A estimativa de covariância c_k^m será :

$$c_k^m = \frac{\frac{1}{N_y}\sum_{i=1}^{N_y}(x_{i,m} - \hat{\mu}_m)(x_{i,m-k} - \hat{\mu}_{m-k})}{\hat{\sigma}_m \hat{\sigma}_{m-k}}, \; k=0,1,...,K \tag{2.30}$$

sendo $\hat{\mu}_m$ a média estimada para o mês m, $\hat{\sigma}_m$ a variância, N_y o número de anos envolvido e k o índice que contabiliza o número de observações utilizadas no cálculo.

Novamente, vale a Equação (2.8), que, para o caso mensal, será reescrita como:

$$r_k^m = \frac{c_k^m}{c_0^m} \tag{2.31}$$

Dessa forma, é possível escrever as equações de Yule-Walker para o caso periódico na forma matricial (Souza, Marcato, et al. 2010):

$$\mathbf{P}_{p_m}^m = \begin{bmatrix} 1 & \rho_1^{m-1} & \dots & \rho_{p_m-1}^{m-1} \\ \rho_1^{m-1} & 1 & \dots & \rho_{p_m-2}^{m-2} \\ \dots & \dots & & \dots \\ \rho_{p_m-1}^{m-1} & \rho_{p_m-2}^{m-2} & \dots & 1 \end{bmatrix}, \quad \mathbf{\Phi}_{p_m}^m = \begin{bmatrix} \phi_1^m \\ \phi_2^m \\ \dots \\ \phi_{p_m}^m \end{bmatrix}, \quad \boldsymbol{\rho}_{p_m}^m = \begin{bmatrix} \rho_1^m \\ \rho_2^m \\ \dots \\ \rho_{p_m}^m \end{bmatrix} \tag{2.32}$$

De modo análogo, a solução ótima para cálculo dos coeficientes do modelo PAR(p_m) no sentido do erro quadrático médio é

$$\mathbf{\Phi}_{p_m}^m = (\mathbf{P}_{p_m}^m)^{-1} \boldsymbol{\rho}_{p_m} \tag{2.33}$$

Comentários

Este capítulo discutiu conceitos fundamentais de séries temporais à luz da teoria de processos estocásticos, forma como podem ser classificadas as séries de vazões médias mensais, as quais possuem como característica comportamento sazonal. Por conta disso, foram abordados modelos de dessazonalização destas séries: padronização, médias móveis e diferenças sazonais.

Foram apresentadas ferramentas de caracterização e análise estatística da série, como as funções de autocorrelação e autocovariância. De posse destes conceitos, é possível sugerir modelos lineares de representação para os dados de vazões.

Abordamos os modelos lineares de previsão, baseados na teoria de Box e Jenkins que serão utilizados neste trabalho: os modelos auto-regressivo (AR) e periódico auto-regressivo (PAR), bem como a forma de calcular seus parâmetros livres de modo determinístico no sentido do erro médio quadrático, via equações de Yule-Walker.

No próximo capítulo, serão discutidas as máquinas desorganizadas e as redes MLP, paradigmas de redes neurais muito úteis para previsões de séries temporais.

Capítulo 3. Redes Neurais MLP e Máquinas Desorganizadas

As redes neurais artificiais (RNAs) são sistemas de processamento da informação paralelos e distribuídos, compostos por neurônios artificiais – unidades funcionais de processamento simples e capazes de suscitar um elevado grau de interconexão (Haykin 2008). A inspiração para essas redes vem do funcionamento do sistema nervoso e do cérebro, sendo uma característica marcante a capacidade de *aprendizagem*, vinculada ao histórico experimental ao qual elas são expostas.

Essa capacidade, ligada, via de regra, à modulação do valor das conexões entre neurônios, confere às RNAs o caráter de ferramentas gerais para a solução de diferentes tipos de problemas. Dessa maneira, elas são frequentemente aplicadas em tarefas como classificação de padrões, mineração de dados, regressão/aproximação de funções, processamento da informação, sendo úteis em diversas áreas do conhecimento, dentre as quais a previsão de séries temporais (Haykin 2008).

É possível entender problemas de previsão como tarefas de mapeamento não linear estático. Nesse contexto, RNAs como o *perceptron de múltiplas camadas* (MLP), são, sem dúvida, opções relevantes, o que se deve a características como capacidade de generalização e de aproximação universal e a existência de metodologias de treinamento sistemáticas e eficientes (Haykin 2008).

Apesar da ampla utilização de ferramentas tradicionais como a MLP, novos modelos de redes vêm ganhando espaço em aplicações semelhantes. Um deles corresponde às Máquinas de Aprendizado Extremo (*Extreme Learning Machines* – ELMs) (Huang, Zhu e Siew 2004), uma rede *feedforward* de múltiplas camadas como a MLP, mas com um processo de ajuste de conexões mais simples e com custo computacional reduzido. A etapa de treinamento supervisionado é baseada em encontrar os coeficientes de um combinador linear, o qual faz o papel de camada de saída, por meio de solução determinística. Este processo, entretanto, não acarreta, necessariamente, em perdas na qualidade da resposta.

Outra proposta interessante são as Redes Neurais de Estado de Eco (*Echo State Networks* – ESNs) (Jaeger 2001), que, por envolverem elementos dinâmicos diferem estruturalmente das ELMs. Esse caráter de recorrência confere a estas redes elementos de

memória intrínseca que podem ser úteis em processos de previsão, nos quais há dependência temporal entre as amostras.

O treinamento de uma ESN é semelhante ao de uma ELM, concentrando-se no ajuste de um combinador linear. Isto estabelece um contraste essencial entre ESNs e as redes neurais recorrentes RNNs clássicas: enquanto estas últimas são treinadas com algoritmos complexos que fazem uso do cálculo do gradiente da função custo a cada iteração, e que frequentemente enfrentam dificuldades como convergência lenta e pouca robustez (Haykin 2008), as ESNs apresentam uma solução de ajuste muito eficiente e rápida.

Por conta da semelhança entre as redes ELMs e ESNs, este trabalho irá adotar o termo *máquinas desorganizadas* - MD - (*Unorganized Machines*) para designá-las, com base na nomenclatura proposta por Boccato et al. (2011b), que evoca os trabalhos de Alan Turing sobre comportamento inteligente de máquinas (Turing 1968).

Cabe ressaltar ainda que se fará uso de novas propostas de camadas de saída para ESNs, uma originária do trabalho de Boccato et al. (2012), na qual se adota um filtro de Volterra precedido por um modelo de compressão de dados tipo Análise de Componentes Principais (PCA), e outra em que se utiliza uma ELM, seguindo Butcher et al. (2010).

3.1 Alguns Aspectos Biológicos

O sistema nervoso dos organismos superiores e a forma de interação dos neurônios que o compõem ainda não foram plenamente compreendidos, o que se explica pela complexidade que envolve a emergência de padrões decorrentes da interconexão massiva de estruturas relativamente simples (os neurônios). Sabe-se, não obstante, que o sistema nervoso é o responsável por captar as informações acerca do ambiente por meio de sensores, torná-las inteligíveis, compará-las com experiências armazenadas na memória e reagir de forma apropriada a estes estímulos (Haykin 2008). A consequência disso é sua capacidade de realizar tarefas como reconhecimento de padrões, realizar controle motor e a percepção sensorial, embora isso não o exima de cometer erros de generalização, imprecisões, etc. (Haykin 2008).

O neurônio, em termos biológicos, tem a função de receber, processar e encaminhar informações por meio de pulsos elétricos. Essas informações são transmitidas de um neurônio para outro através dos chamados neurotransmissores, que estão presentes nas conexões entre eles, conhecidas como sinapses (Haykin 2008).

Uma característica extremamente importante do sistema nervoso é o *aprendizado,* ou a capacidade de adaptar-se a estímulos desconhecidos ou à realização de tarefas inéditas. Este conceito é essencial do ponto de vista da modificação do comportamento do cérebro na solução de problemas. Acredita-se que o aprendizado esteja ligado ao fato de que a efetividade das transmissões de informação das sinapses podem ser moduladas, o que influi diretamente na forma de compreender ou internalizar as diversas situações (Castro 2006).

Com base em todas essas premissas, surge o ramo da ciência conhecido como neurocomputação, que, inspirado no comportamento biológico do cérebro, procura desenvolver modelos e ferramentas computacionais para solução de problemas de diversas naturezas. É neste âmbito que se inserem as redes neurais artificiais, estruturas com capacidade de *aprender* padrões de determinada tarefa para resolvê-la. Sua estrutura construtiva é o neurônio artificial, o qual, essencialmente, representa uma modelagem matemática básica da forma de ação do neurônio biológico.

3.2 Alguns Aspectos da História das RNAs

Considera-se "*A Logical Calculus of Ideas Immanent in Nervous Activity*", do neurofisiologista Warren McCulloch e do matemático Walter Pitts (McCulloch e Pitts 1943), o trabalho pioneiro no campo de redes neurais artificiais. Nele, foi proposto o primeiro modelo lógico-matemático do comportamento do neurônio biológico. Os autores procuravam explicar a atividade neural a partir de unidades elementares de computação que, neste caso, foi modelada como módulos de processamento binário capazes de realizar cálculos lógicos. A resposta de saída era única e função das entradas recebidas.

No tralhado de Rosenblatt (1958), o autor propôs o *perceptron*, uma abordagem até então nova para o problema de reconhecimento de padrões, partindo da premissa que o cérebro trabalha como um associador adaptável de padrões e não como um circuito lógico (Rosenblatt 1958). Além disso, um algoritmo de treinamento para determinação dos pesos

foi proposto, juntamente com a comprovação matemática de sua convergência para padrões linearmente separáveis. Uma noção associada relevante é aquela levantada por Donald Hebb (Hebb 1949), que sugeriu que a efetividade entre as sinapses de dois neurônios é incrementada pela repetida ativação de um neurônio pelo outro.

Em Widrow & Hoff (1960), os autores apresentaram algoritmo LMS (*Least Mean Square*) e o utilizaram para formular o *Adaline* (*Adaptive Linear Element*), que se baseava nesta regra de aprendizagem. A diferença básica, então, entre o *perceptron* e o *Adaline* era o procedimento de treinamento.

O trabalho de Minsky & Papert (Perceptrons 1969) comprovou matematicamente que o *perceptron* linear com modelo de neurônio de Rosenblatt não era um separador universal de classes, já que ele só seria válido para classes linearmente distinguíveis. A demonstração mostrava que esta arquitetura era capaz de executar operações booleanas AND e OR, mas não outras elementares como X-OR (OU-exclusivo) (Haykin 2008).

Na década seguinte, destaca-se o trabalho de Werbos (1974) que propôs uma forma eficiente de computar o gradiente modelos gerais de redes: o algoritmo de retropropagação do erro (*backpropagation*). A aplicação em redes neurais, entretanto, seria um caso especial. Por isso, a relevância maior deste trabalho na área foi posterior, muito por conta do trabalho de Rumelhart et al. (1986), no qual autores propuseram a aplicação do *backpropagation* em aprendizado de máquina, demonstrando como o procedimento funcionaria para este caso. Esta proposta foi muito importante, a ponto de ainda hoje este ser o algoritmo mais comumente aplicado no treinamento da arquitetura *perceptron de múltiplas camadas* – MLP.

Um outro trabalho de destaque para este campo foi o de Hopfield (1982). Para um tipo de rede recorrente (que inclusive recebe seu nome), o autor demonstrou que ela possui um número finito de estados de equilíbrio. A consequência disso é que o sistema invariavelmente evolui para um destes estados ou para uma sequencia periódica de estados a partir de uma condição inicial. Logo, tais pontos, que são de energia mínima, podem ser utilizamos como dispositivos de memória endereçável por conteúdo, e sua localização é controlada pela intensidade das conexões (Castro 1998).

É interessante notar que, na mesma época dos trabalhos iniciais sobre redes neurais, especificamente em 1948, Alan Turing fez uma proposta alternativa de neurônio artificial e de redes, as quais denominou *máquinas desorganizadas* (Turing 1968). O ramo da inteligência artificial dava os primeiros passos, de forma que o autor procurou contribuir neste campo tendo por base modelos neuronais e lógica clássica. As redes de Turing apresentavam paradigmas conexionistas com uma grande relação de proximidade aos trabalhos anteriormente citados, mostrando resultados relacionados à emergência de comportamento inteligente de forma semelhante ao cérebro humano. Até conceitos como aprendizado supervisionado, que ocorre por meio de interferência externa, já eram vislumbrados nestes estudos. Todavia, este trabalho só foi publicado após sua morte, em 1968.

Em anos mais recentes, novas propostas de arquiteturas de redes neurais continuaram surgindo. As redes neurais de estado de eco (Jaeger 2001) e máquinas de aprendizado extremo (Huang, Zhu e Siew 2004), modelos de interesse deste trabalho, são propostas da última década. Por conta do nosso interesse nestas arquiteturas, as discutiremos detalhadamente no decorrer deste capítulo.

3.3 Classificação das RNAs

Este tópico descreverá de forma sucinta algumas classificações de arquiteturas de redes neurais artificiais que serão úteis no decorrer do trabalho. Abordaremos os possíveis contrastes na forma de tópicos?

- **Quanto ao padrão de interconexão**:

i) redes *feedforward*: o sentido do fluxo de informação é sempre adiante, da entrada para a saída, até que se determine a resposta da rede;

ii) redes recorrentes: há conexões de realimentação, que transmitem informações de camadas posteriores para anteriores ou entre neurônios da mesma camada, criando uma espécie de memória interna.

- **Tipos de mecanismo de aprendizado ou formas de ajuste dos pesos**:

i) supervisionado: a alteração no valor numérico dos pesos da rede acontece por meio de um sinal de referência, ou sinal desejado, com o qual a resposta da rede será

comparada. Caso esta não tenha o grau de aproximação desejado (p ex., que se pode quantificar via alguma medida de erro), o algoritmo de treinamento irá promover modificações;

ii) não-supervisionado: o processo de ajuste, neste caso, é tido como auto-organizado, sendo que conexões entre neurônios ativos de maneira correlata se fortalecem, gerando padrões específicos de resposta. Ademais, neurônios podem concorrer para representar determinada amostra, sendo essa competição determinante para a geração de um mapa topologicamente coerente.

iii) por reforço: nesta abordagem, as respostas fornecidas pela rede são avaliadas por meio de um sinal de reforço, ou um sinal indicativo da qualidade do desempenho, que irá direcionar o ajuste dos pesos. Assim, a rede retém as ações que levem a uma maximização do sinal de reforço via um processo de interativo de tentativas e erros.

erros, que visa maximizar um dado índice de desempenho, denominado de sinal de reforço.

Tratemos agora dos neurônios artificiais, que são, de certa forma, a base de uma rede neural.

3.4 O Neurônio Artificial

O neurônio de McCulloch e Pitts, como dito, é uma unidade de processamento capaz de realizar operações lógicas. Seu modo de operação dá-se de acordo com o seguinte conjunto de instruções: a cada instante de tempo *t,* se não houver sinapse inibitória ativa, as entradas são somadas e o neurônio dispara, passando a informação adiante, caso o valor desta soma ultrapasse um determinado limiar θ. Neste caso, sua saída é igual a 1. Se isso não ocorrer, a resposta do neurônio será 0, ou seja:

$$y = f(x) = \begin{cases} 1 & se \quad x \geq \theta \\ 0 & caso \quad contrário \end{cases} \tag{3.1}$$

A função $f(x)$ é representada graficamente na Figura 3.1:

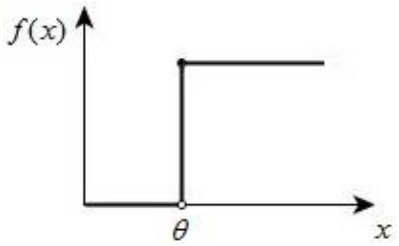


Figura 3.1 – Função de ativação de McCullock e Pitts

O *perceptron* de Rosenblatt é, por sua vez, um classificador de padrões linear, ou seja, separa dados através de um hiperplano no espaço de atributos. A função de ativação tem a forma:

$$y = f(x) = \begin{cases} 1 & se \quad x \geq 0 \\ 0 & caso \quad contrário \end{cases} \tag{3.2}$$

na qual

$$x = \sum_{n=0}^{N} w_{kn} u_n \tag{3.3}$$

O índice n representa a entrada à qual o peso se associa e k é o índice do neurônio, sendo k=0 a ponderação da entrada de polarização.

A Figura 3.2 mostra o esquema do neurônio artificial empregado em redes como MLP, ELM e ESN, uma extensão do modelo *perceptron*. Este dispositivo recebe um conjunto de entradas $\mathbf{u} = [u_1, u_2, ..., u_N]$ provenientes de outros neurônios ou da entrada da rede, processa esta informação e responde com um sinal y. Esta resposta, uma transformação das entradas recebidas, é propagada para outros neurônios

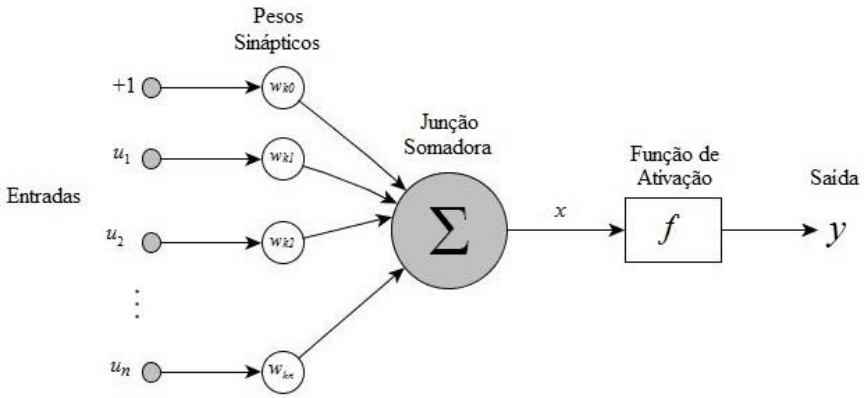


Figura 3.2 – Neurônio Artificial

Nesta figura, é perceptível que os sinais de entrada u_n, n=1, 2,..., N, juntamente com um sinal de *bias* (ou polarização) de valor fixo "+1", são ponderados pelos pesos w_{kn}. Somando-se estes termos, chega-se à ativação x, que passa pela função de ativação $f(.)$, gerando a saída y. Em geral, a função de ativação $f(.)$ é não-linear. A representação matemática deste modelo é dada pela Equação (3.4):

$$y_k = f\left(\sum_{n=0}^{N} w_{kn} u_n\right) \tag{3.4}$$

As funções de ativação mais utilizadas em redes neurais são do tipo sigmoidal (e. g. tangente hiperbólica), por uma série de fatores dentre os quais podemos destacar:

a) Uma função desse tipo é contínua e diferenciável em todos os pontos, o que permite ajuste com algoritmos baseados em derivadas, como o do gradiente;

b) Saturação da saída, o que acaba por evitar que o sinal de saída de cada neurônio divirja;

c) É possível utilizá-la para criar variados tipos de mapeamentos, já que estas funções na região em torno da origem apresentam caráter quase linear, enquanto próximas à saturação são fortemente não-lineares.

O exemplo da Figura 3.3 mostra esta função e sua derivada

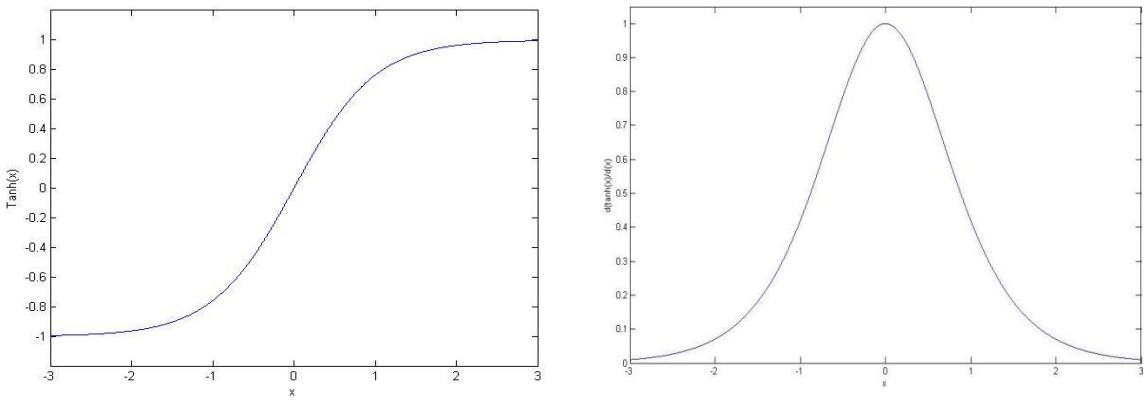


Figura 3.3 – Função tangente hiperbólica e sua derivada

A seguir, passemos às redes neurais construídas a partir de neurônios do tipo acima discutido.

3.5 Perceptron de Mútilplas Camadas - MLP

Uma das mais conhecidas arquiteturas de RNAs *feedforward* é o *perceptron* de múltiplas camadas (*multilayer perceptron* – MLP), que, estruturalmente é uma generalização do *perceptron* de Rosenblatt. Esta rede possui capacidade de aproximação universal, como demonstrado em Cybenko (1989): uma MLP é capaz de aproximar qualquer tipo de função não-linear contínua, limitada, diferenciável e com entradas definidas em um espaço compacto com precisão arbitrária. Isto é possível pela composição aditiva de funções-base, que, para a MLP, são funções de expansão ortogonal (*ridge function*). Contudo, este teorema não especifica a quantidade requerida de neurônios artificiais nem define algum método de ajuste do valor dos pesos para que a configuração ótima da rede seja garantida.

Chamamos de treinamento de uma rede MLP o processo de ajuste de suas ponderações ou pesos sinápticos. O objetivo é encontrar o conjunto de pesos que melhor realiza o mapeamento requerido, com base em alguma métrica que represente o grau de adequação da resposta a cada iteração. Este procedimento pode ser visto como uma tarefa de otimização não-linear irrestrita, com a minimização de uma função custo baseada no erro de aproximação (Von Zuben 1996). Por isso, métodos de otimização muito difundidos de 1ª. ou de 2ª. ordem podem ser utilizados. O mais básico destes é o método do gradiente descendente.

Entretanto, seja qual for o modelo de otimização empregado, será necessária a aplicação do algoritmo de retropropagação de erro, também conhecido como *backpropagation* (Rumelhart, Hinton e Williams 1986), o qual permite, de maneira sistemática, o cálculo do vetor gradiente da função custo, que será parte essencial para a tarefa da otimização pelos modelos citados.

Uma rede neural MLP clássica é composta por neurônios artificiais semelhantes ao da Figura 3.2. Esta arquitetura é dividida em uma camada de entrada, uma ou mais camadas intermediárias (ou escondidas ou, ainda, ocultas), e uma camada de saída. A primeira transmite os sinais de entrada para as camadas intermediárias, em geral sem nenhuma modificação. As camadas escondidas são responsáveis por mapear o sinal de entrada de forma não-linear em outro espaço, de acordo com a demanda do problema a resolver. A

camada de saída recebe este sinal transformado e, através de combinações lineares, produz a resposta da rede. A Figura 3.4 mostra uma das possibilidades de construção, com uma camada intermediária e uma saída.

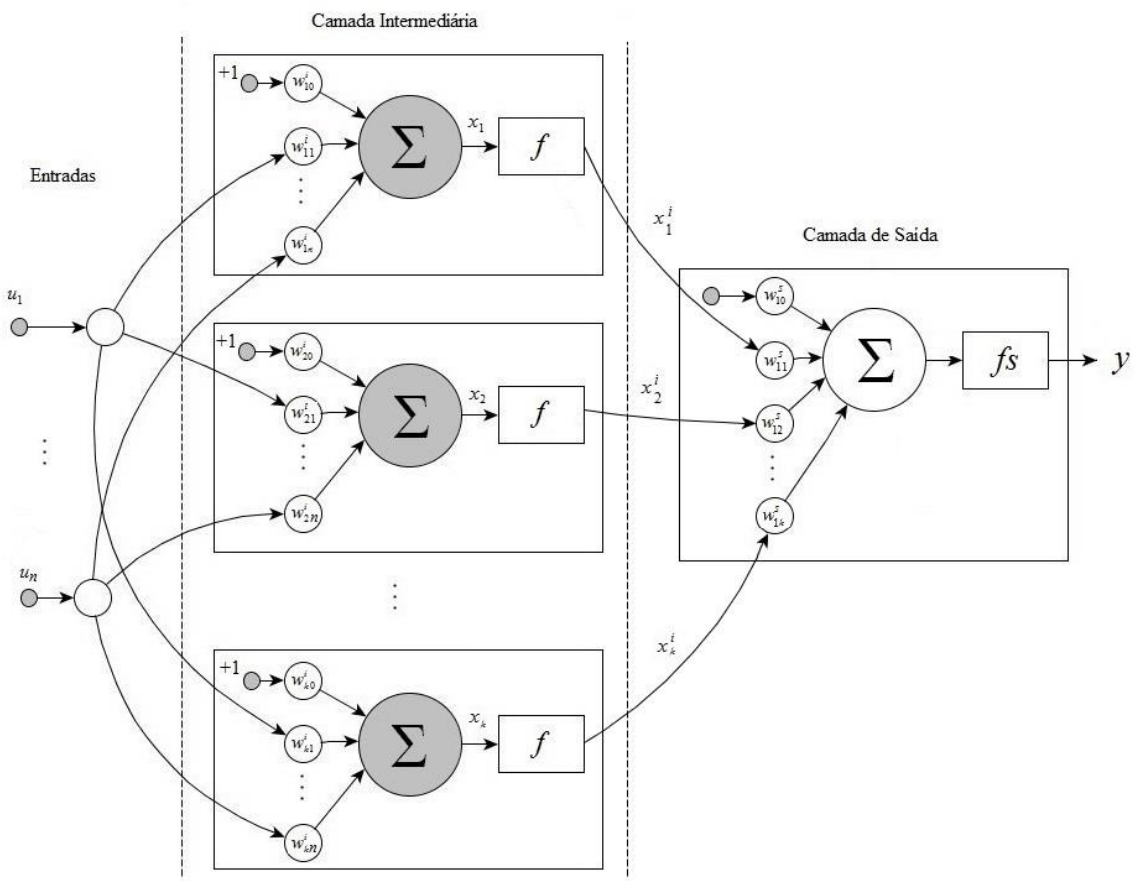


Figura 3.4 – Rede Neural MLP

Nesta figura são mostrados os neurônios da camada intermediária, com os pesos w_{kn}^{i}, sendo k=0,1,...,K a variável que referencia o neurônio e n=0,1,...,Na entrada que chega até o mesmo. Os coeficientes w_{1k}^{s} são as conexões do neurônio de saída. As demais indicações são: "+1"representa a entrada de polarização, x_k^i simboliza a ativação do k-ésimo neurônio da camada intermediária e f é a função de ativação. No neurônio de saída, substituímos f por f_s tendo em vista que essa função costuma ser linear. A saída da rede será:

$$y = f_s\left(\sum_{k=0}^{K} w_{1k}^{s}\left[f\left(\sum_{n=0}^{N} w_{kn}^{i} u_n\right)\right]\right) \tag{3.5}$$

Variáveis com valor "0" de índice se vinculam a entradas de polarização.

Como dito, as conexões entre os neurônios são ponderadas através de pesos que simulam as sinapses reguláveis dos neurônios biológicos. Observa-se ainda que os neurônios de camadas disjuntas se conectam, enquanto neurônios de uma mesma camada não se comunicam, o que confere à rede a condição *feedforward* à rede, ou seja, de rede sem nenhum tipo de realimentação. As arquiteturas de MLPs podem conter diversas camadas intermediárias, mas, em aplicações práticas, é comum que se empregue apenas uma (Haykin 2008). No caso deste trabalho as redes usadas terão sempre uma camada oculta e uma única saída.

- Treinamento

O processo de treinamento de uma MLP é tipicamente supervisionado, e consiste em duas etapas:

i) Fase *forward:* os dados de entrada da rede são introduzidos e propagados, gerando a resposta de saída, com o valor dos pesos fixos. Essa resposta é comparada ao valor observado (esperado) e uma medida de erro é gerada.
ii) Fase *backward:* a etapa seguinte é o ajuste dos valores dos pesos de acordo com a regra de correção de erro assumida. Isto ocorre após o cálculo do gradiente da função custo, gerado a partir do erro de treinamento.

A medida de erro mais utilizada para esta tarefa é o erro quadrático médio (MSE – inglês *mean squared error*), o qual os algoritmos de treinamento irão minimizar com respeito aos pesos. Este tipo de procedimento nada mais é do que um problema de otimização não-linear, e pode ser tratado com qualquer método de otimização linear irrestrito de 1ª. ou 2ª. ordens, como os métodos de gradiente e de Newton.

Uma outra classe de algoritmos de otimização multidimensional de segunda ordem é a dos métodos de gradiente conjugado. Eles podem ser considerados intermediários entre o método de Newton e o do gradiente descendente, pois não realizam buscas unidimensionais para definição do passo de ajuste, o que reduz o número de avaliações da função objetivo.

Neste trabalho, utilizou-se como método padrão de treinamento da MLP o método do Gradiente Conjugado Escalonado (SCG) (Moller 1990), com a modificação proposta por

Von Zuben & Netto (1995). O SCG escalona o passo de ajuste α, com a utilização da abordagem de Levenberg-Marquadt (Bazaraa, Sherali e Shetty 2006). A informação de segunda ordem precisa ser, então, calculada e armazenada a cada iteração, o que acarreta em um custo computacional elevado.

O operador diferencial proposto no trabalho de Pearlmutter (1994), entretanto, pode contribuir para uma redução drástica deste custo por permitir o cálculo exato da informação de segunda ordem, com custo computacional semelhante ao do cálculo da informação e primeira ordem (Von Zuben 1996). Este operador pode ser aplicado diretamente quando uma matriz hessiana aparece multiplicada por um vetor, o que ocorre, invariavelmente, no método SCG, de forma que esta matriz não precise ser calculada ou armazenada. Segundo Castro (1998), o custo computacional envolvido na aplicação do método é (OP+P) operações de ponto flutuante por época de treinamento, sendo O o número de amostras de treinamento e P o número de parâmetros livres da rede. Em termos comparativos, o método de Newton Modificado, que calcula a matriz Hessiana diretamente, tem custo (OP+$3P^2$).

A escolha deste método de treinamento também está relacionada ao trabalho de Castro (1998), no qual este algoritmo mostrou-se bastante eficiente em termos de desempenho e esforço computacional para tarefas como aproximações de funções e classificação de dados.

-Gradiente Conjugado Escalonado Modificado

A função custo a ser minimizada, baseada no MSE entre o sinal desejado e a resposta da rede, é expressa na Equação (3.6):

$$J(\mathbf{w}_t) = \frac{1}{2}(y_t - d_t)^T (y_t - d_t) \tag{3.6}$$

O cálculo do gradiente será dado pela soma dos gradientes parciais calculados a cada iteração t (Santos e Von Zuben 1999).

A descrição do método passa pela escolha da direção de busca $\mathbf{d}_t$, do passo de ajuste α_t e de um coeficiente de momento β_t. As seguintes componentes são definidas:

$\mathbf{w}_t$ - conjunto de pesos;

$\nabla J(\mathbf{w}_t)$ - o vetor gradiente da função custo;

$\nabla^2 J(\mathbf{w}_t)$ - Matriz Hessiana de $\mathbf{w}$;

O termo α_t é definido para o ponto $\mathbf{w}_t$ como

$$\alpha_t = \frac{(\mathbf{d}_t)^{\mathrm{T}} \nabla J(\mathbf{w}_t)}{(\mathbf{d}_t)^{\mathrm{T}} \nabla^2 J(\mathbf{w}_t)\mathbf{d}_t} \tag{3.7}$$

Moller (1990) propõe que o termo que compõe parte do denominador e que envolve a Hessiana, renomeado como $s_t = \nabla^2 J(\mathbf{w}_t)\mathbf{d}_t$, seja aproximado por

$$\mathbf{s}_t = \nabla^2 J(\mathbf{w}_t)\mathbf{d}_t = \frac{\nabla J(\mathbf{w}_t + \sigma_t \mathbf{d}_t) - \nabla J(\mathbf{w}_t)}{\sigma_t} \tag{3.8}$$

sendo $0 < \sigma_t << 1$ um escalar a ser determinado.

Esta aproximação tende, no limite, ao valor de $\nabla^2 J(\mathbf{w}_t)\mathbf{d}_t$ (Castro 1998). Adotando δ_t como o denominador de (3.7), ou seja $\delta_t = (\mathbf{d}_t)^{\mathrm{T}} \nabla^2 J(\mathbf{w}_t)\mathbf{d}_t$, e substituindo-o na Equação (3.8), tem-se:

$$\delta_t = (\mathbf{d}_t)^{\mathrm{T}} \mathbf{s}_t \tag{3.9}$$

O ajuste de λ_t a cada iteração e a análise do sinal de δ_t permitem determinar se a matriz Hessiana é definida-positiva ou não.

Por fim, note que o algoritmo usa uma aproximação quadrática para a função custo $J_{quad}(\mathbf{w}_t)$. Todavia, esta a aproximação nem sempre é adequada, uma vez que λ_t escalona a matriz Hessiana de forma artificial. Dessa forma, um mecanismo que faça o valor de λ_t aumentar ou diminuir automaticamente é conveniente. Assim, Moller define a seguinte variável:

$$\Delta_t = \frac{J(\mathbf{w}_t) - J(\mathbf{w}_t + \alpha_t \mathbf{d}_t)}{J(\mathbf{w}_t) - J_{quad}(\mathbf{w}_t + \alpha_t \mathbf{d}_t)} = \frac{2\delta_t [J(\mathbf{w}_t) - J(\mathbf{w}_t + \sigma_t \mathbf{d}_t)]}{\mu_j^2} \tag{3.10}$$

na qual $\mu_t = -\mathbf{d}_j^{\mathrm{T}} \nabla J(\mathbf{w}_t)$.

A definição da variável Δ_t torna-se importante porque ela representa uma medida da qualidade da aproximação $J_{quad}(\mathbf{w}_t)$ em relação a $J(\mathbf{w}_t + \alpha_t \mathbf{d}_t)$. Quanto mais próximo for da unidade, melhor a aproximação.

O resumo deste método em forma de pseudocódigo é apresentado no Quadro 3.1, no qual ε é o limiar de parada arbitrariamente definido (Moller 1990).

Como é perceptível não se faz necessário nenhum tipo de busca unidimensional. O problema é a necessidade de se calcular $\nabla^2 J(\mathbf{w}_t)$ a cada iteração, o que torna o algoritmo computacionalmente custoso. Todavia, o trabalho de Pearlmutter (1994) trouxe para esta aplicação uma solução simples e eficiente por meio da definição de um operador diferencial que garante que uma matriz Hessiana multiplicada por um vetor pode ter este produto calculado de maneira exata, sem a necessidade do cálculo desta matriz. Como é possível perceber pelo Quadro 3.1, quando $\nabla^2 J(\mathbf{w}_t)$ aparece no algoritmo, ela está multiplicada pelo vetor $\mathbf{d}_t$, em consonância com (3.7). Este operador, chamado no trabalho original por $\Re\{\cdot\}$, garante as seguintes propriedades:

$$\Re^d\{\nabla J(\mathbf{w})\} = \nabla^2 J(\mathbf{w}_t)\mathbf{d} \qquad \text{e} \qquad \Re^d\{\mathbf{w}\} = \mathbf{d} \tag{3.11}$$

Por $\Re(.)$ ser um operador diferencial, ele obedece às regras usuais de diferenciação.

O algoritmo apresentado no Quadro 3.1 será, então, alterado nos seguintes passos:

- Passo 1: não é necessário definir o parâmetro σ

- Passo 3: o parâmetro $\mathbf{s}_t$ será calculado exatamente por: $\mathbf{s}_t = \nabla^2 J(\mathbf{w}_t)\mathbf{d}_t$

Por último, é importante salientar que Moller (1990) aponta que $\nabla J(\mathbf{w}_t)$ e $\nabla^2 J(\mathbf{w}_t)$ precisam ser calculados a cada passo *t*.

Quadro 3-1 - Algoritmo do método Gradiente Conjugado Escalonado

1. *Escolha o vetor de pesos* $\mathbf{w}_1$ *e os escalares* $\sigma > 0$ *e* $\lambda_1 > 0$.
2. *Faça* $\mathbf{d}_1 = \mathbf{r}_1 = -\nabla J(\mathbf{w}_1)$, $t = 1$ *e sucesso=1.*
3. *Se sucesso=1, calcule a informação de segunda ordem:*

$$\sigma_t = \frac{\sigma}{|\mathbf{d}_t|}$$

$$\mathbf{s}_t = \frac{\nabla J(\mathbf{w}_t + \sigma_t \mathbf{d}_t) - \nabla J(\mathbf{w}_t)}{\sigma_t}$$

$$\delta_t = \mathbf{d}_t^{\mathrm{T}} \mathbf{s}_t$$

4. *Faça de* $\mathbf{s}_t$

$$\mathbf{s}_t = \mathbf{s}_t + (\lambda_t - \overline{\lambda}_t)\mathbf{d}_t$$

$$\delta_t = \delta_t + (\lambda_t - \overline{\lambda}_t)|\mathbf{d}_t|^2$$

5. *Se* $\delta_t \leq 0$, *faça a matriz Hessiana ser definida positiva:*

$$\overline{\lambda}_t = 2\left(\lambda_t - \frac{\delta_t}{|\mathbf{d}_t|^2}\right)$$

$$\delta_t = -\delta_t + \lambda_t |\mathbf{d}_t|^2$$

$$\lambda_t = \overline{\lambda}_t$$

6. *Calcule a taxa de ajuste:*

$$\mu_t = \mathbf{d}_t^{\mathrm{T}} \mathbf{r}_t$$

$$\alpha_t = \frac{\mu_t}{\delta_t}$$

7. *Calcule o parâmetro de comparação:*

$$\Delta_t = \frac{2\delta_t [J(\mathbf{w}_t) - J(\mathbf{w}_t + \alpha_t \mathbf{d}_t)]}{\mu_t^2}$$

8. *Se* $\Delta_t \geq 0$ *então atualize o vetor de pesos (o erro pode ser reduzido):*

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \alpha_t \mathbf{d}_t$$

$$\mathbf{r}_{t+1} = -\nabla J(\mathbf{w}_{t+1})$$

$$\overline{\lambda}_t = 0$$

sucesso=1

7a. Se (t mod N)=0 então reinicie o algoritmo: $\mathbf{d}_{t+1} = \mathbf{r}_{t+1}$

senão crie uma nova direção conjugada:

$$\beta_t = \frac{|\mathbf{r}_{t+1}|^2 - \mathbf{r}_{t+1}\mathbf{r}_t}{\mu_t} \qquad e \qquad \mathbf{d}_{t+1} = \mathbf{r}_{t+1} + \beta_t \mathbf{d}_t$$

7b. Se $\Delta_t \geq 0{,}75$ *então reduza o parâmetro escalonado:* $\lambda_t = 0{,}5\lambda_t$

Senão a redução no erro não é possível: $\overline{\lambda}_t = \lambda_t$ *e sucesso=0*

9. *Se* $\Delta_t < 0{,}25$ *então incremente o parâmetro escalonado:* $\lambda_t = 4\lambda_t$
10. *Se o a direção de maior descida* $\mathbf{r}_t \neq 0$ *então adote* $t = t + 1$ *e vá para o passo 2.*

Senão encerre e retorne $\mathbf{w}_{t+1}$ *como sendo o mínimo desejado.*

-Validação Cruzada[1]

Durante o processo de treinamento de uma MLP, é necessário considerar que o ajuste precisa ser adequadamente controlado quanto ao grau de flexibilidade. O que se busca é a máxima capacidade de generalização, ou seja, uma condição que garanta que a rede apresentará a melhor resposta para entradas desconhecidas.

Redes que sejam muito ajustadas ao conjunto de dados a ela apresentado para treinamento podem ter um desempenho muito ruim quando for preciso aproximar dados que não pertençam ao conjunto original. Este fenômeno é conhecido como sobretreinamento (ou *overfitting*), e pode ser causado por um excesso no número de neurônios ou por um treinamento além do necessário. Uma técnica muito utilizada para contornar a segunda causa citada é a validação cruzada.

A metodologia consiste em dividir os dados de forma que haja um conjunto de validação, com amostras que não participem do ajuste dos pesos. A cada iteração, após a fixação no valor dos parâmetros, os dados de validação são inseridos na rede e a resposta de saída irá gerar uma medida de erro independente do de treinamento. O ponto de mínimo do erro de validação é escolhido como o de melhor generalização estimada e os valores nominais dos pesos neste ponto são fixados como os ótimos da rede.

Deve-se considerar que os dois conjuntos sejam suficientemente representativos do mapeamento que se pretende aproximar. Além disso, se os conjuntos forem compostos por amostras muito similares, ambos os erros serão decrescentes. A Figura 3.5 mostra o comportamento do erro quadrático médio de treinamento e de validação durante o ajuste de uma MLP para a série de Furnas. A curva superior mostra que o valor do MSE de treinamento é sempre decrescente, mas o de validação, a linha inferior, não tem essa característica. Há um ponto de inflexão (o ponto marcado com um círculo), alcançado por volta da época 250. Este é o ponto em que o valor dos pesos que levaria capacidade de generalização estimada como ideal.

---

[1] A forma de validação cruzada descrita nesta seção é a mais simples, sendo também conhecida por método *holdout*

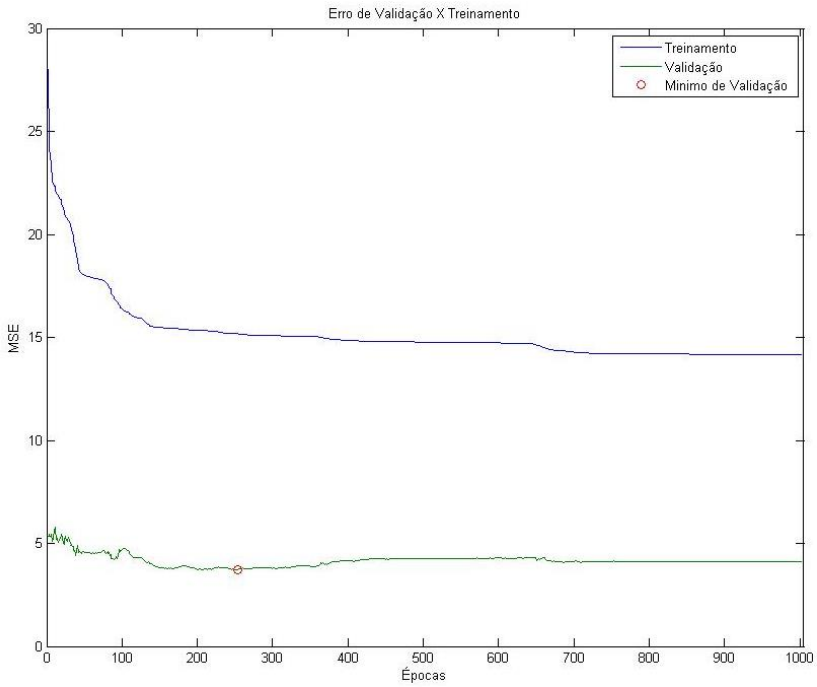


Figura 3.5 – Exemplo de Validação Cruzada

Tratemos agora das arquiteturas de redes neurais tipo máquinas desorganizadas.

3.6 Máquinas Desorganizadas

Os trabalhos de Jaeger (2001) e Maass et al. (2002) inauguraram novos paradigmas de redes neurais, sob uma perspectiva inédita para adaptação de estruturas realimentadas, com um processo de treinamento muito simples e eficaz computacionalmente. Estas premissas deram início a uma nova frente de investigação conhecida como computação de reservatório (*reservoir computing* - RC) (Lukoševičius e Jaeger 2009). As propostas apresentadas foram, respectivamente, as redes neurais de estado de eco (ESNs) e as *liquid state machines* (LSMs).

Nesses paradigmas, as redes são compostas por uma camada intermediária recorrente, chamada de reservatório, que propicia a existência de uma memória interna. Tal camada não passa por processos de ajuste supervisionado das conexões, enquanto a camada de saída é treinada para aproximar o sinal que vem do reservatório do sinal desejado.

Em 2004, Huang et al. (2004) propuseram uma metodologia de projeto de redes *feedforward*, na qual os pesos da camada intermediária são definidos de maneira aleatória e não passam por processos de ajuste. Dessa maneira, o treinamento supervisionado da rede

se limita aos coeficientes ótimos do combinador linear que forma a camada de saída. Essas redes recebem o nome de máquinas de aprendizado extremo (ELMs).

Esta tese lidará com as ELMs e ESNs, que discutiremos em profundidade no decorrer deste capítulo. Há similaridades interessantes entre as duas propostas originais pelo fato de que camadas de neurônios não têm pesos ajustados e as ativações são combinadas linearmente. Assim, o treinamento acaba por ser vantajoso em termos de tratabilidade e simplicidade. Esse fato leva a uma reflexão que nos parece válida.

A aleatoriedade das camadas internas traz implícita a ideia de desorganização, já que uma parte da rede permanece sem ajuste. Isso remete aos trabalhos de Turing (1968), como exposto em Boccato et al. (2011b) e Boccato (2013). As redes propostas por Turing também pressupunham a existência de uma estrutura recorrente formada por neurônios artificiais, na qual os padrões de conexões não eram ajustados. Além disso, conceitos como aprendizado já eram discutidos na proposta, pois unidades aleatoriamente geradas e conectadas eram treinadas por meio de um sinal externo de referência, com o intuito de realizar uma tarefa pré-determinada. Com esta ideia, Turing procurou investigar a emergência de comportamento inteligente em máquinas, à luz do desenvolvimento do sistema nervoso no decorrer da infância de um ser humano. A este tipo de estrutura, Turing deu o nome de *máquina desorganizada*.

As abordagens de ELM e ESN são derivadas da visão atual de redes neurais como aproximadores universais, nas quais se buscou uma solução de compromisso entre simplicidade da rede e a capacidade de mapeamento. Assim, embora a inspiração de Turing e das demais RNAs citadas sejam diversas, pelas semelhanças discutidas entre as propostas das estruturas com conexões arbitrariamente criadas, iremos adotar a terminologia *máquinas desorganizadas* para agrupar as ELMs e ESNs sob a mesma nomenclatura, adotando a proposta de Boccato et al (2011b).

3.6.1 Máquinas de Aprendizado Extremo

As máquinas de aprendizado extremo são RNAs *feedforward* semelhantes às MLPs, e com apenas uma camada intermediária. Todavia, elas se diferenciam pelo processo de treinamento, uma vez que os pesos dos neurônios da camada intermediária são determinados de maneira aleatória e independente. O processo de ajuste, portanto, não

adapta os pesos desta camada, mas apenas os da camada de saída. Os valores ótimos dos pesos são tipicamente calculados de forma analítica, já que o treinamento envolve a solução de um problema de regressão linear (Huang, Chen e Siew 2006). Dessa forma, não há necessidade de cálculo de derivadas, retropropagação de sinais de erro ou utilização de algoritmos iterativos, o que reduz sobremaneira o custo computacional envolvido no processo de treinamento.

Na proposta inicial (Huang, Zhu e Siew 2004), há a demonstração teórica de que uma rede neural tipo *feedforward* de única camada é capaz de representar com uma precisão arbitrária um conjunto de dados, caso as funções de ativação dos neurônios sejam infinitamente diferenciáveis e os pesos da camada oculta sejam gerados de forma aleatória sob qualquer distribuição contínua. Nesta mesma perspectiva, comprova-se que uma ELM possui capacidade de aproximação universal, uma vez que o erro de aproximação pode ser sempre reduzido com a inserção de um novo neurônio na camada escondida, via determinação rigorosa dos pesos da camada de saída (Huang, Zhu e Siew 2004). A opção deste trabalho para a definição deste conjunto de coeficientes foi o operador generalizado de Moore-Penrose, que minimiza o erro quadrático médio entre o sinal que chega da camada intermediária e o desejado.

Um resultado teórico importante foi obtido por Bartlett (1998) para problemas de classificação de padrões. A conclusão do trabalho é que controlar a norma dos pesos sinápticos acaba por ser mais relevante em termos de capacidade de generalização do que controlar o número de neurônios da camada intermediária. Isto leva a evidências importantes de que um aumento na capacidade de generalização ocorre quando o vetor de parâmetros possui norma mínima, de maneira que o número efetivo de neurônios na camada intermediária será definido pela configuração dos pesos da camada de saída.

Diante desta afirmação, a ELM terá garantia de boa generalização efetivamente dada pelos pesos da camada de saída, podendo os pesos da camada intermediária serem definidos de forma aleatória. Por conta disso, o treinamento da rede passa a ser linear em relação aos parâmetros ajustáveis para treinamento supervisionado. O operador generalizado de Moore-Penrose torna-se candidato importante para solução deste problema.

A Figura 3.6 mostra a arquitetura de uma ELM genérica com suas camadas. Note que está representada apenas uma saída, mas a estrutura admite que elas sejam múltiplas. Nesta figura, não estão presentes os pesos das conexões, e mantivemos os nomes das variáveis de acordo com a proposta original (Huang, Zhu e Siew 2004)

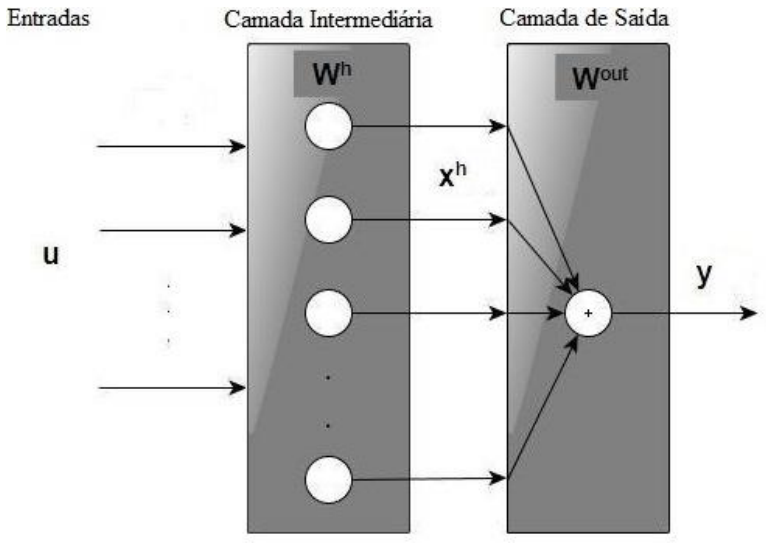


Figura 3.6 – Máquina de Aprendizado Extremo

O vetor de entradas é $\mathbf{u}_t = [u_1, u_2, ..., u_{t-K+1}]^{\mathrm{T}}$ que tem seus componentes passados à camada intermediária, e $\mathbf{W^h} \in \Re^{N_h \times K}$ contém os pesos gerados de forma aleatória. O sinal de saída desta camada é produzido de acordo com a Equação (3.12):

$$\mathbf{x}_t^{\mathbf{h}} = \mathbf{f}^{\mathbf{h}}(\mathbf{W}^{\mathbf{h}}\mathbf{u}_t + \mathbf{b}) \tag{3.12}$$

na qual $\mathbf{f^h}(.)$ é a função de ativação dos neurônios, N_h é o número de neurônios, K é o número de entradas da rede e **b** o vetor contendo as respectivas polarizações. A saída $\mathbf{y}_t$ da rede será composta pela combinação linear das ativações dos neurônios da camada oculta de acordo com a expressão:

$$\mathbf{y}_t = \mathbf{W}^{\mathbf{out}}\mathbf{x}_t^{\mathbf{h}} \tag{3.13}$$

sendo $\mathbf{W^{out}} \in \Re^{N_h \times L}$ os pesos da camada de saída.

O treinamento da ELM é equivalente a resolver o problema de otimização da Equação (3.14):

$$\mathbf{w}_k^* = \arg \min_{\mathbf{w}_k \in \Re^{N_h+1}} \left\| \mathbf{w}_k \right\|^2 + C \times J(\mathbf{w}_k) \tag{3.14}$$

na qual, *k* é o índice que indica a saída, $\|\cdot\|^2$ é a norma euclidiana e *C* um coeficiente de ponderação ou de regularização. É possível considerar o valor de *C* como infinito, o que o retiraria do cálculo do valor ótimo no sentido do mínimo MSE entre a saída desejada $\mathbf{d}_t$ e a saída da rede $\mathbf{y}_t$. Neste caso, a camada de saída pode ser obtida por uma solução fechada pelo operador de Moore-Penrose de acordo com a seguinte expressão:

$$\mathbf{W}^{\mathbf{out}} = (\mathbf{X}_h^{\mathrm{T}}\mathbf{X}_h)^{-1}\mathbf{X}_h^{\mathrm{T}}\mathbf{d} \tag{3.15a}$$

onde $\mathbf{X}_h \in \Re^{T_s \times N_k}$ é a matriz de saídas da camada oculta, T_s é o número de amostras de treinamento, $(\mathbf{X}_h^{\mathrm{T}}\mathbf{X}_h)^{-1}\mathbf{X}_h^{\mathrm{T}}$ é a pseudoinversa de $\mathbf{X}_h$ e $\mathbf{d} \in \Re^{T_s \times 1}$ contém as saídas desejadas. Esta solução é simples e eficiente em termos computacionais.

Todavia, o desempenho da rede é incrementado pela utilização de um coeficiente de regularização *C* (Huang, Zhou, et al. 2012). Neste caso, pode-se adicioná-lo à Equação (3.15a), gerando a expressão (3.15b)

$$\mathbf{W}^{\mathbf{out}} = \left(\frac{1}{C} + \mathbf{X}_h^{\mathrm{T}}\mathbf{X}_h \right)^{-1} \mathbf{X}_h^{\mathrm{T}}\mathbf{d} \tag{3.15b}$$

A proposta de (Huang, Zhou, et al. 2012) para o cálculo do coeficiente de regularização pressupõe um conjunto de validação. Assumindo-se que $C = 2^{\lambda}$, sendo o expoente um dos 52 elementos do vetor $\boldsymbol{\lambda} = \{2^{-25}, 2^{-24}, ..., 2^{25}, 2^{26}\}$. Dessa forma, fixa-se cada valor do coeficiente separadamente e apresentam-se as entradas de validação à rede, gerando uma medida de erro quadrático médio. O *C* que apresentar o menor erro de validação será considerado aquele que permite a melhor generalização da rede.

Uma proposta metodológica complementar a esta foi feita por Kulaif & Von Zuben (2013), e apresentou bons resultados para uma série de problemas, mostrando que este procedimento pode ser muito útil para aprimorar o desempenho destas redes. Nela, uma etapa de busca local é inserida como forma de refinar a solução do problema para casos de regressão. Para isso, deve-se partir de duas premissas: 1) pequenos desvios no valor de *C*

podem significar uma variação grande na solução e 2) para qualquer intervalo de busca pequeno associado a possíveis valores de *C*, a curva formada pelos valores desta variável e pelos erros de validação correspondentes é quase-convexa. Todavia, este comportamento da curva pode ser violado se for utilizada uma busca unidimensional no intervalo $(0,+\infty)$. Portanto, como a quasiconvexidade é condição necessária para aplicação de uma busca deste tipo, os autores decidiram aplicar a proposta inicial de Huang et al. (2012) e, de posse do valor de *C* encontrado, aplicar o método da seção áurea (Bazaraa, Sherali e Shetty 2006) nos valores ao redor deste, fazendo os valores dos vizinhos da direita e da esquerda como os pontos extremos de cada intervalo. Assim, a capacidade de generalização pode ser aumentada, melhorando o desempenho geral da rede.

Para exemplificar a proposta, utilizamos a série de vazões mensais de Sobradinho de 1931 a 2010, na qual as amostras entre os anos de 1931 e 1991 (60 anos ou 720 amostras) foram reservadas aos dados de treinamento e de 1992 a 2001 (10 anos ou 120 amostras) formaram o conjunto de validação. Dessa maneira, realizou-se o ajuste de 12 ELMs (uma para cada mês) e chegou-se aos valores de *C* como sugerido por Kulaif & Von Zuben (2013). O intervalo utilizado para aplicação do método da seção áurea e os valores do erro quadrático médio de validação com e sem a regularização estão presentes na Tabela 3.1:

Tabela 3.1 - Valores de *C* para série de Sobradinho

Mês	Intervalo	Valor de *C*	MSE sem regularização	MSE regularizado
1	2^{4} - 2^{6}	27.9255	**0.5165**	**0.5165**
2	2^{25} - 2^{26}	6.7109e+07	1.1081	1.0732
3	2^{25} - 2^{26}	6.7109e+07	**0.6158**	**0.6158**
4	2^{-4} – 2^{-2}	0.1168	0.7693	0.7433
5	2^{-4} – 2^{-2}	0.1654	0.4485	0.3588
6	2^{-2} – 2^{0}	0.6696	**0.2467**	**0.2467**
7	2^{22} – 2^{24}	6.6589e+06	0.2611	0.2432
8	2^{25} - 2^{26}	6.7109e+07	0.2316	0.2263
9	2^{25} - 2^{26}	6.7109e+07	0.5315	0.5189
10	2^{7} – 2^{9}	205.6868	0.6497	0.6450
11	2^{25} - 2^{26}	6.7109e+07	0.9977	0.8662
12	2^{20} – 2^{22}	2.8721e+06	0.6900	0.6770

Os intervalos estão distantes 2 ordens de grandeza, a menos dos casos em que o limitante superior é o valor máximo inicial. Isto significa que o valor sugerido pelo método de Huang et al. (2012) é aquele que está equidistante dos pontos do intervalo e os dois

valores vizinhos serão aqueles utilizados como base para aplicação do método da seção áurea.

Na Tabela 3.1, é possível notar que apenas os meses de janeiro, março e junho não obtiveram redução no MSE nesta única execução da ELM, o que demonstra a importância da regularização. Se considerarmos ainda que para a forma de abordagem deste trabalho interessa o erro acumulado de todos os meses, este processo torna-se ainda mais relevante.

Observe que os valores de C são de variadas magnitudes, tendo em vista que os dados apresentados ao combinador linear são diferentes, além de que os pesos da camada intermediária foram definidos de maneira aleatória para cada mês.

Para exemplificar o procedimento, utilizamos os valores do MSE para cada um dos 52 valores de $C=2^{\lambda}$ utilizando o mês de maio para a primeira busca. A reta horizontal é o valor do MSE sem regularização, ou seja com valor de C muito grande ou, idealmente, infinito. O intervalo selecionado foi $[2^{-4} - 2^{-2}]$, no qual aplicou-se o método proposto por Kulaif & Von Zuben (2013). No fim, o valor ótimo do expoente foi λ= -2.5960, compatível com a curva mostrada na Figura 3.7, na qual λ está no eixo das abcissas. Este valor é aquele que maximiza a capacidade de generalização da rede.

Observa-se ainda, que há um mínimo local que poderia vir a ser selecionado como ponto de máxima generalização, caso a busca unidimensional fosse aplicada à todo intervalo de busca.

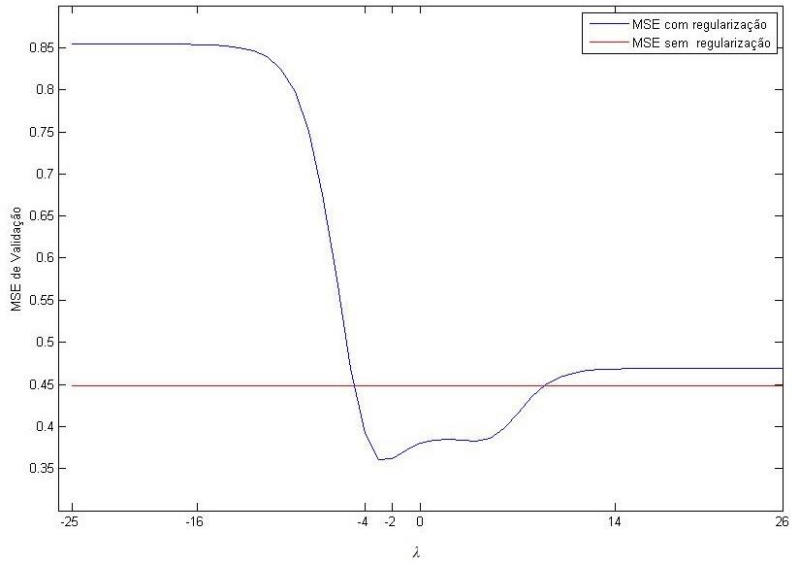


Figura 3.7 – Evolução do MSE de validação em função de *C*

3.6.2 Redes Neurais de Estado de Eco

A outra proposta de máquina desorganizada que utilizaremos neste trabalho é aquela que origina as Redes Neurais de Estado de Eco (*echo state networks* - ESN) (Jaeger 2001). Este tipo de rede neural, diferentemente das ELMs e MLPs, é recorrente, ou seja, possui laços de realimentação entre neurônios, o que lhe permite a formação de uma memória interna. Esta característica é conveniente em se tratando de tarefas dinâmicas ou com características temporais. Além disso, redes recorrentes possuem capacidade de aproximação universal (Schafer e Zimmermann 2007).

Em redes tipo *feedforward,* a informação está distribuída nas coordenadas espaciais que compõem o vetor de entradas, de forma que, para um mesmo padrão apresentado, a resposta da rede sempre será a mesma (Haykin 2008). Mas em problemas em que os dados tenham dependência temporal, como em previsão de séries temporais, pode ser eficaz tratá-los de forma a responder a esta dependência por meio da inserção de propriedades dinâmicas. Dessa maneira, este sistema passa a ter uma memória intrínseca que guarda informações sobre o sinal e pode utilizá-las nas previsões posteriores. Isto é notório se observarmos que respostas diferentes podem ser dadas pela rede para uma mesma entrada, a depender do estado atual em que ela se encontra (Haykin 2008).

Em termos simples, uma rede neural recorrente (RNN) clássica é um sistema dinâmico não-linear complexo criado pela existência dos laços de realimentação. As vantagens que este tipo de rede apresenta são, de certa forma, limitadas pela dificuldade no ajuste dos seus parâmetros livres. A aplicação de métodos de otimização, por exemplo, pode levar a configurações instáveis, além de haver a possibilidade de convergência local e uma dificuldade significativa em se manipular a função custo formada para treinamento supervisionado (Haykin 2008). Os algoritmos utilizados, em geral, têm convergência lenta e parâmetros de difícil definição, além de haver risco de desvanecimento do gradiente da função custo durante o processo (Jaeger 2002b). Dessa forma, apesar do grande poder de processamento que estas estruturas possuem, ajustá-las torna-se um processo dispendioso, o que inibe sua utilização em muitos casos.

As redes de estado de eco, por outro lado, apesar de serem recorrentes, diferem das redes clássicas no tocante à forma de construção da rede e ao processo de treinamento. A

camada intermediária aqui é chamada de reservatório de dinâmicas, na qual neurônios não-lineares são totalmente interconectados e os pesos das conexões são definidos de forma aleatória e mantidos fixos. Cada sinal que sai do reservatório é chamado de estado de eco e passa à camada de saída (ou de leitura), que irá produzir as saídas por meio de uma combinação linear. Dessa forma, o treinamento, semelhantemente ao que ocorre para as ELMs, resume-se a encontrar os coeficientes ótimos do combinador linear da camada de saída, o que pode, inclusive, ser resolvido por uma solução de mínimos quadrados (Jaeger 2001).

Um esquema da arquitetura de uma ESN pode ser visto na Figura 3.8, na qual se omitem, novamente, os pesos das conexões. Observe a relação direta com as ELMs, a menos das realimentações do reservatório de dinâmicas.

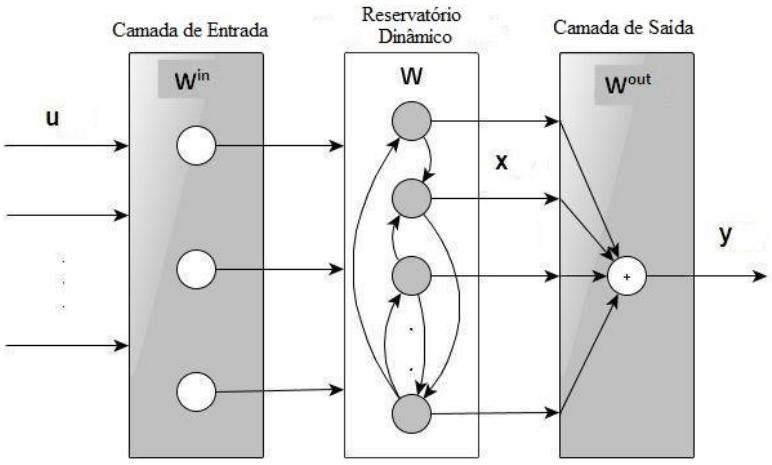


Figura 3.8 – Rede de Estado de Eco

Nesta figura, as entradas estão contidas no vetor $\mathbf{u}_t = [u_1, u_2, ..., u_{t-K+1}]^{\mathrm{T}}$. Elas são ponderadas linearmente pelos coeficientes da camada de entrada $\mathbf{W^{in}} \in \Re^{N \times K}$ e passam ao reservatório $\mathbf{W} \in \Re^{N \times N}$ de unidades não-lineares totalmente interconectadas, gerando as ativações $\mathbf{x}_t = [x_t^1, x_t^2, ..., x_t^N]^{\mathrm{T}}$, que fazem o papel dos estados da rede, sendo atualizados de acordo com a Equação (3.16):

$$\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{W^{in}}\mathbf{u}_{t+1} + \mathbf{W}\mathbf{x}_t) \tag{3.16}$$

na qual $\mathbf{f}(.) = (f_1(.), f_2(.), ..., f_N(.))$ representa as ativações dos neurônios do reservatório, *K* é o número de entradas, *N* é o número de neurônios no reservatório e *L* é o número de saídas da

rede. A inicialização dos estados desta rede é feita neste trabalho sempre com o valor nulo. A saída da rede, por sua vez, será o vetor $\mathbf{y}_{t+1}$:

$$\mathbf{y}_{t+1} = \mathbf{W}^{\mathbf{out}}\mathbf{x}_{t+1} \tag{3.17}$$

sendo $\mathbf{W}^{\mathbf{out}} \in \Re^{L\times N}$ a matriz que contém os pesos da camada de saída.

O processo de treinamento é semelhante ao das ELMs, que baseia-se em minimizar o erro quadrático médio entre a saída da rede e o sinal desejado $\mathbf{d} \in \Re^{L\times Ts}$. Novamente, lançaremos mão do operador de Moore-Penrose a partir dos estado de eco $\mathbf{X} \in \Re^{T_s\times N}$, sendo T_s o número de amostras de treinamento:

$$\mathbf{W}^{\mathbf{out}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{d} \tag{3.18}$$

Observa-se ainda que a matriz $\mathbf{W}$ precisa obedecer à chamada propriedade de estados de eco, a qual iremos discutir na próxima subseção.

3.6.2.1 Propriedade de estados de eco

A possibilidade de um processo de treinamento simples de uma rede recorrente é possível porque a ativação de cada um dos neurônios do reservatório de dinâmicas é, na verdade, uma transformação não-linear do histórico mais recente do sinal de entrada. Esta *propriedade de estados de eco* foi comprovada no trabalho pioneiro de Jaeger (2001), no qual o autor investigou a dinâmica de uma RNN e constatou que os estados $\mathbf{x}_t$ da rede são assintoticamente independentes da condição inicial para uma arquitetura semelhante àquela apresentada na Figura 3.8. Isto quer dizer, grosso modo, que, se esta rede for inicializada em dois estados iniciais diferentes $\mathbf{x}_0$ e $\mathbf{x}'_0$, e houver uma sequência temporal de entradas, os estados que resultam, $\mathbf{x}_n$ e $\mathbf{x}'_n$, convergem para valores similares. Dessa forma, é nítido que o efeito dos estados iniciais deixa de existir se tal propriedade for satisfeita, e a dinâmica do reservatório acaba por ser exclusiva do histórico recente de entradas, fazendo com que a rede possua estados de eco.

Segundo este trabalho, as condições para que seja válida a propriedade são, primeiramente, que os sinais de entrada sejam extraídos de um espaço compacto B e que os estados da rede estejam contidos em um conjunto compacto $\mathrm{A} \subset \Re^N$ de estados

admissíveis, o que quer dizer que a atualização de $\mathbf{x}_n$ sempre mantém os estados dentro de A se esta operação for realizada pela Equação (3.15).

De posse dessas definições, Jaeger demonstrou as condições suficientes para existência da propriedade de estados de eco. A primeira diz que se o módulo do máximo valor singular da matriz $\mathbf{W}$ estiver dentro do círculo real unitário, ou seja, $\left(\left|\sigma_{\max}(\mathbf{W})\right|\right) < 1$, a rede apresentará estados de eco. Esta propriedade é aplicável caso a RNN tenha neurônios com funções de ativação tipo tangente hiperbólica no reservatório e não apresente realimentações da saída da rede para o reservatório.

A segunda condição é definida em termos do maior autovalor em módulo da matriz de pesos internos $\mathbf{W}$, chamado de raio espectral, que deve ser menor do que 1, ou $\left(\left|r_{\max}(\mathbf{W})\right|\right) \leq 1$, sob a penalidade de não existirem estados de eco. Uma recomendação heurística que praticamente garante a existência dos estados de eco é que o autovalor de maior módulo seja menor do que 1 (Jaeger 2001).

Assim, o projeto de uma ESN depende bastante da forma como a matriz de estados de eco é definida, processo este que é realizado de forma independente da tarefa que a rede irá realizar. A camada intermediária desta rede neural tem a função de gerar o comportamento dinâmico que se busca numa rede recorrente, e considera-se desejável que este comportamento seja o mais diversificado possível.

Com isso, percebe-se que é necessário definir $\mathbf{W}$ de forma a obedecer à propriedade de estados de eco, definir $\mathbf{W}^{\mathbf{in}}$ de forma arbitrária e treinar a camada de saída conforme, por exemplo, a Equação (3.17). Assim, tem-se a essência do projeto de redes de estado de eco. Discutiremos agora formas de criação da matriz $\mathbf{W}$.

3.6.2.2 Projeto do reservatório de dinâmicas

O trabalho de Jaeger (2001) apresenta uma maneira simples de criar o reservatório de dinâmicas de maneira diversificada. A ideia básica é gerar uma matriz de pesos com certo grau de esparsidade, já que um padrão esparso de conexões favorece o desacoplamento de grupos de neurônios, induzindo o desenvolvimento de dinâmicas individuais e pouco relacionadas (Boccato 2013).

Uma das possibilidades apontadas pelo trabalho de Jaeger é definir os valores dos pesos das conexões que estão contidos na matriz $\mathbf{W}^{\mathbf{Je}}$ com base na regra a seguir:

$$\mathbf{W}^{\mathbf{Je}} = \begin{cases} 0{,}4 & \textit{com probabilidade} \quad 0{,}025 \\ -0{,}4 & \textit{com probabilidade} \quad 0{,}025 \\ 0 & \textit{com probabilidade} \quad 0{,}95 \end{cases} \tag{3.19}$$

Outra possibilidade de criação desta matriz foi proposta por Ozturk et al. (2007). Os autores partiram da premissa de gerar um reservatório rico do ponto de vista da entropia média dos estados de eco, e, para isso, propuseram uma estratégia na qual os autovalores respeitam uma distribuição uniforme no circulo unitário, criando uma matriz canônica da seguinte forma:

$$\mathbf{W}^{\mathbf{Oz}} = \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 & -r^N \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix} \tag{3.20}$$

na qual o termo *r* é o raio espectral.

A ideia contida nesta proposta é que, na ausência de informações *a priori* a sobre a saída desejada, é razoável espalhar de maneira uniforme estes autovalores na tentativa de buscar boas aproximações para quaisquer mapeamentos arbitrários. Os autores verificaram que, de fato, esta proposta produzia estados de eco $\mathbf{x}_t$ com entropia média maior que os obtidos com a estratégia de Jaeger (2001), para um conjunto de cenários de teste.

Discutidos os projetos do reservatório, trataremos agora de estratégias alternativas para criação da camada de saída. A ideia será aumentar o poder de mapeamento de uma ESN sem que se perca a característica principal, que é a simplicidade do processo de treinamento.

3.6.2.3 Outras propostas para camada de saída de uma ESN

A camada de saída de uma rede de estados de eco mapeia os estados de eco $\mathbf{x}_t$, que saem do reservatório de dinâmicas na resposta de saída da rede $\mathbf{y}_t$, que, sob a ótica do treinamento supervisionado, procurará aproximar o sinal desejado $\mathbf{d}_t$.

A utilização de um combinador linear como proposto por Jaeger (2001) tem os atrativos da simplicidade estrutural e da eficiência do processo de treinamento, por conta da obtenção da solução ótima de forma analítica, segundo a Equação (3.17). Todavia, é possível obter um incremento na capacidade de processamento por meio da utilização de camadas de saída não-lineares. Maass et al. (2002) propuseram a utilização de uma rede MLP como camada de saída de uma ESN por conta de sua maior capacidade de mapeamento, já que esta rede traz maior flexibilidade na aproximação do sinal vindo do reservatório com a saída desejada. O problema é que, neste caso, perde-se uma das características mais marcantes deste tipo de rede, que é exatamente a simplicidade no processo de treinamento, e nem sempre há redução no erro de teste (Boccato 2013).

Foi nesse contexto que Butcher et al. (2010) apresentaram uma arquitetura de ESN com uma rede neural tipo ELM como camada de saída. Ao criar uma rede híbrida com duas máquinas desorganizadas, o treinamento supervisionado continua sendo simples, e baseado no combinador linear da ELM apenas. O modelo que utilizaremos neste trabalho é mostrado na Figura 3.9:

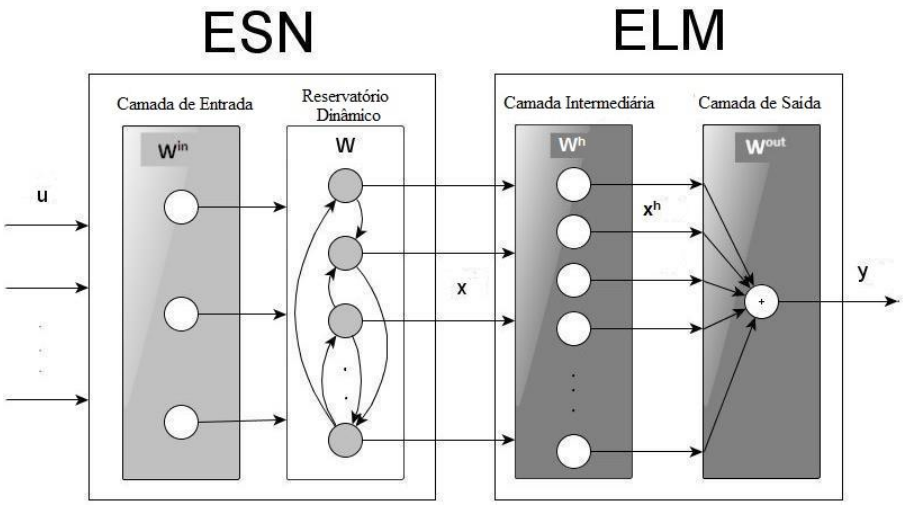


Figura 3.9 – Rede de Estado de Eco com uma ELM como camada de saída

Outra proposta que busca a manutenção da simplicidade no treinamento, mas com a inserção de não-linearidades no treinamento de uma ESN foi feita por Boccato et al. (2011a). Nela, os autores substituíram o combinador linear por um filtro de Volterra (Mathews e Sicuranza 2001), visando minimizar a limitação que a proposta linear apresenta em termos de não aproveitamento de informação estatística de ordem superior a dois. A ideia central baseia-se no fato de que esta estrutura é não-linear, mas emprega combinações polinomiais de modo linear com respeito aos parâmetros livres.

O filtro de Volterra pode ser definido, para uma saída única[2],em termos matemáticos de acordo com a expressão (3.21):

$$y_t = h_0 + \sum_{p=1}^{N} h_{1(p)} x_{p,t} + \sum_{p=1}^{N}\sum_{q=1}^{N} h_{2(p,q)} x_{p,t} x_{q,t} + \sum_{p=1}^{N}\sum_{q=1}^{N}\sum_{r=1}^{N} h_{3(p,q,r)} x_{p,t} x_{q,t} x_{r,t} + \ldots \tag{3.21}$$

onde $x_{i,t}$ denota a saída do *i*-ésimo neurônio da camada intermediária (ou o *i*-ésimo estado de eco) no instante *t* e h_m, m=1,…,M, são os coeficientes do combinador linear. O valor de *M* é um número inteiro que trunca a expansão polinomial na ordem desejada pelo projetista.

Como é possível observar, a simplicidade essencial está mantida no processo de treinamento, já que o objetivo será encontrar os temos $h_m(\cdot)$,ou seja, tem-se um filtro não-linear mas linear nos parâmetros a determinar. A forma de cálculo destes coeficientes é tipicamente baseada em mínimos quadrados, levando a

$$\mathbf{h} = (\mathbf{X}_\mathbf{v}^{\mathrm{T}} \mathbf{X}_\mathbf{v})^{-1} \mathbf{X}_\mathbf{v}^{\mathrm{T}} \mathbf{d} \tag{3.22}$$

sendo $\mathbf{X}_\mathbf{v} \in \Re^{T_s \times N_c}$ a matriz que contém os termos de Volterra associados aos estados de eco de ordem *M* para todas as T_s amostras de treinamento e **d** o vetor contendo as amostras do sinal desejado. Cada coluna da matriz $\mathbf{X}_\mathbf{v} = [\mathbf{X}_v^1, \mathbf{X}_v^2, \ldots \mathbf{X}_v^{Nc}]$, para o caso de *M*=3 e *N*=2, terá a seguinte forma:

$$\mathbf{X}_v^{Nc} = \begin{bmatrix} x_1 & x_2 & x_1 x_2 & x_1^2 & x_2^2 & x_1^3 & x_1^2 x_2 & x_1 x_2^2 & x_2^3 \end{bmatrix} \tag{3.23}$$

[2]É possível generalizar este filtro para uma estrutura com múltiplas saídas

Todavia, esta estrutura tem uma característica que pode atrapalhar seu desempenho: a "maldição da dimensionalidade". Ela pode ocorrer com o aumento do número de coeficientes a serem ajustados à medida que o número de estados de eco cresce. Para exemplificar este comportamento, observemos, na expressão (3.24) a relação entre número de estados de eco N e *kernels* não ambíguos N_{ker} a serem determinados (Boccato, Lopes, et al. 2012):

$$N_{\ker} = 1 + N + \frac{N(N+1)}{2} + \frac{N(N+1)(N+2)}{6} + \ldots \tag{3.24}$$

sendo que cada termo da adição representa uma ordem polinomial.

Para contrabalançar este problema, os proponentes desta lançaram mão de uma técnica muito conhecida de compressão de dados, a análise de componentes principais (*principal componente analisys* – PCA) (Hyvärinen, Karhunen e Oja 2001). Com isso, em vez de transmitir todos os N_{eco} estados de eco que saem do reservatório dinâmico, apenas $N_{pc} < N_{eco}$ combinações não-lineares destes estados fazem o papel de entradas do filtro. Esta ideia também é balizada no fato de que, reconhecidamente, há redundâncias nos estados de eco formados (Ozturk, Xu e Principe 2007). A princípio, esta característica justifica a aplicação do PCA sem grande perda de informação, e com redução significativa no número de coeficientes a serem determinados.

A base teórica de PCA pode ser descrita de forma sucinta. Considere-se a matriz $\mathbf{X} \in \Re^{N \times Ts}$, que contém as ativações dos neurônios do reservatório para T_s amostras de treinamento. Se o vetor que contem os estados de eco tem média zero, a matriz de covariância pode ser estimada de acordo com a expressão

$$\hat{\mathbf{O}} = \frac{\mathbf{X}\mathbf{X}^T}{T_s} \tag{3.25}$$

Tomando os autovalores e autovetores desta matriz, é possível construir uma nova matriz $\mathbf{Q} \in C^{N \times N_{pc}}$ que contém os N_{pc} autovetores organizados de forma que os autovalores $\hat{\mathbf{O}}$ respectivos estejam ordenados de forma crescente. Esta tarefa, na verdade, é equivalente a projetar os dados em direções ortogonais de forma que se reduza o MSE entre a projeção e os dados originais. Dessa maneira, tem-se um processo de compressão dos dados pela sua

redução de dimensionalidade. Esta matriz contém, em última análise, as direções de maior energia dos dados.

Assim, pode-se dizer que a primeira componente principal de um sinal é a projeção que leva à maior variância do sinal projetado, ou seja, aquela que preserva ao máximo o conteúdo de energia do mesmo. A segunda componente principal é necessariamente ortogonal à primeira, e se associa ao segundo maior autovalor. Logo, todas as N_{pc} componentes são ortogonais entre si e são ordenadas de acordo com seus respectivos autovalores. Um exemplo ilustrativo simples de aplicação desta técnica é apresentado na Figura 3.10:

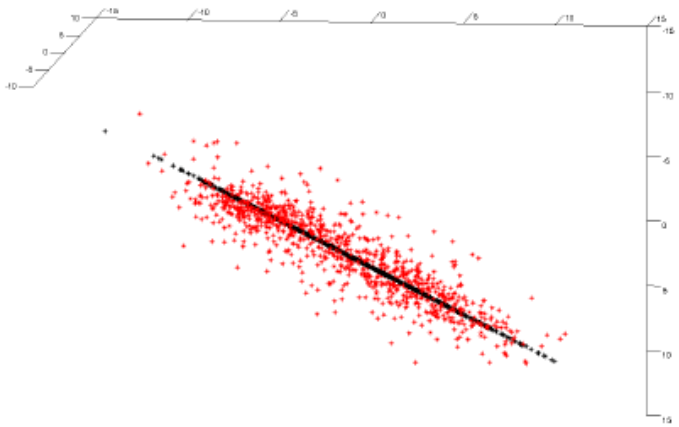


Figura 3.10 – Exemplo de aplicação do PCA

Neste exemplo uma nuvem de dados distribuídos em 3 dimensões foi projetada na direção da primeira componente principal, que segue a direção de maior energia, como é visualmente perceptível.

É possível, inclusive, medir o nível de erro quadrático médio amostral associado às projeções nas N_{pc} componentes principais. Para isso, deve-se partir do princípio de que o número de elementos do vetor de dados máximo é igual a N_{pc}, e $N_{pc}<N$. Pode-se calcular o fator α com os γ_i maiores autovalores de $\hat{\mathbf{O}}$ da seguinte forma:

$$\alpha = \frac{\sum_{i=1}^{Npc} \gamma_i}{\sum_{i=1}^{N} \gamma_i} \tag{3.26}$$

sendo que o valor de α estará no intervalo entre 0 e 1. Como exemplo, note que o valor desta variável, para o caso da Figura 3.10, é $\alpha = 0{,}9537$, o que mostra que apenas uma já preserva, em grande medida, o sinal original.

A arquitetura proposta é mostrada na Figura 3.11:

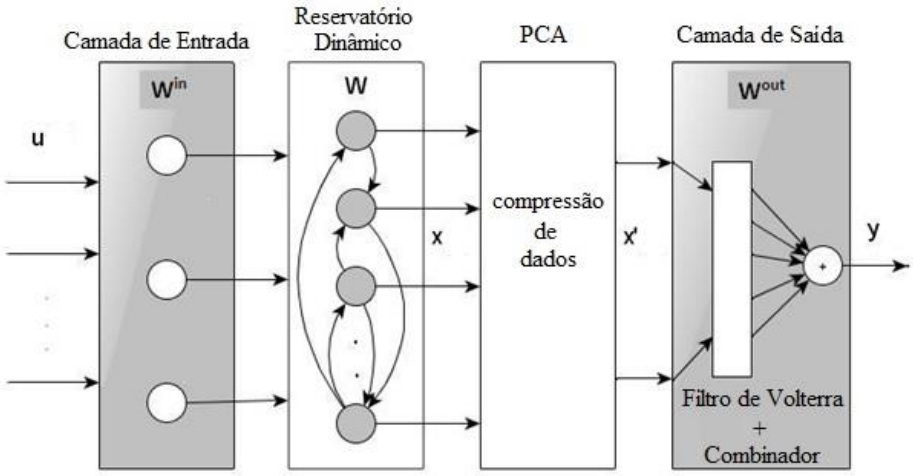


Figura 3.11– ESN com camada de saída baseada em PCA e filtro de Volterra

Comentários

Este capítulo apresentou os conceitos de máquinas desorganizadas, nome dado às arquiteturas de redes neurais do tipo Máquinas de Aprendizado Extremo (ELM) e Redes de Estado de Eco (ESN). A primeira é uma rede do tipo *feedforward,* enquanto a segunda é recorrente. O conceito por trás do epíteto apresentado está no fato de que as camadas intermediárias não são treinadas com sinal de referência, sendo geradas de forma aleatória.

O treinamento deste tipo de rede baseia-se em uma solução de mínimos quadrados, com coeficientes calculáveis por meio de solução analítica, o que confere grande eficiência computacional e simplicidade a este processo.

No caso das ESNs, foram apresentadas duas formas de se criar a camada escondida, denominada reservatório de dinâmicas, as abordagens de Jaeger (2001) e de Ozturk et al. (2007). Todavia é necessário obedecer a chamada propriedade de estados de eco para que a rede seja realmente um sistema dinâmico.

Quanto à camada de saída, foram introduzidas as propostas de Butcher et al. (2010) e de Boccato et al. (2012). A primeira sugere uma rede híbrida com uma ELM sendo a

camada de saída de uma ESN, enquanto a segunda troca o combinador linear por um filtro de Volterra precedido de um compressor de dados do tipo análise de componentes principais (PCA). Ambas as propostas visam elevar o poder de processamento não-linear da rede.

Além disso, também foram brevemente discutidas as conhecidas redes *multilayer perceptron* (MLP), que servirão de base de comparação as abordagens de máquinas desorganizadas.

No próximo capítulo, trataremos de alguns modelos de seleção de variáveis para definição do melhor conjunto de entradas para previsão de séries.

Capítulo 4. Métodos de Seleção de Entradas

A seleção de variáveis é etapa fundamental para obtenção de um modelo eficiente de previsão de séries temporais (Moon, Rajagopalan e Lall 1995). Dentre os benefícios potenciais de sua utilização, pode-se citar a facilitação na visualização e no entendimento dos dados, a redução da ordem e dos requisitos de memória para armazenamento dos mesmos e a redução no tempo de treinamento e no esforço computacional (Guyon e Elisseeff 2003). A seleção trabalha no sentido de identificar um subconjunto de entradas que auxiliem a previsão, o reconhecimento de padrões ou mesmo regressão de dados, fazendo com que o modelo alcance melhores resultados.

Identificar um sistema é tarefa influenciada por fatores como conhecimento prévio de suas características, complexidade, presença de ruídos e métrica de desempenho a ser utilizada. As entradas devem representar a dinâmica do sistema, o que auxilia na escolha de um modelo de previsão que seja adequado ao problema. A estrutura dos modelos lineares para realização de determinada tarefa é, em grande medida, definida pela quantidade de entradas que os compõem. Para o caso das redes neurais, o número de entradas impacta na determinação da sua estrutura, já que, quanto mais entradas houver, mais complexa a rede será e mais custoso o seu treinamento, sem garantia de melhoria de desempenho. Além disso, a quantidade de entradas influencia a superfície da função custo, que, com um número excessivo de entradas, tende a possuir mais mínimos locais.

As metodologias de seleção podem utilizar informações disponíveis *a priori*, por meio de testes empíricos de tentativa e erro, ou ainda de algum critério de informação. Um exemplo simples de como o processo geral funciona é descrito por Villanueva (2006). Tomemos um conjunto $\mathbf{V}$ que representa o espaço de varáveis de entrada, o qual limitaremos aqui a 3. Define-se, dessa forma, o vetor de entradas $\mathbf{V} = [v_1, v_2, v_3]$, com o qual é possível formar 2^3-1 = 7 subconjuntos de entradas de um determinado modelo, como explicitado na Tabela 4.1

Tabela 4.1 – Possíveis subconjuntos do vetor de entradas V

Subconjunto	Entradas pertencentes
1	v_1
2	v_2
3	v_3
4	v_1, v_2
5	v_1, v_3
6	v_2, v_3
7	v_1, v_2, v_3

O papel dos métodos de seleção é definir qual destes subconjuntos é o mais adequado para representar a informação contida nos dados, possivelmente em contraste com a adoção de todas elas. No caso de estudo, selecionar variáveis é escolher o subconjunto que permita a melhor previsão de valores futuros de uma série temporal, ou seja, escolher o vetor $\overline{\mathbf{V}} \in \Re^k$ dentre as possíveis combinações entre as variáveis de $\mathbf{V} \in \Re^l$, tal que $k \leq l$. Este conjunto representará a estrutura de dependência de um processo estocástico ao longo do tempo.

Em Yu & Liu (2004), os autores apresentam alguns critérios que relacionam-se a este procedimento:

i) Relevância: conceito que se associa à importância que determinada variável pode ter para o problema, uma vez que a informação que ela contém será base do processo de seleção. A relevância é forte ou fraca dependendo de quanto a sua remoção degrada o desempenho do preditor;

ii) Redundância: duas ou mais variáveis são redundantes caso seus valores observados sejam altamente correlacionados ou dependentes. O nível desta correlação revela o grau de redundância;

iii) Otimalidade: um subconjunto de variáveis de entrada é chamado de ótimo caso não haja outro subconjunto que produza melhores resultados do que aqueles alcançados poe ele.

Estas características, no entanto, se combinadas, não possuem implicação direta. Por exemplo, uma variável ser relevante não significa que o subconjunto ótimo a contenha. Da mesma forma, estar no subconjunto ótimo não significa que determinada entrada seja relevante (Villanueva 2006).

De acordo com Guyon & Elisseeff (2003), os procedimentos para seleção de variáveis são classificados em *embedded, wrappers* e *filtros.* Cada um deles tem pontos positivos e negativos, e as particularidades de determinado problema indicará qual deles será mais adequado. Vamos discutir cada uma dessas metodologias nas próximas seções.

Antes, porém, é importante pontuar rapidamente a diferença entre selecionar variáveis e atributos. O entendimento de *atributo* é atrelado à ideia de um conjunto de entradas que é formado a partir de uma combinação das variáveis originais ou da extração de alguma característica importante. Um exemplo de seleção de atributos foi exposto no Capítulo 3: trata-se da técnica de PCA que combina sob alguns critérios a os dados do sinal que sai do reservatório dinâmico de uma rede de estado de eco para a camada de saída, conforme proposta de Boccato et al. (2012). Observe que, neste caso, os estados de eco são substituídos por um conjunto de combinações lineares de menor dimensão. Dessa forma, há um conjunto de novas variáveis que estão em um novo espaço que em essência difere da seleção de variáveis, já que, neste caso, o sub-conjunto é formado pelas das entradas originais que não sofrem nenhum tipo de transformação.

4.1 Embedded

Métodos do tipo *embedded* são aqueles nos quais a seleção de variáveis está implícito no processo de síntese do modelo. Esta abordagem é marcante em aplicações de classificadores do tipo máquinas de vetores suporte (*support vector machines* – SVM), nas quais a seleção sai como resultado da resolução do problema de programação quadrática associado. Algumas arquiteturas de ELMs também possuem esta característica. A Figura 4.1 exemplifica esta ideia.

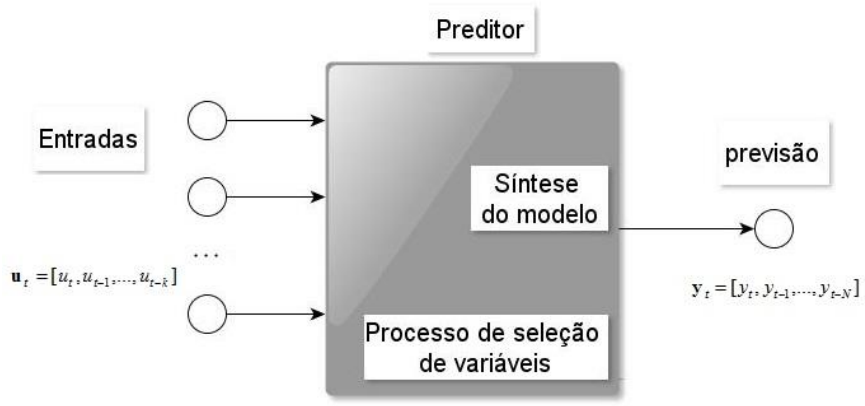


Figura 4.1 - Esquemático do método *embedded*

4.2 Wrappers

Na abordagem tipo *wrapper,* a principal característica é a interação entre o mecanismo de seleção de variáveis e o modelo de previsão. Uma vez que o modelo já tenha sido ajustado, o *wrapper* irá avaliar, por meio de algum critério de desempenho, cada um dos subconjuntos formados para solução do problema (Villanueva, Santos e Von Zuben 2006).

Todavia, o custo computacional envolvido é elevado, pois o modelo precisa ser treinado para cada subconjunto candidato de entradas. Portanto, recomenda-se a seleção via *wrapper* para casos em que o número de amostras seja reduzido. A Figura 4.2 apresenta o esquema do método.

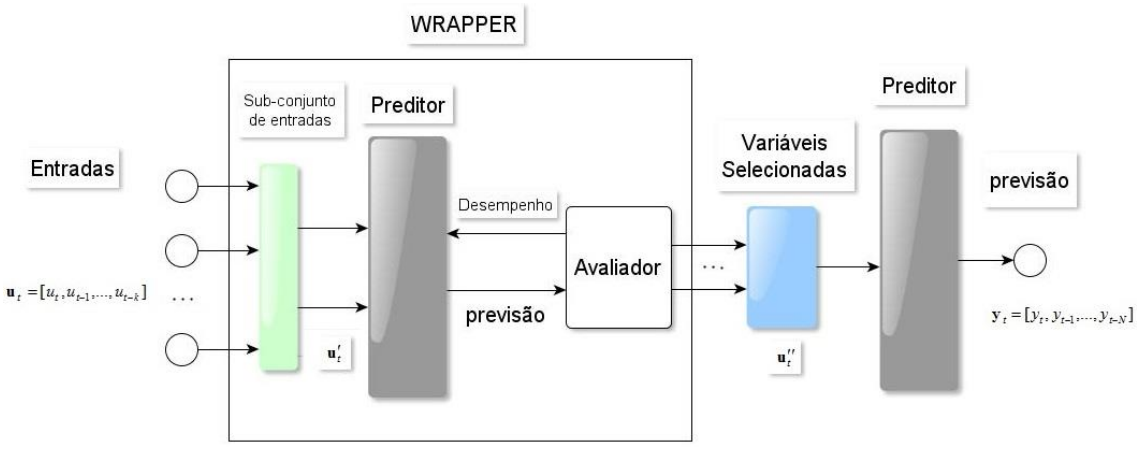


Figura 4.2 – Esquemático do método *wrapper*

O funcionamento do *wrapper*, como mostra a figura, dá-se da seguinte forma: primeiro o conjunto das entradas $\mathbf{u}_t$ é subdividido em conjuntos menores $\mathbf{u}'_t$; em seguida, o preditor é treinado e executado para cada um destes; após a etapa de treinamento e previsão, uma medida de desempenho presente no bloco *avaliador* é gerada para cada subconjunto. Aquele que apresentar melhor resultado de avaliação será utilizado para a previsão final.

É possível não utilizar o último preditor mostrado na figura, bastando que o resultado de cada avaliação seja armazenado, tal qual cada métrica de erro necessária. Todavia, por uma questão de simplicidade no entendimento, optamos por apresentar a seleção como tarefa disjunta da previsão final.

4.2.1 Seleção progressiva

O custo computacional de se fazer uma busca exaustiva sobre todas as possibilidades de conjuntos de entradas pode tornar até mesmo um problema relativamente pequeno impraticável, já que o custo computacional é fatorial. Uma das propostas para contornar este problema é o método de *seleção progressiva*. Esta metodologia estabelece uma forma de construir os subconjuntos de entradas considerando cada uma delas individualmente.

O procedimento inicia-se com um subconjunto vazio, e cada variável será comparada com todas as demais: aquela que apresentar o melhor desempenho medido pelo avaliador será selecionada, ou em termos de melhoria do resultado ou em termos da menor deterioração deste valor. Selecionada a primeira entrada, que será fixada no subconjunto, as que não foram selecionadas são avaliadas para ser a segunda entrada. Este procedimento repete-se até a avaliação de todas as V variáveis. O subconjunto final será aquele que apresentar o melhor resultado geral.

A Figura 4.3 apresenta esta ideia na previsão do mês de outubro da série de Emborcação, com ajuste de uma rede neural tipo Máquina de Aprendizado Extremo com um número fixo de 20 neurônios na camada oculta e máximo de 10 atrasos como entrada do modelo. Como é possível perceber, foram selecionadas 3 entradas nesta ordem: 4, 10 e 6 - já que esta foi a combinação que apresentou menor erro quadrático médio. Para este caso,

é perceptível que as entradas selecionadas não necessariamente são consecutivas e que o aumento do número de entradas não necessariamente acarreta melhora no desempenho.

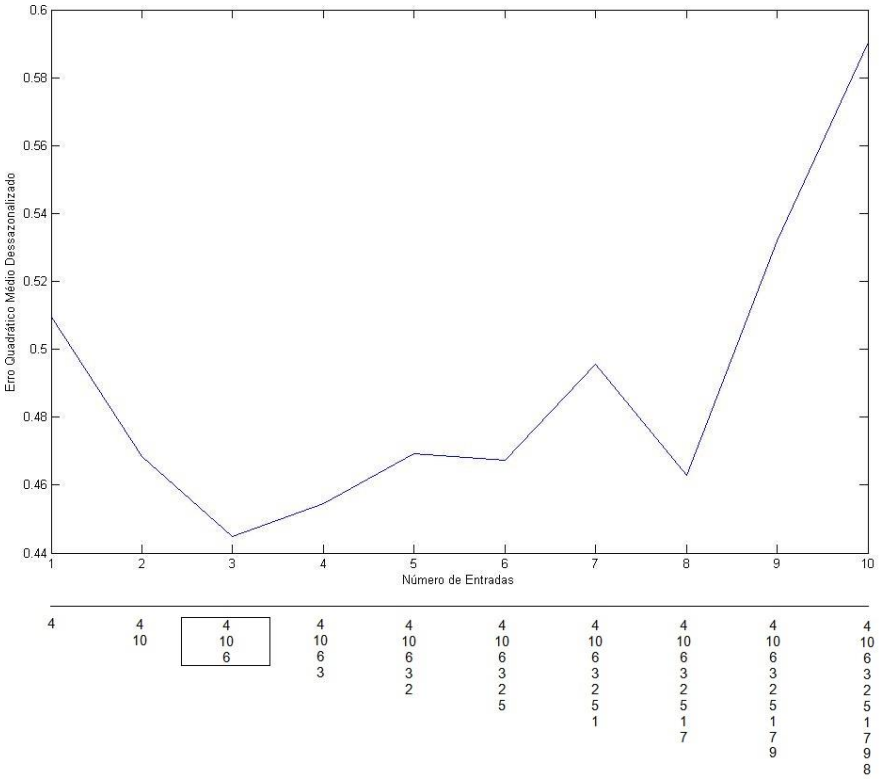


Figura 4.3 – Comportamento do erro com método de Seleção Progressiva

A quantidade de subconjuntos formados neste caso é igual ao número *V* de entradas, e o número de vezes que o preditor precisa ser treinado obedece a $V(V+1)/2$. Como, no exemplo, $V=10$, 55 ELMs foram treinadas.

Também é interessante observar o comportamento do MSE entre as iterações 3 e 4 da Figura 4.3: ao acrescentar-se a entrada v_4, o erro aumenta, deteriorando o valor da função objetivo a ser minimizada. Todavia, com a inserção da variável v_6, o erro diminui. Este comportamento acontece porque a busca corre o risco de cair em mínimos locais, o que pode ser contornado em momentos posteriores. Assim, caso seja possível, é sempre aconselhável executar o procedimento de forma completa (Villanueva 2006).

4.2.2 Funções de Avaliação

Após a discussão da forma geral de funcionamento do método *wrapper*, é preciso definir um critério de avaliação da qualidade do ajuste ou previsão do modelo em estudo,

utilizando os subconjuntos anteriormente definidos. Esta etapa corresponde ao bloco *Avaliador* da Figura 4.2.

O critério mais simples é utilizar alguma métrica de erro que, a cada conjunto ajustado, mostre a média das diferenças entre o dado desejado e o previsto. A proposta utilizada por (Villanueva, Santos e Von Zuben 2006) é o erro absoluto médio (MAE), dado por:

$$MAE = \frac{1}{N_s}\sum_{t=1}^{Ns}\left|x_t - \hat{x}_t\right| \tag{4.1}$$

na qual x_t é o dado observado no instante *t,* $\hat{x}_t$ é o dado previsto pelo subconjunto selecionado e N_s o número de amostras previstas.

Outra possibilidade é utilizar o erro quadrático médio (MSE), métrica mais comumente empregada como função custo no treinamento de redes neurais *feedforward* e na estimação de parâmetros de modelos AR, como discutido nos capítulos anteriores. Esta medida é definida pela Equação (4.2):

$$MSE = \frac{1}{N_s}\sum_{t=1}^{Ns}(x_t - \hat{x}_t)^2 \tag{4.2}$$

Note que estes critérios apenas consideram o resultado final do ajuste, independentemente do número de entradas que levam à performance em questão. Os resultados apresentados na Figura 4.3 foram realizados com o uso do MSE, como comentado.

Todavia, existem outros tipos de funções avaliadoras que procuram penalizar a quantidade de entradas de um modelo com vistas à seleção de subconjuntos de entradas mais parcimoniosos. Critérios muito utilizados em tarefas desse tipo são baseados em medidas de informação (Ehlers 2007).

O Critério de Informação Bayesiano (BIC) foi proposto por Schwarz (1978). Ele é baseado em métricas de correlação linear, e vincula-se às ordens ótimas do modelo de previsão, sendo definido como

$$BIC = N\ln(\hat{\sigma}_a^2) + p\ln(N) \tag{4.3}$$

na qual N é o número de observações, p é a ordem ou número de entradas do modelo (ou ainda a ordem do modelo auto-regressivo) e $\hat{\sigma}_a^2$ a variância estimada do ruído branco (ou resíduo) $a_t = e_t$, como definido no Capítulo 3. Assim, o *wrapper* escolherá o conjunto de entradas que apresentar o menor valor *BIC*. De forma análoga, o Critério de Informação de Akaike (AIC) (Akaike 1978) é definido por

$$AIC = N \ln(\hat{\sigma}_a^2) + 2p \tag{4.4}$$

Ambos podem assumir valores diversos, inclusive negativos.

Nos dois casos, fica patente que há uma penalização em relação ao número de entradas, de forma que a inclusão de uma delas no subconjunto depende não apenas da melhora do desempenho, mas também de quanto ele é incrementado. Dessa forma, procura-se um conjunto que seja tanto eficiente quanto parcimonioso.

A diferença entre os critérios reside no fato de que o BIC penaliza mais fortemente a inclusão de entradas no subconjunto que o AIC, pois o segundo termo da soma em (4.3) o logaritmo natural do número de observações, enquanto em (4.4) há uma multiplicação por 2.

Existem estudos comparativos entre os dois métodos, procurando analisar qual deles seria mais adequado para seleção de variáveis em diferentes contextos. Em Emiliano et al. (2010) eles são utilizados para definir ordens de modelos lineares de séries temporais como o auto-regressivo, médias móveis e auto-regressivo e médias móveis. Em Santurio & Gomes (2011), tais propostas são aplicadas para determinar a ordem de modelos PAR na previsão de séries hidrológicas. A conclusão, em ambos os casos, é que o desempenho acaba por ser parecido, com ligeira vantagem para o BIC.

Neste trabalho, utilizaremos como avaliador o MSE, BIC e AIC para seleção das entradas de redes neurais e modelos lineares. Tratemos agora da seleção de variáveis por meio dos filtros.

4.3 Filtros

O método de seleção tipo filtro é baseado apenas nos dados disponíveis e independe do modelo preditor que posteriormente será utilizado. As variáveis são escolhidas por meio

de medidas de correlação linear ou não-linear entre as observações. A principal vantagem deste método é sua generalidade, por não ser necessário sintetizar o preditor, o que tende a torná-lo mais eficiente computacionalmente que os *wrappers*.

Todavia, como se trata de um passo prévio, o conjunto ótimo de entradas pode não ser selecionado, já que não há interação com o modelo de previsão. Quer dizer, métricas que se baseiam na dependência entre as amostras podem ser úteis, mas insuficientes para garantir que o conjunto escolhido seja o melhor possível. Portanto, recomenda-se sua utilização para problemas com grande quantidade de dados disponível, pois, se por um lado, o critério de otimalidade não for alcançado, o custo computacional deverá ser compensador.

Na Figura 4.4, está o esquema do método tipo filtro, no qual evidenciamos que um conjunto de possíveis entradas contidas no vetor $\mathbf{u}_t$, após serem selecionadas, estarão contidas em um vetor $\mathbf{u}'_t$ de menor ou igual dimensão. A previsão é realizada, então, com este último, sendo o conjunto de entradas do preditor para um determinado dado de saída no instante *t*. Para previsão de séries, em geral, a saída é única.

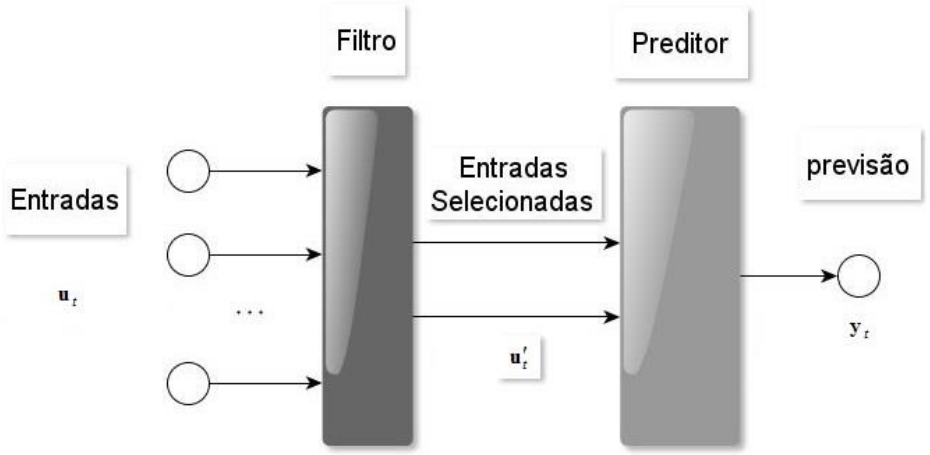


Figura 4.4 – Esquemático do modelo Filtro para seleção de variáveis

Tratemos agora de alguns importantes modelos de filtros.

4.3.1 Função de Autocorrelação Parcial

A função de autocorrelação parcial (FACP) (Maceira 1989) é uma ferramenta muito utilizada na identificação da ordem de modelos lineares de séries temporais, sendo,

inclusive, a técnica implementada para seleção de entradas em modelos lineares de simulação pelo Operador Nacional do Sistema Elétrico (CEPEL 2001a). A definição do coeficiente de correlação parcial tem relação direta com modelos auto-regressivos e com as equações de Yule-Walker, como será discutido a seguir.

Definição 4.1: O coeficiente de autocorrelação parcial de ordem *k* é o último coeficiente de um modelo AR(*k*), ajustado para uma série temporal x_t e denotado por ϕ_{kk} .

Isto quer dizer que um processo auto-regressivo de ordem *p* terá ϕ_{kk} diferente de zero para *k* menor ou igual a *p,* e será igual a zero para *k>p.* Partindo-se deste pressuposto e retomando as equações de Yule-Walker, a relação entre as estimativas da autocorrelação de uma série temporal nestes termos obedecerá o seguinte conjunto de equações:

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{k-2} + ... + \phi_{k(k-1)}\rho_{j-k+1} + \phi_{kk}\rho_{j-k}\,, \qquad j=1, 2, \ldots, k. \tag{4.5}$$

Ou, de forma matricial:

$$\mathbf{P}_{kp} = \begin{bmatrix} 1 & \rho_1 & \ldots & \rho_{p-1} \\ \rho_1 & 1 & \ldots & \rho_{p-2} \\ \ldots & \ldots & & \ldots \\ \rho_{p-1} & \rho_{p-2} & \ldots & 1 \end{bmatrix} \qquad \boldsymbol{\rho}_{kp} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \ldots \\ \rho_p \end{bmatrix} \quad \boldsymbol{\Phi}_{kp} = \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \ldots \\ \phi_{kp} \end{bmatrix} \tag{4.6a}$$

$$\boldsymbol{\Phi}_{kp} = \mathbf{P}_{kp}^{-1}\boldsymbol{\rho}_{kp} \tag{4.6b}$$

Dessa forma, deve-se ajustar *k* modelos AR(*p*) de ordem *p* = 1, 2,...*k* para encontrarmos os coeficientes ϕ_{kk} . Os coeficientes dos dois primeiros modelos ajustados serão dados por:

AR(1): $\rho_1 = \phi_{11}\rho_0$

$$\mathbf{P}_{1p} = [\rho_0], \quad \boldsymbol{\rho}_{1p} = [\rho_1]\,, \quad \boldsymbol{\Phi}_{1p} = [\phi_{11}]$$

Da onde sai $\phi_{11} = \rho_1$, o primeiro coeficiente de autocorrelação parcial. De forma similar

AR(2): $\begin{vmatrix} \rho_1 = \phi_{21}\rho_0 + \phi_{22}\rho_1 \\ \rho_2 = \phi_{21}\rho_1 + \phi_{22}\rho_0 \end{vmatrix}$

$$\mathbf{P}_{2p} = \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}, \qquad \boldsymbol{\rho}_{2p} = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}, \qquad \boldsymbol{\Phi}_{2p} = \begin{bmatrix} \phi_{21} \\ \phi_{22} \end{bmatrix}$$

Isolando ϕ_{21} e igualando as equações, tem-se $\phi_{22} = \dfrac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$, valor que será admitido como o segundo coeficiente de autocorrelação parcial.

Veja que os coeficientes de autocorrelação ρ_p são dados do problema, calculados como disposto no Capítulo 2. A FACP de uma série, então, pode ser estimada através de sucessivos ajustes dos modelos auto-regressivos, determinando as ordens mais apropriadas de um modelo AR. De forma prática: ajusta-se o modelo AR(1), de onde estimamos o coeficiente ϕ_{11}. Em seguida, ajustamos o modelo AR(2), e temos ϕ_{21} e ϕ_{22}, sendo que nos interessa o último. Continuamos dessa maneira sistemática até o ajuste da ordem *k* requerida, de onde saem os coeficientes ϕ_{kk} desejados.

Para uma série temporal, procura-se a maior ordem tal que todas as estimativas ϕ_{kk} para $k>p$ não sejam significativas. A ordem do modelo será igual ao valor correspondente à entrada selecionada, ou seja, se forem selecionados os coeficientes ϕ_{11} e ϕ_{55}, os atrasos 1 e 5 faram parte do subconjunto selecionado para previsão.

Em Quenouille (1949), o autor mostrou que, para um processo AR(*p*), os coeficientes ϕ_{kk} estimados para ordens superiores a *p* têm distribuição gaussiana com média igual a zero, variância igual a $VAR[\phi_{kk}] \cong 1/N$, e, consequentemente, desvio padrão $DP[\phi_{kk}] \cong \sqrt{1/N}$, sendo *N* o número de amostras. Assim, o limiar de confiança para os coeficientes baseado no desvio padrão será $|\phi_{kk}| \cong 2/\sqrt{N}$, considerando que a estimativa é diferente de zero neste intervalo.

Como visto, o método pode selecionar como entradas do modelo atrasos não consecutivos. Por exemplo, se $\mathbf{V} = [v_1, v_2, v_3, v_4]$, ele pode selecionar $\mathbf{V} = [v_1, v_4]$, o que implica dizer que ϕ_{11} e ϕ_{44} foram significativos, enquanto ϕ_{22} e ϕ_{33} não. Todavia, Stedinger (2001) afirma que, em previsões de séries mensais de hidrologia (como as séries de vazões), não faz sentido que uma determinada amostra tenha relação com atrasos não

consecutivos, ou, no nosso caso, que um mês *m* tenha dependência com a vazão *m*-2 e não com *m*-1. Tomando como exemplo um modelo AR(6), se a FACP definir que são significativas apenas as entradas ponderadas pelos coeficientes ϕ_{11} e ϕ_{44}, a última será tratada como dado espúrio, devendo ser descartada. Isso significa dizer que valores intermediários não devem ser considerados. O trabalho de Oliveira & Souza (2011) utilizou técnicas de *boostrapping* para avaliação da melhor ordem de modelos PAR e chegou a mesma conclusão de Stedinger, com ordens semelhantes para séries de vazões.

A Figura 4.5 é relativa ao cálculo da FACP dos meses de janeiro da série de Furnas. O traço horizontal é o limiar de confiança calculado em função do desvio padrão. Observe que, com 12 atrasos, o método sugere que sejam selecionadas as ordens 1, 5 e 7 de um modelo AR. Caso leve-se em consideração a proposta de Stedinger (2001), apenas o atraso 1 será selecionado.

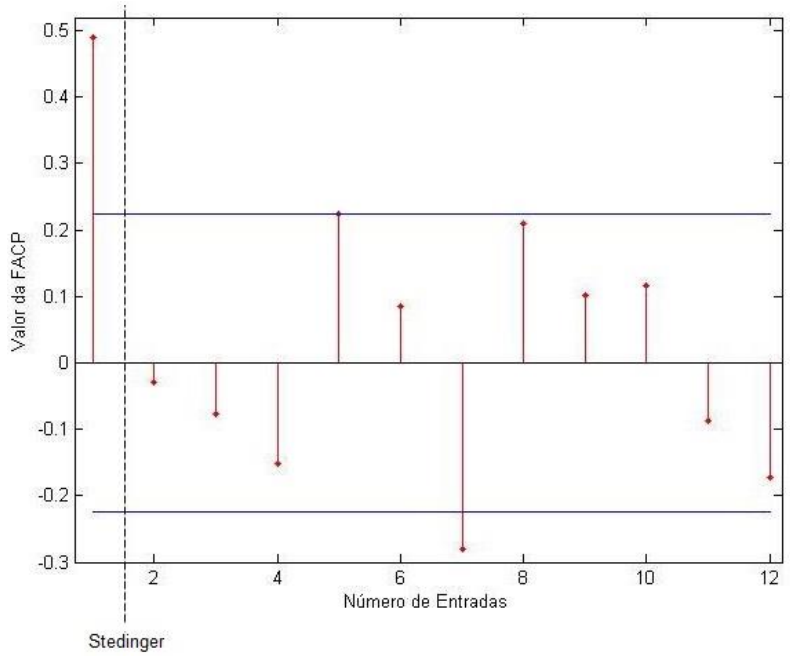


Figura 4.5 – Exemplo de Função de Autocorrelação Parcial

Este método poderia ser encarado como um tipo de *wrapper* se o preditor em questão fosse um modelo AR, mas, como podemos usar os subconjuntos por ele selecionados para quaisquer outros preditores, é mais conveniente encará-los como filtros por uma questão de coerência.

4.3.2 Informação Mútua

A relação de dependência entre duas variáveis é uma base importante para a seleção de entradas de um modelo. Neste contexto, um critério sólido pertencente ao escopo da teoria da informação pode ser utilizado: a informação mútua (*mutual information* – MI) (Bonnlander e Weigend 1994).

A MI é uma métrica que fornece uma medida do grau de dependência entre variáveis, ou seja, reflete a quantidade de que as vincula. Dessa maneira, pode ser aplicada como critério para selecionar as entradas de modelos não-lineares, como as redes neurais.

A definição de informação mútua entre duas variáveis aleatórias pode ser interpretada como uma medida de proximidade entre a distribuição de probabilidade conjunta das variáveis *x* e *y* e o produto das suas distribuições marginais. Matematicamente, temos:

$$MI = \int f_{xy}(x,y) \log\left(\frac{f_{xy}(x,y)}{f_x(x) f_y(y)} \right) dxdy \tag{4.7}$$

na qual $f_{xy}(x,y)$ é a função densidade de probabilidade conjunta (*probability density function* - PDF) e $f_x(x)$ e $f_y(y)$ são as respectivas funções de densidade marginais. Este critério terá valor zero para variáveis independentes e maior que zero caso contrário.

Caso se disponha de amostras dos dados, pode-se estimar (4.7) como (Luna, Ballini e Soares 2006)

$$MI = \frac{1}{N} \sum_{i=1}^{N} \log_e \left[\frac{f_{xy}(x_i, y_i)}{f_x(x_i) f_y(y_i)} \right] \tag{4.8}$$

na qual (x_i, y_i) é o *i*-ésimo par de dados do conjunto de amostras, com $i=1, 2,..., N$.

A dificuldade neste caso é estimar as probabilidades, já que, na prática, as distribuições são quase sempre desconhecidas. Isso cria uma dificuldade pois tal estimativa pode requerer grande quantidade de dados.

Existem variadas maneiras de se aproximar as PDFs. Neste trabalho, utilizaremos as funções de *kernel* do tipo distância absoluta ou *city-block,* que, inclusive, já foram aplicadas

em séries de vazões em (Luna, Ballini e Soares 2006). A opção por esta proposta se justifica em termos de simplicidade computacional, além de não se assumir nenhum tipo de distribuição dos dados *a priori*. Para defini-las, considere o conjunto de dados de entrada e saída $[\mathbf{x}_k, y_k]$, com *k*=1, 2,..., *N*. A aproximação das densidades de probabilidade de uma variável **x** unidimensional via estimadores não paramétricos de *kernel* é:

$$\hat{f}_x = \frac{1}{N\lambda}\sum_{i=1}^{N} K\left[\frac{x - x_i}{\lambda}\right] = \frac{1}{N}\sum_{i=1}^{N} K_\lambda\left(x - x_i\right) \tag{4.9}$$

sendo $K_\lambda(t)$ a função de *kernel* e λ a largura de banda ou parâmetro de dispersão.

A função densidade de probabilidade marginal de **x** aproximada é (Luna 2007):

$$\hat{f}_x(\mathbf{x}) = \frac{1}{N(2\lambda)^p}\sum_{i=1}^{N}\exp\left[-\frac{1}{\lambda}\sum_{i=1}^{p}\left|\mathbf{x}_j - \mathbf{x}_{ij}\right|\right] \tag{4.10}$$

sendo *p* é a dimensão de **x**. Esta equação surge de (4.9) como um caso adaptado para **x** multidimensional e para a função *city-block.* O parâmetro λ é dado por:

$$\lambda = \left(\frac{4}{p+2}\right)^{1/(p+4)} N^{-1/(p+4)} \tag{4.10b}$$

É necessário observar que λ, da forma como está definido, pressupõe distribuição normal dos dados, que, embora não ocorra para todos os casos, é bastante aceita na literatura (Bowden, Maier e Dandy 2005) pela sua simplicidade e eficiência.

Por fim, define-se a probabilidade conjunta de **x** - *y*, sendo a última uma saída unidimensional (Akaho 2002):

$$\hat{f}_{xy}(\mathbf{x}, y) = \frac{1}{N(\lambda)^{p+1}}\sum_{i=1}^{N} K\left(\frac{\mathbf{x} - \mathbf{x}_j}{\lambda}\right)K\left(\frac{y - y_j}{\lambda}\right) \tag{4.11}$$

Ou, ainda,

$$\hat{f}_{xy}(\mathbf{x}, y) = \frac{1}{N(\lambda)^{p+1}}\sum_{i=1}^{N}\exp\left[-\frac{1}{\lambda}s_i\right] \tag{4.11b}$$

De forma que s_i é calculado por:

$$s_i = \sum_{j=1}^{p} \left| \mathbf{x}_j - \mathbf{x}_{ij} \right| + \left| y - y_j \right| \tag{4.11c}$$

Um exemplo que mostra a capacidade de aproximação desta proposta é utilizá-la para construir uma função de distribuição gaussiana bi-variável, conforme explicita (Luna 2007). Esta função é definida pela Equação (4.12a):

$$f_{x,y} = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp(-\Gamma) \tag{4.12a}$$

onde

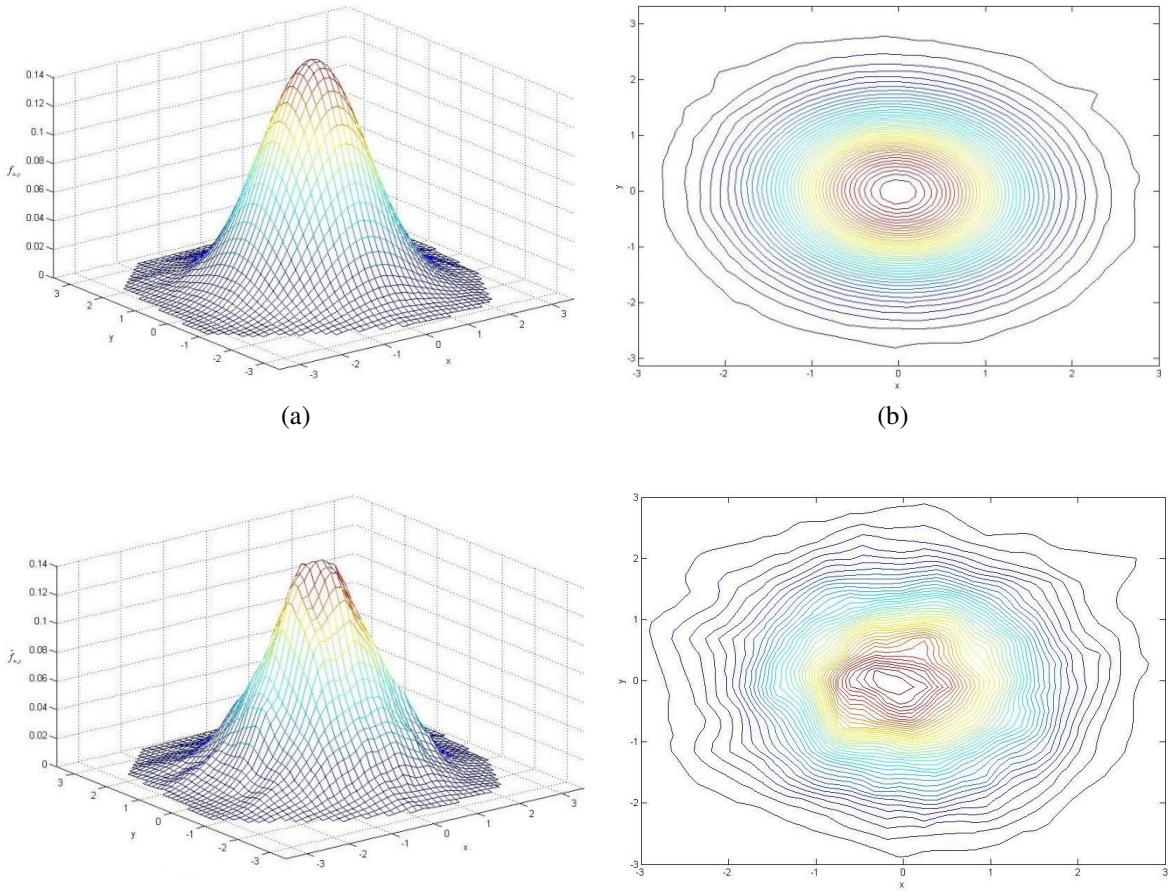
$$\Gamma = \frac{1}{2}(1-\rho^2)\left[\frac{(x-\mu_x)}{\sigma_x^2} + \frac{(y-\mu_y)}{\sigma_y^2} + 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x^2\sigma_y^2}\right] \tag{4.12b}$$

sendo *x* e *y* as variáveis utilizadas, μ_x e μ_y suas respectivas médias, σ_x e σ_y os desvios padrões e ρ o coeficiente de correlação entre elas.

Gerando 2000 amostras das variáveis *x* e *y* com distribuição normal de média zero, e utilizando (4.12), é possível chegar a $f_{x,y}$, representada graficamente na Figura 4.6a juntamente com seu diagrama em curvas de nível (4.6b). Em paralelo, apresentamos as aproximações pela função de *kernel* tipo *city-block* da Equação (4.11) na Figura 4.6c e 4.6d. É nítida a proximidade das curvas, o que ilustra a capacidade de aproximação desta função. O coeficiente de correlação entre elas é de 0,9932 apesar de os círculos não serem perfeitamente concêntricos.

A etapa final é definir um limiar de confiança que determina se uma certa entrada vai pertencer ao subconjunto selecionado. Uma possibilidade é estabelecer um valor mínimo para o MI e rejeitar as entradas que apresentarem um valor abaixo dele. Outra opção é utilizar uma técnica de *bootstrapping* ou reamostragem, que consiste na construção de um teste de hipóteses que utiliza *p* sequências diferentes de *x* em relação a *y,* nas quais se reordena a variável independente e obtém-se um vetor de MIs. Quer dizer, o limiar de confiança que define se determinada entrada é relevante é obtido admitindo-se

independência entre as variáveis de entrada e saída (Luna, Ballini e Soares 2006). Esta última será adotada neste trabalho.



(a) (b)

(c) (d)

Figura 4.6 – Função gaussiana bi-variável original (a e b) e aproximada (c e d)

O processo se dá com a reordenação de forma crescente dos valores de MI calculados e agregados em um vetor. O valor que corresponde ao γ-ésimo percentil será o limiar para aceitar ou não a hipótese de independência entre as variáveis. Assim, determinada entrada será aceita no subconjunto caso seu valor de MI seja superior a um nível de significância pré-estabelecido, em geral igual a α=1-γ. Em outras palavras a entrada será considerada relevante com α% de probabilidade de elas serem independentes.

O exemplo da Figura 4.7 é referente ao mês de janeiro da série de Furnas. Neste caso, adotou-se *p*=100 sequencias, γ=95% e consequentemente α=5%, sendo calculados os

valores do MI respectivos. Na figura, fica evidente que, para 12 entradas, as selecionadas seriam 1, 8 e 9.

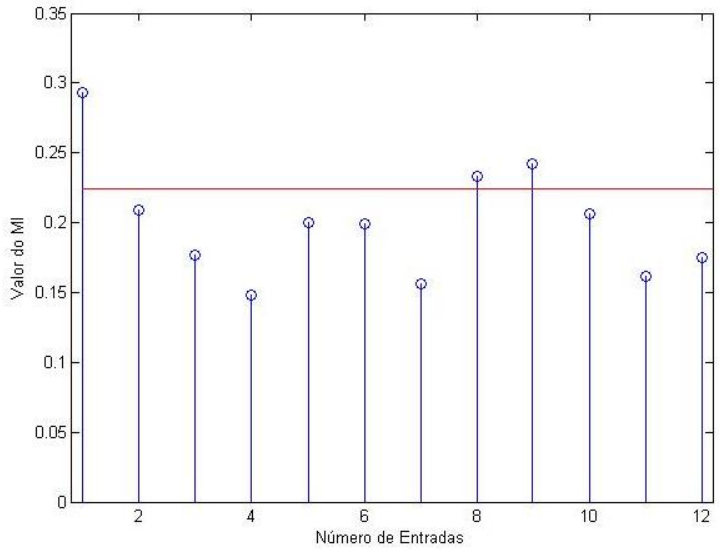


Figura 4.7 - Exemplo de valores da Informação Mútua

Na seção seguinte, apresentaremos alguns resultados utilizando as técnicas de seleção de variáveis tipo *wrapper* e filtros na previsão de séries de vazões.

4.4 Simulação de preditores com modelos de seleção de entradas

Nesta seção, estão sumarizados os resultados computacionais de simulações realizadas com um modelo linear e uma máquina desorganziada tipo Máquina de Aprendizado Extremo, utilizando duas das técnicas de seleção de variáveis discutidas neste capítulo: filtros e *wrappers*.

Primeiramente, é importante mencionar que partimos de alguns pressupostos antes das previsões serem feitas, tendo em vista a forma que as vazões mensais são tratadas pelo Sistema Elétrico Brasileiro. As simulações que servem como entrada, como relatado no Capítulo 1, são realizadas com no máximo 6 atrasos e separa cada uma das séries em 12 modelos diferentes, um para cada mês do ano. Além disso, a técnica de dessazonalização adotada é do tipo padronização, como mostrado no Capítulo 2, levou a bons resultados.

As citadas séries de Furnas, Emborcação e Sobradinho foram novamente utilizadas. Como amostra de treinamento para ajuste de parâmetros, utilizou-se todo o histórico disponível, a menos do período compreendido entre os anos de janeiro de 2001 e dezembro

de 2010, utilizados para validação e regularização da ELM, e os seguintes intervalos como amostra de testes:

Furnas – 1967 a 1976;

Emborcação – 1951 a 1960;

Sobradinho – 1977 a 1986.

Cada um deles totaliza 120 dados mensais. Estes períodos foram escolhidos por apresentarem comportamentos hidrológicos distintos, possibilitando uma análise mais ampla dos resultados. Os preditores utilizados foram o modelo PAR e a rede ELM.

O objetivo destas simulações é procurar indícios de que tipo de modelo de seleção de entradas poderia ser mais adequado a uma metodologia linear e outra não-linear. Observe que os *wrappers* para o modelo PAR levam em conta a avaliação do ajuste do conjunto de treinamento, enquanto as ELMs o de testes uma vez que, no último caso, nos interessa o menor erro de generalização.

Os resultados estão sumarizados nas Tabelas 4.2 e 4.3 e, para a ELM são médias de 20 simulações. As abreviaturas "WR" e "FILT" indicam modelos de *wrapper* e filtros, respectivamente

Tabela 4.2 - Resultados Seleção de variáveis do modelo PAR

		Seleção de entradas	MSE	MAE	MSE dessaz.	MAE dessaz.
FURNAS	WR	BIC	8.2719	194.8392	0.4494	0.5178
		AIC	7.8476	192.1714	0.4211	0.5097
		WRAPPER-MSE	8.3483	201.3464	0.4557	0.5334
	FILT	FACPPe	7.6961	189.7886	0.4124	0.5064
		FACPPe- Sted.	7.8342	190.5114	0.4143	0.5070
		MI	8.0767	195.1510	0.4459	0.5222
EMBORC.	WR	BIC	4.6871	128.0953	0.6447	0.5651
		AIC	4.7530	125.0621	0.6389	0.5577
		WRAPPER-MSE	4.9655	126.2582	0.6539	0.5560
	FILT	FACPPe	5.3416	135.0453	0.6078	0.7671
		FACPPe- Sted.	4.6077	126.6292	0.6231	0.5549
		MI	5.9765	137.7981	0.7265	0.5833
SOBRAD.	WR	BIC	143.2984	704.0259	0.5296	0.5487
		AIC	142.3177	702.2297	0.5277	0.5464
		WRAPPER-MSE	146.7268	717.8556	0.5766	0.5691
	FILT	FACPPe	149.5968	702.0867	0.5921	0.5570
		FACPPe- Sted.	142.9056	704.2105	0.5247	0.5514
		MI	136.0668	678.5628	0.5350	0.5435

Tabela 4.3 – Resultados Seleção de variáveis da ELM

		Seleção de entradas	MSE	MAE	MSE dessaz.	MAE dessaz.
FURNAS	WR	BIC	6.9580	178.6464	0.3601	0.4733
		AIC	6.9539	176.1694	0.3640	0.4691
		WRAPPER-MSE	7.0181	176.5193	0.3665	0.4685
	FILT	FACPPe	10.8711	229.7249	0.6603	0.6380
		FACPPe- Sted.	8.5266	211.8476	0.6200	0.6179
		MI	12.1825	246.8223	0.8524	0.7209
EMBORC.	WR	BIC	3.8968	114.0687	0.5087	0.5092
		AIC	3.9472	115.8435	0.5119	0.5170
		WRAPPER-MSE	3.8944	114.6424	0.5076	0.5147
	FILT	FACPPe	4.9058	130.7474	0.6405	0.7824
		FACPPe- Sted.	4.6180	125.3885	0.7169	0.5985
		MI	5.3454	139.9012	0.9906	0.7411
SOBRAD.	WR	BIC	138.3265	722.3746	0.5568	0.5752
		AIC	140.4758	731.7529	0.5591	0.5791
		WRAPPER-MSE	135.4028	722.5684	0.5498	0.5498
	FILT	FACPPe	178.6328	867.5044	1.1559	0.8095
		FACPPe- Sted.	154.4385	816.1996	1.0931	0.7835
		MI	310.2328	1117.2695	1.7478	1.0706

Os resultados obtidos permitem algumas observações relevantes. A primeira é que o modelo PAR apresentou melhores desempenhos com uso de filtros, enquanto a ELM se sai melhor com o *wrapper* com avaliador tipo MSE, tanto no domínios real quanto no dessazonalizado. Segundo, mesmo o critério MI sendo uma abordagem não-linear, ele apresentou desempenho melhor em apenas 1 caso. Isto pode ter relação direta com o número reduzido de dados quando trata-se do caso mensal, com apenas 60 amostras disponíveis para ajuste. Além disso, o filtro baseado na FACP com a utilização de atrasos consecutivos, de acordo com a proposta de Stedinger, foi sempre melhor que a FACP clássica, como esperado.

Não seria surpreendente se os critérios BIC e AIC apresentassem melhores resultados que o critério MSE. Todavia este comportamento não foi verificado, de forma que para a ELM, este último mostrou-se adequado. Apesar de não ter sido mostrado, em geral o critério MSE não selecionou subconjuntos com muitas entradas e acabou por ser parcimonioso apesar da liberdade do método. Este fato não é surpreendente já que adotamos desde o início um número máximo de entradas pequeno.

De posse das performances apresentadas, iremos utilizar no Capítulo seguinte o filtro de autocorrelação parcial sob o ponto de vista de Stedinger como seletor de entradas para modelos lineares PAR e o *wrapper* com o MSE para as redes neurais.

4.5 Comentários

Este capítulo discutiu algumas formas de seleção das variáveis de entradas para determinação dos melhores subconjuntos para previsão de séries. Foram apresentados os métodos tipo *embbeded, wrappers* e *filtros*. Particularmente os dois últimos casos interessam particularmente a este trabalho.

A técnica do tipo *wrapper* pode avaliar a qualidade deste subconjunto sob diversos critérios, dentre as quais o erro quadrático médio de ajuste ou de previsão, BIC e AIC. Estes últimos penalizam a entrada de variáveis extras caso a melhora na resposta não seja substancial.

Os filtros discutidos neste Capítulo foram: do tipo linear com apresentação funções de autocorrelação parcial tradicional e sob o ponto de vista do trabalho de Stedinger (2001) para séries hidrológicas; do tipo não-linear com a introdução do conceito de informação mútua.

Testes computacionais foram realizados com a previsões feitas para as séries de Furnas, Emborcação e Sobradinho utilizando o modelo PAR e uma rede neural tipo Máquinas de Aprendizado Extremo – ELM. Os resultados mostraram que para o modelo linear o filtro de autocorrelação parcial é mais adequado enquanto para a ELM, o *wrapper* com função de avaliação via MSE apresentou melhores resultados.

No próximo capítulo, estes métodos de seleção serão utilizados para simulações com os modelos de preditores lineares (FACP) e redes neurais (*wrappers*) discutidos nos Capítulos 3 e 4.

Capítulo 5. Estudo de Casos

Nos capítulos anteriores, foram discutidos diversos modelos de previsão de séries temporais: dentre os lineares, analisaram-se os modelos auto-regressivos (AR) e periódicos auto-regressivos (PAR); na categoria das redes neurais, foram apresentados o *perceptron* de múltiplas camadas (MLP) e as máquinas desorganizadas do tipo máquinas de aprendizado extremo (ELM) e redes de estado de eco (ESN).

Neste capítulo, será feita uma análise comparativa da performance de previsão desses modelos, com a utilização das séries de Furnas, Emborcação e Sobradinho. Para isso, lançaremos mão de modelos de seleção de variáveis abordados no Capítulo 4: a função de autocorrelação parcial com atrasos consecutivos (Stedinger 2001) para o modelo PAR e o *wrapper* com avaliador baseado no mínimo erro quadrático médio para as redes neurais.

Outra etapa complementar discutida previamente é a dessazonalização de séries. Como comentado no Capítulo 2, apresentaremos evidências de que o processo de padronização é o mais adequado para esta tarefa.

Os preditores que apresentarem os melhores desempenhos serão novamente testados por meio a série histórica da usina de Passo Real, a qual apresenta vazões médias com amplitudes distintas das demais. Discutiremos e aplicaremos ainda o teste estatístico de Friedman para análise da significância estatística dos erros. Por fim, será realizada uma análise sobre as entradas selecionadas pelos modelos, além da quantidade de neurônios aproveitados pelas redes neurais.

5.1 Previsão de Vazões

Os estudos de caso abordados neste trabalho são baseados nas séries históricas de importantes usinas hidrelétricas brasileiras: Furnas, localizada no Rio Grande, Emborcação, que fica no rio Paranaíba, e Sobradinho, no rio São Francisco.

As séries destas usinas são frequentemente utilizadas em estudos deste tipo por possuírem comportamentos hidrológicos diversos quanto ao volume de água afluente, o que permite verificar a adequação das técnicas de previsão empregadas de maneira ampla (Ballini 2000, Luna 2007, Sacchi 2009). Os dados destas séries compreendem as vazões observadas entre os anos 1931 e 2010 (80 anos ou 960 meses). É possível caracterizá-las

parcialmente por suas médias e desvios padrões calculados segundo as Equações (2.3) e (2.4):

Tabela 5.1 – Médias e Desvio Padrões históricos

Série	Média ($\hat{\mu}$)	D. Padrão ($\hat{\sigma}$)
FURNAS	926.6177	613.1671
EMBORCAÇÃO	486.0781	362.8067
SOBRADINHO	2.6660e+03	1.9590e+03

Conforme comentado no Capítulo 2, uma etapa inicial importante na previsão de série de vazões mensais é a retirada da componente sazonal. Como forma de definir um modelo de dessazonalização adequado a estas séries, procuraram-se evidências entre os três métodos discutidos na Seção (2.4) - padronização, médias móveis e diferenças sazonais - de qual deles deveria ser utilizado para os ensaios deste trabalho.

Para treinamento dos preditores, foram utilizadas todas as demais amostras disponíveis. Os modelos de previsão utilizados foram o Periódico Autoregressivo (PAR) e uma Máquina de Aprendizado Extremo (ELM), para previsão de cada um dos 12 meses. No caso da ELM, os valores do erro quadrático médio (MSE) e do erro absoluto médio (MAE) são uma média de 50 simulações. Um *wrapper* foi o utilizado como método de seleção de entradas.

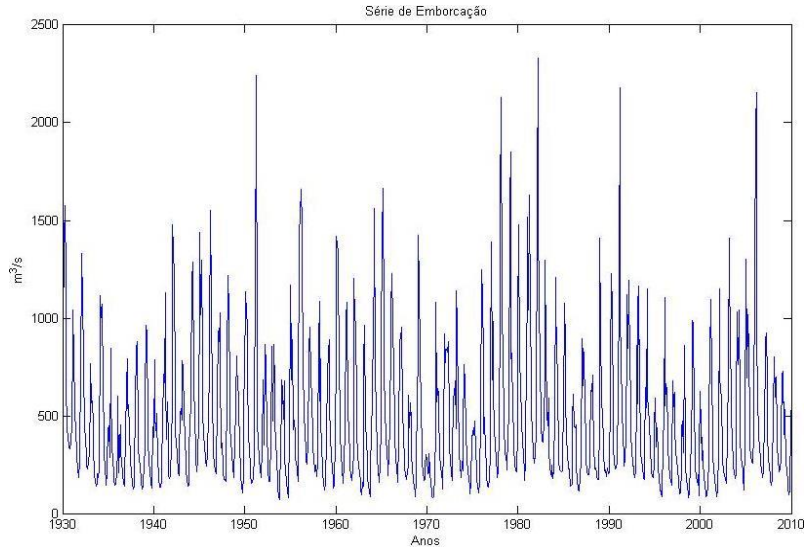
Tabela 5.2 – MSE e MAE para previsão 1 passo à frente

	Série	Modelo	MSE ($\times 10^4$)	MAE
PADRON.	**FURNAS**	**PAR**	8.3483	201.3464
		ELM	**6.9610**	177.3034
	EMBORCAÇÃO	**PAR**	**4.9655**	126.2582
		ELM	**3.8856**	114.6981
	SOBRADINHO	**PAR**	**146.7298**	717.8556
		ELM	**135.2756**	718.7678
MÉDIAS MÓVEIS	**FURNAS**	**PAR**	**8.1445**	202.4745
		ELM	9.1235	201.5305
	EMBORCAÇÃO	**PAR**	5.0416	131.3532
		ELM	5.0389	131.1217
	SOBRADINHO	**PAR**	169.0258	778.4184
		ELM	245.4487	907.4180
DIF. SAZONAIS	**FURNAS**	**PAR**	15.948	279.1860
		ELM	16.040	277.6040
	EMBORCAÇÃO	**PAR**	10.758	215.8107
		ELM	10.754	202.0221
	SOBRADINHO	**PAR**	283.967	1049.2542
		ELM	346.714	1060.8453

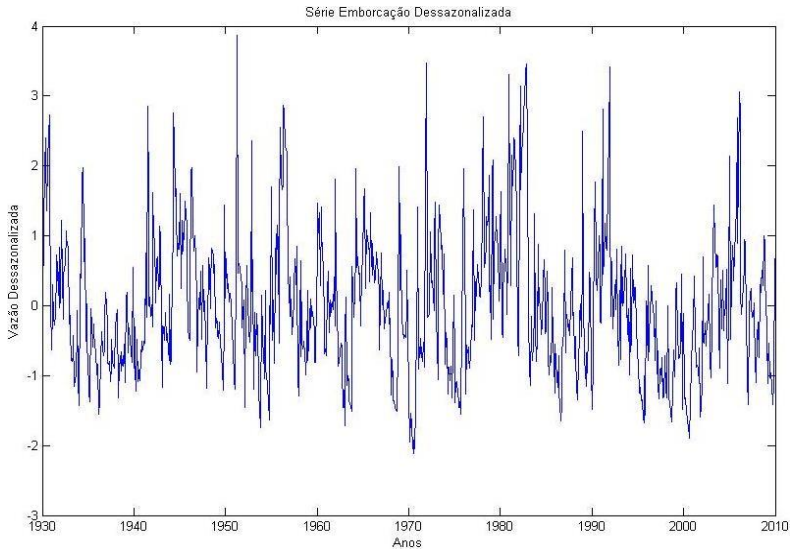
Os resultados indicam que o método de padronização é adequado à retirada da sazonalidade de séries de vazões, em consonância com a opção feita pelo Sistema Elétrico Brasileiro.

As Figuras 5.1 e 5.2 mostram ambas as séries, de Emborcação e Sobradinho, no domínio real e dessazonalizado pelo processo de padronização descrito na Equação (2.12), a qual reapresentamos:

$$z_{i,m}^{PA} = \frac{x_{i,m} - \hat{\mu}_m}{\hat{\sigma}_m} \tag{2.12}$$

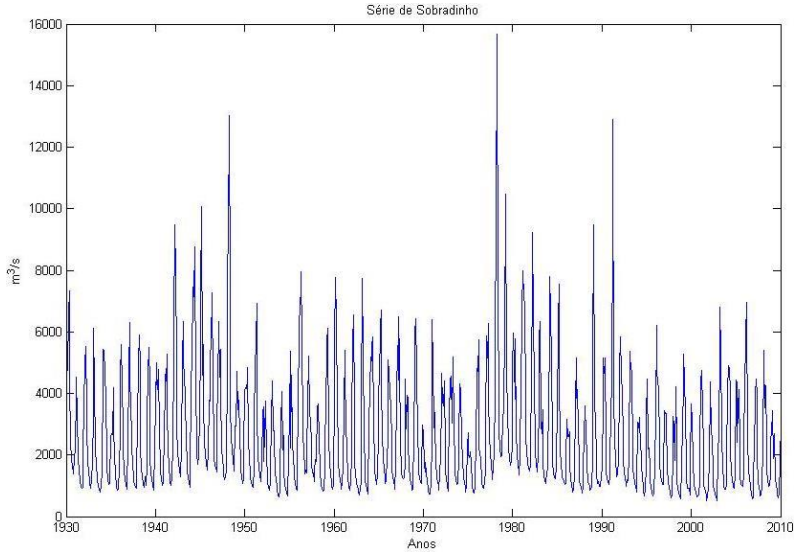


(a)

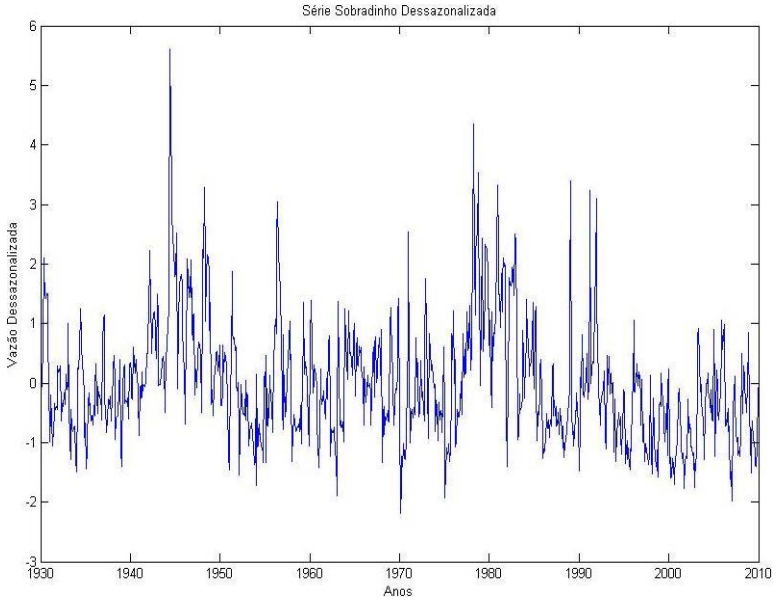


(b)

Figura 5.1 – Série Usina de Emborcação (a) real e (b) dessazonalizada



(a)



(b)

Figura 5.2 – Série Usina de Sobradinho (a) real e (b) dessazonalizada

O método de dessazonalização considerado deixa a série com média zero e desvio padrão unitário, sendo a componente sazonal eliminada. Como as séries são sazonais, a padronização baseia-se nas médias e desvios padrões mensais calculados conforme as já citadas Equações (2.10) e (2.11):

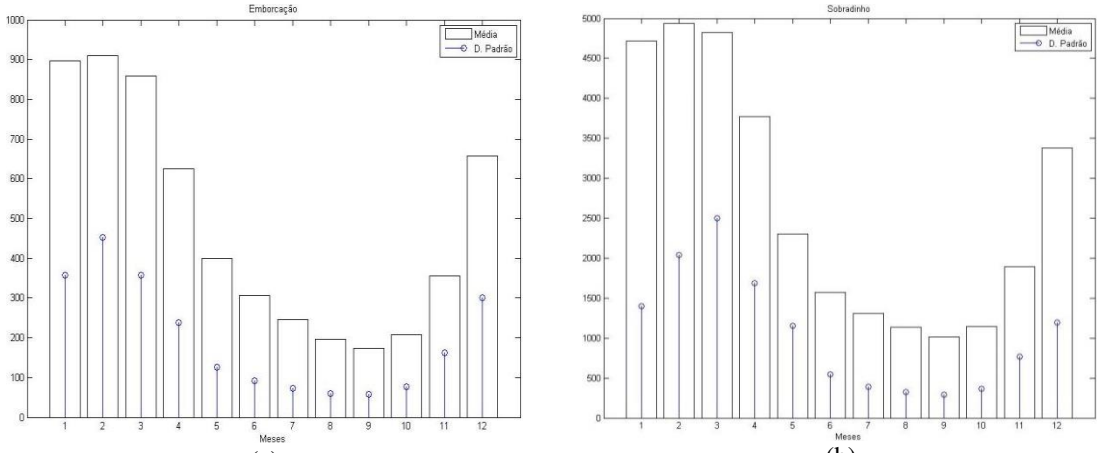
$$\hat{\mu}_m = \frac{1}{n}\sum_{i=1}^{n} x_{i,m} \tag{2.10}$$

$$\hat{\sigma}_m = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{i,m} - \hat{\mu}_m)^2} \qquad (2.11)$$

Os valores de $\hat{\mu}_m$, m=1,...,12, formam a média de longo termo (MLT), limitante superior do erro quadrático médio da previsão. Na Tabela 5.3 estão sumarizados as médias e desvios padrões das três series históricas, enquanto a Figura 5.3 mostra graficamente a relação entre essas medidas para Emborcação e Sobradinho:

Tabela 5.3 – Médias e Desvios Padrões mensais para a Série da Usina de Furnas

	Emborcação		Sobradinho		Furnas	
Mês	$\hat{\mu}_m$	$\hat{\sigma}_m$	$\hat{\mu}_m$	$\hat{\sigma}_m$	$\hat{\mu}_m$	$\hat{\sigma}_m$
Janeiro	897.12	358.23	4.714,06	1.403,08	1755,075	688,969
Fevereiro	909.00	451.80	4.939,20	2.041,24	1665,587	624,769
Março	858.95	358.22	4.822,77	2.498,39	1474,087	583,299
Abril	625.43	238.96	3.772,52	1.690,97	1013,362	348,446
Maio	398.70	126.36	2.301,68	1.152,50	741,037	227,413
Junho	305.83	92.92	1.569,16	542,49	613,787	241,340
Julho	245.58	72.65	1.311,75	390,72	506,112	151,377
Agosto	197.01	59.14	1.136,07	321,44	417,537	120,179
Setembro	174.28	57.01	1.017,02	288,28	438,087	224,886
Outubro	207.70	77.50	1.142,65	362,71	514,962	220,589
Novembro	355.62	161.97	1.888,61	770,20	730,366	302,715
Dezembro	657.67	301.05	3.376,75	1.194,45	1249,475	454,605



(a) (b)

Figura 5.3–Médias e Desvios Padrões mensais para a Série de Emborcação(a) e Sobradinho (b)

Como é possível verificar, a usina de Emborcação apresenta a maior variação dos dados no mês de fevereiro e a menor em setembro. Já em Sobradinho, a média histórica de fevereiro é a que apresenta maior valor, mas o desvio padrão mais elevado ocorre em março. Novamente, setembro é o mês de menor variação. A série de Furnas, por outro lado, apresenta média e desvio padrão maiores em janeiro e menores em agosto. Esta pequena

divergência quanto ao comportamento mensal se dá pela localização de cada posto (Ballini 2000). Além disso, o volume de água afluente é compatível com o regime de chuvas brasileiro, já que, no verão, a densidade pluviométrica é elevada, ao contrário do que ocorre para os meses de inverno.

As séries históricas das usinas hidrelétricas brasileiras em geral apresentam comportamento mensal parecido, de forma que podemos classificar cada período de acordo com o volume afluente, a exemplo dos gráficos apresentados na Figura 5.4:

i) meses de cheia: janeiro a março;

ii) meses de seca: maio a outubro;

iii) meses de transição: abril, novembro e dezembro.

Uma medida conveniente para este estudo, então, é o coeficiente de variação (*CV*), apresentado na Equação (5.1)

$$CV_m = \frac{\hat{\sigma}_m}{\hat{\mu}_m} \tag{5.1}$$

Segundo Andrade et al. (2012), em aplicações de modelos lineares como o PAR, meses que possuem um valor alto para este coeficiente apresentam baixa capacidade preditiva. Isto quer dizer que, com poucos passos à frente, chegam ao limitante superior do erro, o qual é dado pelo valor da MLT. A Tabela 5.4 apresenta os valores do *CV* mensal para as usinas em estudo, que corroboram as observações da Tabela 5.3.

Tabela 5.4 – Coeficiente de Variação Mensal

Mês	Emborcação	Sobradinho	Furnas
Janeiro	0.3993	0.2976	0.3926
Fevereiro	0.4970	0.4133	0.3751
Março	0.4171	0.5180	0.3957
Abril	0.3821	0.4482	0.3439
Maio	0.3169	0.5007	0.3069
Junho	0.3038	0.3457	0.3932
Julho	0.2958	0.2979	0.2991
Agosto	0.3002	0.2829	0.2878
Setembro	0.3271	0.2835	0.5133
Outubro	0.3732	0.3174	0.4284
Novembro	0.4555	0.4078	0.4145
Dezembro	0.4578	0.3537	0.3638

As medidas de erro utilizadas para comparação do desempenho dos preditores serão: erro quadrático médio (MSE), erro absoluto médio (MAE) e eficiência relativa (RE). Esta última é muito usual em trabalhos deste tipo, sendo largamente aplicada pelo setor elétrico (Andrade, et al. 2012). As expressões matemáticas destas métricas são apresentadas a seguir

$$MSE = \frac{1}{N_s}\sum_{t=1}^{Ns}(x_t - \hat{x}_t)^2 \tag{5.2}$$

$$MAE = \frac{1}{N_s}\sum_{t=1}^{Ns}\left|x_t - \hat{x}_t\right| \tag{5.3}$$

$$RE_m = \frac{\sqrt{MSE}}{\hat{\sigma}_m} \tag{5.4}$$

sendo $\hat{\sigma}_m$ a estimativa do desvio padrão calculado por (2.11).

O MAE é uma medida que indica a média do afastamento de todos os valores fornecidos pelos modelos e o valor observado, penalizando o erro de forma igual para todas as amostras. Já o MSE, que se baseia no erro ao quadrado, pondera mais fortemente erros grandes, ou seja, penaliza mais uma previsão com poucos erros grandes do que outra com grande número de erros pequenos. O MSE (ou a raiz do MSE) é a medida de acurácia mais utilizada em previsões de séries temporais (Morettin e Toloi 2006).

Os dados das séries foram divididos em: conjunto de treinamento, para calibração dos modelos; conjunto de testes, para avaliação dos resultados, e conjunto de validação, necessário para ajuste da MLP e das ELMs. Como dito, estão disponíveis 80 anos ou 960 amostras mensais entre os anos de 1931 e 2010. Dessa forma, foram escolhidos três períodos de testes distintos, cada um composto por 10 anos ou 120 amostras: de 1951 a 1960, de 1967 a 1976 e de 1977 a 1986. Esta seleção foi feita tendo em vista a obtenção de diferentes regimes, que compreendem um período de seca, um mediano e outro de cheias, respectivamente. As amostras de validação foram, em todos os casos, providos pelos anos entre 2001 e 2010 (10 anos), enquanto as amostras de treinamento foram compostas por todas as amostras restantes (60 anos).

É necessário notar que há modelos, como o PAR, que não utilizam o conjunto de validação. Todavia, como forma de darmos os mesmos subsídios aos preditores, consideramos o mesmo conjunto de treinamento para todos os modelos.

5.2 Ajustes dos modelos de previsão

Os modelos de previsão discutidos ao longo deste documento foram ajustados para as séries e períodos descritos. São eles:

i) modelo periódico auto-regressivo (PAR), ajustado via equações de Yule-Walker;

ii) rede neural *perceptron* de múltiplas camadas (MLP), treinada com o algoritmo gradiente conjugado escalonado modificado;

iii) máquina de aprendizado extremo (ELM);

iv) rede de estado de eco (ESN).

As diversas arquiteturas de ESNs podem ser caracterizadas de acordo com os diferentes fatores:

a) Quanto ao modelo de reservatório de dinâmicas:

- com reservatório de Jaeger (2001) e um combinador linear (JAE-ESN);

- com reservatório de (Ozturk, Xu e Principe 2007) e um combinador linear (OZT-ESN);

b) Quanto à camada de saída:

- PCA e filtro de Volterra – (JAE-PV-ESN) e (OZT-PV-ESN) (Boccato, Lopes, et al. 2011);

- ELM - (JAE-ESN-ELM) e (OZT-ESN-ELM) (Butcher et al 2013).

O processo de validação cruzada foi aplicado como critério de parada no treinamento das redes MLP. As ELMs foram ajustadas e passaram pelo processo de regularização com o mesmo conjunto de dados destinado à validação. Observe que as ELMs que fazem papel de camada de saída de ESNs também foram regularizadas.

As redes munidas do PCA foram todas determinadas com a utilização de duas componentes principais, N_{pc}=2. A razão para isto reside em resultados preliminares que

mostraram que a inclusão de mais componentes principais não trouxe melhorias no desempenho das redes.

Uma etapa primordial no ajuste de modelos de previsão é definir quais são as suas entradas mais relevantes. Como explanado no Capítulo 5, utilizamos a função de autocorrelação parcial sob a proposta de Stedinger (2001) para otimização do modelo linear PAR. No caso das redes neurais, o *wrapper* com função de avaliação baseada no mínimo MSE foi a opção escolhida. É necessário salientar que o número máximo de entradas permitido foi restrito a seis, seguindo o critério adotado pelo setor elétrico (Souza, Marcato, et al. 2010).

No ajuste de redes neurais, é necessário definir ainda o número de neurônios na respectiva camada intermediária, tarefa determinante no resultado final da previsão. A forma de escolha adotada neste trabalho foi avaliar o MSE de teste para cada número de neurônios definido no vetor $N_{Ne} = [3,5,7,10,15,20,30,40,50,60,70,80,90,100,110,120]$. A Figura 5.4 exemplifica o processo realizado para a rede OZT-ESN-ELM para o mês de novembro da série de Emborcação, usado para teste os dados entre 1951 e 1960. Neste caso, como é possível verificar, optou-se por 7 neurônios

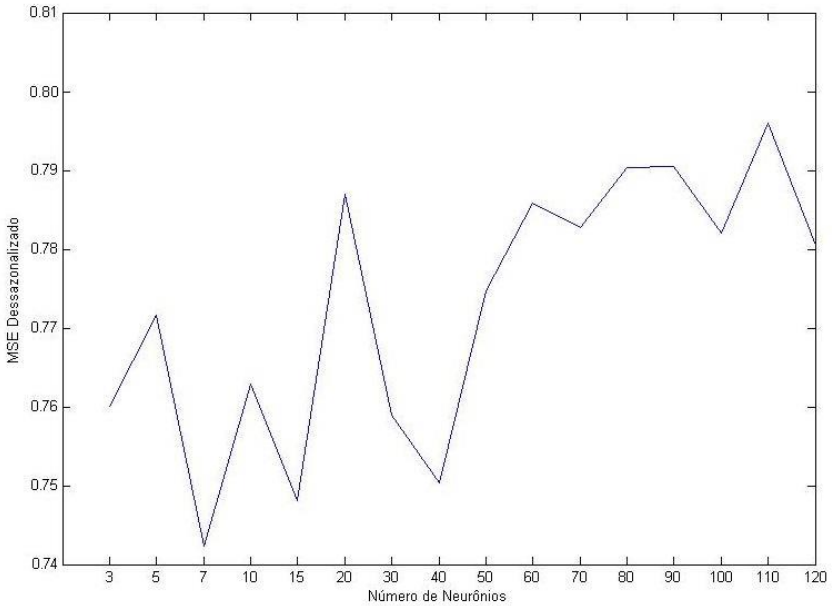


Figura 5.4 – Seleção do número de neurônios

O mesmo procedimento foi realizado para todas as redes. Na definição dos neurônios das ELMs que funcionam como cada de saída de ESN, o processo foi repetido para cada valor presente no vetor N_{Ne}, ou seja, fixou-se um certo número de neurônios na ESN e otimizou-se a ELM. A partir daí, variou-se o número de neurônios da primeira e o processo é refeito. Dessa forma, cada quantidade de neurônios do respectivo reservatório de dinâmicas terá uma ELM diferente ajustada.

5.3 Testes computacionais – observações iniciais e comportamento mensal

Os modelos de previsão foram ajustados e executados 50 vezes, exceto o modelo PAR, que, por ser determinístico, sempre apresenta o mesmo resultado para um mesmo conjunto de dados. Os resultados estão sumarizados nas subseções seguintes e separados por período de testes. As tabelas contêm os valores do MSE, MAE, RE e o desvio padrão destas execuções. Observe que, para cada mês, é ajustado um modelo diferente, de forma que o resultado da série completa será a previsão dos 12 meses ordenados corretamente.

Cada série foi ajustada e predita para os horizontes de P=1, 3, 6 e 12 passos à frente. A forma de previsão para P>1 foi do tipo recursiva, na qual o ajuste é realizado para um passo à frente e as previsões para horizontes maiores do que este são realimentadas à entrada do modelo fazendo papel de um dado real. Esta metodologia segue o receituário adotado pelo Setor Elétrico Brasileiro (Ballini 2000).

O comportamento das previsões mensais esteve próximo do esperado, ou seja, os meses de cheia foram aqueles que acarretaram os maiores erros. Isto pode ser verificado pela Figura 5.5, que apresenta a execução do modelo PAR e o gráfico tipo *box-plot* das 50 simulações das redes MLP, ELM e JAE-ESN para o período de testes de 1967-76 da série de Sobradinho. Percebe-se que os meses de janeiro a março apresentam maiores erros e desvios padrão que os demais

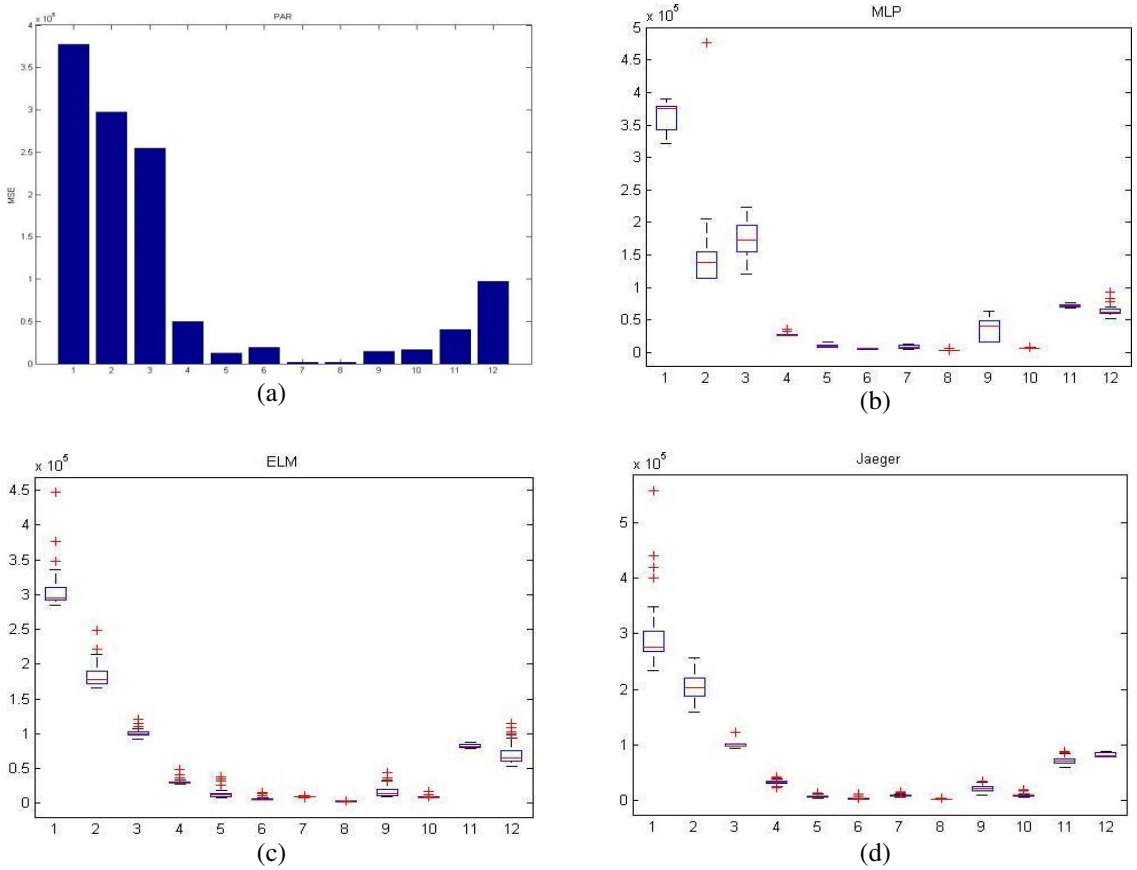


Figura 5.5–Box-plot para 50 execuções de Furnas 67-76 (a) PAR, (b) MLP, (c) ELM, (d) ESN Jaeger

Quanto ao horizonte de previsão, nota-se um comportamento semelhante e independente do número de passos à frente. São raros os casos em que o desvio padrão de janeiro ou fevereiro não são maiores, assim como a média. A Figura 5.6 mostra os gráficos *box-plot* para 50 execuções com a rede ELM, para a série de Furnas, com período de testes 67-76 e os variados valores de *P*.

Na execução em questão, os desvios padrão para as séries completas foram:

a) *P*=1 – 2.8549e+03;
b) *P*=3 – 7.4437e+03;
c) *P*=6 – 5.8876e+03;
d) *P*=12 – 2.1210e+11.

Veja que, no último caso, o desvio foi muito elevado, por conta de uma execução muito ruim do mês de janeiro. Dessa forma, a Figura 5.7 (d) foi construída com 49 amostras e o desvio recalculado da série completa foi *P*=12 – 1.0258e+05.

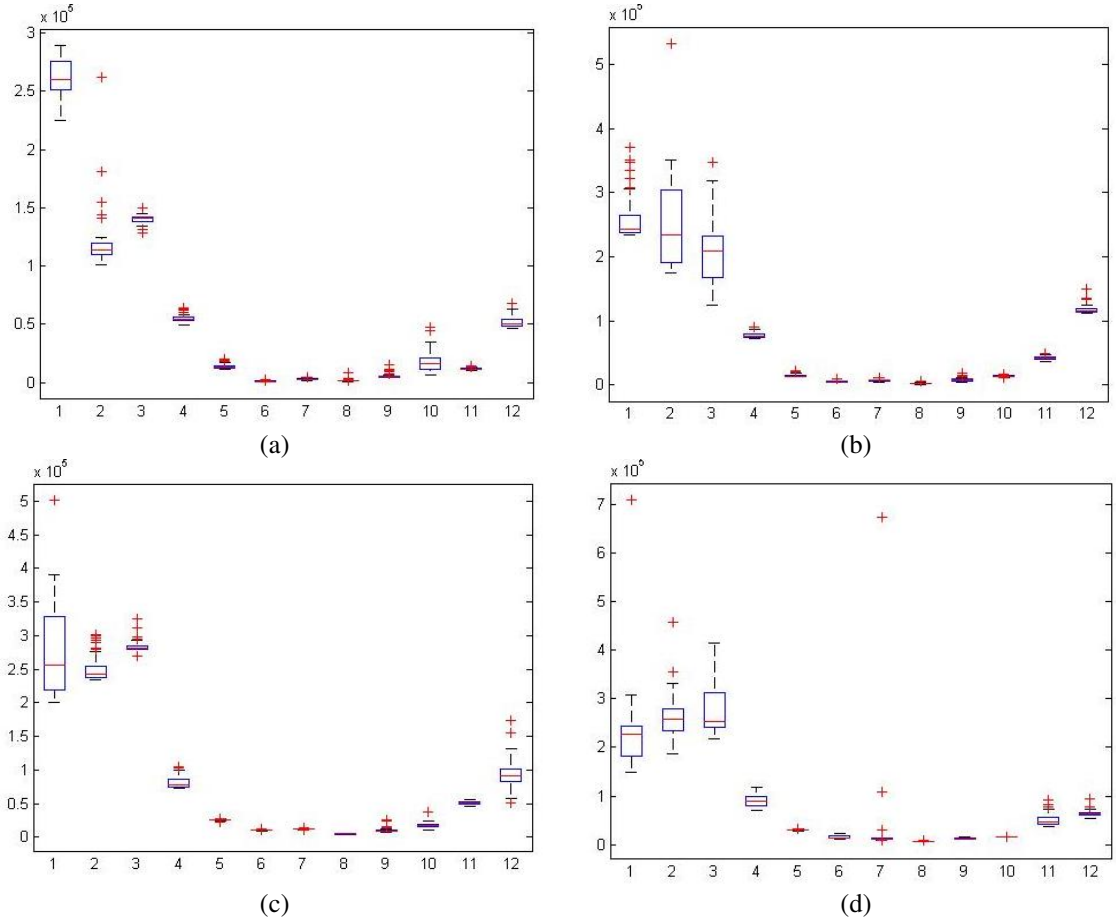


Figura 5.6 – *Box-plot* para 50 execuções da rede ELM para Furnas 67-76 (a) *P*=1, (b) *P* =3, (c) *P* =6, (d) *P* =12

Este tipo de comportamento foi verificado poucas vezes para as redes neurais *feedforward* MLP e ELM. Outro exemplo ocorreu com a ELM para o mês de janeiro de Furnas 51-60 e *P*=12. Em duas execuções, o erro foi extremamente elevado, como mostrado abaixo:

Execução 1 – MSE janeiro = 9.9672e+12 – MSE série completa - 8.3060e+11;

Execução 2 - MSE janeiro = 6.9501e+15 – MSE série completa - 5.7917e+14.

A razão para isto está na aleatoriedade com que são gerados os coeficientes da camada intermediária. Os modelos podem estar sujeitos a uma combinação de pesos desfavorável, não sendo possível para a camada de saída linear (no caso da ELM) minimizar o erro entre os sinais de saída e o desejado. No caso da MLP, deve-se considerar a hipótese de o algoritmo de treinamento ter convergido para um mínimo local ruim.

Todavia, como dito, este comportamento é raro e, em todas as simulações, visto poucas vezes.

Se excluirmos estas duas execuções, teremos o MSE de janeiro igual a 2.3962e+05, com valores de erro próximos, como mostra o histograma das 48 execuções da ELM sem os dois casos extremos citados:

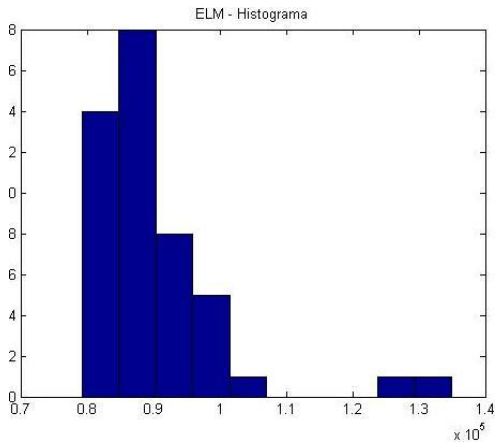


Figura 5.7 – Histograma ELM Furnas sem *outliers*

Apresentaremos agora os resultados computacionais separados por período de testes avaliado e os diferentes horizontes de previsão. Cada coluna de erro considera a série completa e apresenta a média de 50 execuções de cada rede neural.

5.4 Período 1951 a 1960

O período de 1951 a 1960 é conhecido por apresentar uma das maiores secas de todo o histórico disponível das usinas brasileiras, sobretudo para Furnas e Sobradinho. A Tabela 5.5 apresenta as respectivas médias e desvios padrão do período, enquanto a Tabela 5.6 mostra os resultados computacionais dos preditores para P=1.

Tabela 5.5 - Médias e Desvio Padrões período 51/60

Série	Média($\hat{\mu}$)	D. Padrão ($\hat{\sigma}$)
FURNAS 51/60	788.2583	458.7726
EMBORCAÇÃO	501.4333	376.2775
SOBRADINHO	2.4834e+03	1.6812e+03

Analisando os erros das previsões, observa-se um comportamento recorrente em estudos sobre séries de vazões: em diversas ocasiões, o preditor que apresenta o menor MSE no domínio real nem sempre é o melhor no domínio dessazonalizado. Na série

Emborcação, esta afirmação fica evidente, uma vez que, enquanto a rede JAE-ESN foi a de menor MSE real, a OZT-PV-ESN foi a melhor no espaço dessazonalizado. Embora seja este último domínio que o modelo efetivamente "enxerga", o resultado final acaba sendo degradado quando da reincorporação da componente sazonal.

Tabela 5.6 – Resultados de Previsão para 1 passo à frente (P = 1)

	Preditor	RE	Desvio Padrão	MSE	MAE	MSE dessaz.	MAE
FURNAS	**PAR**	0.4128	-	6.4077e+04	165.1262	0.2801	0.4118
	MLP	0.3940	9.8238e+03	5.8366e+04	158.9450	0.2624	0.4007
	ELM	0.3900	3.8056e+03	5.7191e+04	154.7925	0.2603	0.3925
	JAE-ESN	0.3779	4.7705e+03	**5.3696e+04**	**157.2884**	**0.2589**	**0.4105**
	OZT-ESN	0.4011	4.9628e+03	6.0473e+04	161.3877	0.2772	0.4157
	JAE-PV-ESN	0.3927	2.1540e+03	5.7968e+04	159.4470	0.2657	0.4056
	OZT-PV-ESN	0.4146	4.5174e+03	6.4638e+04	167.6837	0.2910	0.4249
	JAE-ESN-ELM	0.4224	5.8381e+03	6.7086e+04	176.0160	0.3248	0.4567
	OZT-ESN-ELM	0.4578	8.5673e+03	7.8813e+04	187.6918	0.3729	0.4890
EMBORCAÇÃO	**PAR**	0.5917	-	4.6077e+04	126.6292	0.6231	0.5549
	MLP	0.5393	1.7453e+04	3.8284e+04	113.0414	0.4932	0.5051
	ELM	0.5441	1.9182e+03	3.8966e+04	114.1961	0.5080	0.5137
	JAE-ESN	0.5234	2.1734e+03	**3.6062e+04**	**113.9533**	**0.4837**	**0.5162**
	OZT-ESN	0.5528	3.7173e+03	4.0217e+04	116.9412	0.5149	0.5203
	JAE-PV-ESN	0.5346	1.2780e+03	3.7621e+04	108.8454	0.4778	0.4838
	OZT-PV-ESN	0.5324	619.9087	3.7314e+04	111.0340	0.4777	0.4932
	JAE-ESN-ELM	0.5740	5.6233e+03	4.3369e+04	121.9824	0.5669	0.5513
	OZT-ESN-ELM	0.5726	1.6166e+04	4.3157e+04	123.7026	0.5917	0.5666
SOBRADINHO	**PAR**	0.4616	-	8.1778e+05	554.4859	0.3709	0.4400
	MLP	0.4179	4.4735e+04	6.7030e+05	518.0475	0.2927	0.4084
	ELM	0.3955	5.1907e+04	**6.0018e+05**	**476.5552**	**0.2795**	**0.3841**
	JAE-ESN	0.4100	4.1366e+04	6.4516e+05	515.1562	0.3038	0.4195
	OZT-ESN	0.4253	5.5779e+04	6.9427e+05	528.6689	0.3274	0.4297
	JAE-PV-ESN	0.4106	2.5758e+04	6.4716e+05	526.6320	0.2975	0.4244
	OZT-PV-ESN	0.4220	3.7783e+04	6.8332e+05	536.4513	0.3238	0.4404
	JAE-ESN-ELM	0.4455	7.9978e+04	7.6183e+05	558.1521	0.3787	0.4671
	OZT-ESN-ELM	0.4675	1.0524e+05	8.3892e+05	596.4406	0.4303	0.5107

Paralelamente a isto, há também casos em que o melhor preditor em termos do MSE não coincide com o de menor MAE. Os resultados de Emborcação são novamente elucidativos deste fato. Entretanto, interessa a este estudo avaliar a previsão como um todo, e o MSE é a métrica mais comumente utilizada neste caso (Haykin 1997).

Para a série de Furnas, a JAE-ESN foi a estrutura que apresentou o menor MSE real e dessazonalizado. A rede OZT-ESN e as máquinas desorganizadas munidas de uma ELM como camada de saída apresentaram erro mais elevado inclusive que o obtido via metodologia linear. Por fim, a de menor MAE foi a ELM.

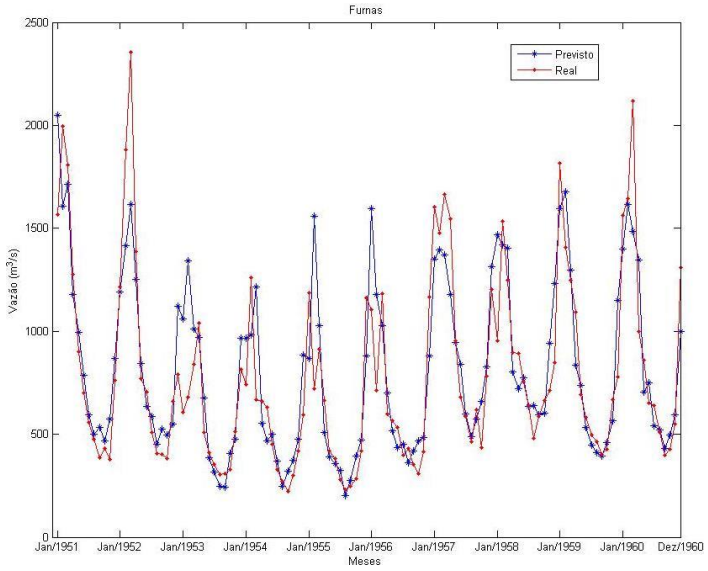
No caso da série de Emborcação, novamente a JAE-ESN foi a de melhor desempenho geral no domínio real. Todavia, no dessazonalizado, a rede OZT-PV-ESN foi a que apresentou menor MSE, enquanto a de menor MAE foi JAE-PV-ESN. Todas as redes

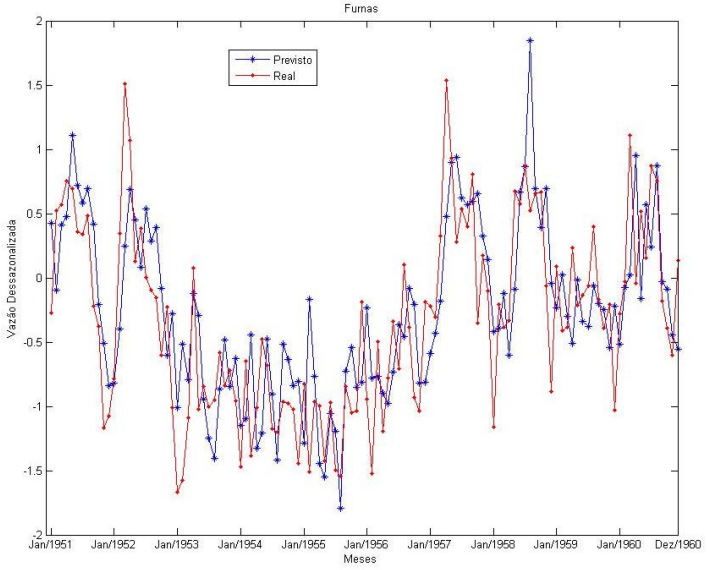
neurais foram superiores ao modelo linear, embora as arquiteturas JAE-ESN-ELM e OZT-ESN-ELM tenham apresentado resultados inferiores aos verificados para as redes *feedforward*.

A série do posto de Sobradinho foi mais bem aproximada pela rede ELM em todas as métricas de erro e domínios analisados. Apenas a OZT-ESN-ELM teve desempenho pior que o modelo PAR.

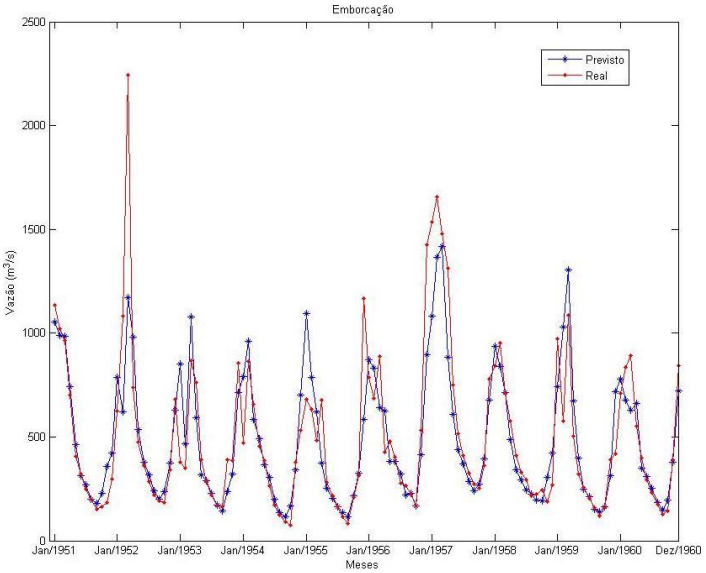
Embora as estruturas com camadas de saída não lineares sejam ferramentas com maior poder de processamento intrínseco, observa-se que a forma de abordagem proposta, com separação de dados mensais, acarreta em um número reduzido de amostras de treinamento disponível. Isto é observado diretamente em alguns resultados que não foram aqui apresentados: em diversos momentos, o erro de treinamento destas redes é menor que as demais abordagens, mas o MSE de teste acaba comprometido. Ou seja, a camada de saída não-linear com a inserção da ELM suscita elevado poder de aproximação, mas, pelo número de amostras, isso nem sempre se reflete em capacidade de generalização. Notaremos este comportamento com alguma frequência nos resultados posteriores deste trabalho.

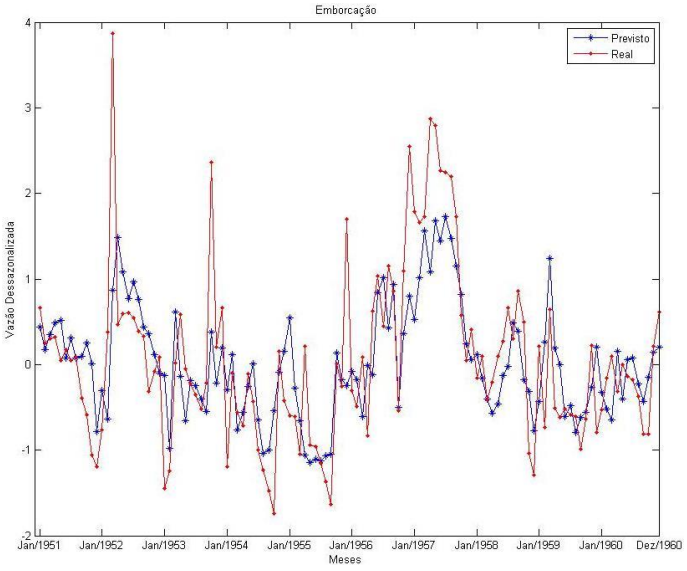
Os gráficos de uma execução dos preditores que apresentaram melhores resultados no espaço real estão sumarizados na Figura 5.8, juntamente com seus respectivos gráficos dessazonalizados.


Furnas
Previsto
Real
Vazão (m3/s)
2500
2000
1500
1000
500
0
Jan/1951
Jan/1952
Jan/1953
Jan/1954
Jan/1955
Jan/1956
Jan/1957
Jan/1958
Jan/1959
Jan/1960
Dez/1960
Meses

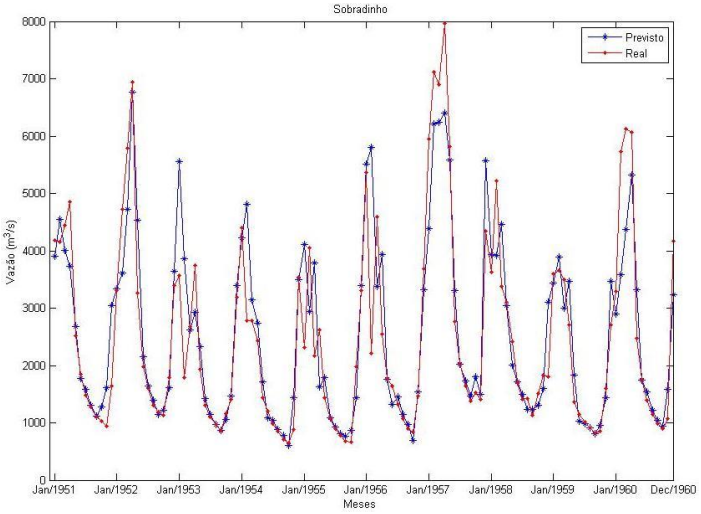

Furnas
Previsto
Real
Vazão Dessazonalizada
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
Jan/1951
Jan/1952
Jan/1953
Jan/1954
Jan/1955
Jan/1956
Jan/1957
Jan/1958
Jan/1959
Jan/1960
Dez/1960
Meses

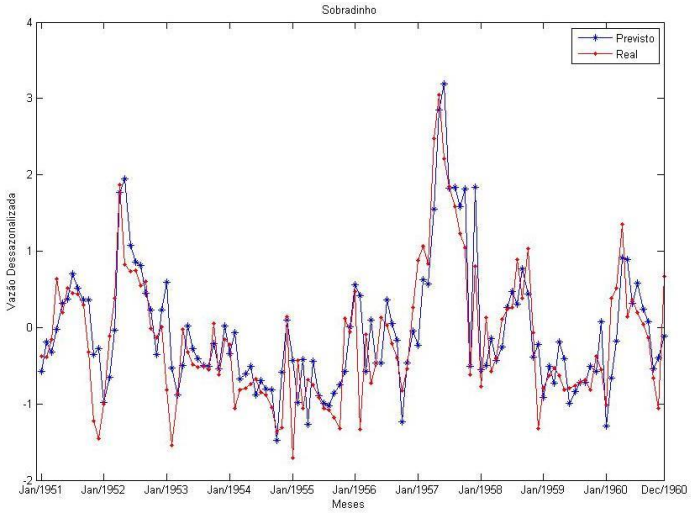
(a)


Emborcação
Previsto
Real
Vazão (m³/s)
2500
2000
1500
1000
500
0
Jan/1951
Jan/1952
Jan/1953
Jan/1954
Jan/1955
Jan/1956
Jan/1957
Jan/1958
Jan/1959
Jan/1960
Dez/1960
Meses


Emborcação
Previsto
Real
Vazão Dessazonalizada
4
3
2
1
0
-1
-2
Jan/1951
Jan/1952
Jan/1953
Jan/1954
Jan/1955
Jan/1956
Jan/1957
Jan/1958
Jan/1959
Jan/1960
Dez/1960
Meses

(b)





(c)

Figura 5.8 – Resultados melhores previsões P=1, real e dessazonalizado – (a) Furnas, (b) Emborcação, (c) Sobradinho

A seguir, na Tabela 5.7, encontram-se os resultados computacionais para *P*=3. A previsão três passos à frente para a série Furnas não apresentou uma relação direta entre MSE real e dessazonalizado nem entre MSE e MAE. O preditor de menor MSE real foi a OZT-ESN. As máquinas desorganizadas munidas de camadas de saída com filtro de Volterra e ELM não conseguiram ser superiores às demais redes neurais. Entretanto, apenas a OZT-ESN-ELM foi pior que o modelo linear.

Já Emborcação teve a JAE-ESN com menores valores de MSE para o espaço real e dessazonalizado e a OZT-ESN teve papel análogo para o MAE. OZT-ESN-ELM e JAE-

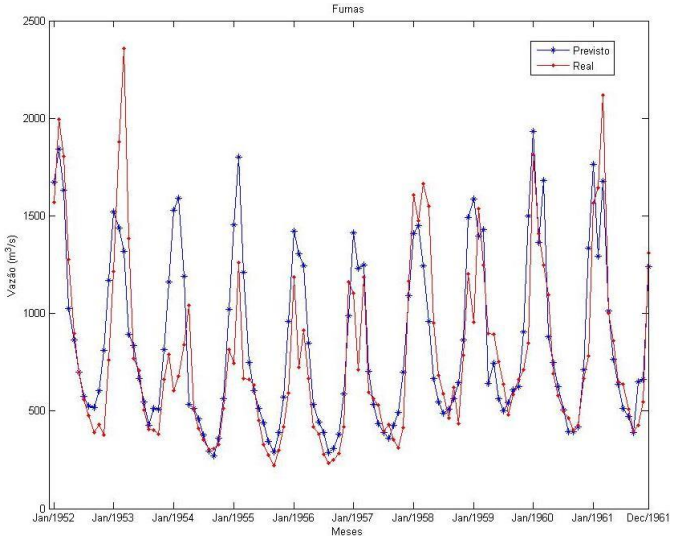
PV-ESN apresentaram resultados inferiores aos do modelo PAR, ao contrário das demais redes neurais.

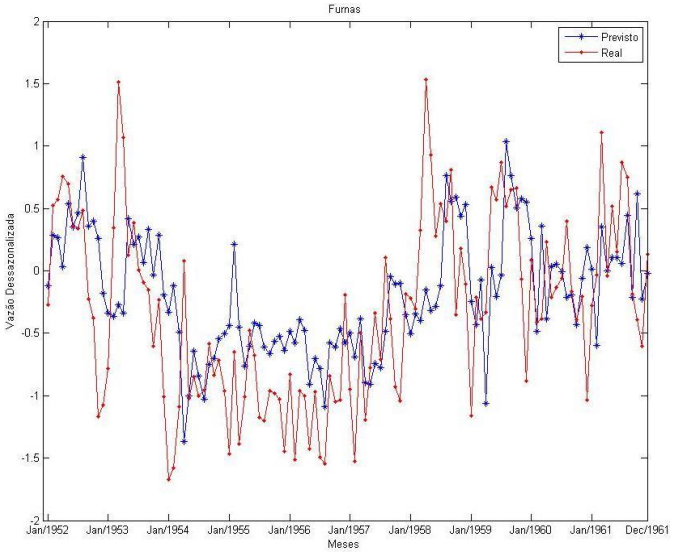
Tabela 5.7 – Resultados de previsão para 3 passos à frente (P = 3)

	Preditor	RE	Desvio	MSE	MAE	MSE dessaz.	MAE
FURNAS	**PAR**	0.5297	-	1.0549e+05	213.1960	0.4864	0.5452
	MLP	0.4949	2.4415e+04	9.2097e+04	197.4613	0.4180	0.5058
	ELM	0.4724	8.2522e+03	8.3915e+04	189.1057	0.3932	0.4889
	JAE-ESN	0.4697	6.3985e+03	8.2957e+04	188.0635	0.4109	0.4942
	OZT-ESN	0.4651	5.2999e+03	**8.1338e+04**	**191.8087**	**0.4070**	**0.5033**
	JAE-PV-ESN	0.5026	7.5438e+03	9.4963e+04	201.2739	0.4542	0.5427
	OZT-PV-ESN	0.5221	5.4531e+03	1.0249e+05	208.7898	0.4986	0.5613
	JAE-ESN-ELM	0.5228	1.3580e+04	1.0279e+05	217.2946	0.5069	0.5662
	OZT-ESN-ELM	0.5444	2.1380e+04	1.1145e+05	223.3987	0.5404	0.5825
EMBORCAÇÃO	**PAR**	0.6539	-	5.6276e+04	134.2924	0.7700	0.6326
	MLP	0.6350	9.5042e+03	5.3069e+04	129.4318	0.6848	0.5969
	ELM	0.5886	1.6533e+03	4.5605e+04	127.7468	0.6322	0.5984
	JAE-ESN	0.5878	4.0043e+03	**4.5481e+04**	**125.0703**	**0.6212**	**0.5814**
	OZT-ESN	0.5991	4.3585e+03	4.7250e+04	124.5593	0.6292	0.5810
	JAE-PV-ESN	0.6597	3.8918e+03	5.7282e+04	136.7967	0.7935	0.6537
	OZT-PV-ESN	0.6306	2.6683e+03	5.2335e+04	133.4052	0.7436	0.6462
	JAE-ESN-ELM	0.6333	4.7380e+03	5.2797e+04	132.7965	0.7093	0.6260
	OZT-ESN-ELM	0.6564	2.1019e+04	5.6721e+04	137.9189	0.7611	0.6442
SOBRADINHO	**PAR**	0.5649	-	1.2245e+06	707.9950	0.5861	0.5986
	MLP	0.5216	1.1160e+05	1.0443e+06	626.0565	0.4819	0.5244
	ELM	0.5181	6.2819e+04	1.0301e+06	640.1818	0.4350	0.5069
	JAE-ESN	0.4828	1.8790e+05	**8.9462e+05**	**603.6756**	**0.4622**	**0.5138**
	OZT-ESN	0.5285	1.0224e+05	1.0718e+06	664.5329	0.5288	0.5556
	JAE-PV-ESN	0.5785	5.6205e+04	1.2844e+06	719.8115	0.5803	0.5919
	OZT-PV-ESN	0.5847	4.3560e+04	1.3121e+06	728.8818	0.6314	0.6167
	JAE-ESN-ELM	0.5545	2.6936e+05	1.1802e+06	657.4957	0.5913	0.5610
	OZT-ESN-ELM	0.5708	1.3503e+05	1.2506e+06	711.6941	0.6287	0.5968

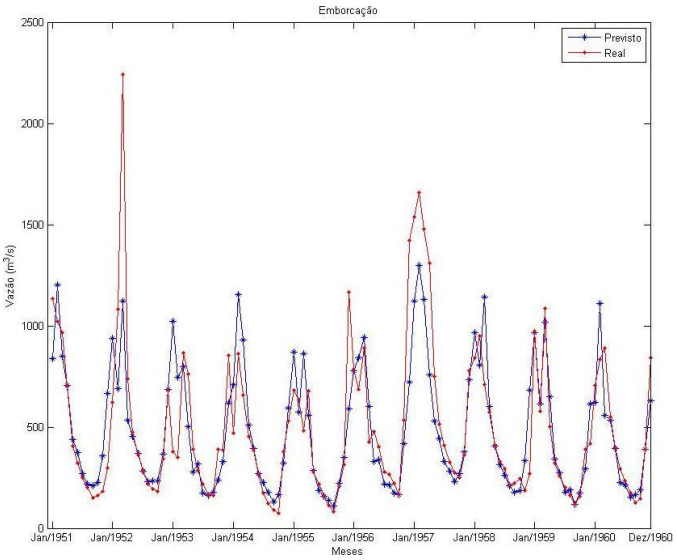
Sobradinho teve novamente a JAE-ESN com melhor MSE e também com o menor MAE no espaço real. No domínio dessazonalizado, o menor valor foi o da ELM. As redes com filtro de Volterra foram as que apresentaram os erros mais elevados.

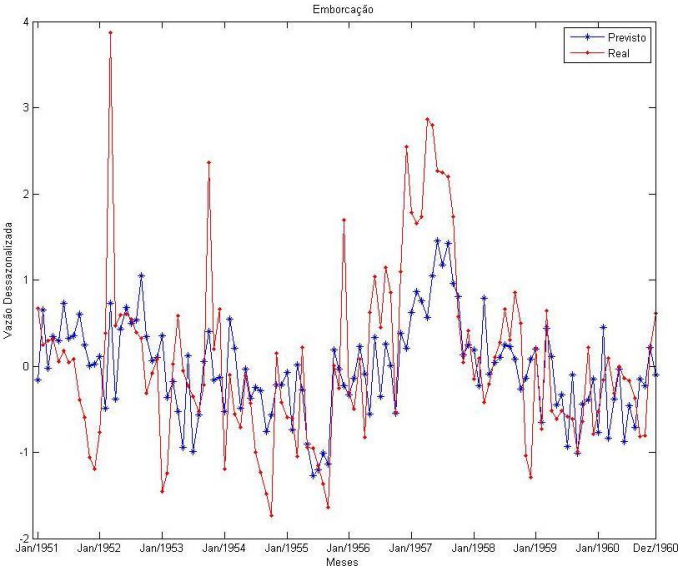
As evoluções temporais dos melhores resultados para cada série no domínio real e dessazonalizado são apresentadas na Figura 5.9.


Furnas
Previsto
Real
Vazão (m3/s)
2500
2000
1500
1000
500
0
Jan/1952
Jan/1953
Jan/1954
Jan/1955
Jan/1956
Jan/1957
Jan/1958
Jan/1959
Jan/1960
Jan/1961
Dec/1961
Meses

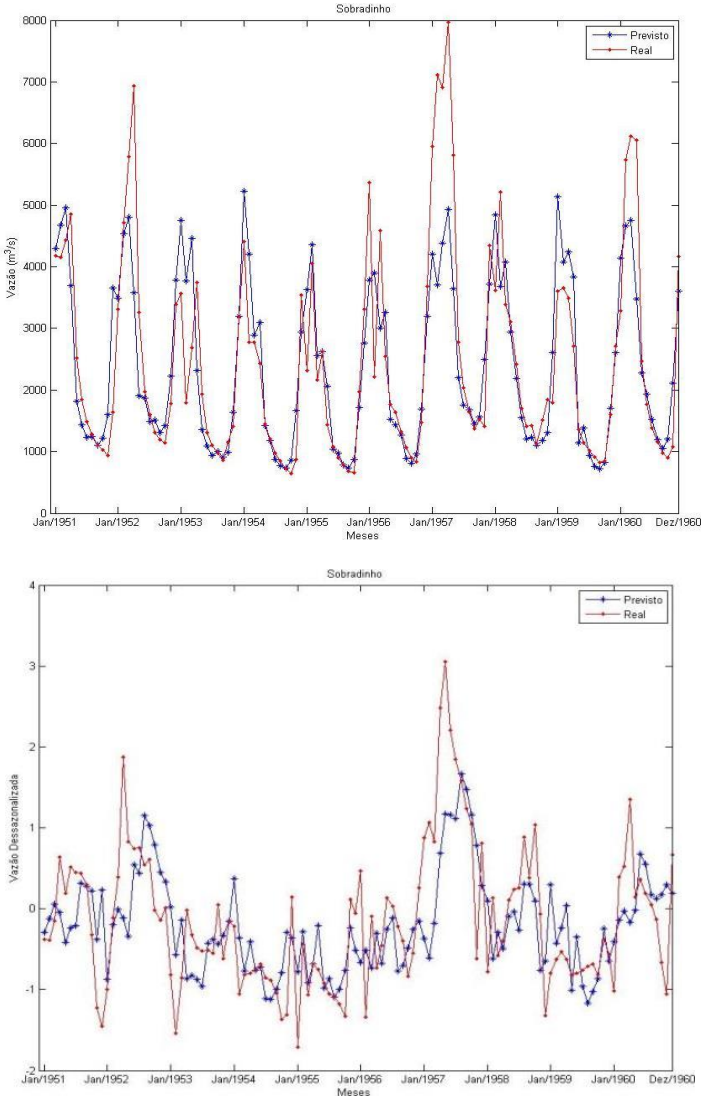

Furnas
Previsto
Real
Vazão Dessazonalizada
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
Jan/1952
Jan/1953
Jan/1954
Jan/1955
Jan/1956
Jan/1957
Jan/1958
Jan/1959
Jan/1960
Jan/1961
Dec/1961
Meses

(a)


Emborcação
Previsto
Real
Vazão (m3/s)
2500
2000
1500
1000
500
0
Jan/1951
Jan/1952
Jan/1953
Jan/1954
Jan/1955
Jan/1956
Jan/1957
Jan/1958
Jan/1959
Jan/1960
Dez/1960
Meses


Emborcação
Previsto
Real
Vazão Dessazonalizada
4
3
2
1
0
-1
-2
Jan/1951
Jan/1952
Jan/1953
Jan/1954
Jan/1955
Jan/1956
Jan/1957
Jan/1958
Jan/1959
Jan/1960
Dez/1960
Meses

(b)



(c)

Figura 5.9 – Resultados melhores previsões P=3, real e dessazonalizado – (a) Furnas, (b) Emborcação, (c) Sobradinho

Os resultados das simulações para o horizonte de 6 passos à frente estão sumarizados na Tabela 5.8.

Os resultados computacionais para *P*=6 foram, em todos os casos, superiores para a arquitetura ELM em termos do MSE real e do dessazonalizado. Em Emborcação, a OZT-ELM apresentou melhor MAE, enquanto, em Sobradinho, isso ocorre para a JAE-ESN.

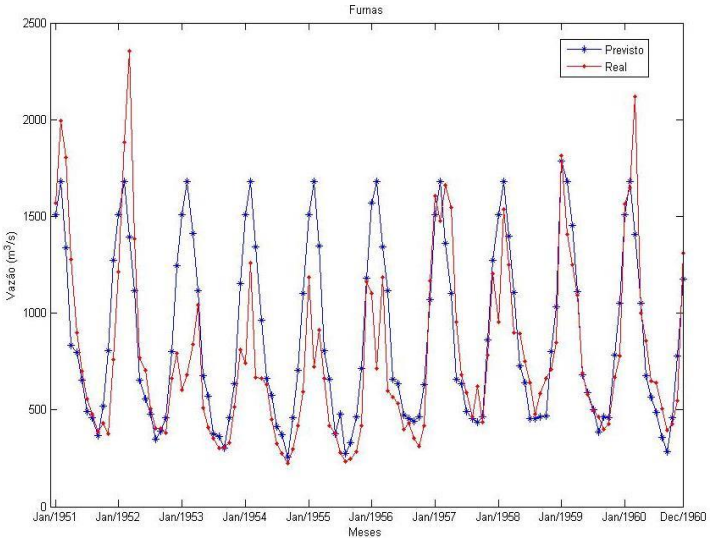
Nota-se que, em Furnas e Sobradinho, as redes ESN munidas da ELM tiveram resultados inferiores ao obtido para o modelo PAR. Em Emborcação, este fato foi verificado apenas para a OZT-ESN-ELM.

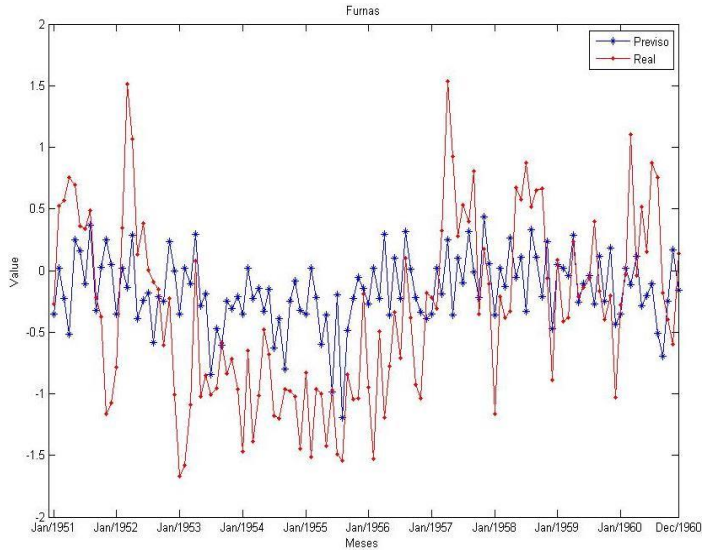
Tabela 5.8 - Resultados de Previsão para 6 passos à frente ($P = 6$)

	Preditor	RE	Desvio Padrão	MSE	MAE	MSE dessaz.	MAE dessaz.
FURNAS	**PAR**	0.5575	-	1.1685e+05	234.5016	0.5954	0.6269
	MLP	0.5439	5.9307e+04	1.1122e+05	216.8192	0.4925	0.5574
	ELM	0.4981	7.0141e+03	**9.3266e+04**	**210.7465**	**0.4873**	**0.5710**
	JAE-ESN	0.5051	5.8105e+03	9.5920e+04	217.0526	0.4875	0.5868
	OZT-ESN	0.5119	2.2235e+04	9.8537e+04	215.4724	0.4934	0.5795
	JAE-PV-ESN	0.5179	9.9187e+03	1.0085e+05	216.3331	0.5421	0.5902
	OZT-PV-ESN	0.5248	4.3342e+03	1.0353e+05	216.7608	0.5178	0.5814
	JAE-ESN-ELM	0.5570	1.4150e+04	1.1668e+05	234.9199	0.5982	0.6348
	OZT-ESN-ELM	0.5647	1.5365e+04	1.1991e+05	240.8781	0.6098	0.6489
EMBORCAÇÃO	**PAR**	0.6789	-	6.0665e+04	145.2065	0.9157	0.7139
	MLP	0.6648	2.1787e+03	5.8174e+04	141.1820	0.8650	0.7098
	ELM	0.6091	1.7465e+03	**4.8838e+04**	**136.6635**	**0.7545**	**0.6753**
	JAE-ESN	0.6452	2.4243e+03	5.4800e+04	135.4325	0.8070	0.6698
	OZT-ESN	0.6479	3.8425e+03	5.5247e+04	135.3472	0.8063	0.6585
	JAE-PV-ESN	0.6628	3.4059e+03	5.7819e+04	139.4214	0.8480	0.6944
	OZT-PV-ESN	0.6548	1.2107e+03	5.6434e+04	141.6630	0.8394	0.7033
	JAE-ESN-ELM	0.6686	7.8601e+03	5.8840e+04	141.7245	0.8728	0.6987
	OZT-ESN-ELM	0.6986	1.5583e+04	6.4247e+04	149.2522	0.9548	0.7297
SOBRADINHO	**PAR**	0.5741	-	1.2650e+06	753.5039	0.7689	0.6929
	MLP	0.5712	7.9364e+04	1.2519e+06	719.1340	0.6843	0.6372
	ELM	0.5487	4.9575e+04	**1.1554e+06**	**720.5922**	**0.6375**	**0.6392**
	JAE-ESN	0.5533	3.1145e+04	1.1749e+06	700.6872	0.6506	0.6203
	OZT-ESN	0.5526	4.5222e+04	1.1719e+06	698.2706	0.6768	0.6259
	JAE-PV-ESN	0.6544	1.7882e+06	1.6433e+06	802.8963	0.8136	0.6970
	OZT-PV-ESN	0.5751	6.0021e+05	1.2691e+06	736.0158	0.7505	0.6737
	JAE-ESN-ELM	0.5897	1.7975e+05	1.3348e+06	729.9462	0.7205	0.6504
	OZT-ESN-ELM	0.6110	2.2679e+05	1.4331e+06	756.6507	0.7739	0.6652

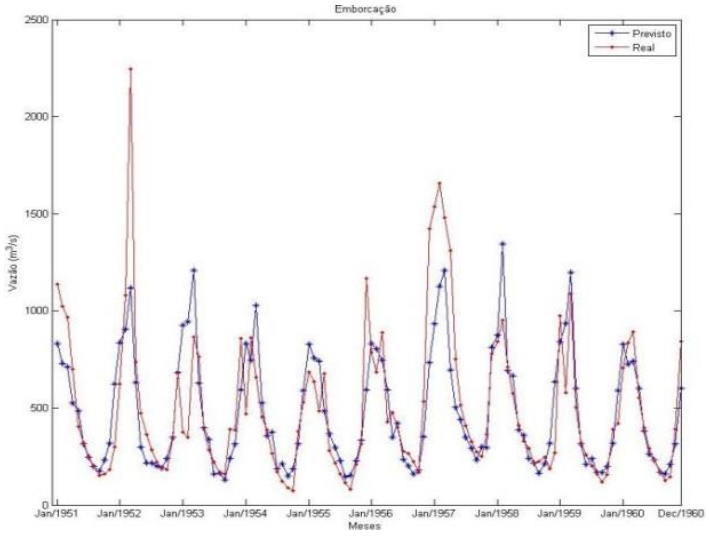
No caso de Sobradinho, a JAE-PV-ESN apresentou um erro muito superior que os dos demais, ocorrendo o mesmo para desvio padrão. Isso se deve ao comportamento discutido na seção anterior, no qual duas das execuções no mês de novembro apresentaram MSE duas ordens de grandeza superiores à média.

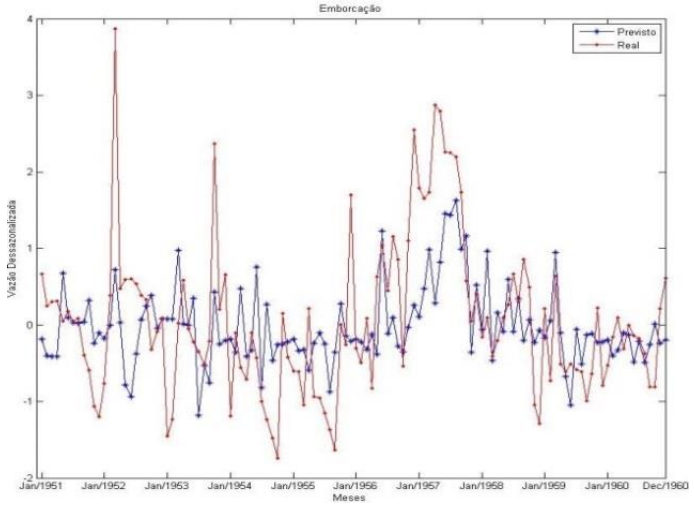
As previsões dos melhores que apresentaram menor MSE real são mostradas na Figura 5.10.


Furnas
Previsto
Real
Vazão (m³/s)
2500
2000
1500
1000
500
0
Jan/1951
Jan/1952
Jan/1953
Jan/1954
Jan/1955
Jan/1956
Jan/1957
Jan/1958
Jan/1959
Jan/1960
Dec/1960
Meses

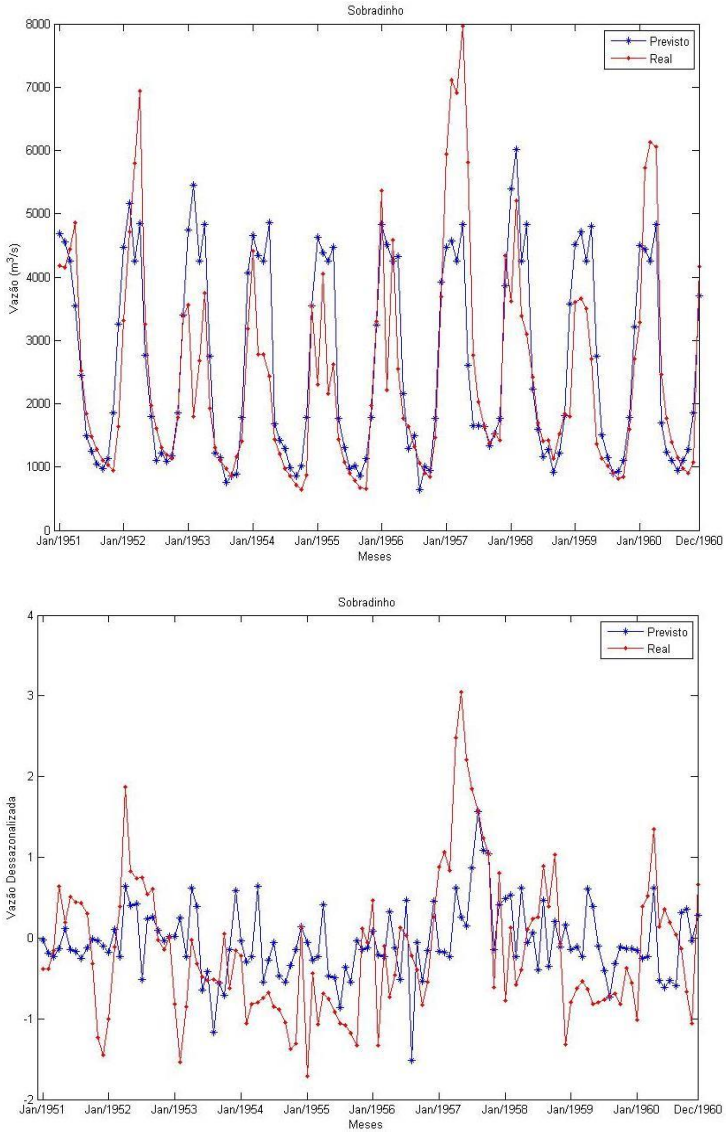

Furnas
Previso
Real
Value
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
Jan/1951
Jan/1952
Jan/1953
Jan/1954
Jan/1955
Jan/1956
Jan/1957
Jan/1958
Jan/1959
Jan/1960
Dec/1960
Meses

(a)


Emborcação
Previsto
Real
Vazão (m³/s)
2500
2000
1500
1000
500
0
Jan/1951
Jan/1952
Jan/1953
Jan/1954
Jan/1955
Jan/1956
Jan/1957
Jan/1958
Jan/1959
Jan/1960
Dec/1960
Meses


Emborcação
Previsto
Real
Vazão Dessazonalizada
4
3
2
1
0
-1
-2
Jan/1951
Jan/1952
Jan/1953
Jan/1954
Jan/1955
Jan/1956
Jan/1957
Jan/1958
Jan/1959
Jan/1960
Dec/1960
Meses

(b)



(c)

Figura 5.10 – Resultados melhores previsões P=6, real e dessazonalizado – (a) Furnas, (b) Emborcação, (c) Sobradinho

Finalmente, para o horizonte de 12 passos à frente, os resultados das simulações estão na Tabela 5.9.

Na previsão com *P*=12, apenas em Furnas não houve correspondência quanto aos menores MSE e MAE, os quais foram obtidos pelas MD's JAE-ESN e ELM, respectivamente. O melhor MSE dessazonalizado também foi alcançado pela ESN de Jaeger. O interessante é que, apesar disso, o desvio padrão da ELM foi bastante elevado, fato que utilizamos como exemplo na subseção anterior. O comportamento geral visto nas previsões dos demais horizontes estudados se manteve, com as ESNs com a ELM como

camada de saída tendo piores resultados que os modelos lineares. Neste caso, as ESNs com filtro de Volterra acabaram por ter erros maiores que suas versões mais simples e que a ELM.

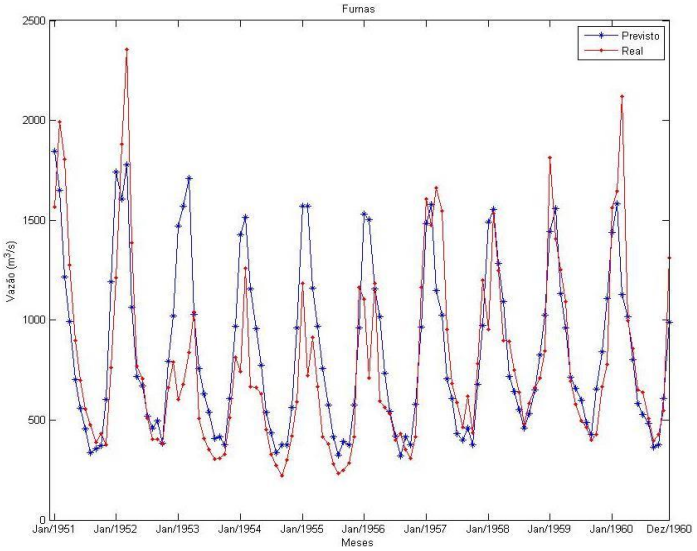
Tabela 5.9- Resultados de Previsão para 12 passos à frente (*P* = 12)

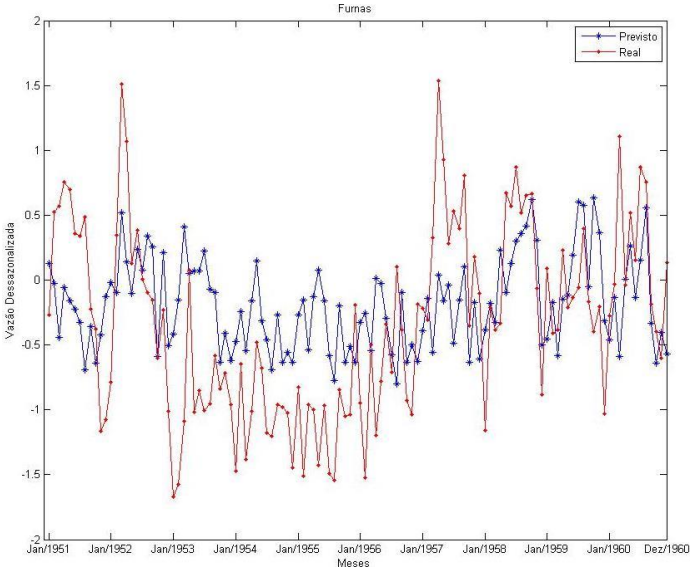
	Preditor	RE	Desvio Padrão	MSE	MAE	MSE dessaz.	MAE dessaz.
FURNAS	**PAR**	0.5591	-	1.1751e+05	238.9580	0.6042	0.6477
	MLP	0.5323	3.4164e+04	1.0653e+05	218.8894	0.5190	0.5840
	ELM	0.4893	8.1905e+13	9.0010e+04	205.8100	0.5065	0.5675
	JAE-ESN	0.4826	7.2450e+03	**8.7553e+04**	**208.3242**	**0.4645**	**0.5697**
	OZT-ESN	0.4893	5.6870e+03	8.9999e+04	213.9805	0.4949	0.5913
	JAE-PV-ESN	0.5219	6.2767e+03	1.0241e+05	228.0470	0.6238	0.6495
	OZT-PV-ESN	0.5411	2.9068e+03	1.1008e+05	231.5001	0.5896	0.6393
	JAE-ESN-ELM	0.5607	2.1975e+04	1.1821e+05	237.4941	0.6181	0.6451
	OZT-ESN-ELM	0.5706	2.0754e+04	1.2242e+05	244.8457	0.6663	0.6741
EMBORCAÇÃO	**PAR**	0.6955	-	6.3676e+04	154.7175	1.2147	0.8265
	MLP	0.6661	7.6614e+03	5.8405e+04	149.1752	0.9600	0.7607
	ELM	0.6430	3.7008e+04	5.4428e+04	144.9070	0.9218	0.7503
	JAE-ESN	0.6331	3.9325e+03	**5.2751e+04**	**133.6270**	**0.9020**	**0.6978**
	OZT-ESN	0.6512	3.0503e+03	5.5822e+04	138.5620	0.9609	0.7259
	JAE-PV-ESN	0.6325	2.1055e+03	5.2662e+04	141.4995	0.9396	0.7428
	OZT-PV-ESN	0.6502	6.2774e+03	5.5655e+04	142.6957	0.9853	0.7535
	JAE-ESN-ELM	0.6903	7.2494e+03	6.2725e+04	151.7704	1.0125	0.7694
	OZT-ESN-ELM	0.7199	1.6028e+04	6.8223e+04	155.5377	1.1418	0.7965
SOBRADINHO	**PAR**	0.5788	-	1.2857e+06	789.4146	1.0262	0.7909
	MLP	0.5471	1.4824e+05	1.1488e+06	724.3030	0.7657	0.6955
	ELM	0.5267	1.8965e+06	**1.0645e+06**	**682.3870**	**0.7058**	**0.6575**
	JAE-ESN	0.5455	6.0187e+04	1.1420e+06	709.5978	0.8548	0.7145
	OZT-ESN	0.5437	1.0949e+05	1.1345e+06	705.5581	0.8300	0.7072
	JAE-PV-ESN	0.5584	3.1120e+04	1.1968e+06	776.0708	1.1464	0.8376
	OZT-PV-ESN	0.5527	4.1870e+04	1.1722e+06	744.7203	0.9462	0.7642
	JAE-ESN-ELM	0.6171	1.5817e+05	1.4618e+06	801.4469	0.9671	0.7686
	OZT-ESN-ELM	0.6145	1.7998e+05	1.4493e+06	800.9222	1.0287	0.7915

A série de Emborcação foi mais bem modelada pela JAE-ESN. Aqui, apenas OZT-ESN-ELM foi pior que o modelo PAR, embora JAE-ESN-ELM tenha alcançado desempenho parecido, inferior a todas as demais metodologias.

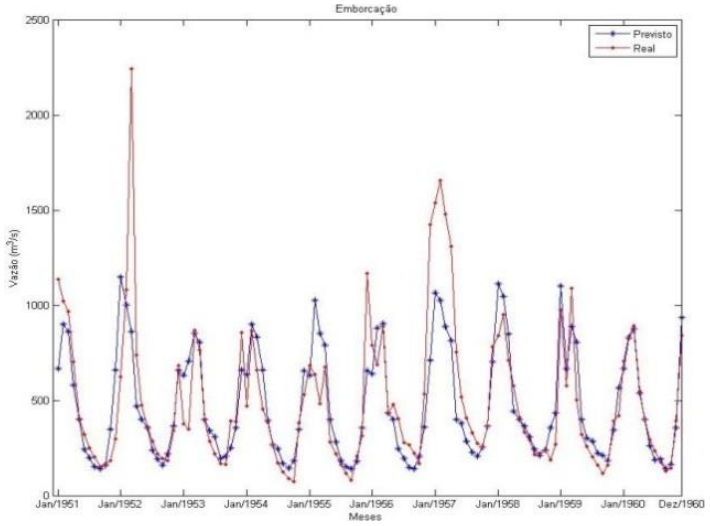
Para a usina de Sobradinho, a ELM foi sempre superior, embora seu desvio padrão tenha sido bastante elevado. Outra vez as ESNs com ELM na camada de saída tiveram resultados piores. Os resultados dos melhores preditores estão na Figura 5.11.

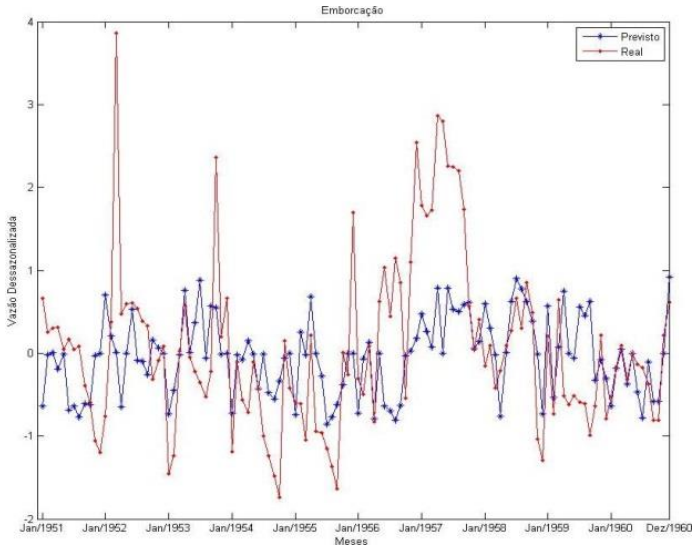
Os resultados finais em termos do MSE real discutidos neste tópico estão sumarizados por meio de gráficos de barras na Figura 5.11. Esta forma de apresentação nos dá uma representação visual muito conveniente, que auxilia na comparação entre os resultados alcançado pelos modelos.


Furnas
Previsto
Real
Vazão (m³/s)
2500
2000
1500
1000
500
0
Jan/1951
Jan/1952
Jan/1953
Jan/1954
Jan/1955
Jan/1956
Jan/1957
Jan/1958
Jan/1959
Jan/1960
Dez/1960
Meses

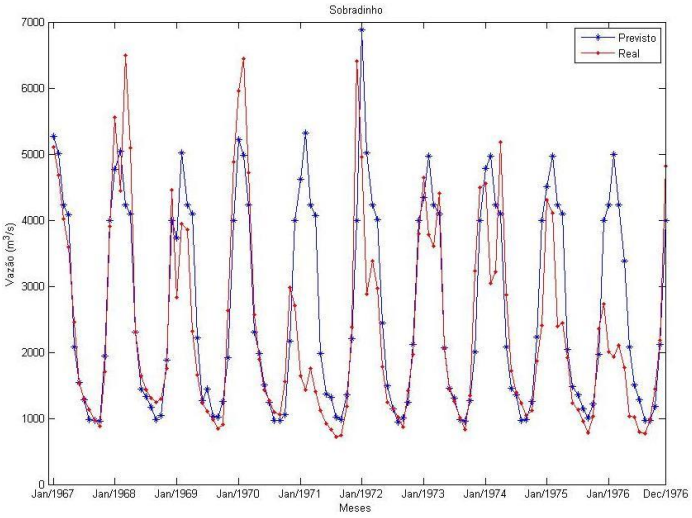

Furnas
Previsto
Real
Vazão Dessazonalizada
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
Jan/1951
Jan/1952
Jan/1953
Jan/1954
Jan/1955
Jan/1956
Jan/1957
Jan/1958
Jan/1959
Jan/1960
Dez/1960
Meses

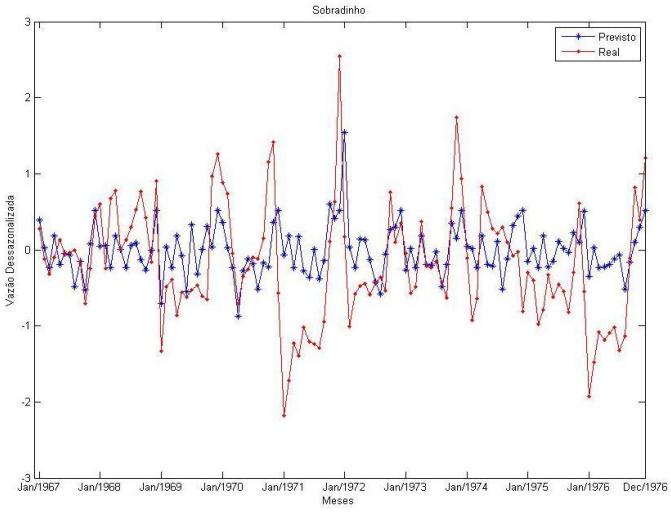
(a)


Emborcação
Previsto
Real
Vazão (m³/s)
Meses
Jan/1951
Jan/1952
Jan/1953
Jan/1954
Jan/1955
Jan/1956
Jan/1957
Jan/1958
Jan/1959
Jan/1960
Dez/1960


Emborcação
Previsto
Real
Vazão Dessazonalizada
Meses
Jan/1951
Jan/1952
Jan/1953
Jan/1954
Jan/1955
Jan/1956
Jan/1957
Jan/1958
Jan/1959
Jan/1960
Dez/1960

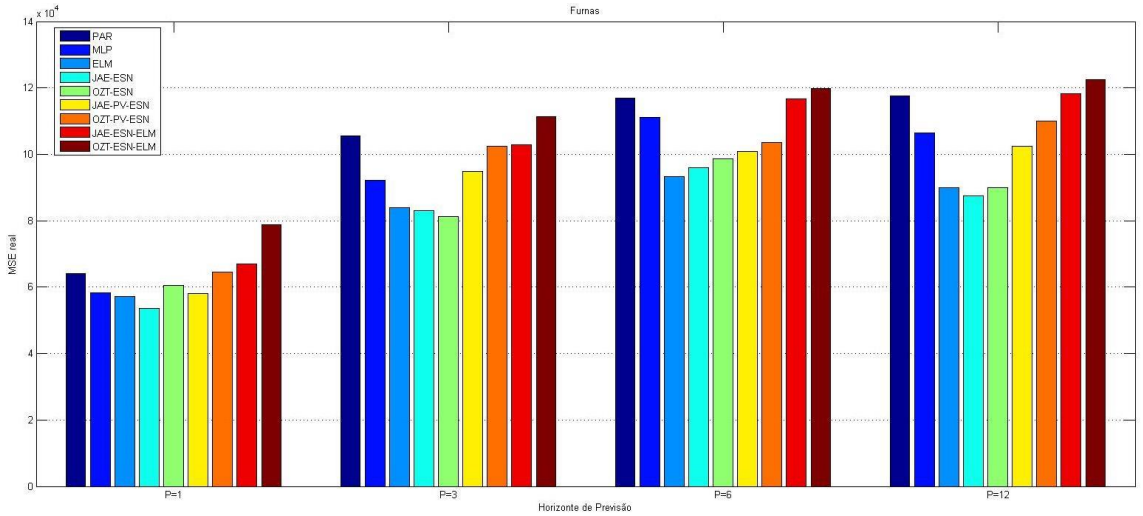
(b)



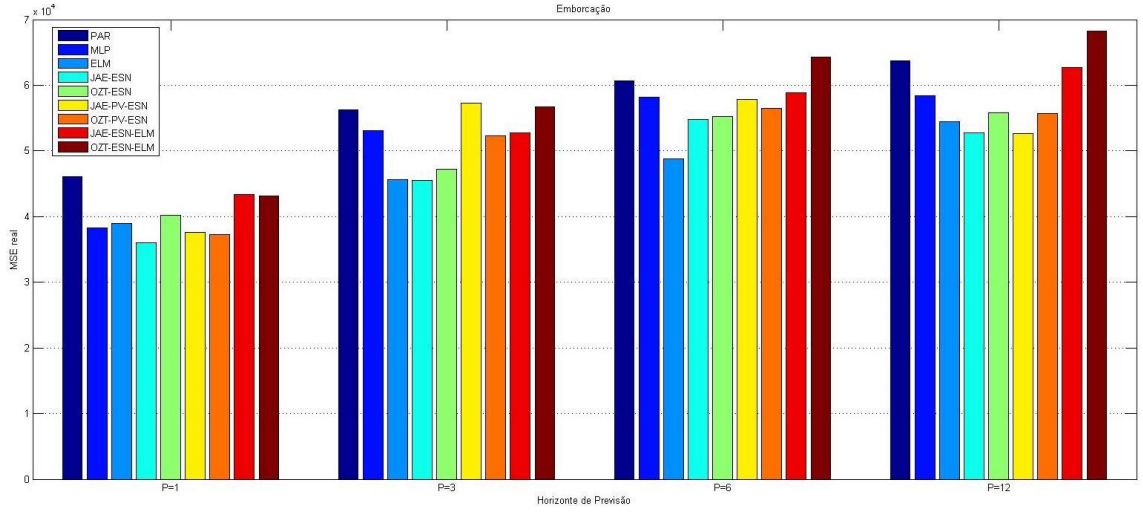


(c)

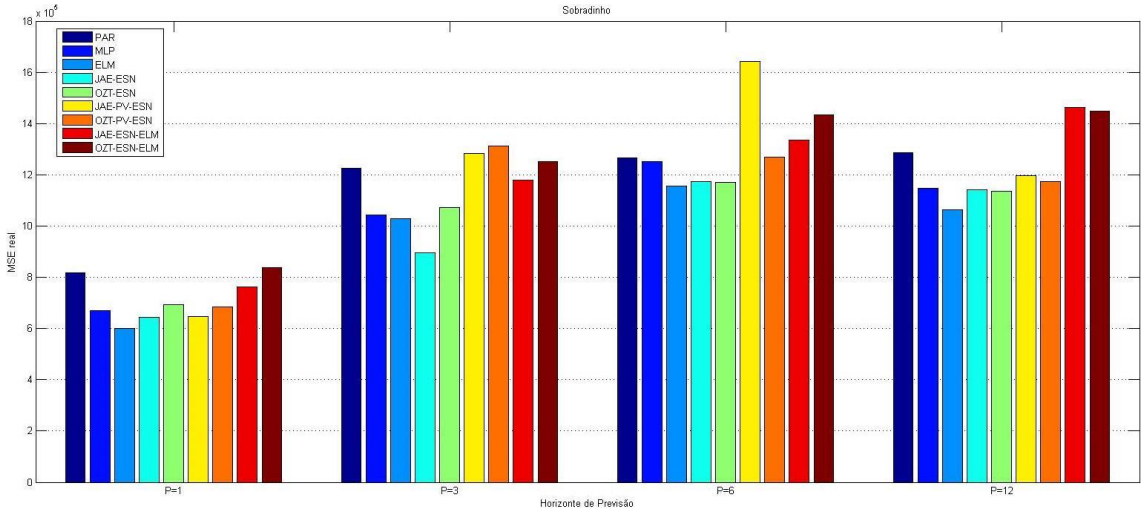
Figura 5.11 – Resultados melhores previsões P=12, real e dessazonalizado – (a) Furnas, (b) Emborcação, (c) Sobradinho



(a)



(b)



(c)

Figura 5.12 – Resultados 1951-1960 – (a) Furnas, (b) Emborcação, (c) Sobradinho

Como esperado, à medida que o horizonte de previsão cresce, a tendência geral dos modelos é de apresentarem piora no desempenho. Este fato é também perceptível ao observar-se as Figuras 5.8 a 5.11 quanto mais elevado for o horizonte, menos os modelos conseguem dar respostas com valores próximos aos desejados. Este comportamento está dentro do esperado em estudos de séries temporais, já que a dependência entre as amostras diminui à medida que o horizonte se eleva (Ballini 2000). Além disso, devido ao uso de previsão recursiva, cada dado reinserido na entrada do modelo já acumula um certo nível de erro que pode deteriorar progressivamente o resultado final.

Um comportamento esperado e verificado é que, quanto mais elevada for a média de longo termo da série, maior será o valor do MSE de previsão. Dessa forma, os erros de Sobradinho superam os de Furnas e estes, por conseguinte, os de Emborcação.

Na Figura 5.12, nota-se que, em geral, as duas primeiras e as duas últimas barras tendem a serem maiores, pois os erros do modelo PAR, da MLP e das ESNs com ELM como camada de saída quase sempre se apresentam menos satisfatórios. No caso geral, a ELM e a ESN de Jaeger se revezam como os melhores preditores: dos 12 casos apresentados, 6 tiveram a JAE-ESN e 5 a ELM com menor MSE real. A comparação dos desempenhos dos reservatórios de dinâmicas, para este período, foi favorável ao de Jaeger, independentemente da camada de saída utilizada.

Logo, como visto, a inserção de uma camada de saída não-linear, para o período 51/60 e com apenas 60 mostras de treinamento, não significou uma maior capacidade de generalização por parte das ESNs.

5.5 Período 1967 a 1976

O período 67-76 é considerado mediano, pois sua média é próxima ao valor calculado para o histórico completo, como mostra a Tabela 5.10.

Tabela 5.10 - Médias e Desvio Padrões período 67/76

Série	Média($\hat{\mu}$)	D. Padrão ($\hat{\sigma}$)
FURNAS 67/76	830.5083	504.7150
EMBORCAÇÃO 67/76	431.8083	305.5643
SOBRADINHO 67/76	2.3768e+03	1.5056e+03

Veja que a média dos 10 anos de Emborcação é menor que em 51-60 por conta da seca de 1970, que derrubou a média do conjunto. Todavia, optou-se por classificar este período ainda como mediano, levando em consideração o comportamento geral das vazões da amostra.

A Tabela 5.11 mostra os resultados computacionais para P=1.

Tabela 5.11- Resultados de Previsão para 1 passo à frente (P = 1)

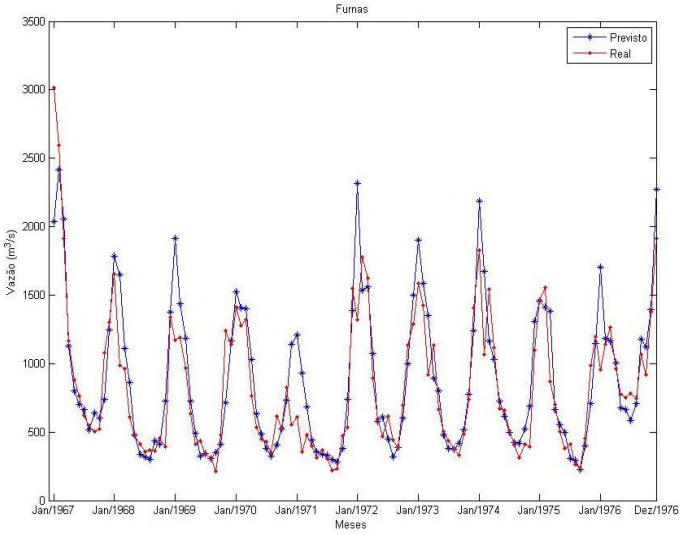
	Preditor	RE	Desvio Padrão	MSE	MAE	MSE dessaz.	MAE
FURNAS	**PAR**	0.4565	-	7.8342e+04	190.5114	0.4143	0.5070
	MLP	0.4351	7.7576e+03	7.1185e+04	174.0230	0.3515	0.4568
	ELM	0.4287	2.9794e+03	6.9085e+04	176.3235	0.3644	0.4702
	JAE-ESN	0.4285	5.1046e+03	6.9031e+04	170.0647	0.3631	0.4526
	OZT-ESN	0.4310	4.8942e+03	6.9845e+04	176.2148	0.3705	0.4719
	JAE-PV-ESN	0.4185	1.4689e+03	**6.5847e+04**	**166.3378**	**0.3199**	**0.4355**
	OZT-PV-ESN	0.4456	1.5635e+03	7.4662e+04	178.5705	0.3645	0.4676
	JAE-ESN-ELM	0.4753	5.8141e+03	8.4952e+04	187.9114	0.4106	0.4908
	OZT-ESN-ELM	0.4823	1.1250e+04	8.7485e+04	197.3610	0.4452	0.5267
EMBORCAÇÃO	**PAR**	0.4855	-	3.1025e+04	108.8272	0.4808	0.4822
	MLP	0.4380	2.8024e+03	2.5254e+04	99.8008	0.4073	0.4490
	ELM	0.4848	1.3688e+03	3.0932e+04	111.9879	0.4507	0.4938
	JAE-ESN	0.4329	2.4208e+03	**2.4669e+04**	**99.4125**	**0.4312**	**0.4641**
	OZT-ESN	0.4495	2.5421e+03	2.6598e+04	104.6349	0.4572	0.4939
	JAE-PV-ESN	0.4495	839.6897	2.6597e+04	103.5376	0.4277	0.4647
	OZT-PV-ESN	0.4590	1.6730e+03	2.7727e+04	104.2252	0.4478	0.4754
	JAE-ESN-ELM	0.4823	5.0217e+03	3.0620e+04	111.2973	0.4810	0.5045
	OZT-ESN-ELM	0.5074	5.3169e+03	3.3884e+04	120.8294	0.5524	0.5634
SOBRADINHO	**PAR**	0.4194	-	6.7514e+05	510.5354	0.3857	0.4339
	MLP	0.4258	2.0069e+05	6.9587e+05	531.1863	0.3950	0.4464
	ELM	0.4133	4.6364e+04	6.5546e+05	500.6412	0.3466	0.4177
	JAE-ESN	0.4118	5.9698e+04	6.5064e+05	505.3692	0.3926	0.4466
	OZT-ESN	0.4262	5.9549e+04	6.9702e+05	538.7252	0.4278	0.4772
	JAE-PV-ESN	0.3999	5.0818e+04	**6.1378e+05**	**488.6399**	**0.3651**	**0.4205**
	OZT-PV-ESN	0.4144	1.6047e+04	6.5901e+05	516.0482	0.3812	0.4415
	JAE-ESN-ELM	0.4546	1.1941e+05	7.9329e+05	565.7017	0.4323	0.4738
	OZT-ESN-ELM	0.4867	1.0923e+05	9.0940e+05	605.3936	0.4847	0.5084

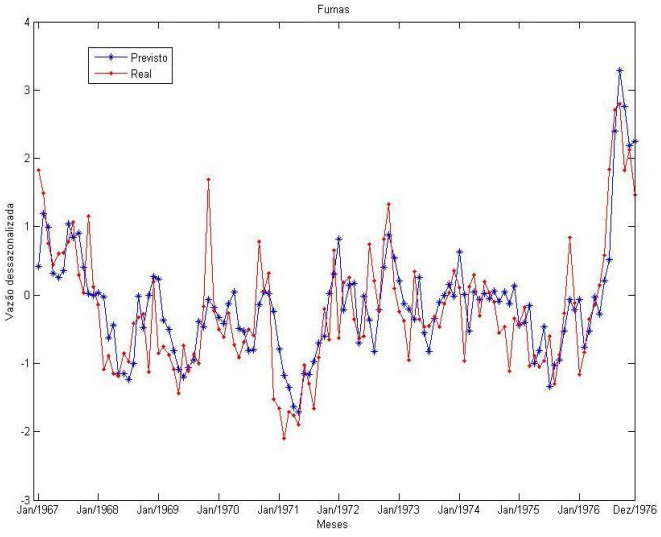
O período 67-76 possui vazão média maior que o de 51-60, de forma que as medidas de erro também terão valor mais elevado. No caso do posto de Furnas, a ESN de Jaeger com filtro de Volterra apresentou os menores valores de MAE, MSE real e MSE dessazonalizado. Como tem sido frequente, as ESNs com ELM tiveram desempenho inferior ao do modelo PAR, ao contrário das demais redes neurais.

Já a série de Emborcação teve como melhor preditor em termos do MSE real a JAE-ESN, assim como em termos do MAE. Neste caso, a MLP apresentou menor MSE dessazonalizado. Apenas a ESN de Ozturk com a ELM como camada de saída teve desempenho inferior ao modelo PAR. A rede JAE-ESN-ELM obteve melhor valor de erro que a ELM no espaço real.

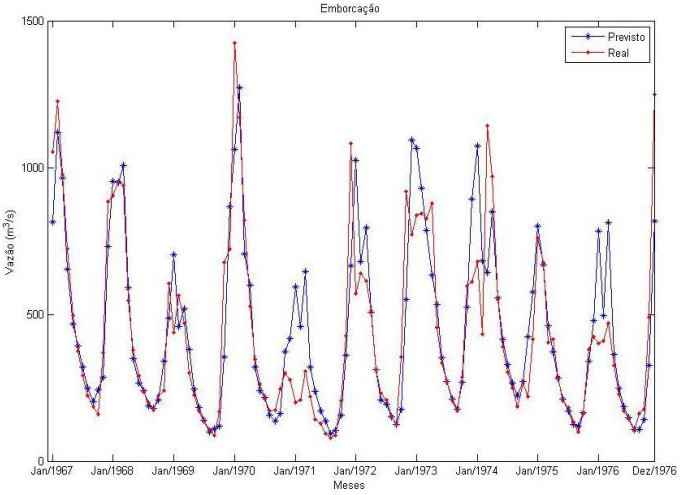
As previsões da usina de Sobradinho tiveram melhores MAE e MSE real para o modelo JAE-PV-ESN, da mesma forma que as de Furnas. Vê-se que a ELM foi o modelo de MSE dessazonalizado mais reduzido, além da observação constante de que as ELM-ESN's foram as de pior desempenho.

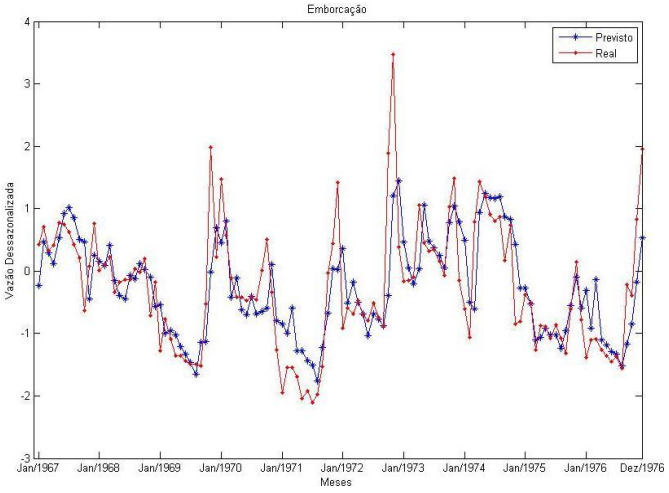
A Figura 5.13 mostra uma execução dos melhores preditores.



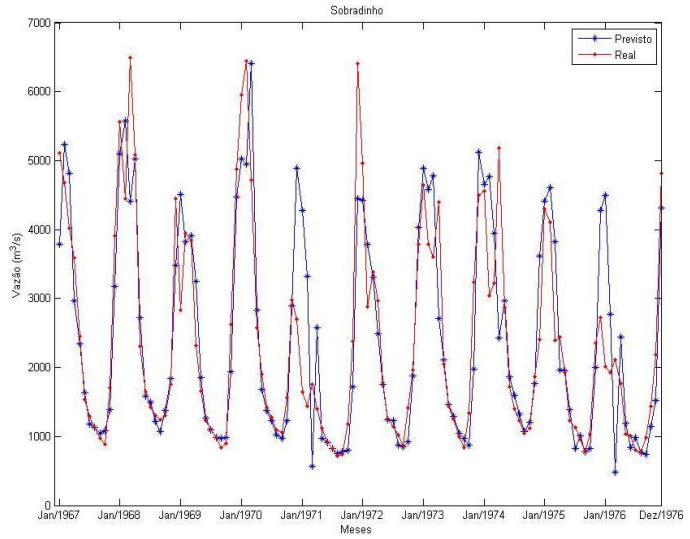


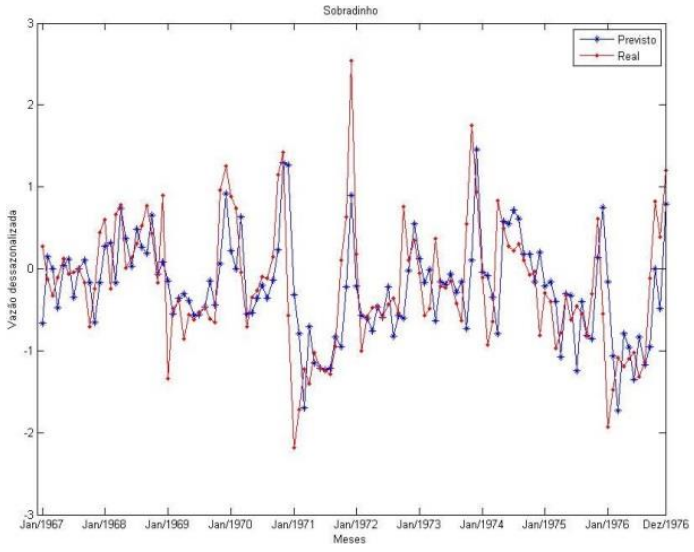
(a)


Emborcação
Previsto
Real
Vazão (m³/s)
1500
1000
500
0
Jan/1967
Jan/1968
Jan/1969
Jan/1970
Jan/1971
Jan/1972
Jan/1973
Jan/1974
Jan/1975
Jan/1976
Dez/1976
Meses


Emborcação
Previsto
Real
Vazão Dessazonalizada
4
3
2
1
0
-1
-2
-3
Jan/1967
Jan/1968
Jan/1969
Jan/1970
Jan/1971
Jan/1972
Jan/1973
Jan/1974
Jan/1975
Jan/1976
Dez/1976
Meses

(b)





(c)

Figura 5.13 – Resultados melhores previsões P=1, real e dessazonalizado – (a) Furnas, (b) Emborcação, (c) Sobradinho

A tabela a seguir apresenta os resultados para três passos à frente.

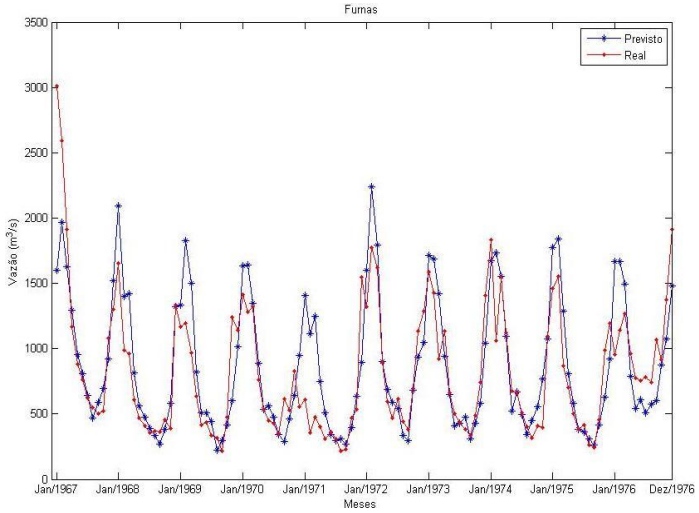
Tabela 5.12 - Resultados de Previsão para 3 passos à frente (*P* = 3)

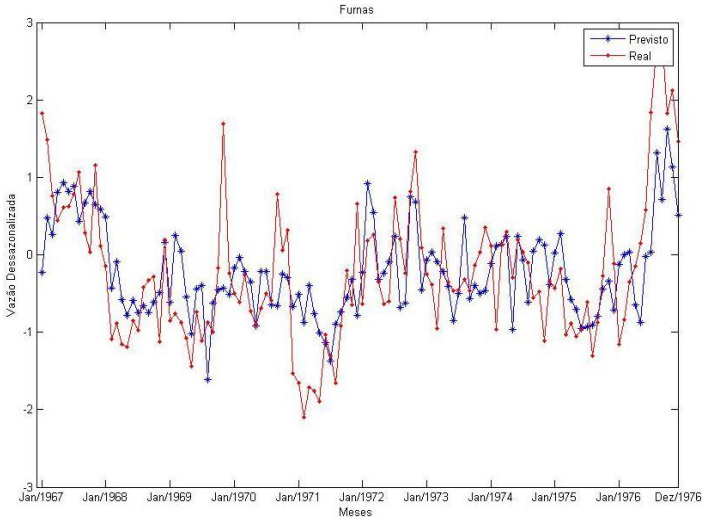
	Preditor	RE	Desvio Padrão	MSE	MAE	MSE dessaz.	MAE
FURNAS	**PAR**	0.5707	-	1.2245e+05	233.4257	0.6716	0.6254
	MLP	0.5530	9.1460e+03	1.1498e+05	224.5139	0.6316	0.6049
	ELM	0.5575	5.9955e+03	1.1684e+05	228.1373	0.6443	0.6081
	JAE-ESN	0.5181	9.4228e+03	**1.0092e+05**	**214.4908**	**0.5614**	**0.5813**
	OZT-ESN	0.5365	7.9858e+03	1.0820e+05	220.3177	0.6039	0.5907
	JAE-PV-ESN	0.5574	4.9147e+03	1.1680e+05	228.6123	0.7002	0.6331
	OZT-PV-ESN	0.5591	1.9620e+04	1.1753e+05	229.2833	0.7085	0.6276
	JAE-ESN-ELM	0.5884	1.6708e+04	1.3017e+05	239.3657	0.7163	0.6414
	OZT-ESN-ELM	0.5927	1.0208e+04	1.3211e+05	242.5528	0.7311	0.6485
EMBORCAÇÃO	**PAR**	0.6258	-	5.1556e+04	144.4390	0.7169	0.6297
	MLP	0.5758	2.7448e+03	4.3640e+04	137.8521	0.6930	0.6195
	ELM	0.5914	2.9567e+03	4.6045e+04	140.0540	0.6911	0.6164
	JAE-ESN	0.5540	3.3183e+03	**4.0400e+04**	**126.7546**	**0.6198**	**0.5779**
	OZT-ESN	0.5752	2.2615e+03	4.3551e+04	129.6751	0.6475	0.5988
	JAE-PV-ESN	0.6070	1.4650e+03	4.8502e+04	145.2032	0.6810	0.6241
	OZT-PV-ESN	0.6219	938.0958	5.0906e+04	147.2547	0.7441	0.6540
	JAE-ESN-ELM	0.6167	6.8820e+03	5.0053e+04	144.1750	0.7450	0.6523
	OZT-ESN-ELM	0.6340	1.2794e+04	5.2915e+04	146.5553	0.8029	0.6796
SOBRADINHO	**PAR**	0.5843	-	1.3102e+06	731.1136	0.6125	0.5854
	MLP	0.5030	2.2249e+05	9.7111e+05	660.3110	0.5445	0.5543
	ELM	0.4588	8.1461e+04	**8.0781e+05**	**584.3865**	**0.4416**	**0.4976**
	JAE-ESN	0.5119	3.5938e+04	1.0057e+06	657.4412	0.5395	0.5545
	OZT-ESN	0.5266	3.7422e+04	1.0642e+06	676.3620	0.5835	0.5717
	JAE-PV-ESN	0.6138	3.3002e+04	1.4458e+06	758.9534	0.6791	0.6246
	OZT-PV-ESN	0.6147	1.3576e+05	1.4502e+06	763.0174	0.6958	0.6216
	JAE-ESN-ELM	0.5534	1.1941e+05	1.1754e+06	694.8062	0.5968	0.5778
	OZT-ESN-ELM	0.5657	1.7766e+05	1.2282e+06	707.5805	0.6332	0.5961

Para *P*=3, todas as métricas de erro analisadas, em casa usina, foram obtidas com o mesmo preditor. No caso de Furnas e Emborcação, o modelo mais adequado foi a JAE-ESN, enquanto, em Sobradinho, foi a ELM. Para Furnas, inclusive, as ESNs foram substancialmente melhores, enquanto as redes com a ELM na camada de saída continuaram não se mostrando adequadas.

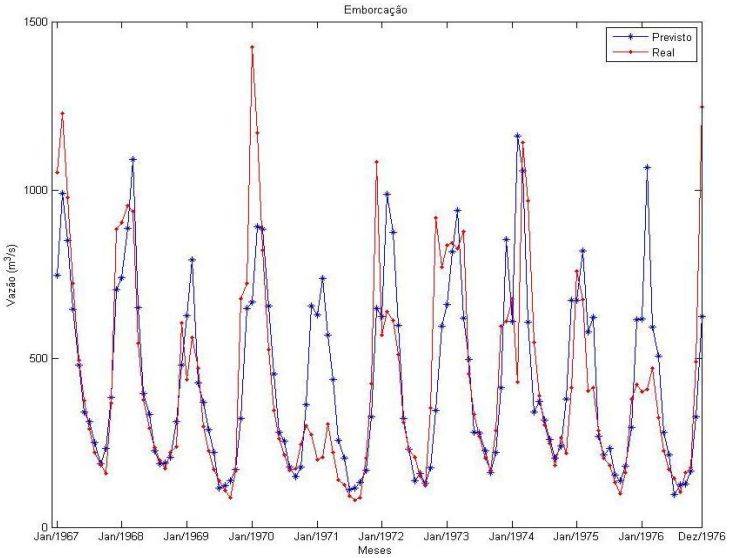
Em Emborcação, a OZT-ESN-ELM foi pior que o modelo PAR, que, por sua vez, teve desempenho inferior aos de todas as outras abordagens. Em Sobradinho, apenas as redes *feedforward* apresentaram erros inferiores à ordem de 10^6. Diferentemente dos casos mostrados até agora, as redes com filtro de Volterra foram as que apresentaram piores resultados, inferiores até aos do modelo PAR, ao contrário das redes com ELM como camada de saída.

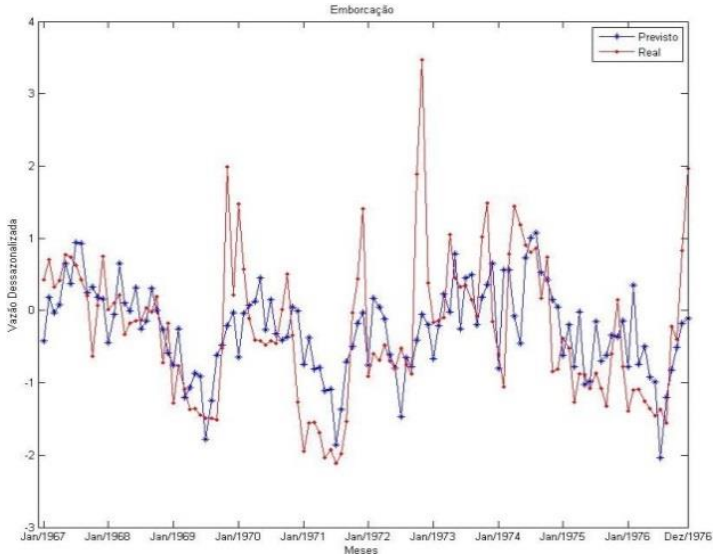
Na Figura 5.14 encontra-se uma realização dos melhores modelos.


Furnas
Previsto
Real
Vazão (m³/s)
3500
3000
2500
2000
1500
1000
500
0
Jan/1967
Jan/1968
Jan/1969
Jan/1970
Jan/1971
Jan/1972
Jan/1973
Jan/1974
Jan/1975
Jan/1976
Dez/1976
Meses

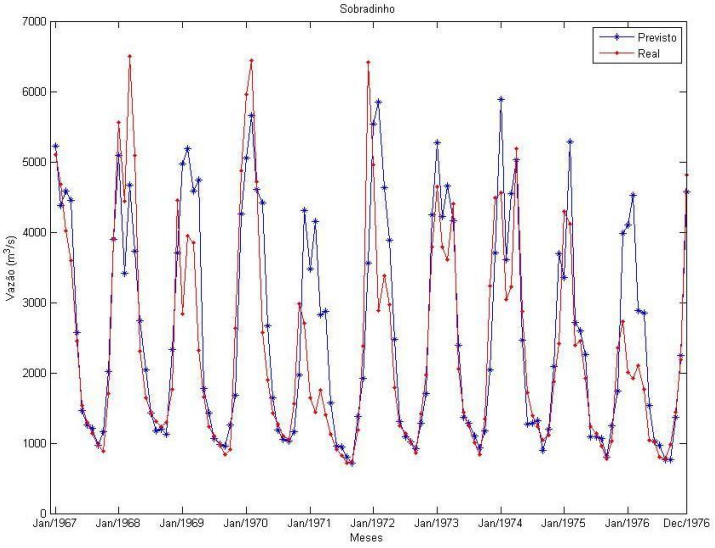

Furnas
Previsto
Real
Vazão Dessazonalizada
3
2
1
0
-1
-2
-3
Jan/1967
Jan/1968
Jan/1969
Jan/1970
Jan/1971
Jan/1972
Jan/1973
Jan/1974
Jan/1975
Jan/1976
Dez/1976
Meses

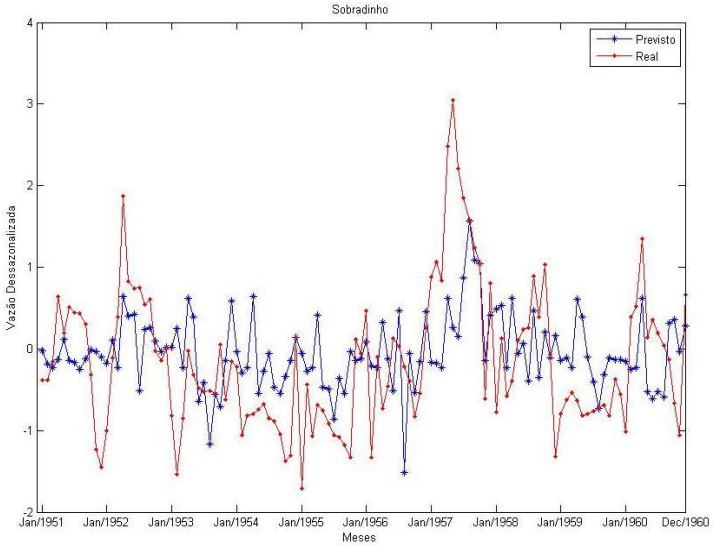
(a)


Emborcação
Previsto
Real
Vazão (m³/s)
1500
1000
500
0
Jan/1967
Jan/1968
Jan/1969
Jan/1970
Jan/1971
Jan/1972
Jan/1973
Jan/1974
Jan/1975
Jan/1976
Dez/1976
Meses


Emborcação
Previsto
Real
Vazão Dessazonalizada
4
3
2
1
0
-1
-2
-3
Jan/1967
Jan/1968
Jan/1969
Jan/1970
Jan/1971
Jan/1972
Jan/1973
Jan/1974
Jan/1975
Jan/1976
Dez/1976
Meses

(b)





(c)

Figura 5.14 – Resultados melhores previsões P=3, real e dessazonalizado – (a) Furnas, (b) Emborcação, (c) Sobradinho

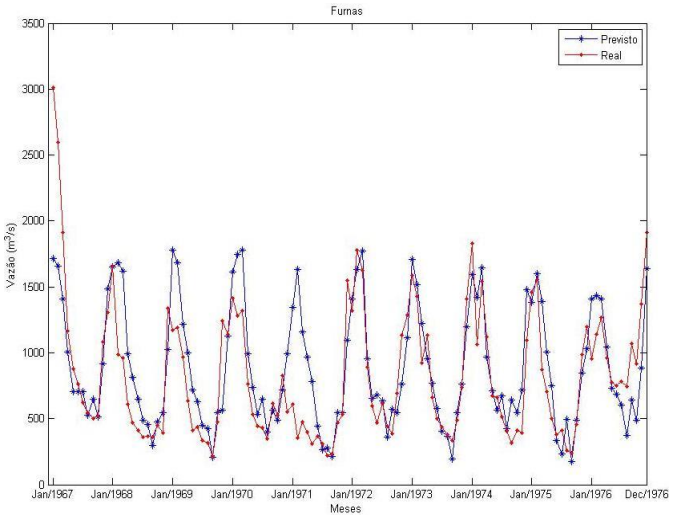
Na Tabela 5.13, estão sumarizados os resultados para P=6.

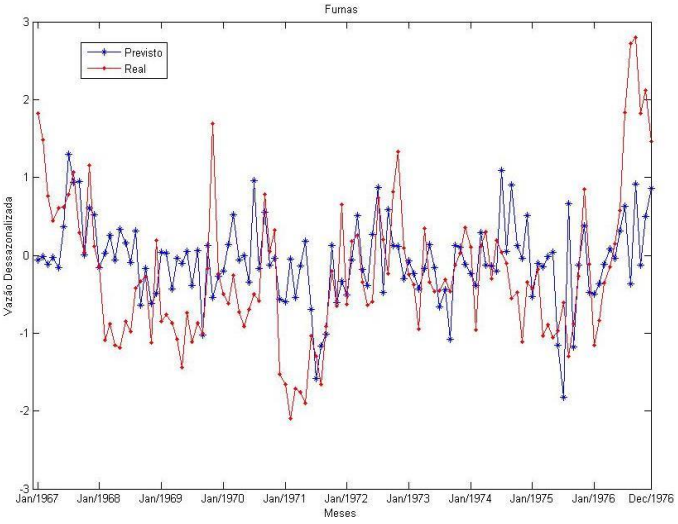
Tabela 5.13- Resultados de Previsão para 6 passos à frente (P = 6)

	Preditor	RE	Desvio Padrão	MSE	MAE	MSE dessaz.	MAE dessaz.
FURNAS	**PAR**	0.6055	-	1.3785e+05	257.2921	0.9549	0.7241
	MLP	0.5967	7.2229e+03	1.3388e+05	252.5452	0.8293	0.7021
	ELM	0.5954	3.3637e+03	1.3330e+05	255.0956	0.8481	0.7245
	JAE-ESN	0.5513	4.6520e+03	1.1427e+05	234.9232	0.7800	0.6811
	OZT-ESN	0.5399	5.2760e+03	**1.0958e+05**	**234.9134**	**0.7611**	**0.6845**
	JAE-PV-ESN	0.5884	4.6699e+03	1.3017e+05	249.1542	0.8817	0.7231
	OZT-PV-ESN	0.5865	4.4359e+03	1.2931e+05	255.4993	0.9012	0.7416
	JAE-ESN-ELM	0.6192	1.4879e+04	1.4416e+05	261.9979	0.9587	0.7401
	OZT-ESN-ELM	0.6164	1.8359e+04	1.4288e+05	262.0502	0.9496	0.7466
EMBORCAÇÃO	**PAR**	0.6281	-	5.1922e+04	155.2827	0.9363	0.7623
	MLP	0.5873	3.6464e+03	4.5404e+04	145.1699	0.8371	0.7064
	ELM	0.5831	2.8734e+03	4.4753e+04	143.4166	0.8095	0.6967
	JAE-ESN	0.5488	2.6868e+03	3.9638e+04	130.7935	0.6940	0.6463
	OZT-ESN	0.5477	3.5810e+03	**3.9481e+04**	**133.8498**	**0.7388**	**0.6717**
	JAE-PV-ESN	0.6516	4.2740e+03	5.5880e+04	160.4752	0.8882	0.7550
	OZT-PV-ESN	0.6410	3.9290e+03	5.4084e+04	157.7922	0.8716	0.7486
	JAE-ESN-ELM	0.6206	7.4749e+03	5.0698e+04	150.4811	0.9089	0.7409
	OZT-ESN-ELM	0.6436	2.0295e+04	5.4524e+04	158.4530	0.9916	0.7842
SOBRADINHO	**PAR**	0.5602	-	1.2042e+06	734.4576	0.7170	0.6439
	MLP	0.4996	1.2383e+05	9.5782e+05	637.3148	0.5413	0.5615
	ELM	0.4872	3.7913e+04	9.1077e+05	621.3228	0.5169	0.5462
	JAE-ESN	0.4726	1.1265e+05	**8.5712e+05**	**595.4681**	**0.5210**	**0.5410**
	OZT-ESN	0.4735	1.2474e+05	8.6054e+05	599.0889	0.5302	0.5447
	JAE-PV-ESN	0.5069	4.4003e+04	9.8620e+05	665.8535	0.6281	0.6261
	OZT-PV-ESN	0.5228	3.2350e+04	1.0488e+06	678.6323	0.6230	0.6113
	JAE-ESN-ELM	0.5413	1.9521e+05	1.1248e+06	683.1421	0.6304	0.5989
	OZT-ESN-ELM	0.5790	2.4618e+05	1.2867e+06	727.3206	0.7015	0.6375

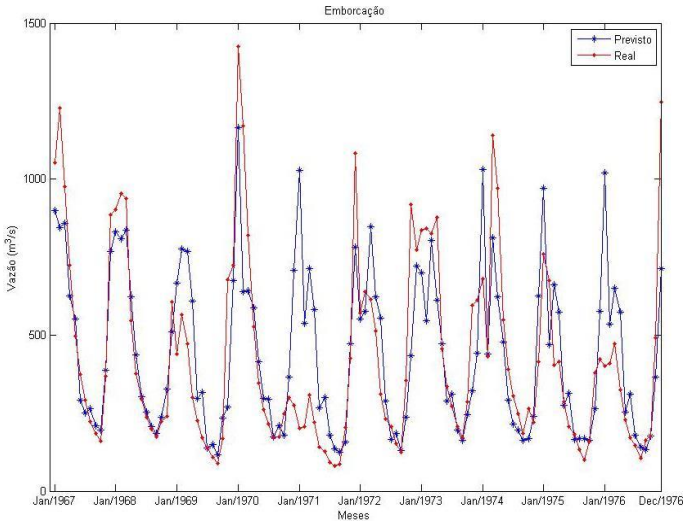
O desempenho computacional associado à usina de Furnas para esse horizonte de previsão teve como melhor representante a rede de estado de eco de Ozturk et al. para todas as métricas de erro. Em Emborcação, esse fato se repetiu para o MSE real, mas, para os MAE e MSE dessazonalizados, o desempenho foi favorável à rede JAE-ESN. Chama a atenção o fato de que as ESNs com camadas de saída não-lineares foram piores que o modelo PAR, à exceção de JAE-ESN-ELM.

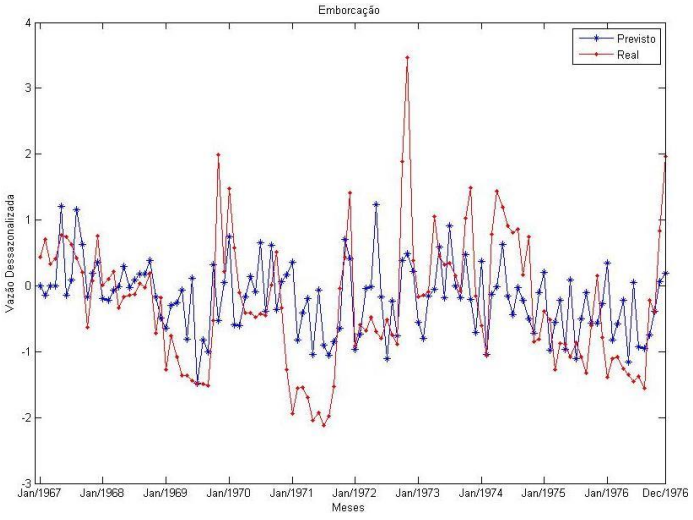
Para o posto de Sobradinho, a JAE-ESN foi a rede de menor MSE real sendo este, muito próximo ao da OZT-ESN. O menor MSE dessazonalizado foi o da ELM. Apenas a OZT-ESN-ELM não superou o modelo linear.


Furnas
Previsto
Real
Vazão (m³/s)
Meses
Jan/1967
Jan/1968
Jan/1969
Jan/1970
Jan/1971
Jan/1972
Jan/1973
Jan/1974
Jan/1975
Jan/1976
Dec/1976

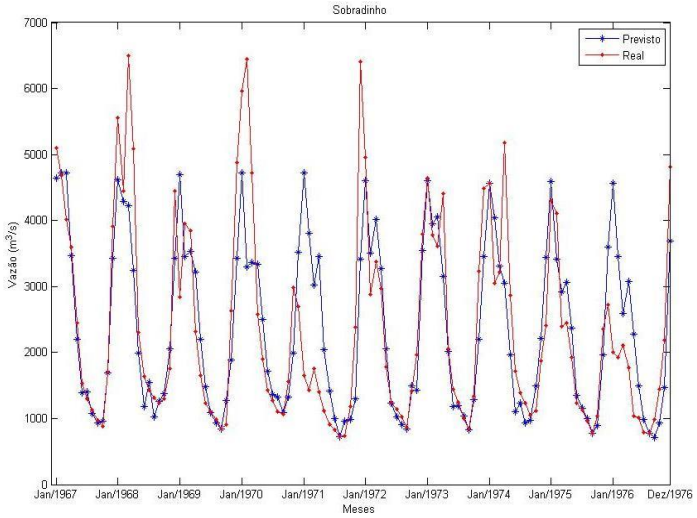

Furnas
Previsto
Real
Vazão Dessazonalizada
Meses
Jan/1967
Jan/1968
Jan/1969
Jan/1970
Jan/1971
Jan/1972
Jan/1973
Jan/1974
Jan/1975
Jan/1976
Dec/1976

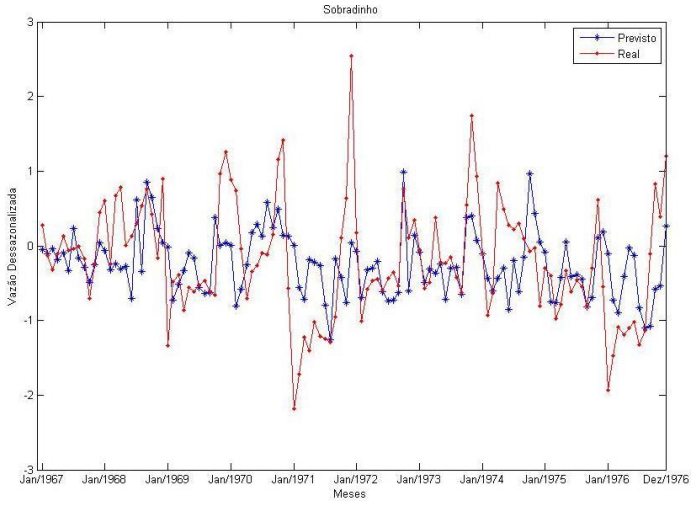
(a)


Emborcação
Previsto
Real
Vazão (m³/s)
1500
1000
500
0
Jan/1967
Jan/1968
Jan/1969
Jan/1970
Jan/1971
Jan/1972
Jan/1973
Jan/1974
Jan/1975
Jan/1976
Dec/1976
Meses


Emborcação
Previsto
Real
Vazão Dessazonalizada
4
3
2
1
0
-1
-2
-3
Jan/1967
Jan/1968
Jan/1969
Jan/1970
Jan/1971
Jan/1972
Jan/1973
Jan/1974
Jan/1975
Jan/1976
Dec/1976
Meses

(b)





(c)

Figura 5.15 – Resultados melhores previsões P=6, real e dessazonalizado – (a) Furnas, (b) Emborcação, (c) Sobradinho

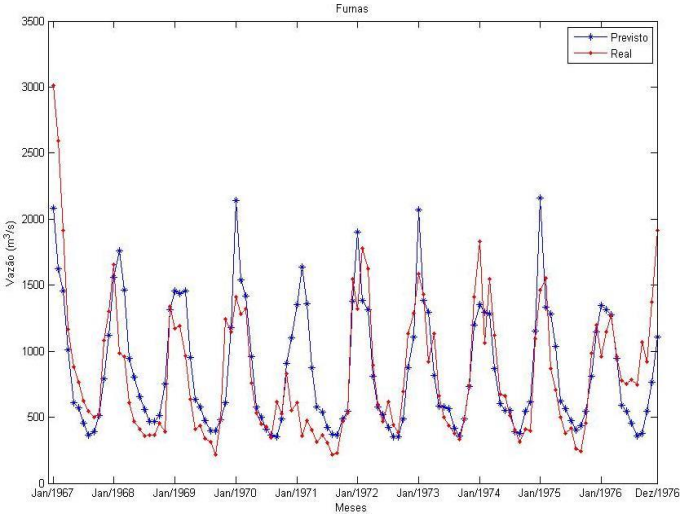
Por fim, na Tabela 5.14, encontram-se os resultados para o horizonte de 12 passos à frente.

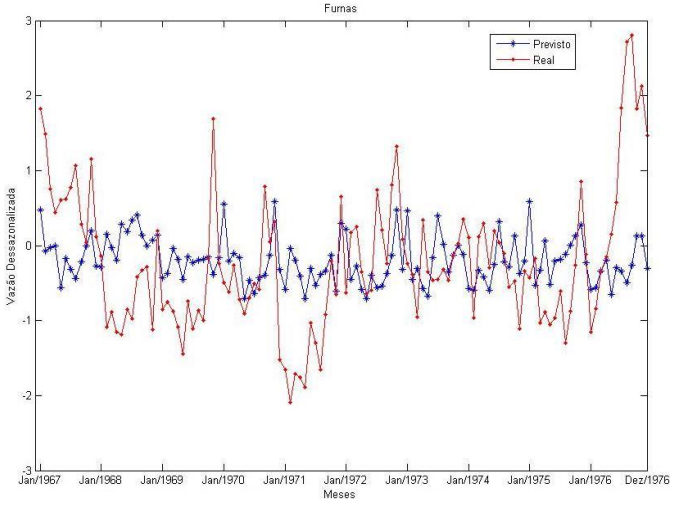
Tabela 5.14 - Resultados de Previsão para 12 passos à frente (*P* = 12)

	Preditor	RE	Desvio Padrão	MSE	MAE	MSE dessaz.	MAE dessaz.
FURNAS	**PAR**	0.6063	-	1.3821e+05	266.1884	1.0348	0.7959
	MLP	0.6064	1.4392e+04	1.3825e+05	262.8031	0.9018	0.7534
	ELM	0.5913	3.4498e+04	1.3147e+05	257.7447	0.8806	0.7464
	JAE-ESN	0.5609	6.5401e+03	**1.1829e+05**	**246.3835**	**0.8781**	**0.7258**
	OZT-ESN	0.5657	4.8424e+03	1.2032e+05	251.2755	0.8766	0.7425
	JAE-PV-ESN	0.6036	2.2038e+04	1.3700e+05	265.0567	1.0347	0.7951
	OZT-PV-ESN	0.5956	9.3584e+03	1.3336e+05	256.2689	0.9047	0.7500
	JAE-ESN-ELM	0.6214	1.6315e+04	1.4519e+05	271.5560	1.0098	0.7886
	OZT-ESN-ELM	0.6385	1.7760e+04	1.5329e+05	281.4755	1.0916	0.8283
EMBORCAÇÃO	**PAR**	0.6384	-	5.3652e+04	165.1184	1.1667	0.8770
	MLP	0.6556	9.4870e+03	5.6580e+04	161.0389	1.0196	0.8194
	ELM	0.6056	2.4096e+05	4.8275e+04	155.9140	1.0064	0.8185
	JAE-ESN	0.5428	3.0098e+03	3.8788e+04	138.6305	0.8870	0.7581
	OZT-ESN	0.5352	2.5049e+03	**3.7700e+04**	**136.6428**	**0.8969**	**0.7498**
	JAE-PV-ESN	0.6572	2.4460e+03	5.6850e+04	162.4813	1.0615	0.8319
	OZT-PV-ESN	0.6498	3.9429e+03	5.5584e+04	164.3458	1.0961	0.8616
	JAE-ESN-ELM	0.6406	7.1236e+03	5.4010e+04	161.7358	1.1031	0.8451
	OZT-ESN-ELM	0.6490	1.3691e+04	5.5439e+04	162.3824	1.1853	0.8643
SOBRADINHO	**PAR**	0.5526	-	1.1718e+06	742.0075	0.6817	0.6811
	MLP	0.4837	1.0664e+05	8.9782e+05	629.8880	0.5881	0.5968
	ELM	0.4880	5.4821e+04	9.1404e+05	617.2839	0.5312	0.5688
	JAE-ESN	0.4603	1.1078e+05	8.1325e+05	596.4277	0.5712	0.5862
	OZT-ESN	**0.4344**	**1.5468e+05**	**7.2407e+05**	**568.8259**	**0.5351**	**0.5734**
	JAE-PV-ESN	0.4888	4.5997e+04	9.1701e+05	634.8431	0.6408	0.6254
	OZT-PV-ESN	0.5120	5.2913e+04	1.0061e+06	671.7733	0.7047	0.6683
	JAE-ESN-ELM	0.5532	1.9660e+05	1.1745e+06	691.5639	0.6823	0.6377
	OZT-ESN-ELM	0.6031	2.9958e+05	1.3960e+06	762.2304	0.7665	0.6855

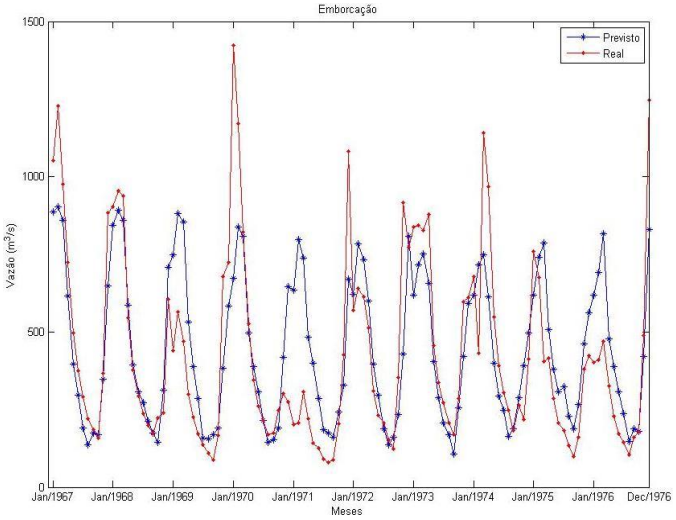
Na previsão de *P*=12 do período 67-76, as máquinas desorganizadas do tipo rede de estados de eco, nas versões com combinador linear como camada de saída, foram as mais consistentes em termos de desempenho e erro. Para a série de Furnas, JAE-ESN foi a de menor MSE real e MAE, enquanto o MSE dessazonalizado de menor magnitude foi obtido pela OZT-ESN. O comportamento inverso foi notado em Emborcação, na qual OZT-ESN teve menores MSE e MAE reais e JAE-ESN o menor MSE dessazonalizado. Resultados parecidos foram observados para a série de Sobradinho, com o adendo de que a ELM foi a de menor MSE dessazonalizado.

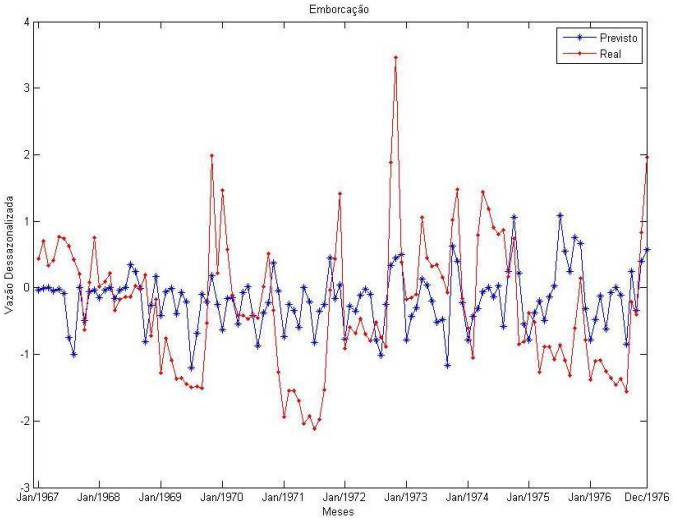
Observa-se que, para as duas primeiras séries, além das arquiteturas com ELM como camada de saída, a MLP também foi inferior ao modelo linear no tocante ao MSE real. Para a série de Emborcação, isso também ocorre com os modelos de ESN com filtro de Volterra. Novamente, apresentamos uma execução dos melhores preditores na Figura 5.16


Furnas
Previsto
Real
Vazão (m3/s)
3500
3000
2500
2000
1500
1000
500
0
Jan/1967
Jan/1968
Jan/1969
Jan/1970
Jan/1971
Jan/1972
Jan/1973
Jan/1974
Jan/1975
Jan/1976
Dez/1976
Meses

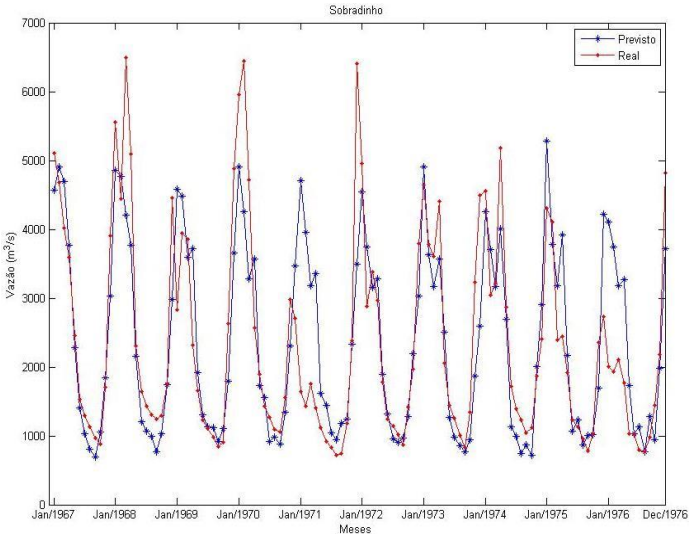

Furnas
Previsto
Real
Vazão Dessazonalizada
3
2
1
0
-1
-2
-3
Jan/1967
Jan/1968
Jan/1969
Jan/1970
Jan/1971
Jan/1972
Jan/1973
Jan/1974
Jan/1975
Jan/1976
Dez/1976
Meses

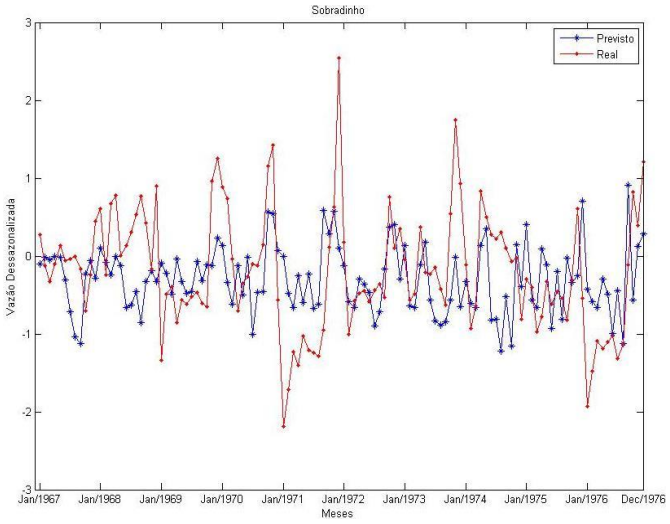
(a)


Emborcação
Previsto
Real
Vazão (m³/s)
1500
1000
500
0
Jan/1967
Jan/1968
Jan/1969
Jan/1970
Jan/1971
Jan/1972
Jan/1973
Jan/1974
Jan/1975
Jan/1976
Dec/1976
Meses


Emborcação
Previsto
Real
Vazão Dessazonalizada
4
3
2
1
0
-1
-2
-3
Jan/1967
Jan/1968
Jan/1969
Jan/1970
Jan/1971
Jan/1972
Jan/1973
Jan/1974
Jan/1975
Jan/1976
Dec/1976
Meses

(b)

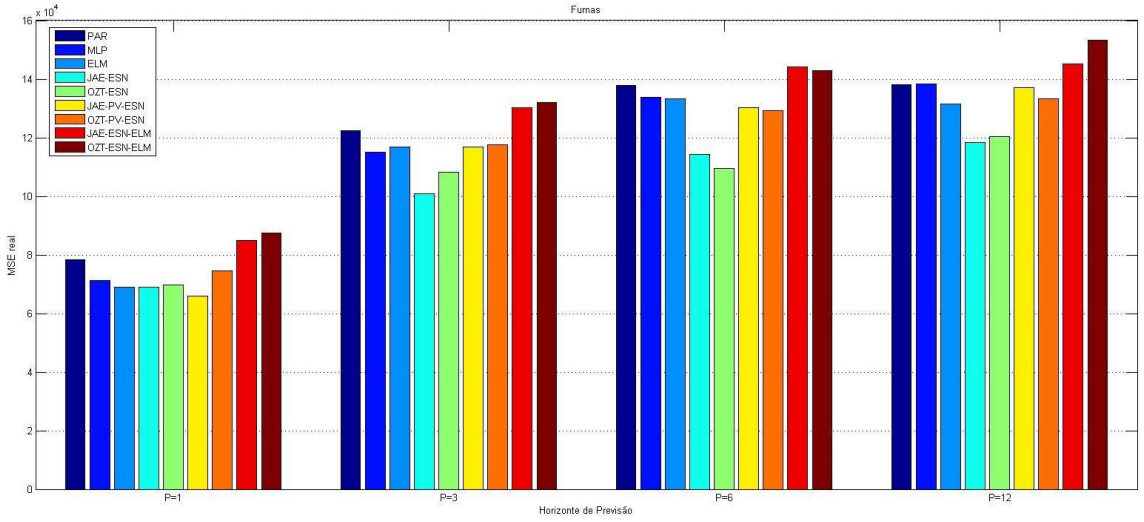




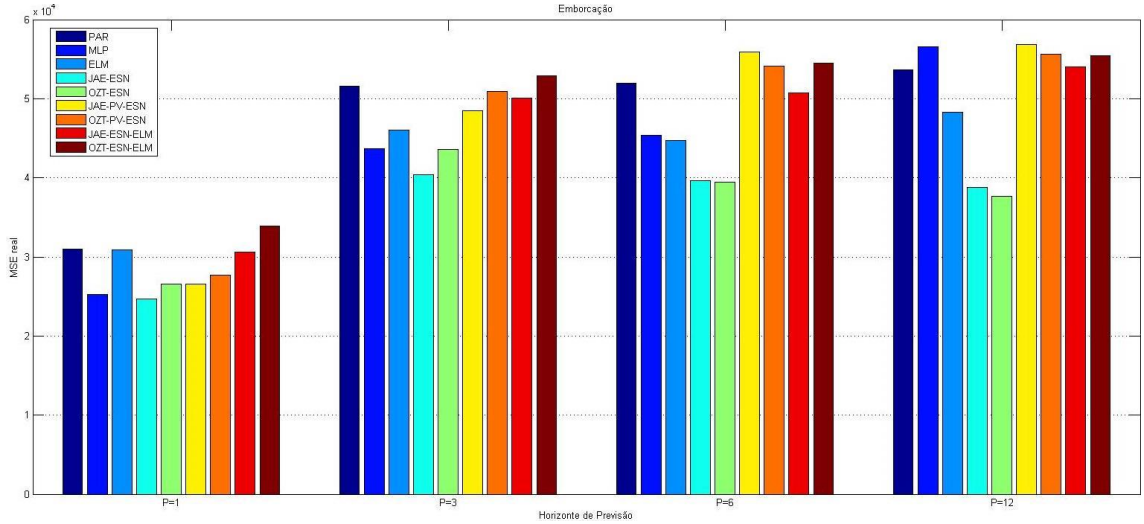
(c)

Figura 5.16 – Resultados melhores previsões P=12, real e dessazonalizado – (a) Furnas, (b) Emborcação, (c) Sobradinho

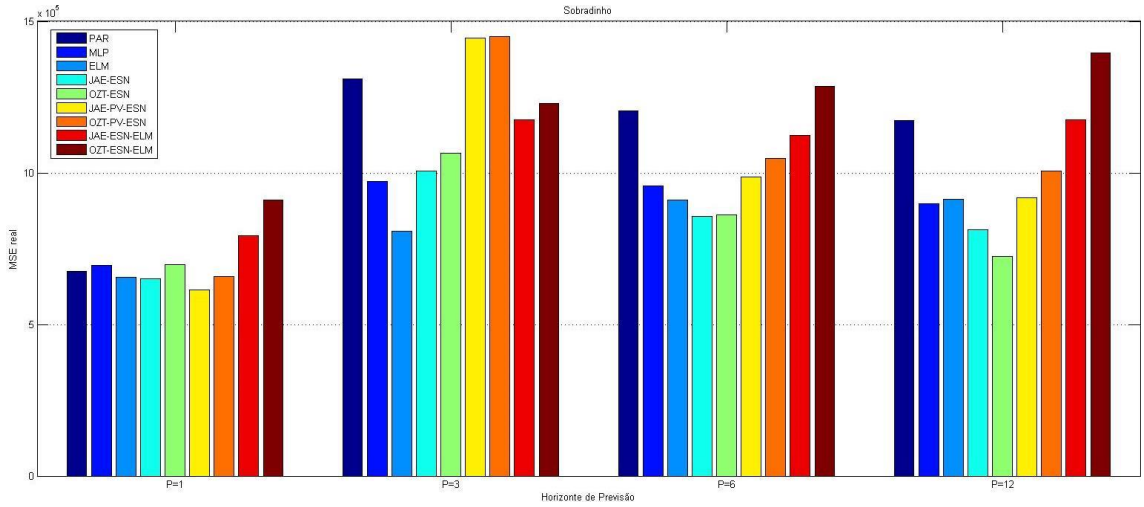
O gráfico em barras que resume o comportamento geral dos modelos de previsão para o período 1967 a 1976 é apresentado na Figura 5.17.



(a)



(b)



(c)

Figura 5.17 - Resultados 1967-1976 – (a) Furnas, (b) Emborcação, (c) Sobradinho

Os gráficos permitem observar que o comportamento do erro de previsão foi, em certa medida, próximo àquele observado no período anterior de 51-60, com a ESN de Jaeger obtendo menores erros em 5 casos e a ESN de Ozturk et al. em 4 casos.

A diferença é que os modelos de ESN munidos do filtro de Volterra chegaram a ser piores que os modelos com a ELM em 4 casos. Por outro lado, em 2 casos, a JAE-PV-ESN alcançou os melhores resultados dentre todos os preditores, para P=1 nas séries de Furnas e Emborcação. Mesmo assim, em geral, as barras relativas ao modelo PAR, MLP e ESNs com ELM foram as que apresentaram MSE real mais elevado.

A seguir apresentaremos o último período que foi objeto de estudo deste trabalho.

5.6 Período 1977 a 1986

O período compreendido entre 1977 e 1986 é muito utilizado em estudos de previsão de vazões, pois, entre estes anos, registram-se volumes muito elevados, de forma que os erros de previsão são normalmente mais altos. Da mesma forma, é ainda mais difícil para os modelos alcançar os picos dos meses de cheia, o que degrada ainda mais o resultado final de qualquer modelo e torna este período bastante desafiador. A Tabela 5.15 apresenta as médias e desvios padrões do período de testes.

Tabela 5.15 - Médias e Desvio Padrões período 77/86

Série	Média($\hat{\mu}$)	D. Padrão ($\hat{\sigma}$)
FURNAS 77/86	1.1678e+03	745.1770
EMBORCAÇÃO 77/86	594.7833	454.6388
SOBRADINHO 77/86	3.4628e+03	2.5430e+03

Na Tabela 5.16, estão sumarizados os resultados para P=1.

Os resultados computacionais para este horizonte apresentaram, pela primeira vez, como melhor preditor em relação ao MSE dessazonalizado, para todos os postos, o modelo linear PAR. Todavia, apenas em Furnas, as demais métricas de erro foram favoráveis a este modelo.

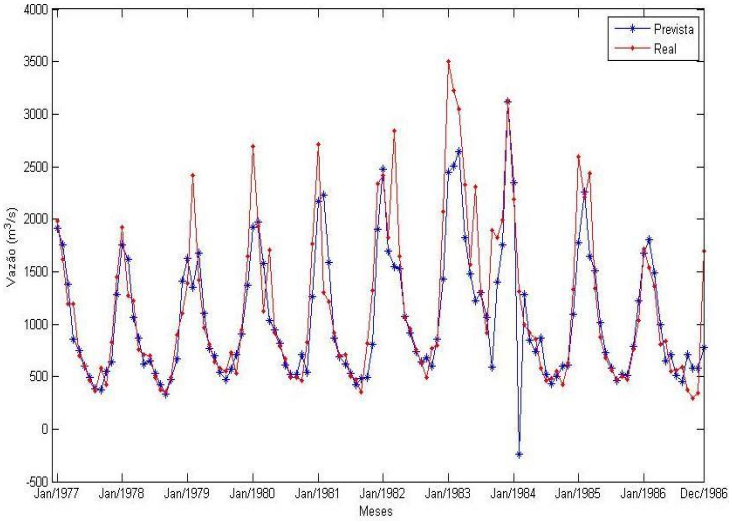
A rede MLP foi a que se mostrou mais adequada para o posto de Emborcação. Já a série de Sobradinho teve como modelo de menor MSE real a JAE-PV-ESN, que, coincidentemente tinha sido a rede com o segundo melhor desempenho para as demais

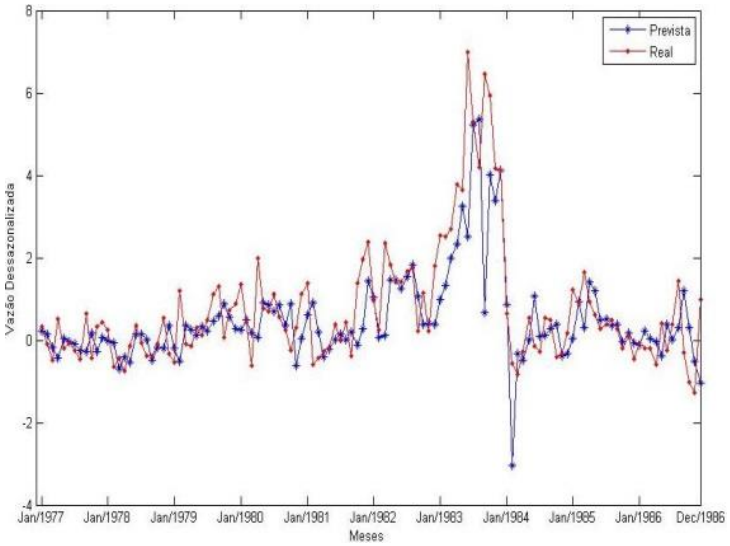
usinas. Aliás, o modelo OZT-PV-ESN foi sistematicamente de pior resultado que este nos 3 casos.

Tabela 5.16 - Resultados de Previsão para 1 passo à frente (*P* = 1)

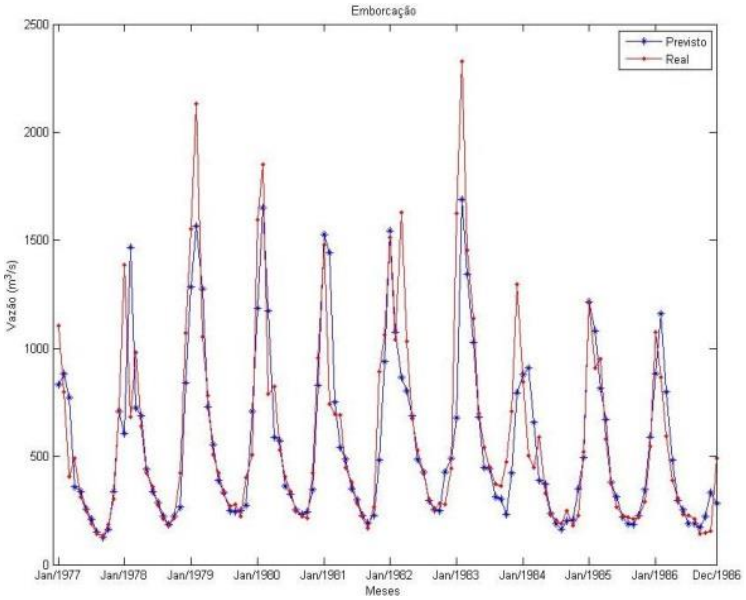
	Preditor	RE	Desvio Padrão	MSE	MAE	MSE dessaz.	MAE dessaz.
FURNAS	**PAR**	0.5936	**-**	**1.3250e+05**	**238.7345**	**0.8497**	**0.6463**
	MLP	0.6210	4.4811e+04	1.4499e+05	254.8895	0.8830	0.6852
	ELM	0.6362	5.1342e+04	1.5217e+05	248.6592	1.0889	0.6882
	JAE-ESN	0.6674	1.4092e+04	1.6748e+05	279.2495	1.4414	0.8274
	OZT-ESN	0.6806	1.2620e+04	1.7418e+05	283.4507	1.4083	0.8153
	JAE-PV-ESN	0.6127	7.8323e+03	1.4115e+05	248.9771	1.1172	0.7187
	OZT-PV-ESN	0.6472	9.7436e+03	1.5748e+05	262.2232	1.2207	0.7383
	JAE-ESN-ELM	0.6753	1.2518e+04	1.7148e+05	273.3874	1.3735	0.7895
	OZT-ESN-ELM	0.7139	1.8552e+04	1.9164e+05	285.1305	1.4415	0.8047
EMBORCAÇÃO	**PAR**	0.6638	-	5.8007e+04	145.1318	0.5982	0.5753
	MLP	0.6301	4.7870e+04	**5.2253e+04**	**132.8129**	**0.6206**	**0.5597**
	ELM	0.6592	4.0064e+03	5.7199e+04	148.5374	0.6600	0.6115
	JAE-ESN	0.6542	4.3864e+03	5.6338e+04	142.8734	0.6611	0.6081
	OZT-ESN	0.6601	9.6238e+03	5.7350e+04	143.9419	0.6862	0.6243
	JAE-PV-ESN	0.6456	4.1747e+03	5.4857e+04	139.0384	0.6245	0.5758
	OZT-PV-ESN	0.6675	1.8208e+03	5.8652e+04	143.8039	0.6582	0.5912
	JAE-ESN-ELM	0.7053	9.0792e+03	6.5480e+04	156.5980	0.7857	0.6671
	OZT-ESN-ELM	0.7169	6.0889e+03	6.7650e+04	159.6321	0.8190	0.6823
SOBRADINHO	**PAR**	0.6102	-	1.4290e+06	704.2105	0.5247	0.5514
	MLP	0.6137	1.0137e+06	1.4456e+06	708.8126	0.5523	0.5492
	ELM	0.6015	1.3522e+05	1.3887e+06	727.6945	0.5556	0.5758
	JAE-ESN	0.6330	1.1733e+05	1.5379e+06	719.0248	0.6084	0.5995
	OZT-ESN	0.6266	9.9235e+04	1.5066e+06	723.2284	0.5870	0.5996
	JAE-PV-ESN	0.5939	2.5059e+05	**1.3538e+06**	**689.8533**	**0.5256**	**0.5505**
	OZT-PV-ESN	0.6140	3.9486e+04	1.4470e+06	742.1569	0.5547	0.5846
	JAE-ESN-ELM	0.6458	1.4817e+05	1.6010e+06	753.1627	0.6568	0.6113
	OZT-ESN-ELM	0.6521	2.2394e+05	1.6323e+06	754.8090	0.6717	0.6174

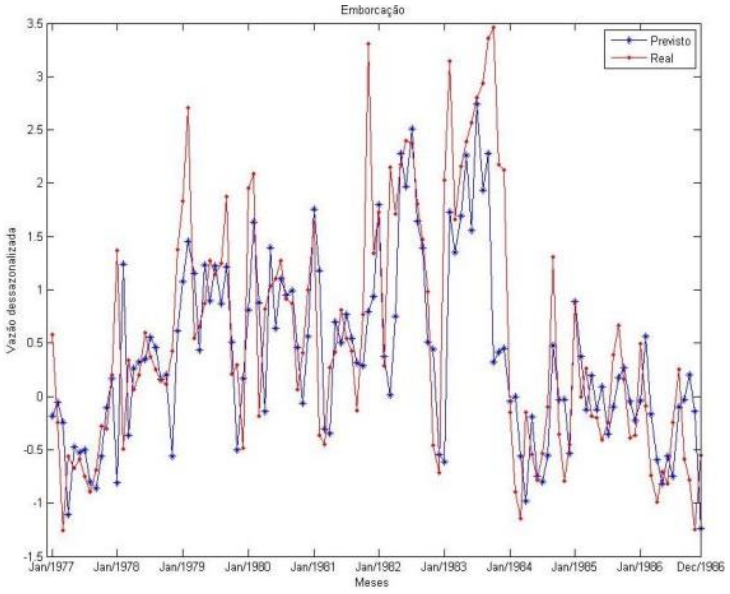
O fato interessante é que, em Furnas e Sobradinho, as ESNs de Jaeger e Ozturk et al. não tiveram bom desempenho comparativo como vinha ocorrendo. Este fato revela quão difícil para qualquer modelo é prever este período. Veja também que, em Sobradinho, a MLP teve um alto valor de desvio padrão, fruto de uma execução com resultado muito ruim.


Prevista
Real
Vazão (m³/s)
Meses
Jan/1977
Jan/1978
Jan/1979
Jan/1980
Jan/1981
Jan/1982
Jan/1983
Jan/1984
Jan/1985
Jan/1986
Dec/1986

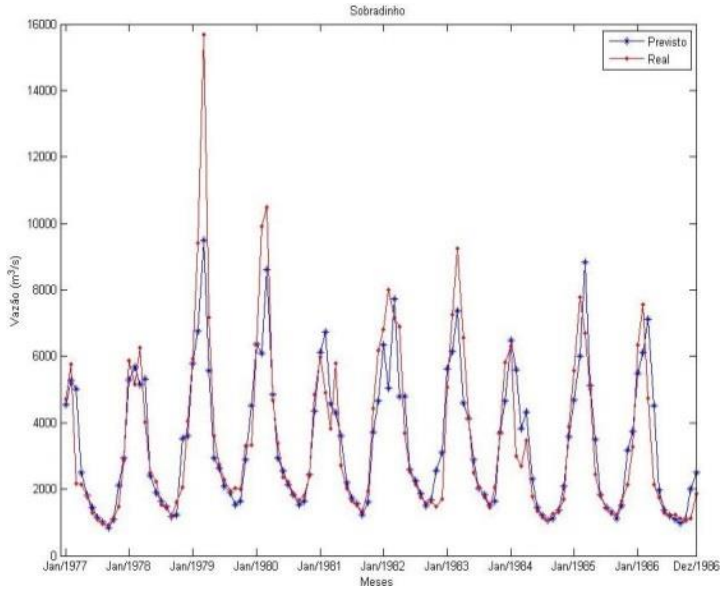

Prevista
Real
Vazão Dessazonalizada
Meses
Jan/1977
Jan/1978
Jan/1979
Jan/1980
Jan/1981
Jan/1982
Jan/1983
Jan/1984
Jan/1985
Jan/1986
Dec/1986

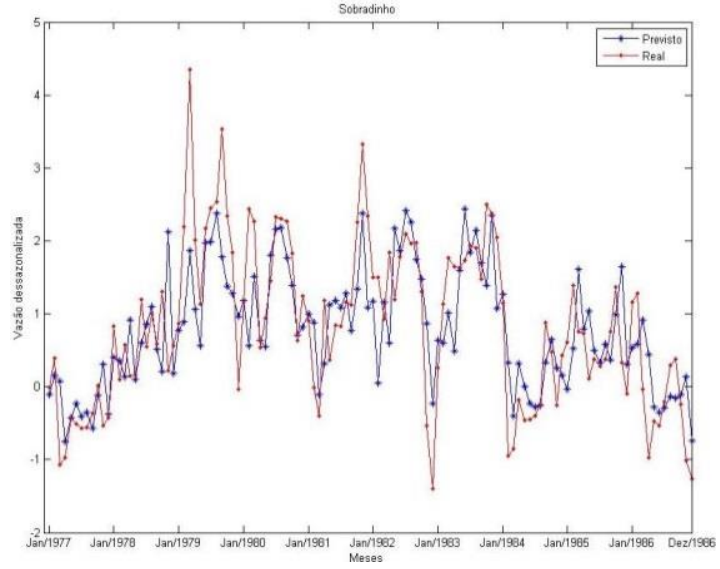
(a)


Emborcação
Previsto
Real
Vazão (m³/s)
2500
2000
1500
1000
500
0
Jan/1977
Jan/1978
Jan/1979
Jan/1980
Jan/1981
Jan/1982
Jan/1983
Jan/1984
Jan/1985
Jan/1986
Dec/1986
Meses


Emborcação
Previsto
Real
Vazão dessazonalizada
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
Jan/1977
Jan/1978
Jan/1979
Jan/1980
Jan/1981
Jan/1982
Jan/1983
Jan/1984
Jan/1985
Jan/1986
Dec/1986
Meses

(b)





(c)

Figura 5.18 – Gráficos das melhores previsões para P=1, real e dessazonalizado – (a) Furnas, (b) Emborcação, (c) Sobradinho

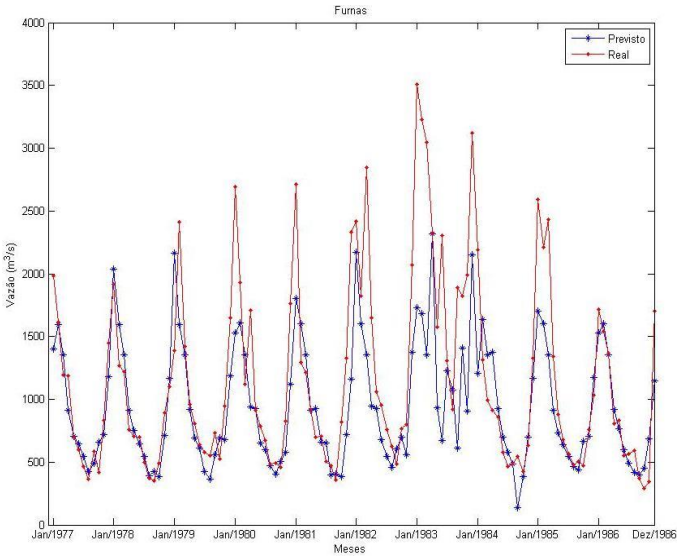
As previsões com 3 passos à frente para o período 1977-1986 estão sumarizadas na Tabela 5.17.

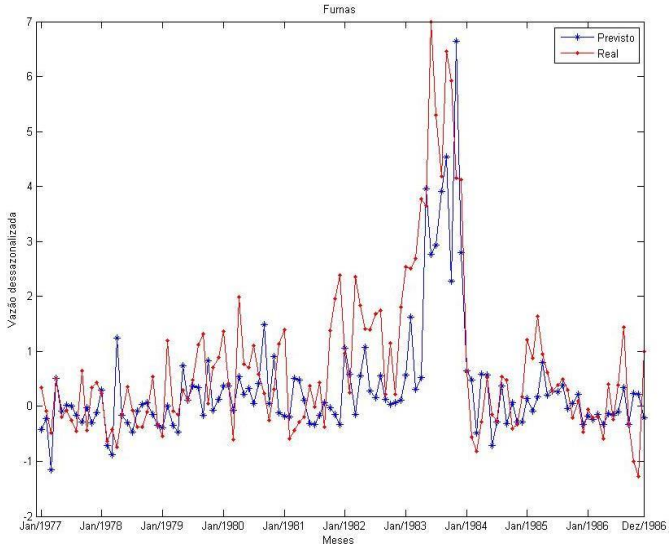
Tabela 5.17 - Resultados de Previsão para 3 passos à frente (*P* = 3)

	Preditor	RE	Desvio Padrão	MSE	MAE	MSE dessaz.	MAE dessaz.
FURNAS	**PAR**	0.7421	-	2.0704e+05	293.6726	1.2878	0.7982
	MLP	0.6891	8.4392e+04	**1.7855e+05**	**272.7713**	**1.0939**	**0.7500**
	ELM	0.7560	1.9849e+04	2.1491e+05	295.4792	1.3344	0.7856
	JAE-ESN	0.7550	1.0060e+04	2.1432e+05	302.4744	1.7822	0.8631
	OZT-ESN	0.7411	1.3404e+04	2.0651e+05	299.8686	1.7607	0.8634
	JAE-PV-ESN	0.7661	5.3111e+03	2.2069e+05	307.0446	2.0801	0.9100
	OZT-PV-ESN	0.7572	7.8458e+03	2.1556e+05	304.0830	1.8099	0.8819
	JAE-ESN-ELM	0.7959	2.4629e+04	2.3818e+05	318.0592	1.8042	0.8971
	OZT-ESN-ELM	0.8180	1.9367e+04	2.5162e+05	328.9985	1.9183	0.9327
EMBORCAÇÃO	**PAR**	0.8431	-	9.3559e+04	182.0518	1.1091	0.7847
	MLP	0.8369	2.0329e+04	9.2200e+04	170.8596	1.0084	0.7288
	ELM	0.7984	6.8359e+03	8.3898e+04	171.7859	0.9902	0.7471
	JAE-ESN	0.7487	2.5338e+03	**7.3775e+04**	**162.1759**	**0.9437**	**0.7168**
	OZT-ESN	0.7549	5.9736e+03	7.5013e+04	169.2796	0.9773	0.7434
	JAE-PV-ESN	0.7585	2.2522e+03	7.5733e+04	169.5040	0.9969	0.7474
	OZT-PV-ESN	0.7861	4.6268e+03	8.1344e+04	173.5630	1.0534	0.7647
	JAE-ESN-ELM	0.7913	7.3824e+03	8.2417e+04	173.4386	1.0778	0.7776
	OZT-ESN-ELM	0.8102	1.1628e+04	8.6411e+04	174.0789	1.1209	0.7853
SOBRADINHO	**PAR**	0.8919	-	3.0525e+06	1.0030e+03	1.1850	0.8394
	MLP	0.7287	4.8740e+05	**2.0376e+06**	**911.6199**	**0.9271**	**0.7626**
	ELM	0.8414	2.9517e+05	2.7170e+06	931.5936	0.9956	0.7515
	JAE-ESN	0.7802	1.5469e+05	2.3363e+06	903.8853	0.9573	0.7690
	OZT-ESN	0.8018	2.2661e+05	2.4674e+06	909.4861	0.9777	0.7699
	JAE-PV-ESN	0.8579	5.6047e+04	2.8247e+06	947.0467	1.1057	0.7934
	OZT-PV-ESN	0.8463	6.7306e+04	2.7486e+06	944.9069	1.0308	0.7681
	JAE-ESN-ELM	0.8350	3.5520e+05	2.6760e+06	933.1208	1.0149	0.7685
	OZT-ESN-ELM	0.8523	2.9001e+05	2.7883e+06	956.4088	1.0912	0.7984

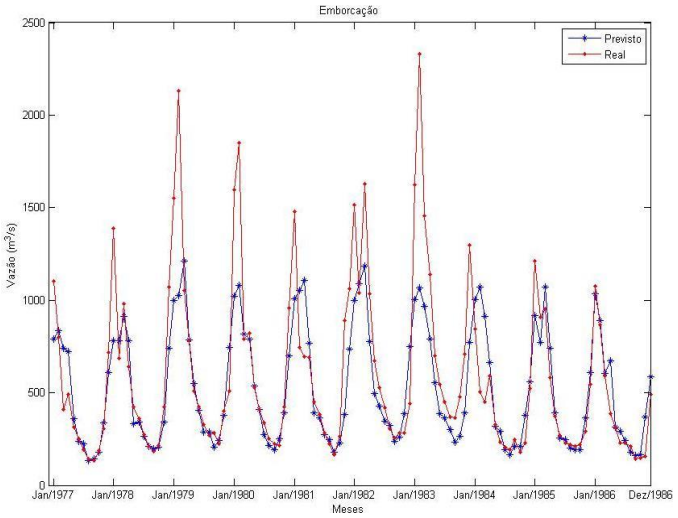
O desempenho computacional no período em questão indicou como método de melhores resultados para Furnas e Sobradinho, em relação ao MSE real e ao MAE, a rede MLP. No tocante à última usina, JAE-ESN foi a estrutura de menor MAE. Esta arquitetura, aliás, foi a que apresentou os menores erros para a série do posto de Emborcação. Neste mesmo caso, observa-se que as redes com ELM como camada de saída apresentaram melhores resultados que o modelo PAR e a MLP.

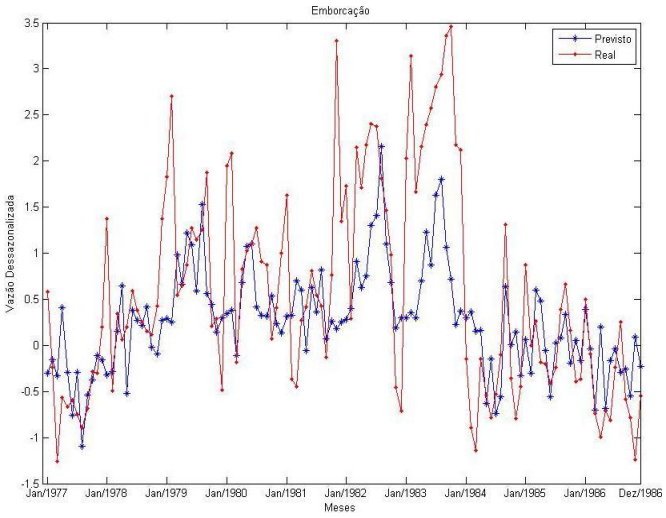
A Figura 5.19 mostra uma execução dos melhores preditores para cada usina.


Furnas
Previsto
Real
Vazão (m³/s)
Meses
Jan/1977
Jan/1978
Jan/1979
Jan/1980
Jan/1981
Jan/1982
Jan/1983
Jan/1984
Jan/1985
Jan/1986
Dez/1986

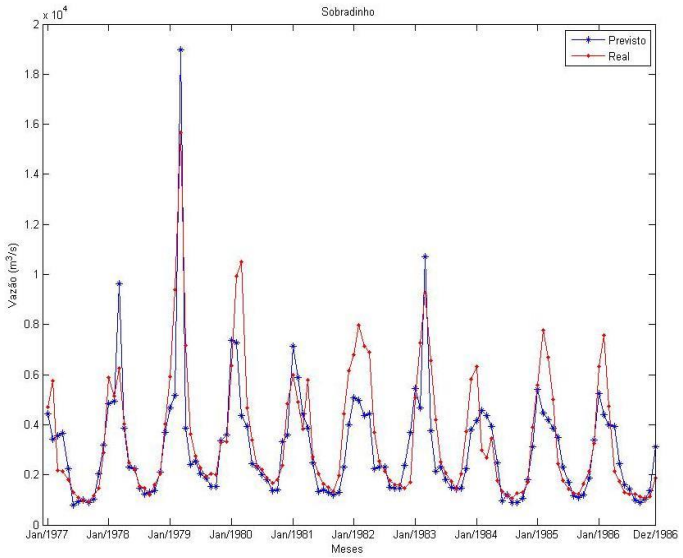

Furnas
Previsto
Real
Vazão dessazonalizada
Meses
Jan/1977
Jan/1978
Jan/1979
Jan/1980
Jan/1981
Jan/1982
Jan/1983
Jan/1984
Jan/1985
Jan/1986
Dez/1986

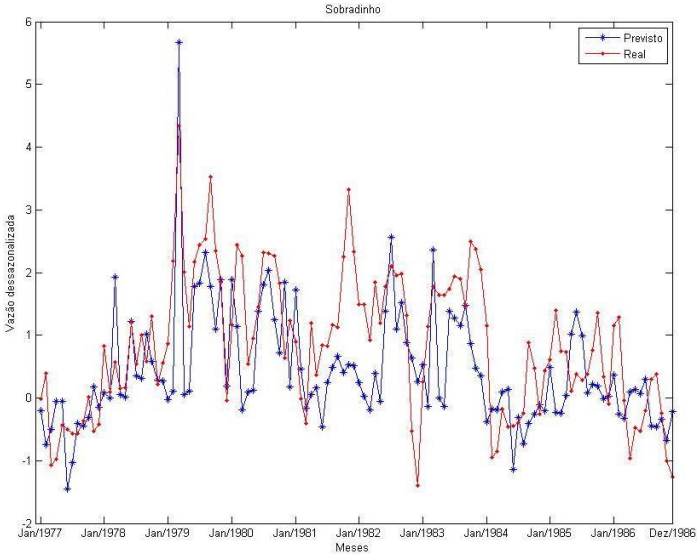
(a)


Emborcação
Previsto
Real
Vazão (m³/s)
2500
2000
1500
1000
500
0
Jan/1977
Jan/1978
Jan/1979
Jan/1980
Jan/1981
Jan/1982
Jan/1983
Jan/1984
Jan/1985
Jan/1986
Dez/1986
Meses


Emborcação
Previsto
Real
Vazão Dessazonalizada
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
Jan/1977
Jan/1978
Jan/1979
Jan/1980
Jan/1981
Jan/1982
Jan/1983
Jan/1984
Jan/1985
Jan/1986
Dez/1986
Meses

(b)





(c)

Figura 5.19 – Resultados melhores previsões P=3, real e dessazonalizado – (a) Furnas, (b) Emborcação, (c) Sobradinho

Na Tabela 5.18, estão os resultados computacionais da previsão 6 passos adiante.

Tabela 5.18- Resultados de Previsão para 6 passos à frente (*P* = 6)

	Preditor	RE	Desvio Padrão	MSE	MAE	MSE dessaz.	MAE dessaz.
FURNAS	**PAR**	0.8815	-	2.9213e+05	346.8537	2.5449	1.0130
	MLP	0.7829	6.6239e+04	**2.3046e+05**	**318.5804**	**1.6019**	**0.8858**
	ELM	0.8005	2.1302e+04	2.4091e+05	322.4530	1.9059	0.9175
	JAE-ESN	0.7960	8.9828e+03	2.3823e+05	318.1795	2.0880	0.9286
	OZT-ESN	0.7844	1.7770e+04	2.3131e+05	319.9102	2.0607	0.9586
	JAE-PV-ESN	0.7839	1.1429e+04	2.3105e+05	318.1763	2.0707	0.9432
	OZT-PV-ESN	0.7951	8.5626e+03	2.3769e+05	323.0817	2.0112	0.9442
	JAE-ESN-ELM	0.8465	2.9011e+04	2.6946e+05	350.3513	2.2068	1.0177
	OZT-ESN-ELM	0.8646	3.1277e+04	2.8109e+05	357.2896	2.2896	1.0507
EMBORCAÇÃO	**PAR**	0.8389	-	9.2636e+04	187.1562	1.4355	0.9013
	MLP	0.8366	1.9660e+04	9.2116e+04	188.9080	1.1927	0.8586
	ELM	0.7788	1.6540e+08	7.9836e+04	178.3216	1.1607	0.8333
	JAE-ESN	0.7550	3.9818e+03	7.5040e+04	171.1892	1.0869	0.8046
	OZT-ESN	0.7734	5.3364e+03	7.8740e+04	174.8383	1.1382	0.8222
	JAE-PV-ESN	0.7461	3.7079e+03	**7.3275e+04**	**173.9452**	**1.1609**	**0.8274**
	OZT-PV-ESN	0.7817	8.4644e+03	8.0430e+04	179.0620	1.2166	0.8454
	JAE-ESN-ELM	0.8033	9.6942e+03	8.4948e+04	182.5309	1.2570	0.8672
	OZT-ESN-ELM	0.8339	1.2774e+04	9.1525e+04	189.4135	1.3917	0.9083
SOBRADINHO	**PAR**	0.8800	-	2.9720e+06	1.0731e+03	1.5149	0.9755
	MLP	0.8655	2.7699e+07	2.8746e+06	1.0179e+03	1.2171	0.8861
	ELM	0.8397	1.5993e+05	2.7057e+06	991.6764	1.1871	0.8746
	JAE-ESN	0.8076	1.1712e+05	**2.5033e+06**	**967.0486**	**1.1938**	**0.8803**
	OZT-ESN	0.8166	1.1607e+05	2.5590e+06	988.5868	1.2349	0.8985
	JAE-PV-ESN	0.8186	6.5796e+04	2.5718e+06	991.9813	1.2360	0.8907
	OZT-PV-ESN	0.8320	1.4931e+05	2.6567e+06	994.1694	1.2625	0.8965
	JAE-ESN-ELM	0.8610	3.9172e+05	2.8454e+06	1.0065e+03	1.2991	0.8945
	OZT-ESN-ELM	0.8842	3.9417e+05	3.0007e+06	1.0544e+03	1.3705	0.9365

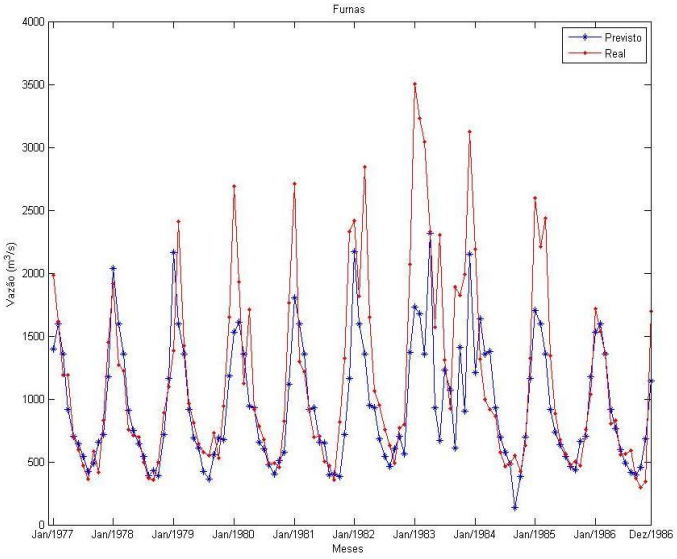
A análise da usina de Furnas mostra que os erros dessazonalizados, para a maioria dos preditores, ultrapassou o valor 2, ressaltando a dificuldade de acerto dos modelos aos dados observados. Uma causa possível para isso é que, como os dados de teste pertencem a um período de cheias, o comportamento estatístico em relação à média das demais amostras da série pode destoar deles. Assim, a rede tem a disposição dados com relativa diferença para ajuste do modelo daqueles que ela efetivamente precisa prever, o que não é desejável e torna a tarefa ainda mais desafiadora para o modelo.

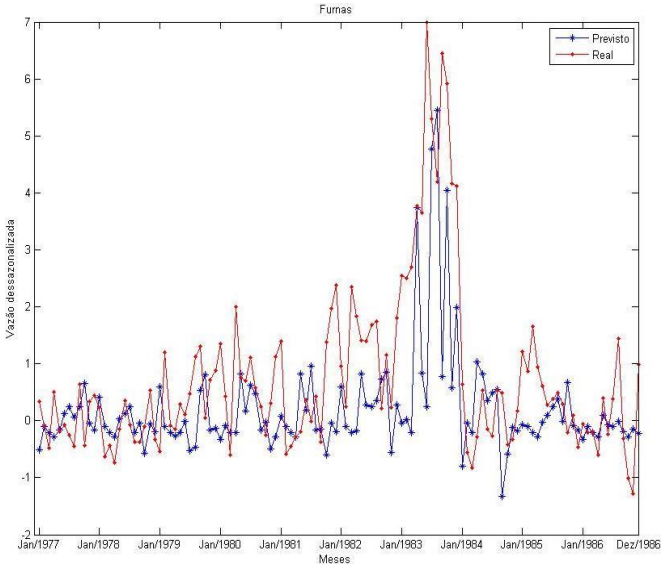
Neste caso, a MLP alcançou um resultado sistematicamente melhor que das os demais redes neste domínio, sendo seguida mais de perto pela rede ELM. Aliás, a MLP foi a que apresentou melhor valor de MSE real.

A usina de Emborcação teve como melhor desempenho aquele alcançado pelo modelo JAE-PV-ESN, enquanto as demais métricas foram favoráveis à JAE-ESN. Esta última máquina desorganizada foi aquela que mostrou-se melhor para a usina de Sobradinho, exceto do ponto de vista da métrica de MAE, que foi menor para a ELM.

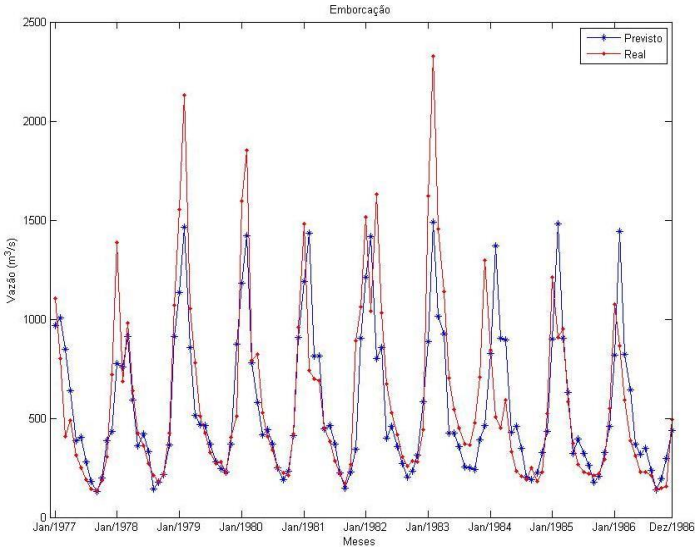
Nota-se, mais uma vez, que JAE-ESN-ELM e OZT-ESN-ELM não foram inferiores ao PAR para as séries de Furna e Emborcação. Além disso, se, por um lado, a MLP foi superior em Furnas, em Emborcação ela só superou o modelo PAR e em Sobradinho o PAR e OZT-ESN-ELM.

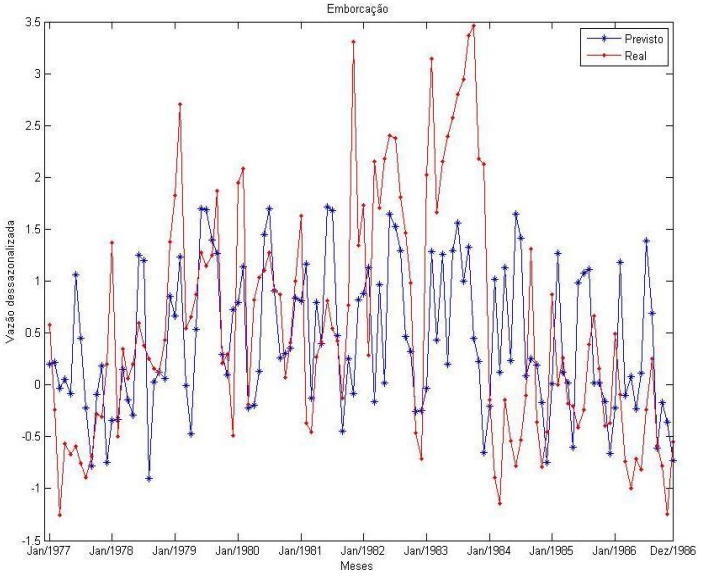
Seguem na Figura 5.20 previsões dos melhores modelos.



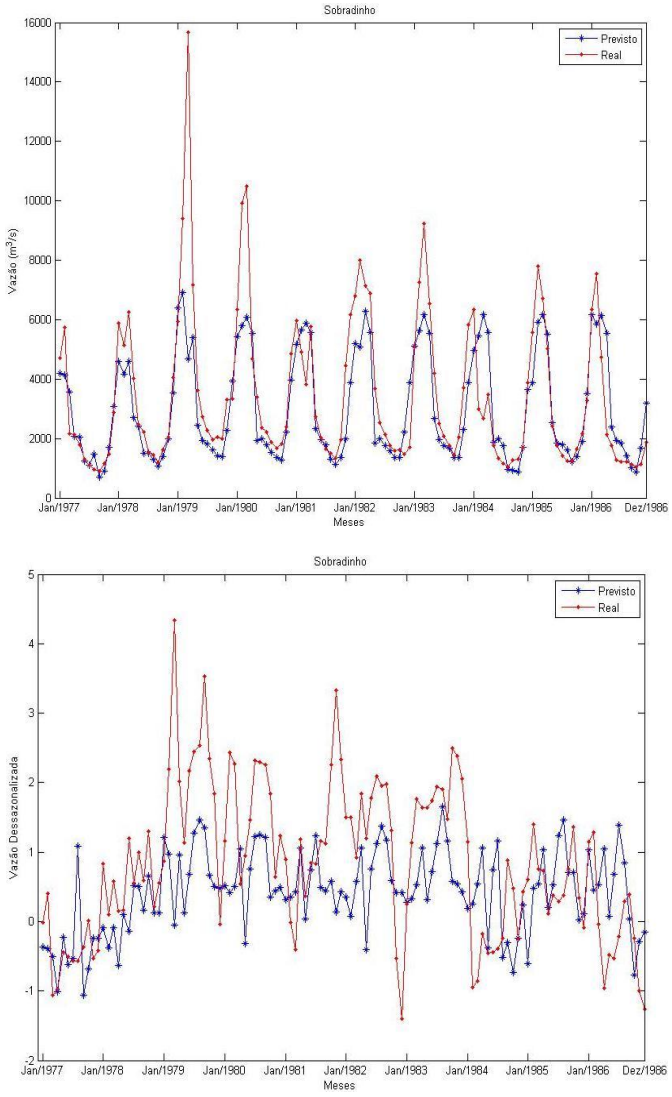


(a)


Emborcação
Previsto
Real
Vazão (m³/s)
2500
2000
1500
1000
500
0
Jan/1977
Jan/1978
Jan/1979
Jan/1980
Jan/1981
Jan/1982
Jan/1983
Jan/1984
Jan/1985
Jan/1986
Dez/1986
Meses


Emborcação
Previsto
Real
Vazão dessazonalizada
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
Jan/1977
Jan/1978
Jan/1979
Jan/1980
Jan/1981
Jan/1982
Jan/1983
Jan/1984
Jan/1985
Jan/1986
Dez/1986
Meses

(b)



(c)

Figura 5.20 – Resultados melhores previsões P=6, real e dessazonalizado – (a) Furnas, (b) Emborcação, (c) Sobradinho

Por fim, chegamos aos resultados para 12 passos à frente, sumarizados na Tabela 5.19.

O modelo de máquina desorganizada tipo JAE-PV-ESN obteve os menores erros em todos os casos, a menos do MAE e MSE dessazonalizado da série de Emborcação, que foram favoráveis à JAE-ESN.

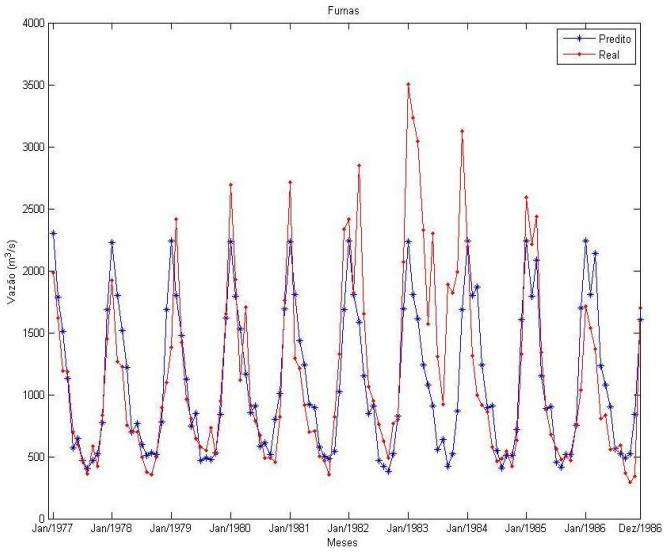
Vê-se, em Furnas, que os resultados do modelo PAR, redes *feedforward* e ESN's com ELM foram muito semelhantes em relação ao MSE real. Também é notório que os

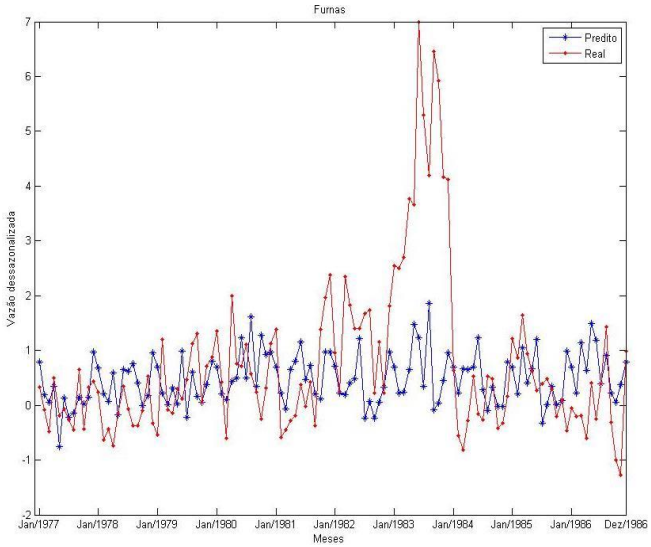
desvios padrões da MLP foram os mais elevados e, no caso de Sobradinho, deteriorados novamente por uma execução muito ruim.

Tabela 5.19 - Resultados de Previsão para 12 passos à frente (*P* = 12)

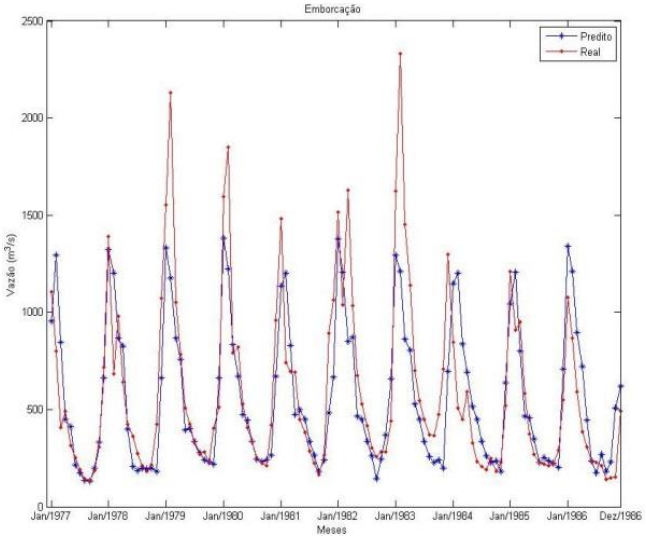
	Preditor	RE	Desvio Padrão	MSE	MAE	MSE dessaz.	MAE dessaz.
FURNAS	**PAR**	0.8919	-	2.9911e+05	356.2449	2.6462	1.0586
	MLP	0.8898	1.4057e+05	2.9769e+05	347.8291	2.4876	1.0105
	ELM	0.8897	5.5676e+04	2.9762e+05	352.2514	2.5254	1.0072
	JAE-ESN	0.8280	1.1606e+04	2.5777e+05	331.6919	2.2570	0.9771
	OZT-ESN	0.8351	1.7428e+04	2.6217e+05	327.2231	2.2825	0.9576
	JAE-PV-ESN	0.7777	7.4016e+03	**2.2738e+05**	**319.5557**	**1.9656**	**0.9417**
	OZT-PV-ESN	0.8045	1.0927e+04	2.4334e+05	329.2908	2.0171	0.9606
	JAE-ESN-ELM	0.8930	2.2939e+04	2.9986e+05	364.8189	2.5484	1.0749
	OZT-ESN-ELM	0.9036	3.8023e+04	3.0704e+05	369.4982	2.6418	1.0904
EMBORCAÇÃO	**PAR**	0.8529	-	9.5762e+04	191.0493	1.5875	0.9447
	MLP	0.7948	2.5023e+04	8.3157e+04	174.3135	1.1627	0.8179
	ELM	0.8224	1.3110e+04	8.9032e+04	182.2417	1.2921	0.8607
	JAE-ESN	0.7542	7.1287e+03	7.4876e+04	170.7526	1.1420	0.8346
	OZT-ESN	0.7918	4.6192e+03	8.2525e+04	178.6005	1.2488	0.8585
	JAE-PV-ESN	0.7509	5.2469e+03	**7.4217e+04**	**179.2694**	**1.1445**	**0.8419**
	OZT-PV-ESN	0.7929	6.8217e+03	8.2757e+04	186.6196	1.2097	0.8699
	JAE-ESN-ELM	0.8596	1.7278e+04	9.7261e+04	190.8521	1.4142	0.9142
	OZT-ESN-ELM	0.8649	2.1966e+04	9.8463e+04	194.1209	1.4901	0.9460
SOBRADINHO	**PAR**	0.9139	-	3.2053e+06	1.1097e+03	1.6588	1.0167
	MLP	0.8835	3.0163e+07	2.9959e+06	1.0401e+03	1.3916	0.9410
	ELM	0.8514	7.6340e+05	2.7821e+06	1.0004e+03	1.3156	0.9058
	JAE-ESN	0.8601	2.1200e+05	2.8389e+06	1.0543e+03	1.4023	0.9595
	OZT-ESN	0.8837	3.3752e+05	2.9969e+06	1.0961e+03	1.5082	1.0080
	JAE-PV-ESN	0.8102	1.1038e+05	**2.5192e+06**	**985.9917**	**1.2199**	**0.8853**
	OZT-PV-ESN	0.8395	8.8106e+04	2.7047e+06	1.0115e+03	1.3137	0.9047
	JAE-ESN-ELM	0.8917	4.8241e+05	3.0521e+06	1.0534e+03	1.5247	0.9643
	OZT-ESN-ELM	0.9329	3.0285e+05	3.3400e+06	1.1116e+03	1.6440	1.0143

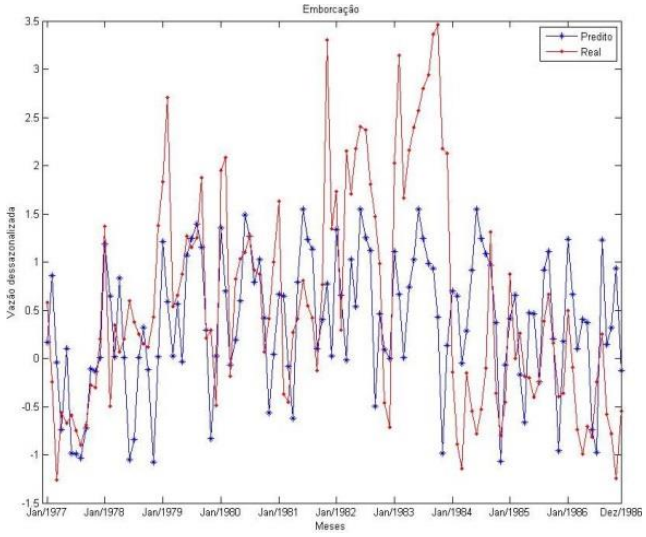
A Figura 5.21 apresenta uma execução para cada usina do modelo JAE-PV-ESN.


Furnas
Predito
Real
Vazão (m³/s)
Meses
Jan/1977
Jan/1978
Jan/1979
Jan/1980
Jan/1981
Jan/1982
Jan/1983
Jan/1984
Jan/1985
Jan/1986
Dez/1986

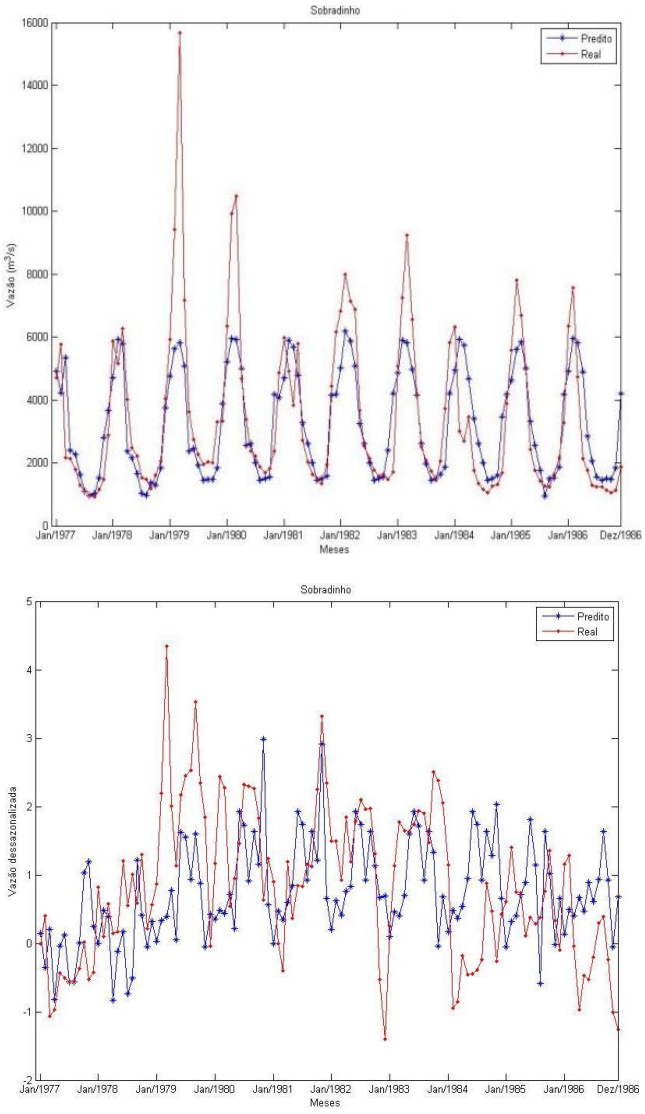

Furnas
Predito
Real
Vazão dessazonalizada
Meses
Jan/1977
Jan/1978
Jan/1979
Jan/1980
Jan/1981
Jan/1982
Jan/1983
Jan/1984
Jan/1985
Jan/1986
Dez/1986

(a)


Emborcação
Predito
Real
Vazão (m³/s)
2500
2000
1500
1000
500
0
Jan/1977
Jan/1978
Jan/1979
Jan/1980
Jan/1981
Jan/1982
Jan/1983
Jan/1984
Jan/1985
Jan/1986
Dez/1986
Meses


Emborcação
Predito
Real
Vazão dessazonalizada
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
Jan/1977
Jan/1978
Jan/1979
Jan/1980
Jan/1981
Jan/1982
Jan/1983
Jan/1984
Jan/1985
Jan/1986
Dez/1986
Meses

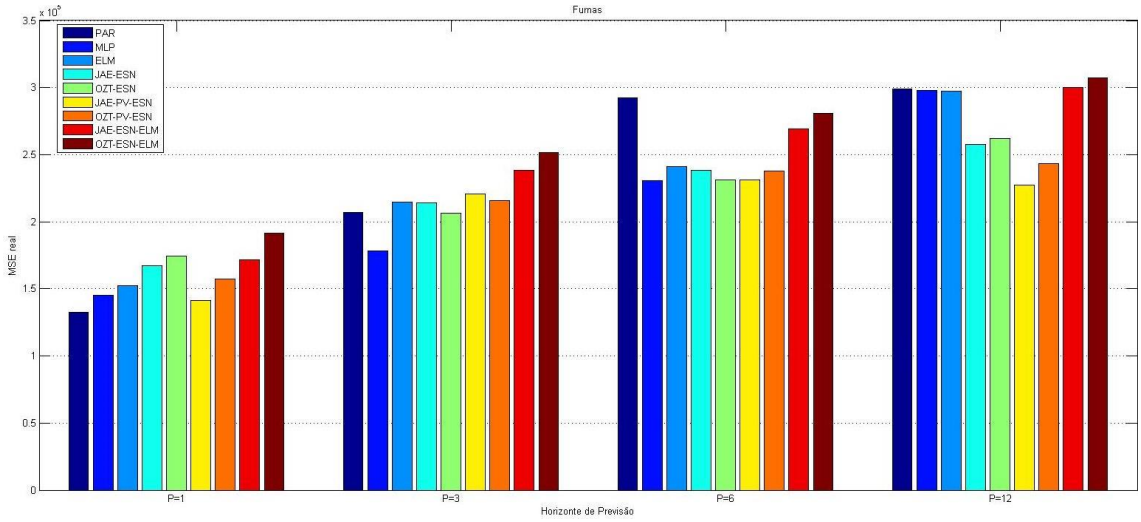
(b)



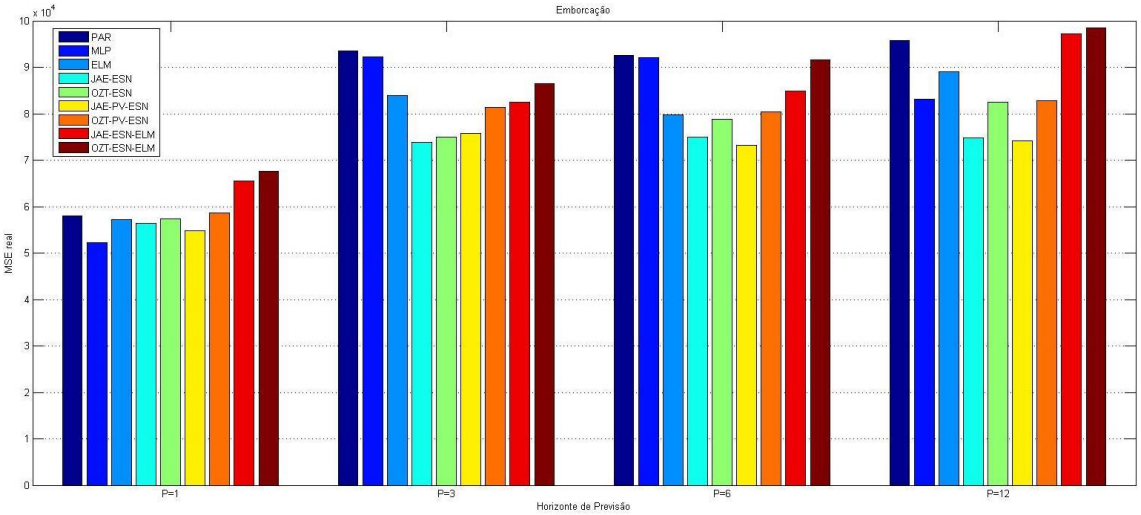
(c)

Figura 5.21 – Resultados melhores previsões P=12, real e dessazonalizado – (a) Furnas, (b) Emborcação, (c) Sobradinho

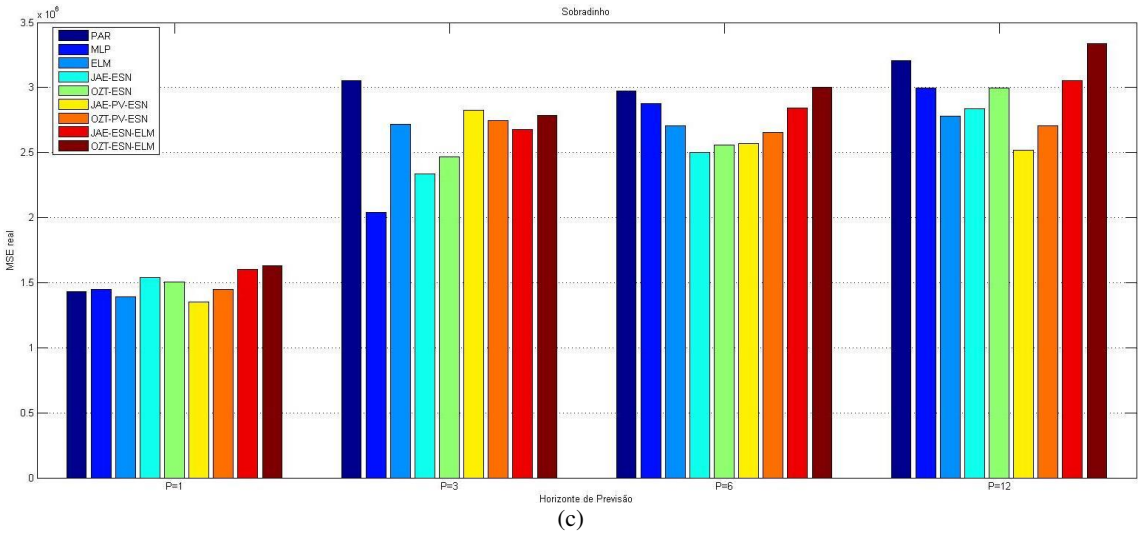
Os gráficos em barras da Figura 5.22 mostram, de forma resumida, os erros alcançados pelos modelos nas previsões das 3 usinas abordadas.



(a)



(b)



(c)

Figura 5.22 - Resultados 1977-1986 – (a) Furnas, (b) Emborcação, (c) Sobradinho

No período 77-86, como já comentado no início, houve uma distribuição diferente entre aqueles que foram os modelos de menor erro final de previsão. Não é surpreendente que isto tenha ocorrido, uma vez que este período possui médias muito elevadas, provocadas pelo maior período de cheias do histórico. Se, em 51-60 e 67-76, as ESNs com camada de saída linear apareciam dentre aquelas com menores erros, aqui em mais de uma vez, o desempenho da MLP, dos modelos com filtro de Volterra e até das redes com ELM com camada de saída apresentaram desempenhos superiores. De fato, em 4 casos, a MLP teve menor MSE real e, em 5 casos, a JAE-PV-ESN foi a ganhadora.

Além disso, a variação do horizonte de previsão para um mesmo período de testes alterou frequentemente os modelos de melhor performance. O modelo PAR para a usina de Furnas é um bom exemplo: se, para P=1, ele alcançou o menor MSE real, para P=6, seu MSE foi o mais elevado.

5.7 Sobre os resultados sumarizados

O comportamento geral dos resultados quanto ao desvio padrão da série completa no conjunto de teste foi coerente com o esperado pois, excluindo-se os casos em que ocorreram *outliers,* todos estiveram dentro da margem de uma ordem de grandeza do valor do MSE. Além disso, as ESNs com PCA e filtro de Volterra estão quase sempre entre os modelos de menor desvio padrão, enquanto as ESN's com ELM como camada de saída e a MLP se revezaram como as redes de maior desvio. A Figura 5.23 é ilustrativa deste fato: ela apresenta o gráfico tipo *box-plot* das 50 execuções das redes neurais para série de Sobradinho 67-76

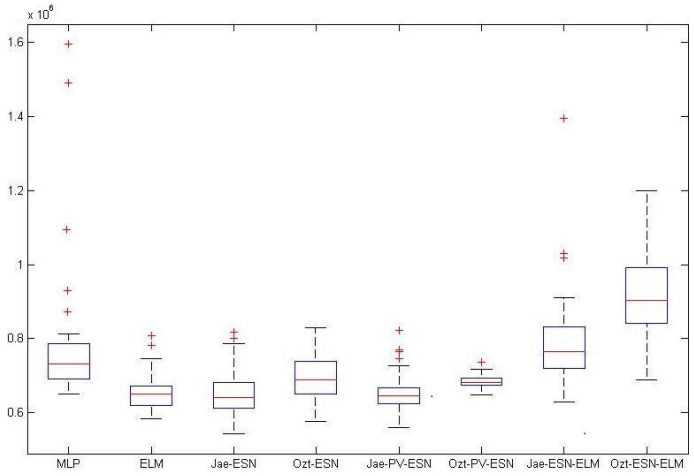


Figura 5.23 – Box-plot – Sobradinho 67-76 - 50 execuções

Esta discussão torna-se importante pelo fato de que o modelo oficial do setor elétrico, o PAR, é determinístico. Dessa forma, caso o desvio fosse elevado para as arquiteturas de máquinas desorganizadas, isto poderia ser um fator desencorajador da sua aplicação para previsão de vazões. Além disso, foi observado, em todos os testes, é que apenas em um único caso o modelo PAR apresentou o melhor resultado em termos do MSE real, o que da suporte ao uso de técnicas não-lineares neste tipo de tarefa.

Sobre as camadas de saída não-lineares, esperava-se que sua introdução fosse reduzir o erro de teste, como evidenciado em alguns estudos (Ballini 2000, Sacchi 2009), visto que estas redes tendem a possuir maior capacidade de aproximação. Apesar disso, o comportamento desejado é o de melhor generalização, o que, mesmo com os processos de regularização, não foi alcançado na medida necessária para as ESNs com ELM. Por outro lado, as redes de estado de eco com o filtro de Volterra, em alguns casos, conseguiram ser superiores, como discutido.

A diferença de abordagem com respeito a trabalhos como Siqueira et al. (2012a) e Siqueira et al. (2012b) é que, nestes, utilizou-se todo o histórico e apenas uma RNA para previsão da série completa. Aqui há uma redução no número de amostras para apenas 60 de treinamento e 10 de teste.

Não foi sempre observada uma relação direta entre o menor MSE de treinamento e de teste. Tal comportamento não chega surpreendente, já que o que se busca durante o treinamento de uma RNA, é a melhor capacidade de generalização. Na Tabela 5.20, são

apresentadas estas medidas para os casos de Furnas 77-86 e 51-60, o que evidencia este fato. Reiteramos que as precauções necessárias para não haver sobretreinamento (*overfitting*) dos modelos não-lineares foram tomadas, como a utilização de validação cruzada para a MLP e regularização no caso das ELMs.

Tabela 5.20 – Relação entre MSE de teste e de treinamento

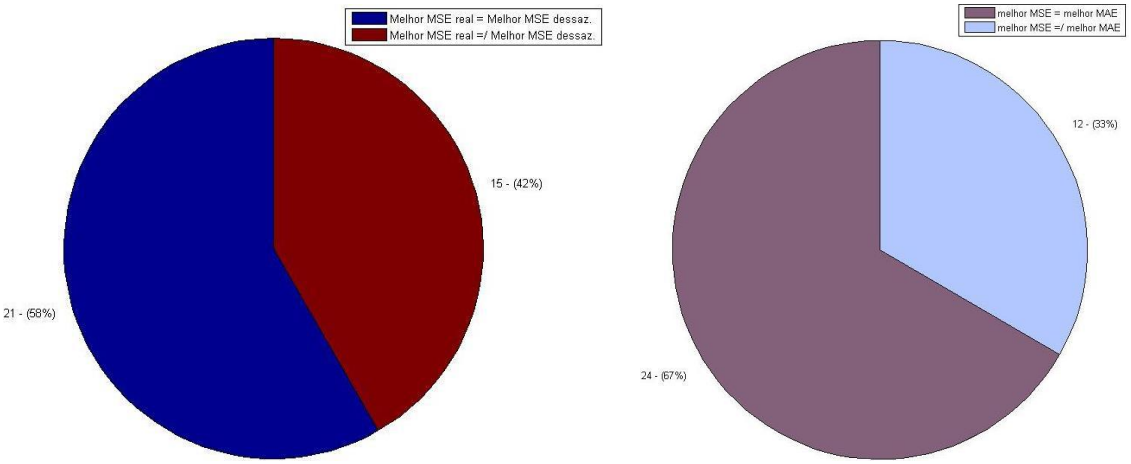
	Preditor	MSE teste	MSE treinamento
FURNAS 77-86	**PAR**	**1.3250e+05**	13.2682e+04
	MLP	1.4499e+05	7.4670e+04
	ELM	1.5217e+05	9.5567e+04
	JAE-ESN	1.6748e+05	7.0514e+04
	OZT-ESN	1.7418e+05	7.1093e+04
	JAE-PV-ESN	1.4115e+05	9.7015e+04
	OZT-PV-ESN	1.5748e+05	7.1093e+04
	JAE-ESN-ELM	1.7148e+05	8.2860e+04
	OZT-ESN-ELM	1.9164e+05	8.8152e+04
FURNAS 51-60	**PAR**	6.4077e+04	9.9153e+04
	MLP	5.8366e+04	6.9532e+04
	ELM	5.7191e+04	8.5868e+04
	JAE-ESN	**5.3696e+04**	10.7591e+04
	OZT-ESN	6.0473e+04	11.3802e+04
	JAE-PV-ESN	5.7968e+04	10.6674e+04
	OZT-PV-ESN	6.4638e+04	10.9533e+04
	JAE-ESN-ELM	6.7086e+04	11.0721e+04
	OZT-ESN-ELM	7.8813e+04	11.4563e+04

Outro comportamento frequente é relativo à discrepância entre as medidas de erro adotadas. Os 36 cenários mostraram que, em 12 casos, o melhor preditor em termos do MAE e o melhor em termos do MSE não corresponderam à mesma estrutura. De forma similar, em 15 ocasiões, o menor MSE real e o menor MSE dessazonalizado foram favoráveis a modelos distintos. Este comportamento é frequente em estudos de vazões e está relacionado diretamente com a dinâmica de dessazonalização por meio da técnica de padronização.

Dessa forma, faz-se necessário o desenvolvimento de alternativas para a retirada da componente sazonal, de forma que as melhores respostas dos preditores no espaço em que eles efetivamente "enxergam" sejam efetivamente aproveitadas.

Uma possibilidade seria penalizar os meses da série que acarretam menores erros e consequentemente valorizar os pontos extremos, pois são estes que elevam o valor do MSE final. É possível que o caminho para isto esteja em ponderar a função custo pelo desvio padrão mensal, fazendo com que os modelos acertem mais os meses de cheias. A Figura

5.24 mostra o gráfico tipo “pizza” do número de casos em que estas proporcionalidades são alcançadas.



(a) (b)

Figura 5.24 – Proporcionalidade entre (a) MSE real e dessazonalizado e (b) entre MSE e MAE dos melhores resultados

5.8 Discussão sobre os preditores que apresentaram melhores resultados

Nesta seção, serão discutidos os resultados computacionais dos preditores que alcançaram os menores valores de erro quadrático médio no domínio real, os quais são considerados os mais bem sucedidos.

A Tabela 5.21 sumariza os modelos que alcançaram menor MSE, separados por horizonte de previsão e período de testes, enquanto a Figura 5.25 resume os desempenhos por modelo preditor e número de passos à frente.

Tabela 5.21 - Melhores preditores por horizonte e período de testes

	1951-1960	1967-1976	1977-1986	Total
P=1	JAE-ESN = 2, ELM =1	JAE-ESN = 1, JAE-PV-ESN= 2	PAR = 1, MLP = 1, JAE-PV-ESN= 1	**PAR = 1, ELM = 1, MLP = 1, JAE-ESN = 3, JAE-PV-ESN = 3**
P=3	OZT-ESN = 1, JAE-ESN = 2	JAE-ESN = 2, ELM = 1	MLP = 2, JAE-ESN = 1	**MLP = 2, ELM = 1, OZT-ESN = 1, JAE-ESN = 5**
P=6	ELM = 3	OZT-ESN= 2, JAE-ESN = 1	MLP = 1, JAE-ESN = 1, JAE-PV-ESN = 1	**ELM = 3, MLP = 1, OZT-ESN = 2, JAE-ESN = 2, JAE-PV-ESN = 1**
P=12	JAE-ESN = 2, ELM = 1	JAE-ESN = 1, OZT-ESN= 2	JAE-PV-ESN= 3	**ELM = 1, OZT-ESN = 2, JAE-ESN = 3, JAE-PV-ESN = 3**

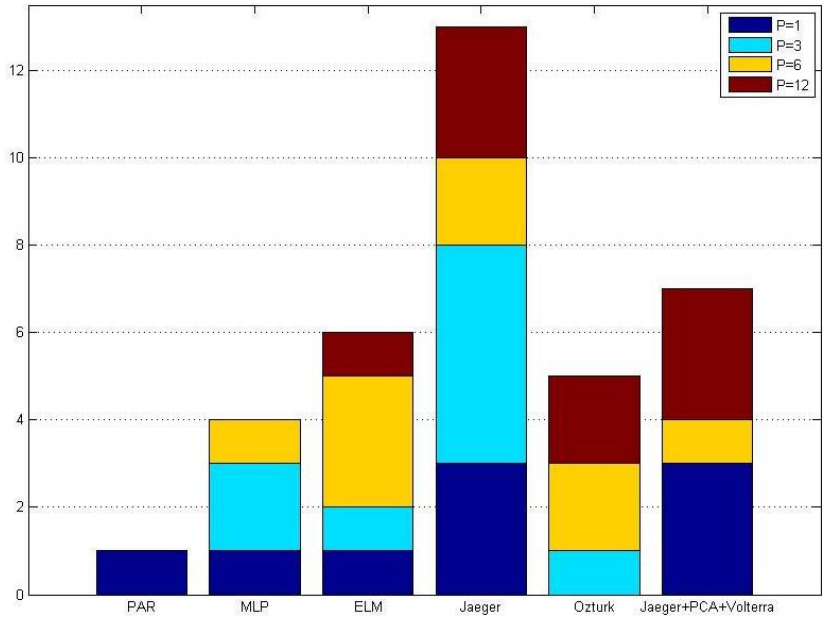


Figura 5.25- Melhores preditores por horizonte de previsão

O número de vezes que cada preditor alcançou o melhor resultado nas 36 ocasiões estudadas foi: PAR=1, MLP=4, ELM=6, JAE-ESN=13, OZT-ESN=5, JAE-PV-ESN=7.

O modelo linear PAR é o modelo oficial do Setor Elétrico Brasileiro (SEB) (Souza, Marcato, et al. 2010). O ONS já utiliza em alguns outros tipos de séries de vazões as redes MLP (ONS 2009). Portanto, a análise comparativa principal das máquinas desorganizadas leva em conta seu desempenho frente às estas propostas que já são aceitas pelo SEB.

Diante disso, algumas das propostas aqui apresentadas se mostraram alternativas consistentes na solução deste problema, como evidenciado pelos resultados computacionais. Embora não seja possível apontar apenas uma das arquiteturas de MDs, em geral, aquelas listadas na Figura 5.26 conseguiram superar o modelo PAR.

Os horizontes de previsão mais curtos mostraram resultados favoráveis às redes de Jaeger. Nas previsões com P=1, vê-se que há um predomínio das ESNs de Jaeger com e sem o filtro de Volterra, as quais alcançaram, em 3 ocasiões cada uma, o menor erro. Já para P=3, a JAE-ESN teve melhor desempenho em mais da metade (5) dos possíveis casos. Para este tipo de mapeamento, em que o erro agregado pela recursão de saídas para formar

as respostas tem impacto menor, a proposta com uma matriz esparsa e o componente de memória se adequou melhor ao problema.

Os cenários com horizontes de 6 e 12 passos à frente são mais desafiadores, pois a correlação entre as amostras diminui e é necessário inserir como entrada do modelo resultados de previsão que já possuem uma componente de erro. Para estes casos, o modelo linear não superou as redes neurais. Para P=6, o que se nota é que houve uma maior distribuição dos melhores valores entre todas as arquiteturas de máquinas desorganizadas, com leve vantagem para as ELMs. Em P=12, quase todos os melhores resultados foram alcançados com redes ESNs, evidenciando a importância da memória em horizontes ainda mais longos.

Na Figura 5.26, são apresentados os gráficos tipo "pizza" da distribuição dos melhores resultados, independemente da série ou horizonte de previsão.

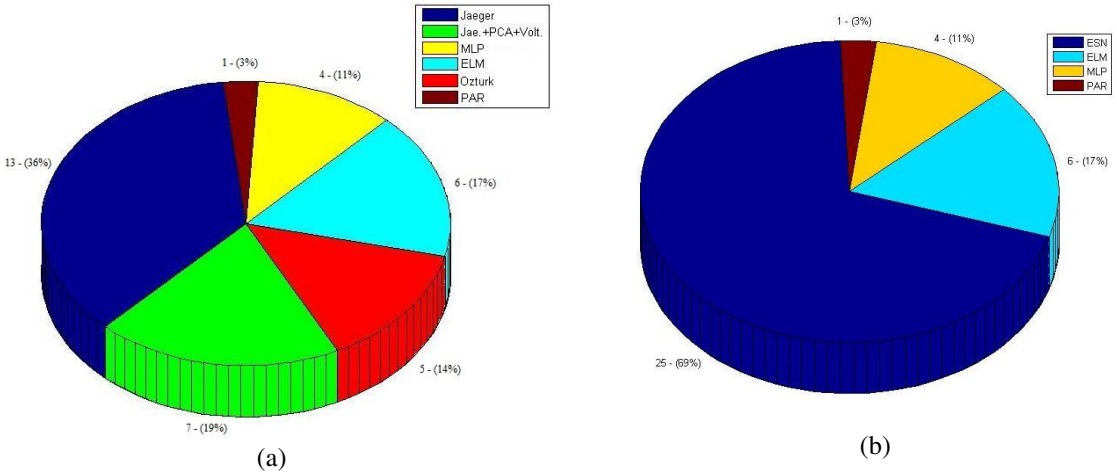


Figura 5.26 – Melhores resultados gerais (a) cada modelo, (b) destacando as ESNs

Como a Figura 5.26 indica, há um predomínio das MDs, sobretudo das ESNs, nos melhores resultados gerais. A razão disso pode estar ligada a alguns fatores, sendo que nos parece correta a afirmação de que a memória de uma rede recorrente, aliada ao potencial de processamento dinâmico, é altamente benéfica em trabalhos em que sejam necessárias informações de amostras passadas, as quais possuem correlação com o dado que se deseja prever. Além disso, apesar de as redes MLPs serem totalmente treinadas, as arquiteturas desorganizadas conseguiram alcançar menores erros, mesmo com um processo de treinamento simples e não necessariamente com a presença de uma camada de saída não-linear.

Uma das justificativas para utilização de modelos PAR pelo setor elétrico é sua velocidade de treinamento e a simplicidade de cálculo de seus parâmetros. Estes resultados, apesar de não mostrarem nenhuma arquitetura como absolutamente melhor dentre as estudadas, mostram que é possível manter esta simplicidade com a introdução de técnicas não-lineares, reconhecidamente mais capazes de aproximar funções e generalizá-las.

Uma outra forma de apresentarmos os melhores resultados da Tabela 5.19 é separando-os por período de testes, conforme mostra a Figura 5.27.

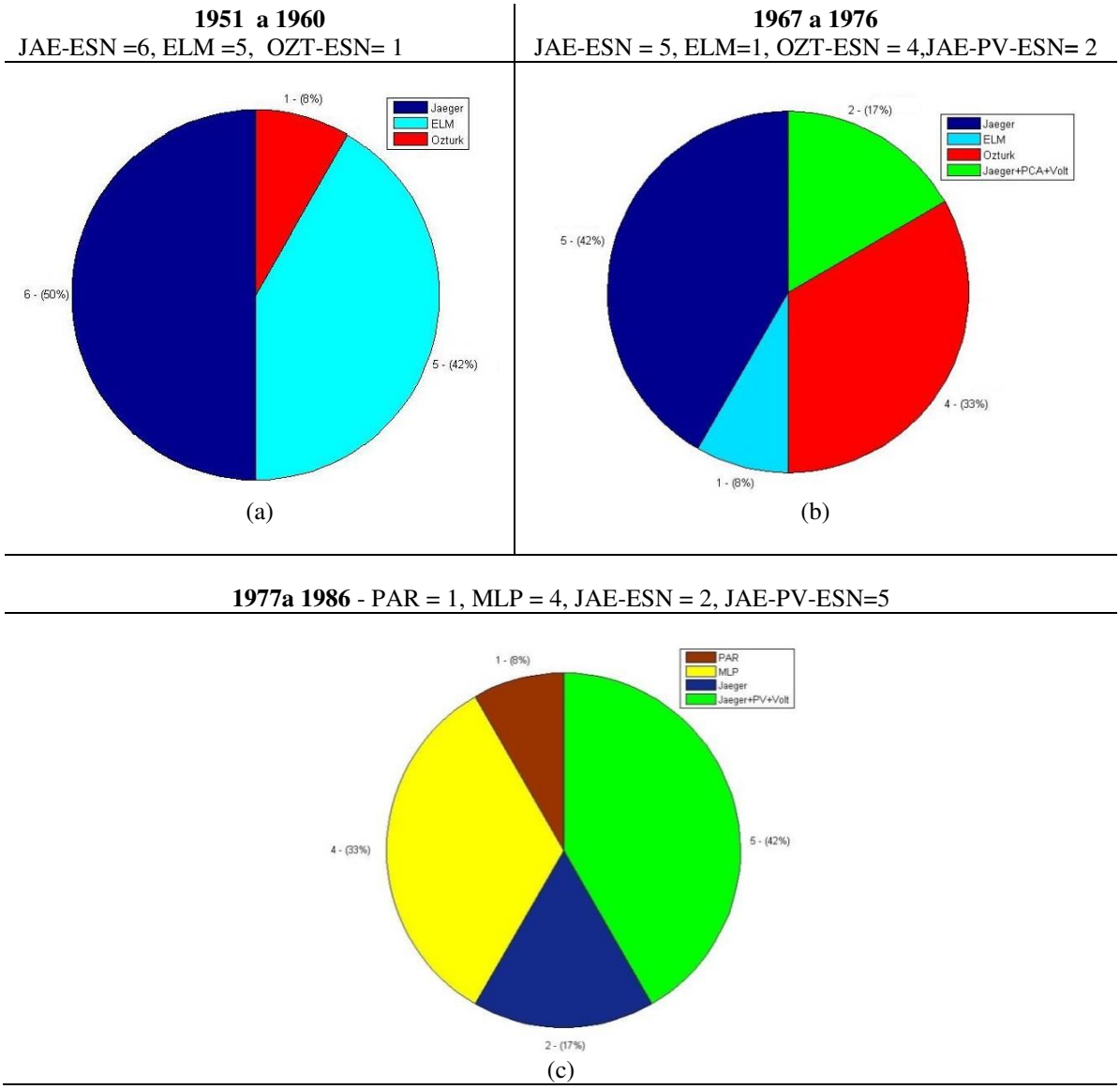


Figura 5.27 – Melhores resultados por período de testes

O período 51-60 é aquele que apresenta menores magnitudes nas vazões registradas, e, portanto, os menores erros de previsão finais. Neste caso, apenas máquinas desorganizadas em suas versões mais simples apresentaram melhores resultados, sendo a metade dos casos favorável à JAE-ESN e 5 outros à ELM.

Já em 67-76, período mediano, as ESNs de Jaeger e Ozturk et al. com combinador linear como camada de saída foram superiores. Todavia, a JAE-PV-ESN apresenta melhores resultados em duas ocasiões, enquanto a ELM em uma. Veja que, nestes dois períodos, houve predominância absoluta das máquinas desorganizadas.

O último conjunto de testes, de 1977 a 1986, como dito, é o de maior amplitude e, portanto, tende a apresentar maiores erros. Foi neste caso que a MLP e o modelo linear PAR obtiveram os melhores resultados pelo menos uma vez. Ainda assim, há predominância das ESNs com reservatório de Jaeger, seja o modelo com combinador linear ou a arquitetura com PCA e filtro de Volterra.

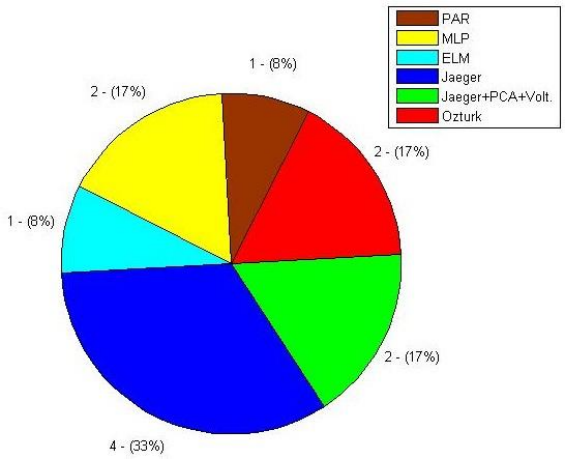
Ainda é possível a análise dos melhores resultados em relação a cada usina hidrelétrica considerada. Para isso, sumarizamos os resultados e os separamos na Figura 5.28.

Como é sabido, a usina de Emborcação é aquela que possui o menor volume de água afluente no seu histórico de vazões. Neste caso, fica clara a predominância da ESN de Jaeger frente aos demais modelos. Esta conclusão pode ser aplicada com menor intensidade à série de Furnas, em que, apesar de ganhar mais vezes, esta arquitetura não é predominante. A série de Sobradinho, de maior amplitude, é a que ressalta o poder de previsão das ELMs, que superaram as demais estruturas quatro vezes. Todavia, as ESNs ainda conseguem bons resultados finais.

Não é possível com estes resultados apontar apenas uma arquitetura de rede neural que seja a melhor solução geral para prever vazões. O que foi possível demonstrar é que a rede proposta por Butcher et al. (2010) para este tipo de tarefa de previsão (recorrente e com 6 entradas) não se mostrou adequada. Da mesma forma, é prudente afirmar que as máquinas desorganizadas, sobretudo as redes de estado de eco, alcançaram boas performances frente aos modelos PAR e MLP, os mais bem-estabelecidos pela literatura neste tipo de problema.

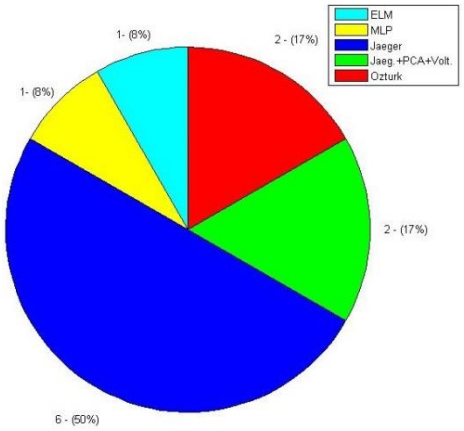
Assim sendo, caso o projetor tenha que escolher apenas uma rede, a sugestão apontada pelos resultados deste trabalho é utilizar a JAE-ESN, que obteve os melhores resultados de forma distribuída entre os períodos, usinas e número de passos à frente.

FURNAS
PAR =1, MLP=2, ELM=1, JAE-ESN=4, OZT-ESN= 2, JAE-PV-ESN=2



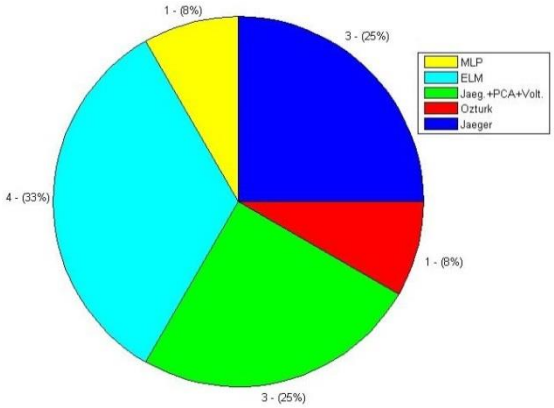
(a)

EMBORCAÇÃO
MLP=1, ELM=1, JAE-ESN= 6, OZT-ESN=2, JAE-PV-ESN=2



(b)

SOBRADINHO
MLP=1, ELM=4, JAE-ESN=3, OZT-ESN= 1, JAE-PV-ESN=3



(c)

Figura 5.28 – Melhores resultados por usina

5.9 Melhores resultados quanto à rede *feedforward*

Esta seção sumariza os resultados computacionais apenas das redes neurais *feedforward* – ELM e MLP. O objetivo é promover um estudo comparativo sobre qual das

arquiteturas foi mais eficiente na previsão das vazões mensais. Os melhores desempenhos separados por horizonte de previsão foram como segue:

P=1 – ELM=6, MLP=3;

P=3 – ELM=6, MLP=3;

P=6– ELM=8, MLP=1;

P=12 – ELM=7, MLP=2;

TOTAL = ELM=27, MLP=9,

sendo o número à frente de cada sigla a quantidade de vezes em que a rede foi superior à outra. Estes números também estão resumidos graficamente na Figura 5.29:

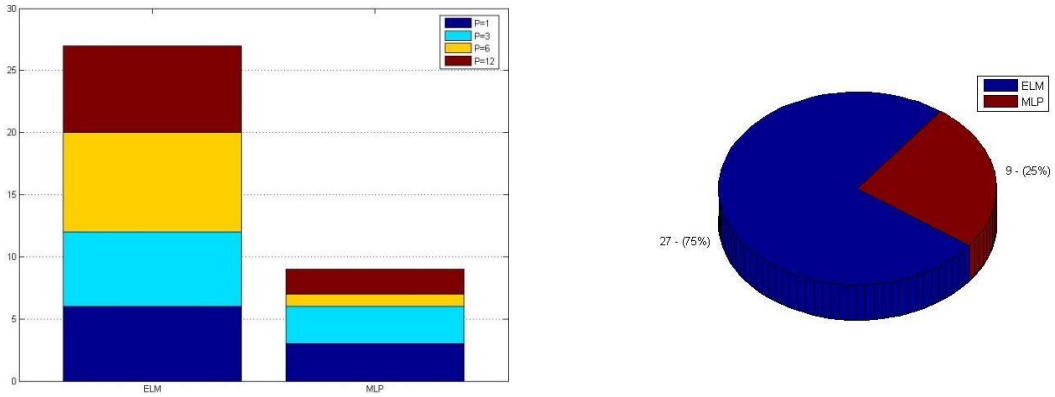


Figura 5.29 - Melhores resultados quanto a rede *feedforward*

Apesar de a ELM ser uma versão "desorganizada" da MLP, não havendo ajuste supervisionado de pesos da camada oculta, ela atingiu melhores resultados no âmbito geral. Isso provavelmente se deve tanto à capacidade de aproximação engendrada pelas projeções não-lineares aleatórias da ELM quanto a uma convergência não ideal pela MLP.

5.10 Melhores resultados por quanto à camada de saída da ESN

Uma outra análise possível é relativa as propostas de camada de saída de uma ESN: a de Boccato et al. (2012), que utiliza o filtro de Volterra e o PCA, e a de Butcher et al. (2010), que sugere uma máquina de aprendizado extremo. Na Tabela 5.22, indica-se qual camada de saída teve melhor desempenho.

Ao longo deste capítulo, já era evidente que a rede com ELM atingira resultados inferiores a quase todas os modelos de previsão, mesmo com o processo de regularização

sendo realizado. No caso geral, em 86% dos testes realizados, a camada de saída com filtro de Volterra foi superior, como mostra a Figura 5.30.

Tabela 5.22 – Melhores resultados quanto a camada de saída

Jaeger- PCA+Volterra X ELM	**Ozturk - PCA+Volterra X ELM**
P=1 – PV=9, ELM= 0; P=3– PV=5, ELM= 4; P=6 – PV=7, ELM= 2; P=12 – PV=8, ELM= 1; TOTAL - PV=29, ELM =7	P=1 – PV=9, ELM= 0; P=3 – PV=7, ELM= 2; P=6 – PV=9, ELM= 0; P=12– PV=8, ELM= 1; TOTAL - PV=33, ELM =3

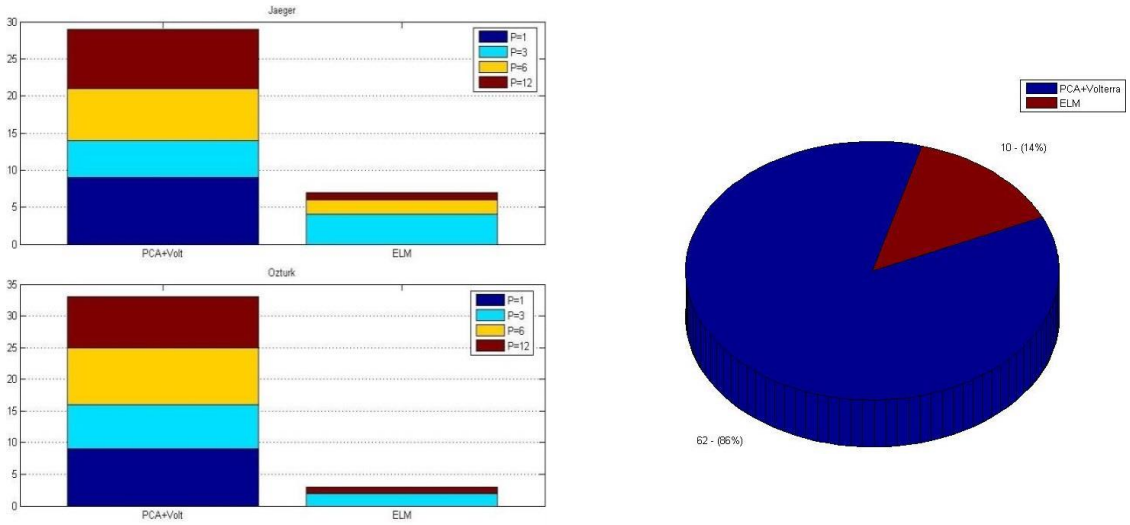


Figura 5.30 – Melhores resultados quanto a camada de saída da ESN

Apesar de a ELM ter potencial de gerar mapeamentos não-lineares, sendo mesmo um aproximador universal, nesta aplicação, é possível que as não-linearidades introduzidas no sinal que sai do reservatório precisem ser limitadas, por conta do número reduzido de amostras de treinamento. Veja que cada rede trata de um mês específico, de maneira que as informações estatísticas do sinal tendem a ser mais homogêneas do que quando se trata da previsão da série toda de uma única vez. Note que até o combinador linear da proposta inicial de Jaeger (2001) foi capaz de aproximar e generalizar melhor o sinal desejado tendo, teoricamente, menos informação disponível para isto.

Um paralelo com outros modelos é conveniente. É possível encontrar na literatura casos de modelos AR que chegam a melhores previsões do que os ARMA, embora o primeiro seja um caso particular do segundo (Guilhon 2002). Observe que mesmo com o

método realimentado dispondo de mais subsídios informacionais, a relação entre quantidade de informação e desempenho não é direta.

5.11 Comparação quanto ao número de neurônios dos melhores resultados e atrasos selecionados

Nesta seção, iremos discutir a quantidade de neurônios utilizada para formação das melhores respostas do conjunto dos preditores, bem como o número de atrasos necessários para este fim. Nas tabelas adiante, são mostrados os preditores de melhor desempenho separados por usina e número de passos á frente. Veja que, para cada caso, foram implementados 12 modelos diferentes, sendo um para cada mês. Lembremos ainda que o número máximo de entradas permitido é 6.

Na Tabela 5.23 estão os resultados do período 1951 a 1960, com as seguintes abreviações: N é o número de neurônios do modelo, N_k os neurônios da ESN como camada de saída e $\boldsymbol{N_{entr}}$ a quantidade de entradas utilizada, com os atrasos discriminados entre parênteses.

Os modelos que apresentaram os melhores resultados deixam evidente um comportamento interessante: o atraso 1 é muito utilizado para P=1, mas ocorre em poucos cenários para horizontes maiores. Além disso, são poucos os casos em que o número de entradas é maior que 2, mostrando que para os modelos foram parcimoniosos mesmo com o *wrapper* não os limitando. Em 4 oportunidades, o número de neurônios foi maior ou igual a 60, que é exatamente o número de amostras de treinamento. Este fato foi observado para duas ELMs na usina de Sobradinho. Nestes casos, o autor verificou que não ocorre nenhum tipo de sobretreinamento, já que o número de entradas é reduzido. As redes foram capazes de balancear adequadamente as entradas mais relevantes de acordo com o número de neurônios.

Tabela 5.23- Resultados para 1951-1960 (*P*=1, 3, 6 e 12)

MÊS		*CV*	P=1		P=3		P=6		P=12	
			N/N_h	N_{entr}	N/N_h	N_{entr}	N/N_h	N_{entr}	N/N_h	N_{entr}
			JAE- ESN		**OZT-ESN**		**ELM**		**JAE-ESN**	
FURNAS	JAN	0.3926	15	2 (1, 3)	5	1 (3)	10	1 (1)	5	2(4,6)
	FEV	0.3751	15	2 (1, 2)	5	1 (1)	7	1 (1)	15	1 (6)
	MAR	0.3957	3	1 (1)	3	1 (2)	7	1 (3)	7	2(1,5)
	ABR	0.3439	15	2 (1,3)	20	2(3,4)	10	1 (5)	10	1 (4)
	MAI	0.3069	10	3 (1,3,5)	3	1 (3)	15	1 (3)	5	2(5,6)
	JUN	0.3932	3	1 (1)	3	1 (4)	10	1 (2)	3	1 (6)
	JUL	0.2991	3	1 (1)	3	1 (5)	7	1 (2)	5	1 (6)
	AGO	0.2878	10	3 (1,2,6)	3	1 (3)	5	1 (4)	3	1 (4)
	SET	0.5133	5	1 (3)	3	1 (3)	7	1 (6)	3	1 (4)
	OUT	0.4284	3	1 (1)	3	2(2,3)	3	1 (5)	5	2(1,3)
	NOV	0.4145	7	2(1,2)	3	1 (5)	10	1 (6)	7	2(4,5)
	DEZ	0.3638	3	1 (1)	5	1 (3)	15	1 (2)	5	2(2,6)
			JAE-ESN		**JAE-ESN**		**ELM**		**JAE-ESN**	
EMBORCAÇÃO	JAN	0.3993	7	1 (2)	15	1 (5)	5	1 (4)	3	4(1,2,4,6)
	FEV	0.4970	5	2(1,4)	10	1 (1)	5	1 (6)	5	2(3,5)
	MAR	0.4171	20	3(1,2,3)	20	2(4,5)	10	2(1,5)	15	1 (5)
	ABR	0.3821	5	2(1,5)	10	1 (3)	10	4(2,3,4,5)	3	3(3,4,5)
	MAI	0.3169	10	3(2,3,4)	7	1 (3)	15	1 (3)	3	1 (5)
	JUN	0.3038	5	1 (1)	10	3(2,3,4)	3	1 (3)	3	1 (4)
	JUL	0.2958	10	1 (1)	15	1 (5)	5	1(6)	3	1 (6)
	AGO	0.3002	7	1 (1)	15	3(3,5,6)	7	3(2,4,6)	3	1 (6)
	SET	0.3271	3	1 (1)	3	1 (1)	5	2(3,4)	3	2(3,4)
	OUT	0.3732	7	1 (6)	5	1 (6)	3	1 (6)	15	1 (5)
	NOV	0.4555	3	1 (6)	3	1 (6)	7	1 (3)	5	1 (5)
	DEZ	0.4578	3	1 (1)	10	1 (2)	3	1 (6)	20	1 (4)
			ELM		**JAE-ESN**		**ELM**		**ELM**	
SOBRADINHO	JAN	0.2976	100	1 (1)	10	4(3,4,5,6)	5	1 (5)	15	2(1,5)
	FEV	0.4133	3	1 (1)	15	2(2,3)	5	1 (4)	7	1 (4)
	MAR	0.5180	10	1 (1)	10	2(2,3)	5	1 (1)	70	1 (1)
	ABR	0.4482	30	1 (1)	20	2(4,5)	10	2(1,3)	5	1 (2)
	MAI	0.5007	7	1 (1)	7	3(1,3,6)	7	1 (5)	5	1 (4)
	JUN	0.3457	3	1 (1)	20	1 (3)	15	1 (6)	7	1 (5)
	JUL	0.2979	3	1 (1)	20	1 (4)	3	1 (3)	10	2(5,6)
	AGO	0.2829	7	1 (1)	3	1 (3)	15	1 (6)	7	1 (6)
	SET	0.2835	7	1 (1)	3	1 (4)	15	1 (6)	5	1 (1)
	OUT	0.3174	90	1 (3)	7	1 (5)	3	1 (4)	60	1 (6)
	NOV	0.4078	10	3(1,2,3)	5	3(2,3,6)	5	2(1,2)	7	1 (1)
	DEZ	0.3537	10	1 (3)	7	4(3,4,5,6)	5	1 (5)	3	1 (4)

A seguir, na Tabela 5.24 estão os modelos selecionados para o período 1967-1976.

Tabela 5.24 - Resultados para 1967-1976 (*P*=1, 3, 6 e 12)

MÊS		*CV*	P=1		P=3		P=6		P=12	
			N/N_h	N_{entr}	N/N_h	N_{entr}	N/N_h	N_{entr}	N/N_h	N_{entr}
			JAE-PV-ESN		JAE-ESN		OZT-ESN		JAE-ESN	
FURNAS	JAN	0.3926	5	2(1,3)	7	4(2,3,4,6)	5	1 (3)	3	2(1,2)
	FEV	0.3751	100	2(1,3)	15	3(2,3,5)	5	1 (5)	15	2(3,6)
	MAR	0.3957	40	2(1,4)	5	1 (1)	3	1 (1)	3	1 (2)
	ABR	0.3439	70	2(1,6)	3	1 (1)	3	1 (5)	5	2(3,5)
	MAI	0.3069	100	1 (1)	3	1 (1)	5	2(5,6)	5	1 (2)
	JUN	0.3932	50	2(2,3)	15	1 (3)	10	1 (6)	3	1 (5)
	JUL	0.2991	15	2(1,4)	15	1 (4)	20	3(2,3,4)	5	1 (6)
	AGO	0.2878	20	2(1,2)	30	2(3,6)	15	1 (4)	3	1 (6)
	SET	0.5133	30	2(1,6)	5	2(1,3)	5	1 (1)	3	1 (5)
	OUT	0.4284	15	2(1,3)	5	2(2,3)	100	1 (3)	80	1 (1)
	NOV	0.4145	5	1(1)	7	2(4,6)	5	1 (3)	7	1 (3)
	DEZ	0.3638	7	1(2)	3	1 (2)	3	1 (2)	3	1 (5)
			JAE-ESN		JAE-ESN		OZT-ESN		OZT-ESN	
EMBORCAÇÃO	JAN	0.3993	5	4(1,2,3,5)	10	1 (2)	10	2(1,3)	5	1 (2)
	FEV	0.4970	15	1 (1)	7	2(1,2)	5	1 (3)	5	1 (3)
	MAR	0.4171	15	2(2,6)	10	3(1,2,6)	3	1 (4)	5	1 (4)
	ABR	0.3821	20	2(1,6)	5	4(1,3,4,5)	3	1 (4)	3	1 (5)
	MAI	0.3169	7	1 (1)	10	4(2,4,5,6)	3	1 (1)	5	1 (5)
	JUN	0.3038	15	3(1,2,3)	3	1 (3)	3	2(2,5)	3	1 (6)
	JUL	0.2958	10	1 (1)	15	3(3,5,6)	20	3(2,4,6)	3	1 (5)
	AGO	0.3002	15	3(1,3,5)	15	1 (3)	3	1 (2)	5	1 (6)
	SET	0.3271	3	1 (1)	3	1 (6)	5	1 (4)	100	1 (2)
	OUT	0.3732	20	4(3,4,5,6)	3	1 (1)	5	1 (5)	5	1 (6)
	NOV	0.4555	10	4(1,2,3,4)	3	2(1,2)	3	1 (1)	10	2(5,6)
	DEZ	0.4578	15	5(1,2,3,5,6)	15	4(2,4,5,6)	7	1 (2)	3	1 (2)
			JAE-PV-ESN		ELM		JAE-ESN		OZT-ESN	
SOBRADINHO	JAN	0.2976	10	2(4,5)	40	1 (3)	5	1 (5)	3	1 (3)
	FEV	0.4133	70	1 (1)	20	1 (2)	10	2(1,3)	7	1 (2)
	MAR	0.5180	60	2(1,3)	10	1 (1)	10	1 (2)	5	1 (2)
	ABR	0.4482	3	2(1,3)	10	1 (2)	7	1 (4)	3	1 (3)
	MAI	0.5007	7	1 (1)	3	2(4,6)	3	1 (5)	3	1 (5)
	JUN	0.3457	60	1 (1)	3	1 (3)	7	1 (5)	5	1 (5)
	JUL	0.2979	3	3 (1,2,3)	7	2(3,5)	7	1 (6)	5	1(5)
	AGO	0.2829	60	2 (2,4)	5	1 (1)	20	1 (6)	5	1 (6)
	SET	0.2835	90	2 (3,6)	5	1 (6)	7	3(2,4,6)	5	1 (5)
	OUT	0.3174	3	1 (3)	5	2(2,6)	3	2(5,6)	100	1 (6)
	NOV	0.4078	50	3 (1,3,5)	20	3(3,4,5)	20	1(4)	7	1 (6)
	DEZ	0.3537	5	2 (1,2)	10	3(1,3,4)	10	3(2,3,4)	20	1(5)

As máquinas desorganizadas selecionadas como melhores preditores para este período tiveram comportamento parecido com o anterior em relação ao número de neurônios na camada oculta, embora a quantidade de vezes que se observam redes com mais de 60 neurônios seja maior, inclusive em Furnas e Emborcação. A maior concentração de um número elevado de neurônios está em *P*=1.

Da mesma forma, para Emborcação, um número maior de vezes são requeridos mais do que 2 atrasos para formação da resposta final. Vale ainda a observação anterior de que o primeiro atraso está é muito utilizado em *P*=1, embora apareça mais vezes em *P*=3.

A Tabela 5.25 apresenta os melhores resultados para 77-86.

Tabela 5.25 -Resultados para 1977-1985 (*P*=1, 3, 6 e 12)

MÊS		*CV*	P=1		P=3		P=6		P=12	
			N/N_h	N_{entr}	N/N_h	N_{entr}	N/N_h	N_{entr}	N/N_h	N_{entr}
			PAR		**MLP**		**MLP**		**JAE-PV-ESN**	
FURNAS	JAN	0.3926	1		3	2(3,4)	60	1 (5)	15	2(1,2)
	FEV	0.3751	1		3	2(5,6)	3	1 (1)	30	2(2,3)
	MAR	0.3957	1		90	1 (3)	3	2(1,5)	20	1 (2)
	ABR	0.3439	2		3	2(1,2)	5	1 (1)	5	1 (1)
	MAI	0.3069	3		5	2(2,3)	3	2(1,5)	20	2(3,4)
	JUN	0.3932	2		5	2(4,6)	100	1 (3)	15	2(2,5)
	JUL	0.2991	2		80	1 (5)	10	1 (1)	5	3(2,3,5)
	AGO	0.2878	1		60	1 (3)	7	2(1,5)	80	3(1,3,6)
	SET	0.5133	4		5	1 (3)	3	3(1,2,6)	30	5(1,2,3,5,6)
	OUT	0.4284	4		20	1 (5)	10	1 (4)	50	3(1,2,4)
	NOV	0.4145	1		60	1 (2)	40	1 (2)	60	2(1,6)
	DEZ	0.3638	2		7	1 (6)	5	1 (6)	40	4(1,2,3,4)
			MLP		**JAE-ESN**		**JAE-PV-ESN**		**JAE-PV-ESN**	
EMBORCAÇÃO	JAN	0.3993	5	4(1,3,4,6)	10	4(1,4,5,6)	7	2(1,3)	40	2(1,2)
	FEV	0.4970	3	2(1,3)	15	6(1,2,3,4,5,6)	70	2(1,2)	40	1 (1)
	MAR	0.4171	80	1 (1)	7	2(1,2)	60	1 (5)	100	1 (6)
	ABR	0.3821	15	1 (1)	15	4(1,2,3,5)	3	1 (5)	70	1 (1)
	MAI	0.3169	3	2(1,4)	30	1 (1)	80	2(5,6)	70	1 (4)
	JUN	0.3038	3	2(1,6)	15	1 (1)	90	3(1,2,5)	100	3(1,2,3)
	JUL	0.2958	3	2(1,3)	40	1 (1)	5	3(1,4,5)	80	2(1,2)
	AGO	0.3002	10	1 (1)	30	1 (1)	60	1 (1)	60	1 (1)
	SET	0.3271	30	2(1,4)	30	1 (1)	90	1 (1)	5	1 (1)
	OUT	0.3732	15	1 (2)	20	1 (1)	20	2(1,5)	60	2(2,5)
	NOV	0.4555	80	1 (6)	30	1 (1)	100	2(3,5)	7	2(3,4)
	DEZ	0.4578	50	1 (1)	30	1 (1)	40	3(1,2,5)	5	1 (2)
			JAE-PV-ESN		**MLP**		**JAE-ESN**		**JAE-PV-ESN**	
SOBRADINHO	JAN	0.2976	50	3(3,5,6)	5	1 (4)	20	1 (6)	60	2(2,5)
	FEV	0.4133	3	2(1,2)	60	1 (6)	20	1 (6)	5	2(1,5)
	MAR	0.5180	3	2(1,2)	7	1 (4)	3	1 (1)	50	1 (1)
	ABR	0.4482	15	2(1,6)	90	1 (3)	5	1 (1)	90	2(1,5)
	MAI	0.5007	30	2(1,2)	100	1 (4)	3	3(4,5,6)	100	1 (4)
	JUN	0.3457	10	2(1,5)	50	1 (1)	3	1 (4)	30	1 (2)
	JUL	0.2979	90	3(1,2,6)	3	4(2,3,5,6)	10	2(3,6)	40	2(1,5)
	AGO	0.2829	100	3(1,2,6)	3	3(3,5,6)	7	2(3,4)	5	4(1,2,3,5)
	SET	0.2835	60	1 (2)	90	1 (1)	15	1 (6)	15	2(1,2)
	OUT	0.3174	90	2(1,6)	50	1 (4)	3	1 (1)	40	4(1,2,5,6)
	NOV	0.4078	70	2(1,2)	20	1 (3)	3	2(5,6)	80	2(1,6)
	DEZ	0.3537	90	1 (1)	20	1 (5)	5	1 (1)	40	1 (1)

O único caso em que o modelo linear tem melhores resultados não pode ser analisado do ponto de vista do primeiro atraso, pois o filtro de autocorrelação parcial com atrasos consecutivos parte do pressuposto de que ele sempre estará presente.

Vê-se na tabela que é muito mais recorrente o uso de mais de 60 neurônios para formação das melhores respostas das 3 séries. Além disso, continua a predominância do primeiro atraso para *P*=1, mas, curiosamente ele é utilizado em todos os meses de *P*=3 da

usina de Emborcação. A utilização de mais de 2 pelos modelos aconteceu em poucos casos, embora, desta vez, isto ocorra em praticamente todos os horizontes para pelo menos um dos meses.

O aproveitamento do atraso 1 para formação das melhores respostas está sumarizado na Figura 5.31, que mostra os gráficos em barras separados por período e por horizonte de previsão.

É nítido que, para formar o as respostas um passo adiante, os modelos utilizam com frequência o primeiro atraso, já que é este que possui maior correlação com o dado desejado. Entretanto, quando o horizonte de previsão é maior, a utilização deste atraso cai consideravelmente. Isto pode ter relação com o fato de que a informação do sinal que mais contribua para a formação da resposta esteja mais relacionada com atrasos mais distantes, mesmo com o ajuste tendo sido feito para P=1.

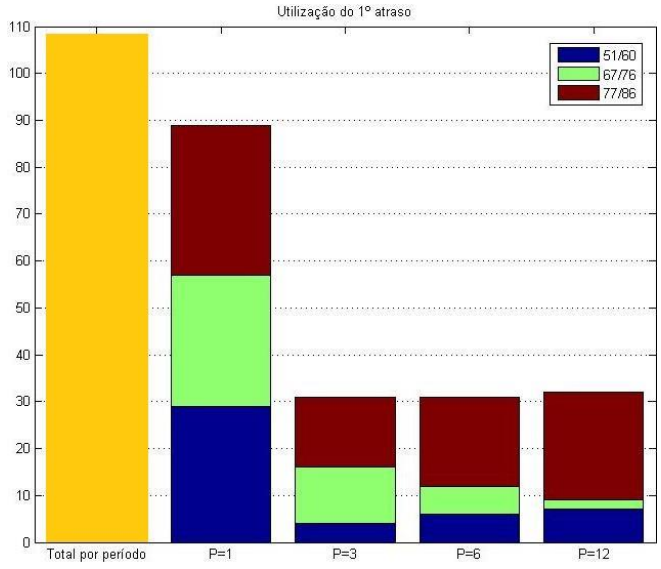
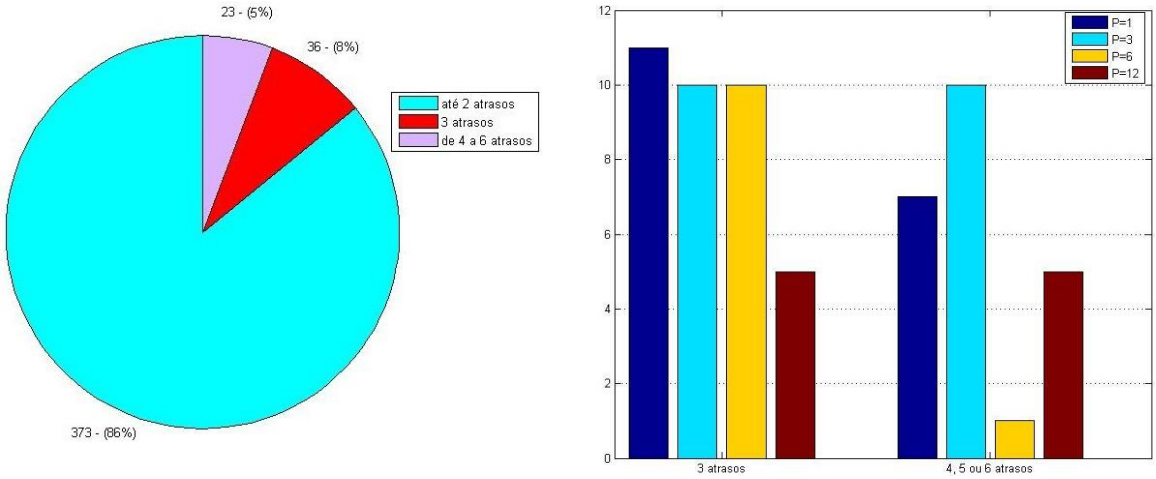


Figura 5.31 – Utilização do primeiro atraso

Cabe aqui uma discussão acerca dos métodos de seleção de variáveis para séries de hidrologia. Segundo o trabalho de Stedinger (2001), não faz sentido considerar atrasos não consecutivos para este tipo de problema. Os resultados mostram que esta afirmação pode ser verdadeira para P=1, mas não em todos os casos. Além disso o autor trata mais especificamente da seleção de entradas em modelos PAR via FACP. Como dito no Capítulo 3, redes neurais são mapeadores universais não lineares, e a forma como a escolha

de suas entradas é feita, baseada no erro de teste, nem sempre escolhe o primeiro atraso para estar presente no conjunto ótimo. Em diversos casos, a informação mais relevante para o preditor pode estar em amostras mais próximas temporalmente do dado que se pretende prever. Isto pode justificar modelos de seleção como o *wrapper,* que não necessariamente usam de atrasos consecutivos.

Uma outra discussão possível pelas tabelas dessa seção é a respeito do número de atrasos que os modelos utilizaram para formar as melhores respostas. O que se observa é que, mesmo com um número máximo permitido de 6 entradas, a maior parte dos casos mostra que os modelos utilizam 1 ou 2 atrasos. Assim eles foram parcimoniosos mesmo sem penalizações, como sugerem os critérios BIC e AIC. Se considerarmos um máximo de 3 entradas, chegamos a 95% dos casos. A Figura 5.32 ilustra este comportamento para a totalidade dos casos em forma de "pizza" e o gráfico de barras para os casos em que se usa mais de 2 entradas:

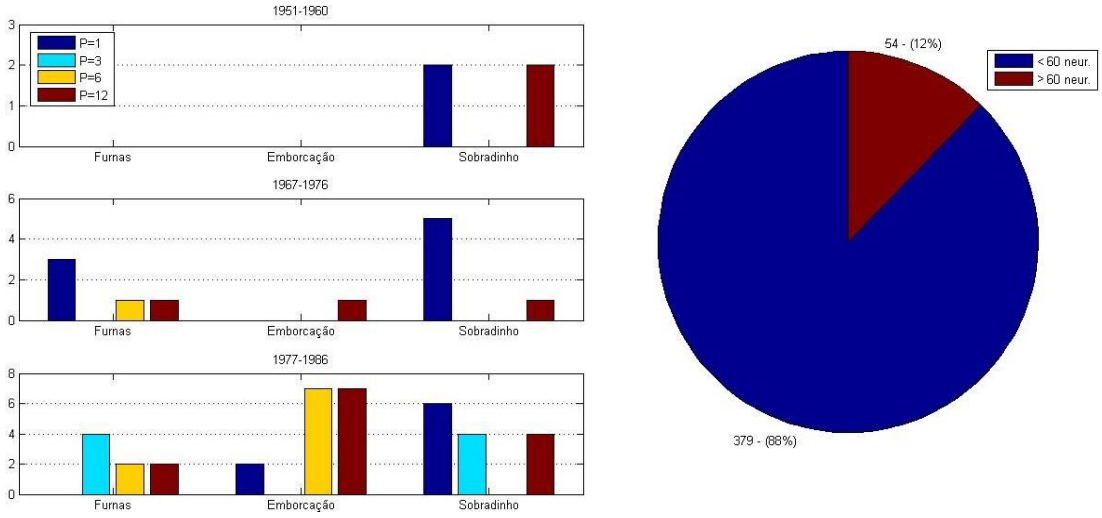


(a) (b)

Figura 5.32 – Número de atrasos utilizados (a) total, (b) de 3 a 6 atrasos

Uma última análise quantitativa pode ser feita observando as vezes que as redes utilizaram mais de 60 neurônios na camada intermediária para formação das melhores respostas. Como dito, este número é exatamente a quantidade de entradas disponível. Nota-se que este fato aconteceu em apenas 54 meses, ou 12% do total de testes realizados. Além disso, a grande maioria deles acontece no período 1977-1986, o de maior dificuldade de ser previsto por conta da sua elevada amplitude.

A Figura 5.33 resume como foi este comportamento para o total de casos e separado por horizonte e período de testes.



(a) (b)

Figura 5.33 – Utilização de mais de 60 neurônios – (a) por período, (b) do total de casos

A seguir, vamos analisar os resultado apresentados pelas máquinas desorganizadas, arquiteturas que são a base deste trabalho.

5.12 Comparação quanto ao número de neurônios e atrasos selecionados pelas MDs

Esta Seção tratará de uma análise similar àquela realizada na Seção 5.11 anterior mas com foco nos resultados alcançados pelas máquinas desorganizadas tipo ELM e pela diversas arquiteturas com o reservatório de Jaeger. Esta análise mostra-se relevante se considerarmos que a maior parte dos melhores resultados foi obtida por essas redes.

Para este estudo, separamos um período de testes de cada série: Furnas, de 1967 a 1976, Emborcação, de 1967 a 1976, e Sobradinho, de 1977 a 1986. Observamos então o modelo selecionado de ELM, JAE-ESN, JAE-PV-ESN e JAE-ELM-ESN para os quatro horizontes de previsão propostos. Interessa-nos verificar novamente a quantidade de neurônios utilizada bem como os atrasos selecionados.

A Tabela 5.26 mostra as MDs para série de Furnas.

Tabela 5.26 – Resultados para FURNAS 1967-1976

MÊS	*CV*	P=1		P=3		P=6		P=12	
		N/N_h	N_{entr}	N/N_h	N_{entr}	N/N_h	N_{entr}	N/N_h	N_{entr}
ELM									
JAN	0.3926	10	1(1)	7	1(6)	3	2(4,6)	7	2(3,5)
FEV	0.3751	7	1(1)	5	3(4,5,6)	7	1(1)	7	2(4,5)
MAR	0.3957	3	1(1)	5	1(3)	7	1(2)	5	1(3)
ABR	0.3439	5	1(1)	10	2(3,4)	15	1(1)	10	2(4,5)
MAI	0.3069	7	1(1)	3	2(3,4)	20	2(2,4)	10	2(3,5)
JUN	0.3932	10	1(1)	3	1(3)	10	1(2)	7	1(5)
JUL	0.2991	5	1(1)	3	2(3,4)	20	1(6)	40	1(6)
AGO	0.2878	10	1(1)	15	3(3,4,5)	3	1(4)	3	1(5)
SET	0.5133	3	1(1)	15	1(3)	5	1(6)	20	1(5)
OUT	0.4284	7	1(2)	10	1(3)	7	1(3)	7	1(1)
NOV	0.4145	5	1(1)	3	1(2)	7	1(4)	7	1(3)
DEZ	0.3638	7	1(2)	7	1(4)	7	1(3)	5	1(5)
JAE-ESN									
JAN	0.3926	5	3(1,3,6)	7	4(2,3,4,6)	3	2(1,2)	3	2(1,2)
FEV	0.3751	10	3(1,4,6)	15	3(2,3,5)	15	2(3,6)	15	2(3,6)
MAR	0.3957	5	1(1)	5	1 (1)	3	1(2)	3	1 (2)
ABR	0.3439	7	2(1,5)	3	1 (1)	5	2(3,5)	5	2(3,5)
MAI	0.3069	15	3(1,2,4)	3	1 (1)	5	1(2)	5	1 (2)
JUN	0.3932	3	1(1)	15	1 (3)	3	1(5)	3	1 (5)
JUL	0.2991	10	2(1,2)	15	1 (4)	5	1(6)	5	1 (6)
AGO	0.2878	15	1(1)	30	2(3,6)	3	1(6)	3	1 (6)
SET	0.5133	20	2(1,6)	5	2(1,3)	3	1(5)	3	1 (5)
OUT	0.4284	5	2(3,5)	5	2(2,3)	80	1(1)	80	1 (1)
NOV	0.4145	15	1(1)	7	2(4,6)	7	1(3)	7	1 (3)
DEZ	0.3638	3	1(2)	3	1 (2)	3	1(5)	3	1 (5)
JAE-PV-ESN									
JAN	0.3926	5	2(1,3)	10	2(5,6)	100	1(3)	10	2(2,5)
FEV	0.3751	100	2(1,3)	20	2(5,6)	5	2(4,6)	3	5(1,2,4,5,6)
MAR	0.3957	40	2(1,4)	90	2(1,6)	5	2(3,4)	50	2(1,6)
ABR	0.3439	70	2(1,6)	40	2(1,4)	70	2(1,4)	10	2(3,4)
MAI	0.3069	100	1 (1)	60	2(1,6)	30	2(1,3)	7	2(1,3)
JUN	0.3932	50	2(2,3)	7	2(1,2)	7	3(1,4,5)	100	1(6)
JUL	0.2991	15	2(1,4)	5	3(2,3,4)	10	2(4,5)	15	2(3,4)
AGO	0.2878	20	2(1,2)	3	2(1,4)	40	1(5)	15	2(4,5)
SET	0.5133	30	2(1,6)	30	2(1,5)	10	4(1,2,3,5)	80	3(4,5,6)
OUT	0.4284	15	2(1,3)	5	1(2)	80	3(2,3,6)	70	2(1,2)
NOV	0.4145	5	1(1)	60	1(2)	40	2(1,3)	100	1(3)
DEZ	0.3638	7	1(2)	3	2(3,5)	90	1(1)	70	3(2,4,5)
JAE-ESN-ELM									
JAN	0.3926	3	1(1)	90	2(2,5)	20	2(3,4)	3	1(6)
FEV	0.3751	3	1(1)	7	2(4,6)	10	3(4,5,6)	15	1(6)
MAR	0.3957	7	1(1)	90	1(1)	80	3(2,4,5)	50	3(4,5,6)
ABR	0.3439	5	1(1)	30	1(3)	30	2(3,5)	3	2(3,5)
MAI	0.3069	3	1(1)	10	2(3,4)	80	1(6)	120	2(4,6)
JUN	0.3932	3	1(1)	7	1(3)	5	1(5)	3	1(5)
JUL	0.2991	7	1(1)	5	1(4)	60	1(6)	7	1(6)
AGO	0.2878	3	1(1)	15	1(5)	80	1(6)	3	1(5)
SET	0.5133	5	1(1)	5	2(3,5)	40	1(6)	15	1(5)
OUT	0.4284	3	1(3)	3	1(3)	3	1(2)	10	1(6)
NOV	0.4145	40	1(1)	7	1(2)	120	1(3)	110	1(3)
DEZ	0.3638	15	1(2)	3	1(2)	80	2(3,4)	20	1(3)

De posse deste estudo comparativo, interessa-nos mostrar as redes quando o problema de previsão é similar. A primeira observação é que, novamente, os modelos utilizaram o primeiro atraso para prever um passo à frente, como já visto na seção anterior. As redes JAE-ESN e JAE-PV-ESN também fazem uso deste atraso para P=3 em alguns casos.

As redes ELM e JAE-ESN-ELM, em dois e três casos, respectivamente, ultrapassam os 2 atrasos. As outras arquiteturas já fazem uso deles em cinco e sete casos.

A rede JAE-PV-ESN é aquela que mais vezes ultrapassa o limite de 60 neurônios, fato que é notado várias vezes, independentemente do horizonte de previsão. A rede JAE-ESN-ELM também faz uso de muitas destas unidades, mas apenas para P>1.

A Tabela 5.27 apresenta os resultados para a série da usina de Emborcação. Na análise do número de entradas deste posto, fica nítido que, novamente, o primeiro atraso é muito utilizado para o horizonte de um passo à frente e menos para os demais. O que ocorre de diferente é que a quantidade de entradas maior do que duas aparece diversas vezes para todos os modelos: na ELM em 5 casos, na JAE-ESN e JAE-PV-ESN em 14 e na JAE-ESN-ELM em 6.

Como na série de Furnas, a JAE-PV-ESN extrapola os 60 neurônios em muitas ocasiões e em todos os horizontes. Em menor medida isto também ocorre com a JAE-ESN-ELM.

Tabela 5.27 - Resultados para EMBORCAÇÃO 1967-1976

MÊS	*CV*	P=1		P=3		P=6		P=12	
		N/N_h	N_{entr}	N/N_h	N_{entr}	N/N_h	N_{entr}	N/N_h	N_{entr}
ELM									
JAN	0.3993	3	1(1)	10	2(4,6)	10	1(5)	5	1(5)
FEV	0.4970	3	1(1)	10	2(2,5)	10	3(1,4,6)	15	3(1,4,6)
MAR	0.4171	3	1(1)	7	1(6)	7	1(2)	7	1(2)
ABR	0.3821	3	1(1)	3	2(3,5)	5	1(1)	5	1(2)
MAI	0.3169	5	1(1)	20	3(2,3,5)	5	3(2,4,5)	3	1(4)
JUN	0.3038	3	1(1)	15	3(3,4,5)	40	1(3)	7	1(5)
JUL	0.2958	3	1(1)	7	2(3,6)	15	4(2,4,5,6)	7	2(5,6)
AGO	0.3002	3	1(1)	3	1(3)	7	2(3,6)	3	1(6)
SET	0.3271	3	1(3)	5	1(6)	10	1(1)	10	3(2,4,6)
OUT	0.3732	5	6(1,2,3,4,5,6)	5	1(1)	7	1(5)	10	1(1)
NOV	0.4555	7	1(1)	3	3(1,2,3)	5	1(2)	5	1(6)
DEZ	0.4578	10	1(1)	10	2(2,3)	10	2(2,4)	7	1(2)
JAE-ESN									
JAN	0.3993	5	4(1,2,3,5)	10	1 (2)	15	1(2)	10	1(2)
FEV	0.4970	15	1 (1)	7	2(1,2)	15	1(3)	10	1(3)
MAR	0.4171	15	2(2,6)	10	3(1,2,6)	10	1(1)	10	1(1)
ABR	0.3821	20	2(1,6)	5	4(1,3,4,5)	10	1(4)	20	2(4,5)
MAI	0.3169	7	1 (1)	10	4(2,4,5,6)	7	2(1)	15	1(5)
JUN	0.3038	15	3(1,2,3)	3	1 (3)	7	3(2,3,6)	3	2(4,6)
JUL	0.2958	10	1 (1)	15	3(3,5,6)	10	3(2,5,6)	3	1(5)
AGO	0.3002	15	3(1,3,5)	15	1 (3)	10	3(2,4,5)	5	2(2,6)
SET	0.3271	3	1 (1)	3	1 (6)	3	1(3)	3	1(6)
OUT	0.3732	20	4(3,4,5,6)	3	1 (1)	3	1(5)	20	1(6)
NOV	0.4555	10	4(1,2,3,4)	3	2(1,2)	10	2(1,3)	10	2(5,6)
DEZ	0.4578	15	5(1,2,3,5,6)	15	4(2,4,5,6)	20	1(2)	15	1(4)
JAE-PV-ESN									
JAN	0.3993	10	3(1,4,6)	60	1(5)	40	3(3,4,6)	7	2(4,6)
FEV	0.4970	90	2(1,6)	100	1(6)	80	1(6)	15	2(2,4)
MAR	0.4171	60	2(2,5)	60	2(1,6)	10	2(3,6)	60	4(2,3,4,5)
ABR	0.3821	70	1(1)	7	1(1)	15	2(4,6)	50	4(3,4,5,6)
MAI	0.3169	7	2(1,6)	60	3(2,3,4)	15	5(1,2,4,5,6)	70	1(5)
JUN	0.3038	60	2(1,6)	10	2(1,6)	80	3(2,3,6)	100	1(3)
JUL	0.2958	50	1(1)	90	1(2)	7	3(1,3,4)	7	2(5,6)
AGO	0.3002	80	1(2)	50	1(1)	60	3(2,4,6)	15	1(3)
SET	0.3271	80	2(3,6)	60	3(1,2,6)	30	3(1,3,6)	20	1(6)
OUT	0.3732	70	1(6)	50	3(1,2,3)	30	1(5)	40	1(1)
NOV	0.4555	70	1(1)	80	3(1,2,3)	7	5(1,2,3,4,5)	3	1(1)
DEZ	0.4578	10	2(1,6)	20	2(5,6)	7	1(2)	40	3(2.4,6)
JAE-ESN-ELM									
JAN	0.3993	3	1(1)	15	1(2)	5	1(5)	7	1(5)
FEV	0.4970	80	1(1)	80	2(3,6)	3	1(1)	5	1(1)
MAR	0.4171	7	1(1)	7	1(2)	50	1(6)	10	1(3)
ABR	0.3821	7	1(1)	40	3(3,4,5)	3	1(1)	7	1(3)
MAI	0.3169	110	1(1)	100	2(3,6)	60	3(2,4,6)	7	1(4)
JUN	0.3038	5	1(1)	3	1(3)	70	3(3,4,6)	40	2(4,5)
JUL	0.2958	5	1(1)	3	1(4)	7	1(1)	15	1(6)
AGO	0.3002	7	1(1)	10	1(3)	110	3(3,5,6)	100	1(6)
SET	0.3271	3	1(1)	3	1(3)	10	3(1,5,6)	10	1(3)
OUT	0.3732	60	2(1,6)	7	1(1)	5	1(5)	5	1(6)
NOV	0.4555	5	1(1)	70	2(1,2)	40	3(1,5,6)	20	1(6)
DEZ	0.4578	80	1(1)	10	2(2,4)	7	2(2,3)	40	1(4)

O último caso de estudo está descrito na Tabela 5.28:

Tabela 5.28 - Resultados para SOBRADINHO 1977-1986

MÊS	*CV*	P=1		P=3		P=6		P=12	
		N/N_h	N_{entr}	N/N_h	N_{entr}	N/N_h	N_{entr}	N/N_h	N_{entr}
ELM									
JAN	0.2976	5	1(2)	7	2(3,5)	7	2(2,3)	15	1(4)
FEV	0.4133	3	1(1)	3	1(1)	5	5(1,3,4,5,6)	7	3(1,5,6)
MAR	0.5180	10	1(1)	15	5(2,3,4,5,6)	3	2(5,6)	7	3(2,5,6)
ABR	0.4482	3	1(1)	3	2(3,6)	7	2(2,6)	7	1(2)
MAI	0.5007	3	1(1)	3	1(3)	5	2(4,5)	5	1(3)
JUN	0.3457	10	1(1)	5	2(3,4)	7	1(4)	7	1(4)
JUL	0.2979	10	1(1)	15	3(3,5,6)	5	1(2)	3	1(4)
AGO	0.2829	7	1(1)	7	2(3,6)	5	1(2)	5	2(4,6)
SET	0.2835	7	1(1)	7	2(3,4)	7	2(4,6)	5	1(4)
OUT	0.3174	3	1(1)	3	1(3)	7	1(1)	5	3(4,5,6)
NOV	0.4078	3	1(1)	10	3(3,4,6)	7	1(6)	3	4(1,2,3,6)
DEZ	0.3537	5	1(1)	10	1(6)	7	1(3)	10	1(3)
JAE-ESN									
JAN	0.2976	15	3(1,4,5)	15	1(1)	20	1(6)	5	2(1,2)
FEV	0.4133	20	1(1)	15	2(3,5)	20	1(6)	3	2(1,6)
MAR	0.5180	30	2(1,5)	10	3(2,4,6)	3	1(1)	7	2(1,6)
ABR	0.4482	10	3(1,3,6)	3	1(3)	5	1(1)	3	4(1,2,3,6)
MAI	0.5007	10	1(2)	3	1(3)	3	3(4,5,6)	3	1(3)
JUN	0.3457	15	1(1)	15	1(3)	3	1(4)	10	1(3)
JUL	0.2979	15	2(1,6)	15	1(4)	10	2(3,6)	10	2(1,5)
AGO	0.2829	30	1(1)	30	1(3)	7	2(3,4)	10	2(1,6)
SET	0.2835	30	1(1)	30	1(3)	15	1(6)	3	1(1)
OUT	0.3174	20	2(1,6)	20	2(3,6)	3	1(1)	7	3(4,5,6)
NOV	0.4078	20	1(1)	10	2(3,5)	3	2(5,6)	15	1(1)
DEZ	0.3537	20	2(1,6)	3	1(1)	5	1(1)	20	1(5)
JAE-PV-ESN									
JAN	0.2976	50	3(3,5,6)	40	2(1,2)	10	2(1,2)	60	2(2,5)
FEV	0.4133	3	2(1,2)	40	1(1)	5	2(1,5)	5	2(1,5)
MAR	0.5180	3	2(1,2)	100	2(2,3)	15	2(1,2)	50	1 (1)
ABR	0.4482	15	2(1,6)	90	3(1,2,3)	90	2(1,3)	90	2(1,5)
MAI	0.5007	30	2(1,2)	5	1(4)	70	2(4,5)	100	1 (4)
JUN	0.3457	10	2(1,5)	70	3(1,2,4)	5	1(5)	30	1 (2)
JUL	0.2979	90	3(1,2,6)	90	1(2)	20	2(5,6)	40	2(1,5)
AGO	0.2829	100	3(1,2,6)	100	2(1,5)	60	1(5)	5	4(1,2,3,5)
SET	0.2835	60	1 (2)	30	1(2)	40	1(1)	15	2(1,2)
OUT	0.3174	90	2(1,6)	7	2(1,6)	60	2(1,5)	40	4(1,2,5,6)
NOV	0.4078	70	2(1,2)	90	2(1,2)	60	2(1,4)	80	2(1,6)
DEZ	0.3537	90	1 (1)	7	2(1,4)	80	3(1,2,6)	40	1 (1)
JAE-ESN-ELM									
JAN	0.2976	3	1(4)	7	2(4,5)	10	1(4)	5	1(3)
FEV	0.4133	7	1(1)	80	2(2,6)	3	1(4)	100	2(3,6)
MAR	0.5180	5	1(3)	50	2(3,6)	7	1(1)	10	1(1)
ABR	0.4482	3	1(1)	3	1(3)	110	1(5)	15	1(2)
MAI	0.5007	10	1(1)	3	1(4)	110	1(5)	3	1(4)
JUN	0.3457	10	1(1)	7	1(5)	5	1(5)	30	1(6)
JUL	0.2979	7	1(1)	3	1(6)	5	1(6)	5	1(5)
AGO	0.2829	5	1(1)	10	1(3)	3	1(6)	10	1(6)
SET	0.2835	3	1(1)	7	2(3,6)	5	1(6)	10	1(5)
OUT	0.3174	100	3(1,3,6)	10	2(2,6)	15	2(5,6)	7	1(4)
NOV	0.4078	40	1(1)	80	2(2,5)	110	1(4)	10	1(6)
DEZ	0.3537	7	1(1)	10	2(1,4)	10	2(1,2)	20	3(1,2,6)

O comportamento observado para a série de Emborcação é similar àquele da série de Sobradinho: uso frequente do primeiro atraso para P=1 e uma grande quantidade de neurônios para a JAE-PV-ESN. O modelo que mais diferiu foi o JAE-ESN-ELM, que apresentou apenas 7 casos com mais de 60 neurônios e 2 outros com 3 entradas.

O resumo gráfico dos resultados desta seção será agora apresentado. Inicialmente, vamos discutir o número de vezes que os neurônios extrapolam o número de amostras de treinamento conforme mostra a Figura 5.34:

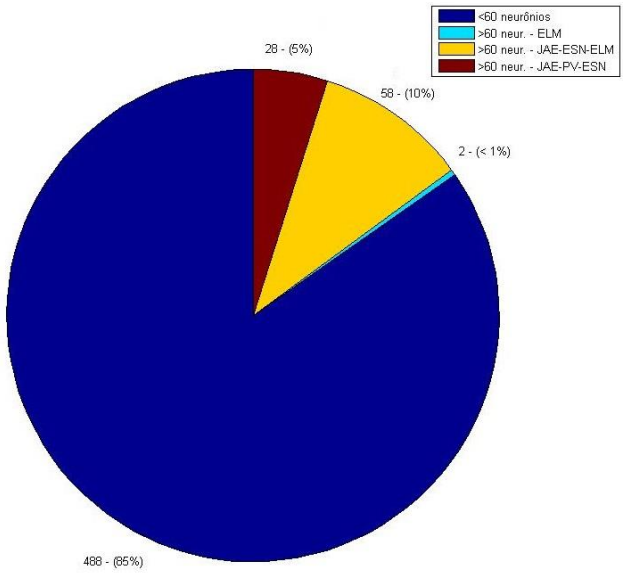


Figura 5.34 – Número de neurônios

Em 85% dos meses previstos pelas máquinas desorganizadas, o número de neurônios foi menor que 60. Os poucos casos em que isso não ocorreu foram em sua maioria da rede JAE-PV-ESN e, em menor medida, de JAE-ESN-ELM. Ou seja, considerando os três períodos selecionados, os modelos foram parcimoniosos quanto aos neurônios na camada intermediária, já que o número deles não ultrapassou o de amostras de treinamento.

Além disso, a JAE-PV-ESN, em geral, utiliza mais neurônios no seu reservatório de dinâmicas que os demais modelos de MDs. Isto tem relação com a utilização do PCA: como há redução no número de dados, a rede pode precisar de mais estados de eco para manter uma adequada quantidade de informação repassada à camada de saída.

Testes preliminares foram realizados pelo autor com o intuito de verificar em que condições as redes poderiam sobretreinar durante o ajuste devido à quantidade de

neurônios. O que se observa é que isto só ocorre quando, juntamente com um elevado número de neurônios, também há pelo menos três entradas selecionadas. Além disso, estas entradas precisam ser temporalmente próximas (p. ex. 1, 2 e 4). Ou seja, o processo de construção da rede acaba por fazer um balanço entre entradas/neurônios que impede que o sobretreinamento (*overfitting*) ocorra.

Outra análise foi do uso do primeiro atraso para formação da resposta de P=1. Praticamente a totalidade das máquinas desorganizadas utilizou este dado para previsão, como mostra a Figura 5.35.

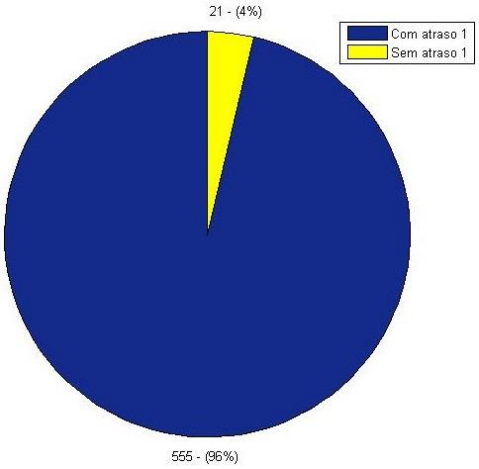


Figura 5.35 – Utilização do primeiro atraso para P=1

Outra verificação necessária é a do grau de parcimônia das MDs tendo em vista que o modelo de seleção de variáveis tipo *wrapper* permite, neste caso, até 6 entradas. O que se percebe é que, novamente, a grande maioria dos casos utiliza apenas 2 atrasos. Se extrapolarmos para 3 entradas, chega-se a 96% dos resultados obtidos, conforme mostra a Figura 5.36.

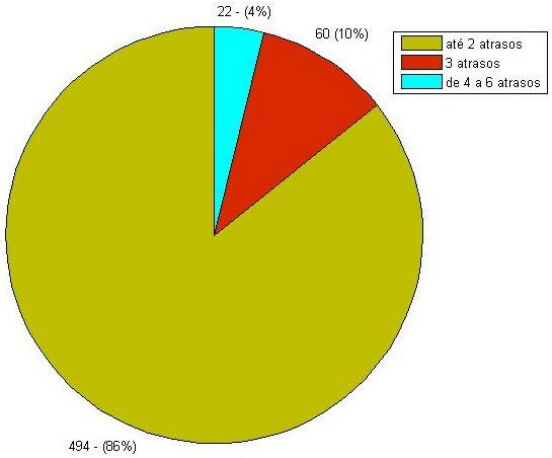


Figura 5.36 - Número de atrasos utilizados pelas MDs

O gráfico em barras da Figura 5.37 apresenta a distribuição deste número de entradas por arquitetura de rede.

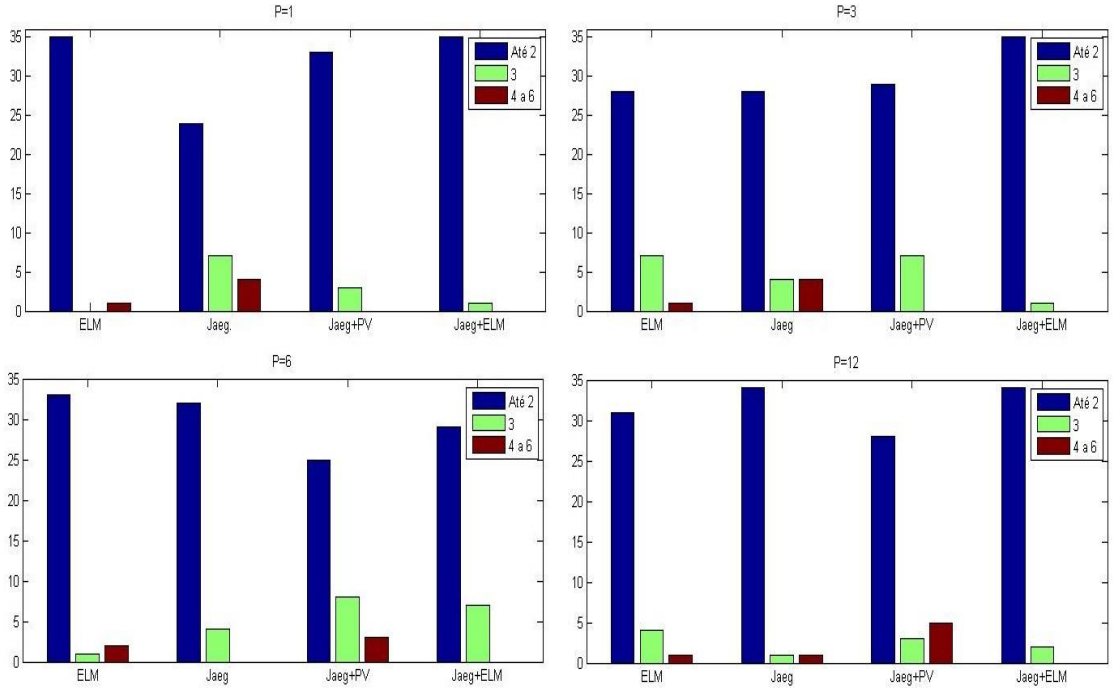


Figura 5.37 - Número de entradas por modelo

Novamente, é perceptível que são poucos os casos em que são utilizadas mais de 2 entradas. Isso é mais frequente com a rede JAE-PV-ESN e, em menor escala, com JAE-ESN.

Depois destas discussões, apresentaremos o teste estatístico de Friedman para verificação da diferença entre os resultados obtidos.

5.13 Teste de Friedman

Testes de hipóteses são ferramentas úteis na averiguação de quão significativamente são diferentes os resultados fornecidos por modelos de previsão. Este tipo de análise faz-se importante, pois somente a verificação de métricas de erro pode induzir o pesquisador a conclusões estatisticamente incorretas (Ferreira 2010).

Os testes são classificados como paramétricos e não paramétricos (Viali 2012). No primeiro caso, os métodos utilizam parâmetros da distribuição de probabilidade dos dados para cálculo das suas estatísticas. Por isso, tendem a ser mais rigorosos e com resultados mais confiáveis. No segundo, os pressupostos são atribuídos apenas aos dados ordenados, sem inclusão de informações sobre distribuições, simplificando o tratamento matemático envolvido. Dessa forma, os testes não paramétricos são menos "exigentes" e podem ser aplicados com menos cálculos e aos mais diversos conjuntos de dados. Além disso, são convenientes nos casos em que a distribuição é desconhecida (Muller, Kruger e Kaviski 1998).

Neste trabalho, a opção foi pela utilização do teste não-paramétrico de Friedman (Friedman 1937), que é útil para comprovar se as amostras foram extraídas da mesma população ou seja, estatística de Friedman (X^2) vai mensurar a probabilidade de que as amostram venham do mesmo processo gerador (Snedecor e Cochran 1980, Luna e Ballini, 2011).

Como é comum em testes estatísticos, duas hipóteses (H) são formuladas para aceitação ou rejeição do que é proposto:

- H_0 (nula) - os tratamentos são iguais;

- H_1 (alternativa) – pelo menos um tratamento em toda a população produz grandes efeitos.

Além disso, o teste de Friedman parte dos seguintes pressupostos:

i) Os blocos são independentes;
ii) Não existe interação entre os blocos e tratamentos;
iii) As observações dentro de cada bloco podem ser ranqueadas.

A estatística do teste é calculada pela expressão (5.5) (Friedman 1937):

$$X^2 = \frac{12}{bc(c+1)} \sum_{j=1}^{c} R_j^2 - 3b(c+1) \tag{5.5}$$

na qual c é número de níveis dos tratamentos (ou, em nosso caso, o número de preditores), b o número de blocos (ou o número de amostras no conjunto de teste) e R_j é a soma dos números de ordem para um particular nível de tratamento j.

A forma de representação dos dados consiste em ordená-los em b linhas e c colunas. O cálculo da variável R_j pode ser exemplificado pela Tabela 5.29 (Ferreira 2010).

Tabela 5.29 – Exemplo de previsão

	Tratamento 1	**Tratamento 2**	**Tratamento 3**	**Tratamento 4**
Bloco 1	7.0 (3)	5.3 (2)	4.9 (1)	8.8 (4)
Bloco 2	9.9 (4)	5.7 (1)	7.6(2)	8.9 (3)
Bloco 3	8.5 (4)	4.7 (1)	5.5(2)	8.1 (3)
Bloco 4	5.1 (4)	3.5 (3)	2.8(1)	3.3 (2)
Bloco 5	10.3 (4)	7.7 (1)	8.4 (2)	9.1 (3)
R_j	R_1=19	R_2=8	R_3=8	R_4=15

Como é possível perceber, para cada elemento das b=5 linhas, dá-se um número de ordem (entre parênteses), que é crescente de acordo com a dimensão do dado previsto. Em seguida, somam-se os números de ordem de cada um dos c=4 preditores, chegando ao valor de R_j para cada caso. Nesse exemplo $X_r^2 = 10{,}68$.

A Equação (5.5) pode ser reescrita como (A. M. Ferreira 2010):

$$X^2 = \frac{12}{bc(c+1)} \sum_{j=1}^{c} b(\overline{R}_j - \overline{R})^2 \tag{5.5b}$$

sendo $\overline{R}_j$ a média dos números de ordem do j-ésimo preditor e $\overline{R}$ a média global dos números de ordem. A variável X_r^2 vale zero quando todos os preditores apresentam a mesma média para os números de ordem e é maior que zero e crescente à medida que eles diferem.

Os valores de X_r^2 são aproximados pela estatística de chi-quadrado, os quais são calculados mais facilmente pela tabela da distribuição. O número de graus de liberdade (gl) é adotado como gl=(c-1). Assim, a hipótese nula deve ser assumida como verdadeira, com um nível crítico de significância α, caso o valor encontrado de X_r^2 seja menor que o seu correspondente na tabela da distribuição chi-quadrado (Viali 2012).

Uma abordagem complementar é calcular a probabilidade limite, ou o *p-valor* relativo ao X^2 encontrado. Esta medida equivale à área da curva formada pela distribuição chi-quadrado (vide Figura 5.39) à direta do valor calculado para X_r^2, ou seja, a área cauda superior (Snedecor e Cochran 1980). Caso o *p-valor* seja menor que o α pré-estabelecido, rejeita-se a hipótese nula. Este número também é obtido mais facilmente pela tabela da distribuição.

É possível interpretar o *p-valor* de duas outras formas. A primeira é abordá-lo em termos da função de distribuição cumulativa (CDF), na qual a área formada pelo valor X_r^2 é (*p-valor* = 1-CDF(X_r^2)), já que o valor máximo desta distribuição é 1. Sob outro ponto de vista, é possível afirmar que o *p-valor* é, na verdade, o nível de significância real de determinado valor de X_r^2.

Para exemplificar o processo, tomemos a Figura 5.38, a qual mostra a distribuição chi-quadrado para 8 graus de liberdade. Suponhamos que, para uma aplicação qualquer de previsão, tenha-se 9 preditores e o nível de significância exigido seja $\alpha = 0{,}05$. Para este caso, o valor crítico da estatística chi-quadrado é $\chi^2_{(0.05;8)} = 15{,}5073$, mostrado na linha vertical intermediária. Isto quer dizer que a área formada à direita deste valor é 0,05. Nesta mesma aplicação, calcula-se o valor da estatística de Friedman para um determinado ponto. Imaginemos que o valor encontrado seja $X_B^2 = 17{,}5345$. Este valor é maior que aquele necessário para rejeitar-se a hipótese nula, portanto assume-se que a mudança nos preditores altera significativamente o resultado final. Se calculado o *p-valor* relativo a X_B^2, este será igual a 0,025, ou seja a área hachurada à direta. Pela figura, fica claro que este valor é menor que o mínimo desejável e, portanto, reafirma-se a rejeição a H_0.

Suponha, por outro lado, que o valor da estatística de Friedman encontrado seja $X_A^2 = 13{,}3616$, menor que $\chi^2_{(0.05;8)} = 15{,}5073$. Isto é suficiente para aceitarmos a hipótese nula de que as previsões não são significativamente diferentes. Vê-se que a área formada pela distribuição é maior que 0,05. Neste caso, o *p-valor* vale 0,10.

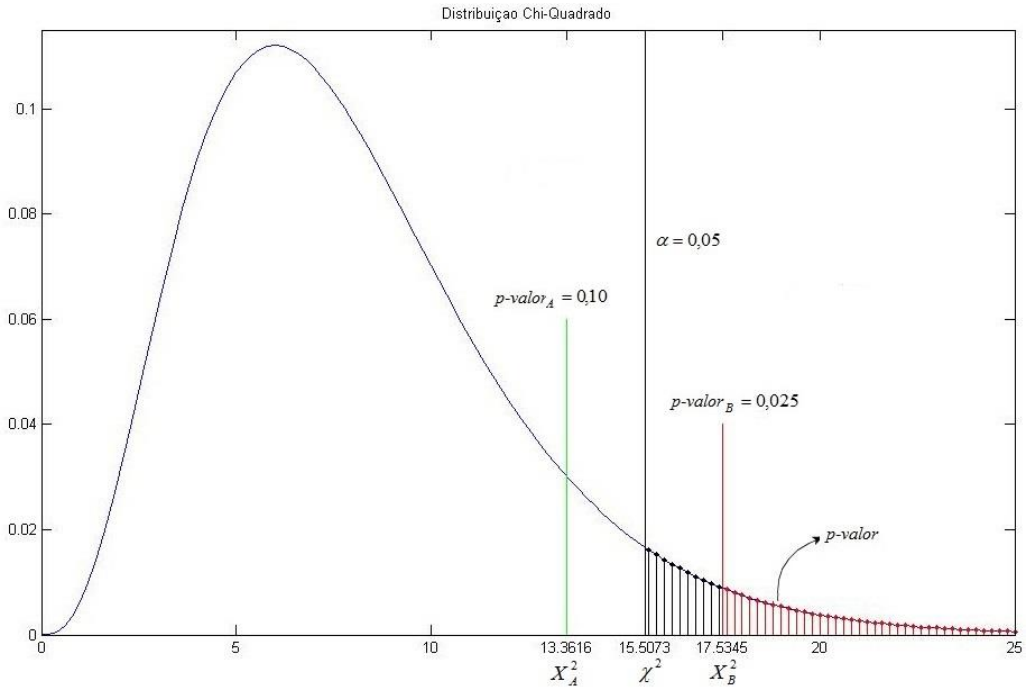


Figura 5.38 – Distribuição chi-quadrado e *p*-valores

No caso de estudo deste trabalho, os tratamentos serão correspondentes aos modelos de previsão envolvidos: PAR, MLP, ELM, ESN – Jaeger, ESN-Ozturk, ESN Jaeger e Ozturk com PCA e filtro de Volterra, ESN de Jaeger e Ozturk com uma ELM como camada de saída. Os blocos serão os meses previstos ou cada dado de teste para o período de estudos.

Aplicou-se o teste de Friedman para as previsões de 1 passo à frente (o horizonte de ajuste) com as três séries – Furnas, Emborcação e Sobradinho - e os três períodos de testes propostos – 51/60, 67/76 e 77/86 - de uma execução que apresentou MSE aproximado à média encontrada para 50 simulações. Todos os *p-valores* foram iguais a ou muito próximos de zero, sendo o de maior magnitude da ordem de 10^{-3}, muito menor que qualquer valor de nível de significância para 8 graus de liberdade. Este experimento comprova que a mudança no modelo altera significativamente as previsões.

Uma outra perspectiva foi analisar os 50 valores de MSE gerados pelos 9 modelos. Procedeu-se também com essa abordagem, e os *p-valores* encontrados novamente foram muito baixos, quase todos iguais a zero.

5.14 Série da usina de Passo Real

Após os testes utilizando as séries históricas das usinas de Furnas, Emborcação e Sobradinho, procedemos com a previsão dos mesmos horizontes abordados para a usina hidrelétrica de Passo Real, localizada no Rio Jacuí, no Rio Grande do Sul. Optamos por averiguar o comportamento das máquinas desorganizadas neste contexto por conta de seu comportamento hidrológico distinto dos demais postos utilizados até aqui. A vazão média é pequena, conforme mostra a Tabela 5.30, a qual apresenta ainda a média e o desvio padrão do período 2001-2010, que será utilizado como conjunto de teste

Tabela 5.30–Média e desvio padrão

	Média ($\hat{\mu}$)	D. Padrão ($\hat{\sigma}$)
Série completa	205.9396	169.3471
Período de teste	223.0333	158.9234

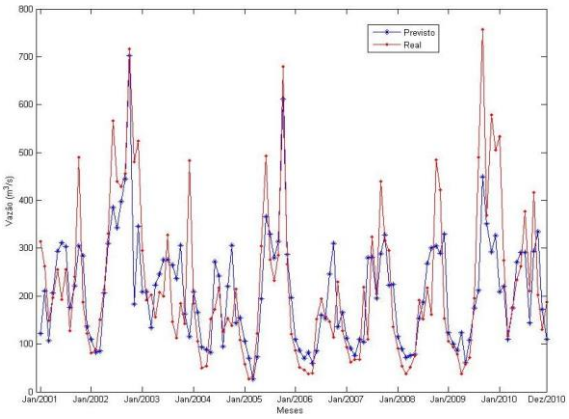
Comparamos o desempenho do modelo linear PAR com uma máquina de aprendizado extremo e a rede de estados de eco de Jaeger, já que estas foram as redes que apresentaram o maior número de melhores resultados para as demais usinas hidrelétricas. A Tabela 5.31 apresenta as performances computacionais para as previsões.

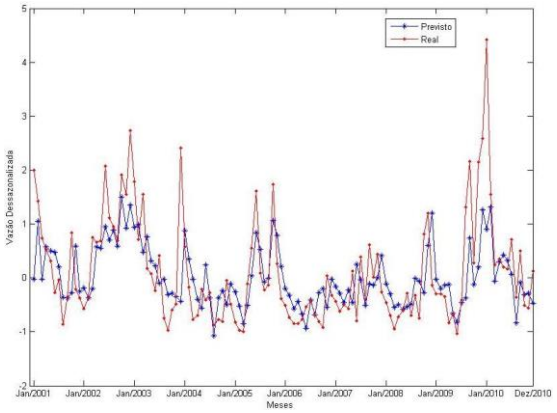
Tabela 5.31 – Resultados para Passo Real 2001-2010

	Preditor	RE	MSE	MAE	MSE dessaz	MAE dessaz
P=1	**PAR**	0.7080	1.4376e+04	85.4504	0.6319	0.5802
	ELM	0.6172	**1.0924e+04**	**75.3018**	**0.5574**	**0.5248**
	JAE-ESN	0.6199	1.1019e+04	74.6153	0.5333	0.5109
P=3	**PAR**	0.8170	1.9141e+04	101.8654	0.9179	0.7085
	ELM	0.7053	**1.4265e+04**	**88.0388**	**0.7257**	**0.6140**
	JAE-ESN	0.7322	1.5377e+04	88.7047	0.7658	0.6238
P=6	**PAR**	0.8395	2.0213e+04	106.1428	0.9722	0.7371
	ELM	0.7711	**1.7053e+04**	**97.0832**	**0.8624**	**0.6790**
	JAE-ESN	0.7995	1.8332e+04	99.6773	0.8544	0.6826
P=12	**PAR**	0.8327	1.9883e+04	105.5147	0.9683	0.7373
	ELM	0.7897	1.7884e+04	99.2847	0.8877	0.6950
	JAE-ESN	0.7863	**1.7733e+04**	**100.0625**	**0.8813**	**0.7016**

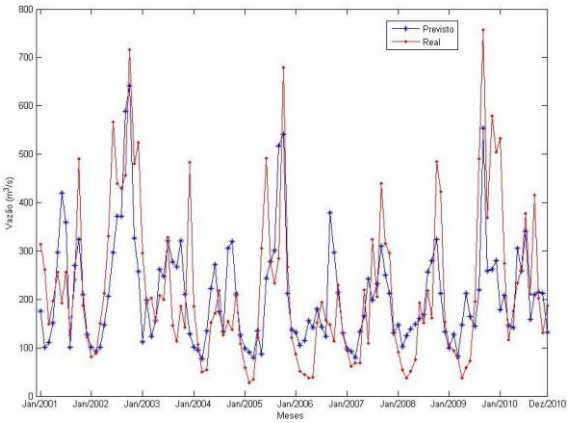
Como é perceptível, as máquinas desorganizadas atingiram desempenho superior, sobretudo para horizontes menores. Em três oportunidades, a ELM foi a solução de menor MSE real. Todavia, para P=1, o menor MSE dessazonalizado foi alcançado pela JAE-ESN. Isto ajuda a corroborar as análises realizadas até aqui, de que estas arquiteturas de redes neurais propostas são viáveis para solução deste problema, mesmo frente a usinas com comportamento tão distinto quanto esta.

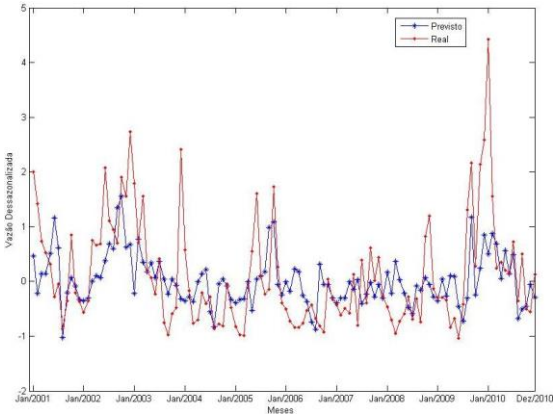
A Figura 5.39 mostra as melhores previsões.



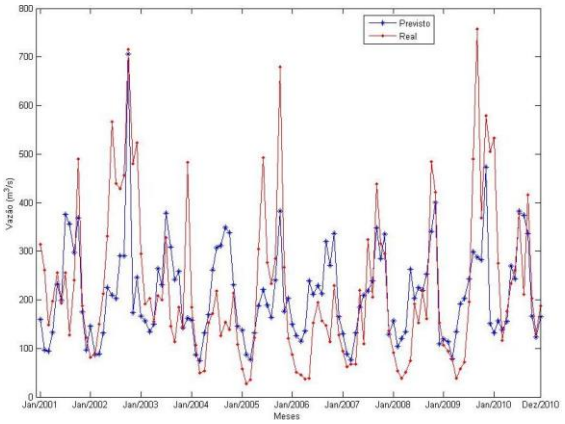


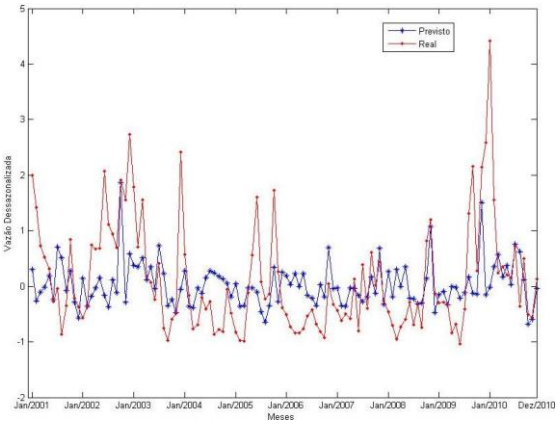
(a)



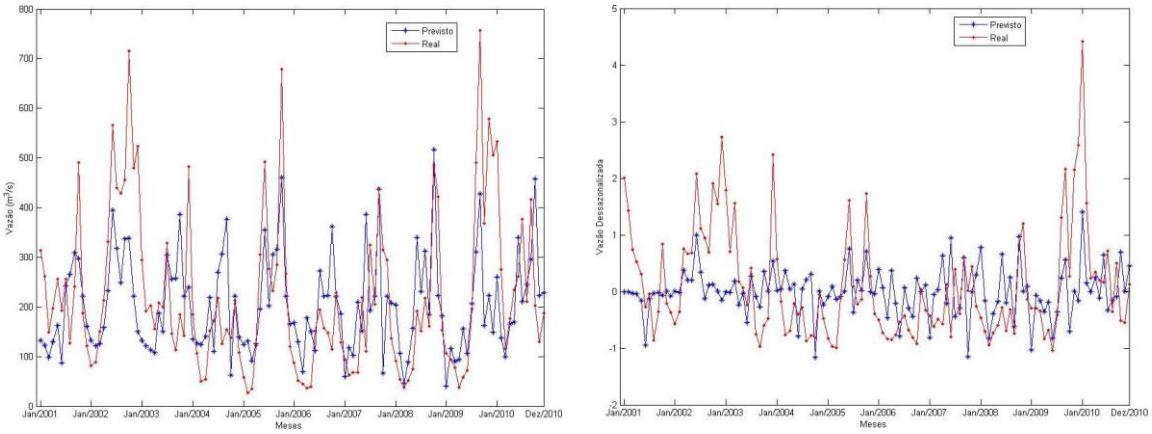


(b)





(c)



(d)

Figura 5.39–Previsões série de Passo Real – (a) P=1, (b) P=3, (c) P=6, (d) P=12

5.15 Formas alternativas de previsão

O último conjunto de testes computacionais que apresentaremos neste trabalho foi desenvolvido no sentido de confrontar as formas de previsão mais utilizadas neste trabalho. De posse do modelo PAR e das ELM's, partiremos de alguns pressupostos diferentes dos assumidos anteriormente. O primeiro é considerar que os atrasos sejam necessariamente consecutivos, conforme Stedinger (2001), mesmo para a seleção de variáveis de um RNA. Continuaremos limitando-os a 6.

Depois, para P>1, iremos proceder também com a previsão de forma direta, em vez da forma recursiva usada até agora. Nesta forma de previsão, as entradas do modelo em determinado instante t geram diretamente a amostra x_{t+P} com o horizonte P requerido, sem que seja necessário prever as amostras intermediárias para formação da reposta final. Se, por um lado, há perda de informação por conta da redução da correlação entre os dados de entrada e o horizonte previsto, por outro não há acúmulo de erro por cada passo intermediário que preciso reinserir como entrada da rede.

Verificaremos a diferença entre prever-se as séries com 12 modelos diferentes e apenas um modelo para todo histórico. Dessa forma, o modelo linear é na verdade o auto-regressivo (AR), descrito no Capitulo 3. No caso das redes neurais, queremos analisar se isto é vantajoso, já que um número maior de dados de entrada pode favorecer o mapeamento mais fiel das estatísticas da série.

Por fim, seguindo o repertório adotado em (Siqueira, Attux e Lyra Filho 2010), retiramos mais uma vez a média dos conjuntos. O período de testes utilizado foi de 1967 a 1976, para as usinas de Furnas, Emborcação e Sobradinho.

A Tabela 5.30 sumariza os resultados obtidos para P=1. As formas de previsão estão descritas como:

Teste original – resultados que foram retirados das tabelas das Seções anteriores e que foram previstas conforme descrito na Seção 5.3;

Recursiva Mensal – forma de previsão semelhante à anterior, mas com a retirada da média dos conjuntos;

Direta mensal – neste caso, há 12 preditores, mas para P>1 a forma de previsão é direta;

Recursiva - Série Completa – um único modelo de previsão para toda a série, e horizontes maiores que 1 passo à frente são previstos de forma recursiva;

Direta - Série Completa – um modelo de previsão para toda série e previsão direta para P>1.

A análise de erros um passo à frente permite algumas considerações interessantes. No caso da série de Furnas, o melhor resultado geral foi conseguido pela ELM no teste original, ou seja, sem a consideração de atrasos consecutivos. Todavia, embora não mostrado, apenas para os meses de outubro e dezembro, o atraso selecionado não foi o primeiro. Já o modelo linear atingiu resultados superiores em todas as novas abordagens propostas, com destaque para o modelo AR recursivo.

Para usina de Emborcação, todos os resultados foram melhorados com as formas de previsão alternativas, sobretudo os modelos mensais diretos. Nesta configuração, a ELM alcançou o menor MSE geral. A observação é que o MSE dessazonalizado de menor magnitude foi da ELM no teste original.

Por outro lado, os resultados da série de Sobradinho foram o oposto do verificado no caso anterior: os resultados dos modelos se degradam com a mudança na forma de prever. É sempre importante lembrar que esta série é a de maior amplitude média e, por isso, muitas vezes seu comportamento é mais difícil de ser previsto.

Tabela 5.32 – Previsões para P=1

		Modelo	MSE	MAE	MSE dessaz.	MAE dessaz.
Furnas	**Teste original**	**PAR**	7.8342e+04	190.5114	0.4143	0.5070
		ELM	**6.9085e+04**	**176.3235**	**0.3644**	**0.4702**
	Recursiva Mensal	**PAR**	7.2232e+04	184.3359	0.4136	0.5004
		ELM	7.3072e+04	180.1248	0.3998	0.4832
	Direta mensal	**PAR**	7.2406e+04	183.9834	0.4132	0.4996
		ELM	7.1217e+04	177.6635	0.3950	0.4787
	Recursiva - Série Completa	**AR**	6.9217e+04	176.8058	0.4069	0.4891
		ELM	6.9783e+04	182.4239	0.4152	0.5013
	Direta - Série Completa	**AR**	6.9217e+04	176.8058	0.4069	0.4891
		ELM	6.9652e+04	182.1650	0.4136	0.5006
Emborcação	**Teste original**	**PAR**	3.1025e+04	108.8272	0.4808	0.4822
		ELM	3.0932e+04	111.9879	0.4507	0.4938
	Recursiva Mensal	**PAR**	3.0282e+04	110.6524	0.4956	0.4980
		ELM	2.7341e+04	102.3563	0.4818	0.4666
	Direta mensal	**PAR**	3.0282e+04	110.6524	0.4956	0.4980
		ELM	**2.6887e+04**	**101.6617**	**0.4846**	**0.4677**
	Recursiva - Série Completa	**AR**	3.0751e+04	110.4946	0.5326	0.5245
		ELM	2.9208e+04	108.0458	0.4985	0.5093
	Direta - Série Completa	**AR**	3.0751e+04	110.4946	0.5326	0.5245
		ELM	2.8572e+04	107.1274	0.4939	0.5080
Sobradinho	**Teste original**	**PAR**	6.7514e+05	510.5354	0.3857	0.4339
		ELM	**6.5546e+05**	**500.6412**	**0.3466**	**0.4177**
	Recursiva Mensal	**PAR**	6.8208e+05	524.4519	0.4132	0.4565
		ELM	6.6551e+05	500.5678	0.4207	0.4428
	Direta mensal	**PAR**	6.9431e+05	528.7092	0.4173	0.4595
		ELM	6.6148e+05	496.3488	0.4193	0.4399
	Recursiva - Série Completa	**AR**	6.8732e+05	508.9615	0.4113	0.4458
		ELM	6.9333e+05	521.4091	0.4219	0.4575
	Direta - Série Completa	**AR**	6.8732e+05	508.9615	0.4113	0.4458
		ELM	6.9805e+05	522.4694	0.4221	0.4568

Na Tabela 5.33, estão os resultados para $P = 3$. A previsão com o horizonte de 3 passos à frente apresentou resultados distintos para a série de Furnas. Neste caso, a forma de previsão recursiva mensal com a retirada da média melhorou substancialmente o resultado do modelo PAR. Entretanto, a ELM reduziu o seu MSE de previsão com a forma mensal direta, com qual chegou ao menor valor de MSE real para as redes neurais.

Tabela 5.33 - Previsões para *P*=3

		Modelo	MSE	MAE	MSE dessaz.	MAE dessaz.
Furnas	**Teste original**	**PAR**	1.2245e+05	233.4257	0.6716	0.6254
		ELM	1.1684e+05	228.1373	0.6443	0.6081
	Recursiva Mensal	**PAR**	1.0652e+05	225.0681	0.6684	0.6207
		ELM	1.2044e+05	233.1324	0.7549	0.6510
	Direta mensal	**PAR**	1.1617e+05	232.0730	0.6957	0.6337
		ELM	**1.1311e+05**	**225.4016**	**0.6463**	**0.6196**
	Recursiva - Série Completa	**AR**	1.1822e+05	228.8348	0.7007	0.6314
		ELM	1.2587e+05	231.7942	0.7045	0.6291
	Direta - Série Completa	**AR**	1.2754e+05	229.9444	0.6936	0.6162
		ELM	1.2203e+05	228.8550	0.6897	0.6201
Emborcação	**Teste original**	**PAR**	5.1556e+04	144.4390	0.7169	0.6297
		ELM	**4.6045e+04**	**140.0540**	**0.6911**	**0.6164**
	Recursiva Mensal	**PAR**	5.1532e+04	142.7482	0.7888	0.6412
		ELM	4.7724e+04	139.4684	0.7921	0.6360
	Direta mensal	**PAR**	5.5093e+04	147.5985	0.8587	0.6704
		ELM	4.6434e+04	135.5318	0.7619	0.6144
	Recursiva - Série Completa	**AR**	5.3520e+04	153.1248	0.9131	0.7330
		ELM	5.3372e+04	152.8435	0.9162	0.7331
	Direta - Série Completa	**AR**	5.9410e+04	157.5318	0.9413	0.7292
		ELM	5.6122e+04	153.8629	0.9351	0.7223
Sobradinho	**Teste original**	**PAR**	1.3102e+06	731.1136	0.6125	0.5854
		ELM	**8.0781e+05**	**584.3865**	**0.4416**	**0.4976**
	Recursiva Mensal	**PAR**	1.1710e+06	710.4393	0.6645	0.6106
		ELM	1.1885e+06	694.5674	0.6988	0.6066
	Direta mensal	**PAR**	1.2139e+06	721.0104	0.6775	0.5998
		ELM	1.1087e+06	681.3580	0.6687	0.5849
	Recursiva - Série Completa	**AR**	1.4677e+06	729.7287	0.7052	0.5964
		ELM	1.7171e+06	766.4611	0.8063	0.6260
	Direta - Série Completa	**AR**	1.5503e+06	739.2503	0.7305	0.6009
		ELM	1.4942e+06	785.1372	0.7755	0.6523

As séries de Emborcação e Sobradinho continuaram com os melhores resultados com base no teste original. A mudança na forma de previsão não surtiu melhorias em nenhum momento.

Na tabela seguinte estão as previsões para *P*=6

Tabela 5.34 - Previsões para *P*=6

		Modelo	MSE	MAE	MSE dessaz.	MAE dessaz.
Furnas	**Teste original**	**PAR**	1.3785e+05	257.2921	0.9549	0.7241
		ELM	1.3330e+05	255.0956	0.8481	0.7245
	Recursiva Mensal	**PAR**	1.2482e+05	246.1398	0.9665	0.7251
		ELM	1.3249e+05	258.2777	1.0509	0.7777
	Direta mensal	**PAR**	1.2690e+05	250.1634	0.9178	0.7208
		ELM	1.2648e+05	248.6957	0.9059	0.7221
	Recursiva - Série Completa	**AR**	**1.2239e+05**	**239.2384**	**0.8640**	**0.7000**
		ELM	1.2754e+05	249.1068	0.8865	0.7304
	Direta - Série Completa	**AR**	1.3205e+05	245.1501	0.9048	0.7040
		ELM	1.3005e+05	245.5125	0.8896	0.7123
Emborcação	**Teste original**	**PAR**	5.1922e+04	155.2827	0.9363	0.7623
		ELM	**4.4753e+04**	**143.4166**	**0.8095**	**0.6967**
	Recursiva Mensal	**PAR**	5.4070e+04	157.7645	1.0313	0.7937
		ELM	5.0509e+04	152.1237	1.0234	0.7759
	Direta mensal	**PAR**	5.6272e+04	157.9445	0.9947	0.7750
		ELM	4.9996e+04	152.6435	0.9496	0.7556
	Recursiva - Série Completa	**AR**	4.8736e+04	152.0719	0.9654	0.7809
		ELM	4.7554e+04	151.7085	0.9782	0.7942
	Direta - Série Completa	**AR**	5.5508e+04	157.9448	1.0026	0.7855
		ELM	5.7312e+04	160.7182	1.0173	0.7909
Sobradinho	**Teste original**	**PAR**	1.2042e+06	734.4576	0.7170	0.6439
		ELM	**9.1077e+05**	**621.3228**	**0.5169**	**0.5462**
	Recursiva Mensal	**PAR**	1.1379e+06	761.4634	0.9183	0.7430
		ELM	9.9495e+05	704.9079	0.8467	0.7192
	Direta mensal	**PAR**	1.1461e+06	723.9106	0.6953	0.6302
		ELM	1.0428e+06	669.3968	0.6492	0.6492
	Recursiva - Série Completa	**AR**	1.0748e+06	695.4799	0.6263	0.6084
		ELM	1.0994e+06	700.8808	0.7029	0.6328
	Direta - Série Completa	**AR**	1.1726e+06	726.6751	0.6887	0.6309
		ELM	1.1880e+06	728.3722	0.7152	0.6387

Na previsão do horizonte *P*=6, vemos que a série de Furnas teve menores erros para o período de testes para todos os casos diferentes do teste original. Enquanto o modelo PAR chegou ao seu mínimo MSE pela forma direta, a ELM alcançou o menor de todos os resultados para a série completa recursiva.

O posto de Emborcação, novamente, teve o menor de todos os valores de MSE no teste original para a rede ELM. Por outro lado, o modelo AR recursivo apresentou erro menor que o teste original, para o caso linear. Isto também foi observado para a usina de

Sobradinho. Aliás, nesta última, todas as novas propostas lineares foram superiores ao modelo PAR original no espaço real.

Por fim, os resultados para *P*=12 são mostrados na Tabela 5.35.

Tabela 5.35 - Previsões para *P*=12

		Modelo	MSE	MAE	MSE dessaz.	MAE dessaz.
Furnas	**Teste original**	**PAR**	1.3821e+05	266.1884	1.0348	0.7959
		ELM	1.3147e+05	257.7447	0.8806	0.7464
	Recursiva Mensal	**PAR**	1.3489e+05	268.2572	1.2270	0.8456
		ELM	1.3931e+05	267.6506	1.1703	0.8224
	Direta mensal	**PAR**	1.4278e+05	262.0720	0.9692	0.7549
		ELM	**1.1968e+05**	**241.9324**	**0.8823**	**0.7284**
	Recursiva - Série Completa	**AR**	1.2059e+05	239.5543	0.8713	0.7088
		ELM	1.2878e+05	253.2022	0.9037	0.7478
	Direta - Série Completa	**AR**	1.2570e+05	245.4769	0.9143	0.7254
		ELM	1.2957e+05	251.4787	0.9489	0.7488
Emborcação	**Teste original**	**PAR**	5.3652e+04	165.1184	1.1667	0.8770
		ELM	4.8275e+04	155.9140	1.0064	0.8185
	Recursiva Mensal	**PAR**	5.2515e+04	167.4596	1.4273	0.9576
		ELM	4.8801e+04	159.1008	1.2701	0.8908
	Direta mensal	**PAR**	6.4933e+04	167.0575	1.2901	0.8832
		ELM	**4.4098e+04**	**141.4417**	**0.9412**	**0.7585**
	Recursiva - Série Completa	**AR**	4.6749e+04	151.0214	0.9811	0.7918
		ELM	4.7111e+04	151.2434	0.9955	0.7972
	Direta - Série Completa	**AR**	4.9917e+04	150.3794	0.9873	0.7836
		ELM	5.1469e+04	152.2029	1.0215	0.7996
Sobradinho	**Teste original**	**PAR**	1.1718e+06	742.0075	0.6817	0.6811
		ELM	**9.1404e+05**	**617.2839**	**0.5312**	**0.5688**
	Recursiva Mensal	**PAR**	1.1463e+06	769.6689	1.1206	0.8252
		ELM	9.9726e+05	694.9405	0.8144	0.7146
	Direta mensal	**PAR**	1.4407e+06	826.2217	0.7629	0.7287
		ELM	9.7293e+05	665.2051	0.6461	0.6405
	Recursiva - Série Completa	**AR**	9.9529e+05	660.3355	0.5899	0.5942
		ELM	9.6322e+05	646.0442	0.6203	0.6021
	Direta - Série Completa	**AR**	1.0362e+06	661.0337	0.6179	0.6068
		ELM	1.0388e+06	678.9325	0.6167	0.6339

Por último, na previsão 12 passos adiante, o modelo AR recursivo foi aquele que conseguiu menor MSE real para os modelos lineares em todas as usinas estudadas. Já a ELM em Furnas e Emborcação foi de menor erro a direta mensal. Em Sobradinho, isto foi verificado no teste original.

O gráfico em barras dos melhores resultados gerais é mostrado na Figura 5.40.

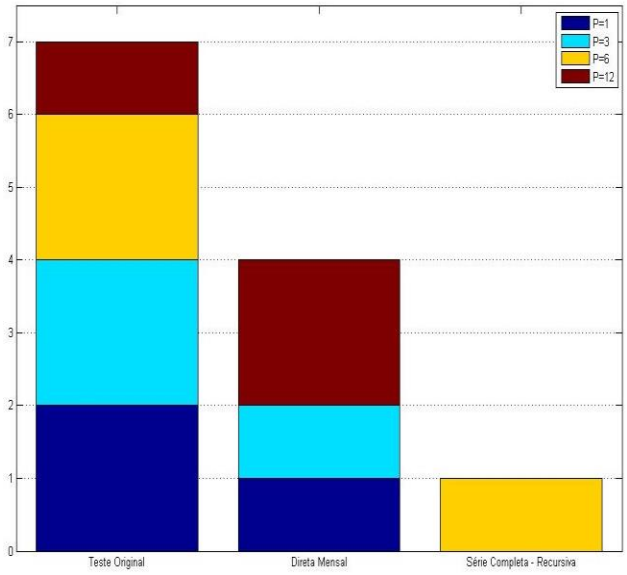


Figura 5.40 – Melhores resultados das formas de Previsão

Como é perceptível, o teste original foi o único que teve o menor MSE real nos quatro horizontes, o que pode indicar que o formato com entradas não consecutivas, recursivo, mensal e sem a subtração da média é o mais adequado.

De forma similar, comparamos apenas as versões alternativas, retirando o teste original. Esta forma de prever tem relação direta com a oficialmente aceita pelo setor elétrico, já que os atrasos são sempre consecutivos e de modo que todas as formas de previsão estão sujeitas as mesmas condições. Assim sendo, para cada valor de *P,* os melhores resultados foram:

P=1:

Direta mensal – 2, Recursiva Mensal – 0, Direta - Série Completa – ½, Recursiva - Série Completa – 1/2

P=3:

Direta mensal – 2, Recursiva Mensal – 1, Direta - Série Completa – 0, Recursiva - Série Completa – 0

P=6:

Direta mensal – 0, Recursiva Mensal – 1, Direta - Série Completa – 0, Recursiva - Série Completa – 2

P=12:

Direta mensal – 1, Recursiva Mensal – 0, Direta - Série Completa – 1, Recursiva - Série Completa – 1

TOTAL:

Direta mensal – 5, Recursiva Mensal – 2, Direta - Série Completa – 1+1/2, Recursiva - Série Completa – 3+1/2.

Onde há o valor ½ estamos nos referindo ao caso em que houve empate, pois para 1 passo à frente, o modelo PAR apresenta mesmo valor tanto para o método recursivo como para o direto. A Figura 5.41 mostra os resultados finais separados por horizonte e pelo número total de vezes que cada uma chegou ao menor MSE.

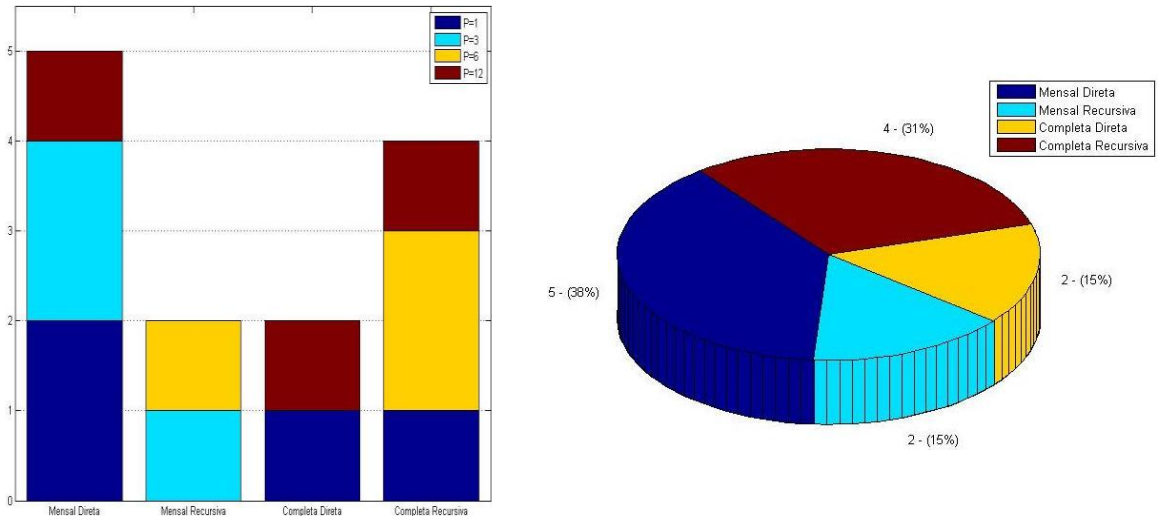


Figura 5.41 – Melhores resultados formas alternativas de previsão

O que se nota pelos gráficos é que nenhuma forma de previsão consegue ser, pelo menos uma vez, a de menor MSE geral para todos os horizontes. Podemos organizar grupos diferentes e compará-los, mas a conclusão permanece a mesma, pois há quase um empate em todos os casos, seja PREVISÃO MENSAL X SÉRIE COMPLETA, ou DIRETA X RECURSIVA. Dessa forma, não é possível, com este número reduzido de testes, ser categoricamente conclusivo sobre qual a forma de se prever vazões mais adequada com as abordagens estudadas. Todavia, estes resultados preliminares abrem uma nova possibilidade de investigação, sobretudo porque aqui não foram avaliados modelos recursivos como as ESNs.

Comentários

O presente Capítulo apresentou diversos estudos de casos para previsão de séries de vazões médias mensais com as metodologias descritas nos capítulos anteriores: modelos lineares PAR, redes neurais MLP e máquinas desorganizadas - máquinas de aprendizado extremo (ELM) e redes neurais de estados de eco (ESN) com reservatórios de Jaeger e de Ozturk et al., e camadas de saída com uma ELM, conforme proposta de Butcher et al. (2010) ou um filtro de Volterra precedido da técnica de PCA, como sugere Boccato et al. (2011).

Foram utilizadas as séries das usinas hidrelétricas de Furnas, Emborcação e Sobradinho, com períodos de testes selecionados entre os anos de 1951 e 1960 (seco), 1967 e 1976 (mediano) e 1977 e 1986 (úmido). A forma de previsão adotada foi limitar a 6 atrasos o número de entradas dos modelos e 12 modelos diferentes para cada série, sendo um para cada mês. Além disso, para horizontes maiores do que um, foi utilizada a previsão do tipo recursiva.

A forma de seleção de entradas para o modelo linear foi o filtro de autocorrelação parcial com atrasos consecutivos, conforme proposta de Stedinger (2001), e para as redes neurais o *wrapper* com função avaliadora via análise de erro quadrático médio.

Os resultados finais foram favoráveis às máquinas desorganizadas, sobretudo às diversas arquiteturas de ESNs. Na comparação entre os modelos *feedforward,* a ELM foi superior à rede MLP na grande maioria dos resultados. De forma similar, a camada de saída tipo filtro de Volterra + PCA alcançou melhores resultados em 75% dos casos, se comparados à rede com ELM como camada de saída.

Por outro lado, vê-se que, nem sempre, o preditor que chegou ao menor MSE real foi também o mais adequado para o domínio dessazonalizado. O mesmo vale para a comparação com o erro absoluto médio.

O teste estatístico de Friedman foi aplicado aos resultados e verificou-se que estes são significativamente diferentes.

Além disso, outras execuções foram realizadas para a usina de Passo Real e, comparando os resultados obtidos, foi possível notar que as máquinas desorganizadas foram superiores aos modelos oficiais PAR.

Por fim, algumas formas alternativas de previsão foram testadas de posse do modelo linear PAR e da ELM, como utilizar apenas um modelo para cada série ou a previsão direta para $P>1$. A comparação com a primeira forma adotada levou a conclusão de que a primeira é mais adequada, segundo mostrou a maior parte dos resultados obtidos. Todavia, caso sejam consideradas as demais maneiras de se prever as séries, com atrasos consecutivos, não é possível ser categoricamente conclusivo, embora seja verificado que a maior parte dos casos tenda à forma mensal direta.

Terminado este Capítulo, passemos agora às conclusões do trabalho.

Capítulo 6. Conclusão

No Brasil, cerca de 77% da matriz de geração de energia elétrica é composta por usinas hidrelétricas. Por conta disso, muitos estudos têm sido realizados ao longo das últimas três décadas, com o intuito de desenvolver técnicas de previsão das vazões afluentes aos postos geradores do País uma vez que todo planejamento da operação e expansão do sistema elétrico depende de tais previsões.

Os modelos de simulação mais comumente empregados pelo Setor Elétrico Brasileiro são baseados na metodologia linear de Box & Jenkins (Souza, Marcato, et al. 2010). Entretanto, modelos não-lineares como redes neurais artificiais podem ser candidatos para a solução do problema de previsão. Modelos de redes neurais artificiais (RNA) são metodologias com capacidades de aprendizado, de generalização, além de serem aproximadores universais. São capazes de reproduzir qualquer mapeamento contínuo, diferenciável e restrito a uma região compacta, com uma precisão arbitrária. Potencialmente, o uso desta rede pode reduzir os erros e assim elevar a confiabilidade das previsões.

Este trabalho investigou a aplicação das redes neurais do tipo *redes de estado de eco* e *máquinas de aprendizado extremo* na para previsão de vazões médias mensais de importantes usinas hidrelétricas brasileiras. Foi adotado o epíteto *máquinas desorganizadas* para agrupar estas redes, conforme proposta de Boccato et al. (2011b).

Em 2001, Hebert Jaeger iniciou uma nova área de investigação das arquiteturas de redes neurais chamada de *computação de reservatório,* com a proposta pioneira das redes de estado de eco (ESN). Estas são recorrentes e possuem como característica marcante o fato dos pesos da sua camada intermediária serem gerados de forma aleatória, respeitando a chamada *propriedade de estados de eco*. O processo de treinamento, então, limita-se a encontrar os coeficientes de um combinador linear.

A camada oculta de uma ESN é denominada reservatório de dinâmicas, que pode ser gerado de formas variadas. Neste trabalho, uma das alternativas investigadas foi a proposta inicial de Jaeger, na qual é gerada um matriz de pesos esparsa com o intuito de favorecer o desacoplamento de grupos de neurônios artificiais, induzindo o desenvolvimento de dinâmicas individuais pouco relacionadas.

Outra forma de interesse foi desenvolvida por Ozturk et al. (2007), que sugerem um reservatório rico do ponto de vista da entropia dos estados de eco. Para isso, os autovalores da matriz de pesos respeitam uma distribuição uniforme no círculo unitário.

Em 2004, Huang et al. propuseram as Máquinas de Aprendizado Extremo (ELM), arquitetura *feedforward* que apresenta semelhanças com uma rede MLP, mas assim como as ESNs, não precisam de treinamento da camada intermediária.

Outras arquiteturas vêm sendo propostas na tentativa de aumentar o poder de mapeamento não-linear da camada de saída de uma ESN, em substituição ao combinador linear inicialmente empregado. Este trabalho investigou as propostas de Butcher et al., empregaram uma ELM como camada de saída de uma ESN, e de Boccato et al., que utilizaram um filtro de Volterra precedido de um compressor de dados do tipo análise de componentes principais (PCA).

Um estudo comparativo foi realizado entre os desempenhos do modelo periódico auto-regressivo (PAR), da rede neural MLP, das ESNs e das ELMs. No caso das ESNs, foram abordadas as duas propostas citadas de projeto de reservatório de dinâmicas e outras as ouras duas alternativas de camadas de saída não-lineares mencionadas no parágrafo anterior.

Em complemento à análise dos preditores, foi ainda investigada a aplicação de modelos de seleção de variáveis, com intuito de aumentar o poder de aproximação das redes. Os métodos abordados foram: os filtros tipo função de autocorrelação parcial (FACP) - na versão clássica e a proposta de Stedinger (2001) - e critérios de informação mútua. O trabalho investigou também o *wrapper,* com funções de avaliação baseadas nos critérios BIC (Critério de Informação Bayesiano), AIC (Critério de Informação de Akaike) e mínimo erro quadrático médio (MSE).

Séries de vazões mensais possuem uma componente sazonal que faz com que o volume de água afluente seja variável no decorrer do ano, a depender da densidade pluviométrica nas proximidades dos rios. Dessa maneira, para que os modelos de previsão possam ser melhor utilizados, esta componente precisa ser subtraída da série, deixando-a aproximadamente estacionária. Esta é uma premissa necessária para aplicação de modelos PAR, mas mesmo redes neurais não alcançam bons desempenhos sem este procedimento.

Para extração da componente sazonal os modelos de médias móveis, de médias móveis sazonais e de padronização foram analisados.

Os resultados obtidos nas simulações apontaram que o modelo linear tem entradas melhor escolhidas pela filtro baseado na FACP, enquanto as redes neurais pelo *wrapper* avaliado pelo menor erro quadrático médio (MSE). Da mesma forma, a dessazonalização foi melhor realizada pela padronização.

As séries históricas utilizadas foram as das usinas hidrelétricas de Furnas, Emborcação e Sobradinho. Os períodos de testes selecionados para todas elas foram de 1951 a 1960 (seco), de 1967 a 1976 (mediano) e de 1977 a 1986 (úmido). Ademais, os horizontes de previsão verificados foram 1, 3, 6 e 12 passos à frente, previstos de forma recursiva. O número máximo de atrasos permitido foi 6. Ao final, o teste de Friedman avalizou a diferença significativa dos resultados.

Os resultados computacionais permitiram afirmar que não há correspondência direta entre os preditores que conseguem chegar ao menor erro no domínio real e no dessazonalizado, assim como na comparação entre o erro quadrático médio e erro absoluto médio.

Os resultados mostraram também que as redes neurais superaram em praticamente todos os cenários o desempenho do modelo PAR. No caso geral, a proposta de ESN de Jaeger com o combinador linear foi a que apresentou melhores resultados, seguida da mesma ESN com o filtro de Volterra como camada de saída. Ou seja, as máquinas desorganizadas são alternativas viáveis e de baixo custo computacional para a abordagem do problema de previsão.

De posse destes resultados, outros estudos foram realizados. O primeiro foi comparar os modelos de rede de estado de eco com a camada de saída proposta por Butcher et al. e aquela sugerida por Boccato et al. Em 75% dos casos, esta última apresentou melhores desempenhos. Após, foram examinadas as propostas de redes *feedforward*: ELMs e a MLP. Os resultados mostraram que as ELMs realizam previsões de melhor qualidade.

Todas as alternativas investigadas mostraram-se parcimoniosas quanto ao número de entradas selecionadas, já que 86% dos melhores resultados foram obtidos com até duas entradas.

Sobre a disposição temporal dos atrasos selecionados, entre os melhores resultados (quase todos com redes neurais), o que se percebe é que a previsão 1 passo à frente é altamente dependente do primeiro atraso, mas à medida que o horizonte se eleva, este comportamento não se verifica necessariamente. A razão disso está na dependência temporal do mês alvo que se pretende prever, a qual nem sempre está ligada ao primeiro atraso.

Avaliou-se também, de forma comparativa, o modelo PAR, a ELM e a ESN para previsão da usina de Passo Real, localizada na região Sul do Brasil. Esta usina, possui vazão afluente média inferior as outras aqui estudadas. Por isso, seu regime hidrológico é mais susceptível a variações pluviométricas, o que a torna um cenário distinto das demais. As conclusões foram semelhantes, favoráveis às máquinas desorganizadas, sobretudo a ELM.

Por último, de posse do modelo PAR e da ELM, foram analisadas formas de previsão diferentes daquela previamente adotada (máximo de 6 atrasos e previsão recursiva), para verificar se poderiam trazer algum ganho de desempenho. Para isso, novos ensaios foram feitos com os modelos preditores com as seguintes especificações:

a) limite de 6 atrasos, mas respeitando a proposta de Stedinger (2001) de que eles deveriam ser consecutivos;

b) adoção de um único modelo para previsão de toda série;

c) formas de previsão direta e recursiva, para previsão com horizontes maiores que 1 passo à frente.

Dessa maneira, tem-se quatro formas de previsão distintas:

i) Mensal direta – 12 modelos, sendo um para prever cada mês do ano, e previsão direta para horizontes mais longos que P=1 passo à frente;
ii) Mensal recursiva – novamente 12 modelos mensais, mas previsão recursiva para P>1;

iii) Série completa direta – apenas um modelo para prever toda a série e previsão direta;
iv) Série completa recursiva – modelo preditor único com previsão recursiva para P>1.

Os resultados apontaram que a forma inicialmente adotada neste trabalho alcançou melhores resultados, na maior parte dos casos. Todavia, investigando-se as outras possibilidades em separado: mensal recursiva, mensal direta, série completa recursiva e série completa direta, não é possível chegar a uma conclusão clara, pois cada uma destas propostas apresenta bons resultados, a depender do número de passos à frente e da série.

6.1 Perspectivas Futuras

Os bons resultados obtidos com a aplicação de máquinas desorganizadas na previsão de vazões médias mensais podem ser estendidos para séries de vazões de natureza distinta, como as séries de vazões semanais e diárias, fundamentais para o planejamento energético de curto prazo.

Por outro lado, novas propostas de reservatório de dinâmicas estão sendo desenvolvidas, como a de Boccato et al. (2013) e podem ser testadas e comparadas com os resultados aqui apresentados.

Redes não-recorrentes, como discutido, são aproximadores universais, de forma que sua capacidade de generalização deve ser maximizada tanto quanto possível. Para modelos que tratam a tarefa de previsão como um mapeamento não-linear estático, como MLPs e ELMs, uma possibilidade é aplicar o método *k-fold* de validação cruzada, no qual os dados reservados aos conjuntos de treinamento e validação são divididos em *k* partições menores, sendo o processo de validação repetido para cada uma destas partes, enquanto os demais dados são utilizados para treinamento. Isto evita a polarização do modelo, embora seja necessário mais esforço computacional.

Observou-se no trabalho uma degradação da resposta dos preditores ao reinserir-se a componente sazonal. Assim, outra investigação necessária é o desenvolvimento de técnicas de dessazonalização que permitam que os melhores preditores no espaço dessazonalizado seja também o de menor erro no espaço real. Inicialmente, modelos de dessazonalização como o X-12 ARIMA e de Holt-Winters devem ser investigados.

Bibliografia

Akaho, S. "Conditionally independent component analysis for supervised feature extraction,." *Neurocomputing*, 2002: 139–150.

Akaike, H. "A new look at the statistical model identification." *IEEE Transactions on Automatic Control*, 1978: 716–723.

Andrade, G. M., R. L. Reis, S. Soares, and D. Silva Filho. "Análise do Erro de Previsão de Vazões Mensais com Diferentes Horizontes de Previsão." *Controle e Automação*, 2012: 294-305.

Atiya, A. F., S. M. El-Shoura, S. I. Shaheen, and M. S. El-Sherif. "A comparison between neural-network forecasting techniques - case study: River flow forecasting." *IEEE Trans. on Neural Networks*, 1999: 402-409.

Attux, R. R. F. "Novos Paradigmas para Equalização e Identificação de canais Baseados em Estruturas Não-Lineares e Algoritmos Evolutivos." *Tese de Doutorado - FEEC - UNICAMP*. 2005.

Ballini, R. "Análise e Previsão de Vazões Utilizando Modelos de Série Temporais Redes Neurais e Redes Neurais Nebulosas." *Tese de Doutorado, FEEC-UNICAMP*. 2000.

Bartlett, P. L. " The Sample Complexity of Pattern Classification with Neural Networks: the Size of the Weights is More Important than The Size of the Network." *IEEE Trans. on Information Theory*, 1998: 525-536.

Bazaraa, M. S., H. D. Sherali, and C. M. Shetty. *Nonlinear Programming - Theory and Algorithms.* 3ª. Wiley, 2006.

Boccato, L. "Novas Propostas e Aplicações de Redes Neurais com Estados de Eco." *Tese de Doutorado - FEEC - UNICAMP*. 2013.

Boccato, L., A. Lopes, R. Attux, and F. J. Von Zuben. "An Echo State Network Architecture Based on Volterra Filtering and PCA with Application to the Channel Equalization Problem." *IEEE Proc. of Intern. Joint Conf. on Neural Networks*, 2011a: 580-587.

—. "An Extended Echo State Network Using Volterra Filtering and Principal Component Analisys." *Neural Networks*, 2012, 32 ed.: 292-302.

Boccato, L., E. S. Soares, M. M. L. P. Fernandes, D. C. Soriano, and R. Attux. "Unorganized Machines: from Turing's Ideas to Modern Connectionist Approaches." *International Journal of Natural Computing Research*, 2011b: 1-16.

Boccato, L., et al. "Error Entropy Criterion in Echo State Network Training." *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning.* Bruges - Bélgica, 2013. 35-40.

Bonnlander, B. V., and A. S. Weigend. "Selecting input variables using mutual information and nonparametric density estimation." *Proceedings of the 1994 Int. Symp. on Artificial Neural Networks (ISANN'94).* Tainan, Taiwan, 1994. 42–50.

Bouzada, M. A. C. "Aprendendo Decomposição Clássica: Tutorial para um Método de Análise de Séries Temporais." *Tecnologias de Administração e Contabilidade*, 2012: 1-18.

Bowden, G. J., H. R. Maier, and G. C. Dandy. "Input determination for neural network models in water resources applications. Part 1–background and methodology." *Journal of Hydrology*, 2005: 75–92.

Box, G., G. Jenkins, and G. C. Reinsel. *Time Series Analysis, Forecasting and Control.* 4ª. Wiley, 2008.

Braga, B. E. P. "Geração e Previsão de Vazões Através de Modelos ARMA e ARIMA." *Curso de Engenharia Hidrológica - Hidrologia Operacional*, 1983: Q.1-Q.59.

Butcher, J. B., D. Verstraeten, B. Schrauwen, C. R. Day, and P. W. Haycock. "Reservoir computing and extreme learning machines for non-linear time-series data analysis." *Neural Networks*, 2013: 76–89.

Butcher, J., D. Verstraeten, B. Schrauwen, C. Day, and P. Haycock. "Extending Reservoir Computing with Random Static Projections: a Hybrid Between Extreme Learning and RC." *Proc. of European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning*, 2010: 303-308.

Castro, L. N. "Análise e Síntese de Estratégias de Aprendizagem para Redes Neurais Artificiais." *Dissertação de Mestrado - FEEC - UNICAMP*. 1998.

—. *Fundamentals of Natural Computing: Basic Concepts, Algorithms and Applications.* Chapman & May, 2006.

Castro, L. N., and F. J. Von Zuben. "Optimized Trainig Techniques For Feedforward Neural Networks." *Technical Report - DCA-RT - FEEC-UNICAMP*. 1998.

CEPEL, Centro de Pesquisas Energéticas Estratégicas. "Modelo GEVAZP - Manual de Referência." 2001b.

CEPEL, Centro de Pesquisas Energéticas Estratégicas. "NEWAVE - Manual de referência." 2001a.

Chiang, Y. M., L. C. Chang, and F. J. Chang. "Comparison of static-feedforward and dynamic-feedback neural networks for rainfall–runoff modeling." *Journal of Hydrology*, 2004: 297-311.

Cover, T., and J. Thomas. *Elements of Information Theory.* John Wiley & Sons, 1991.

Cybenko, G. "Aproximation by Superpositions of a Sigmoidal Function, Mathematics and Control." *Signals, and Systems*, 1989.

Deihimi, A., and H. Showkati. "Application of echo state networks in short-term electric load forecasting." *Energy*, 2012: 327-340.

Deng, W., Q. Zheng, and L. Chen. " Regularized extreme learning machine." *Proc. IEEE Symp. CIDM.* 2009. 389–395.

Ehlers, R.S. *"Análise de Séries Temporais"*. Departamento de Estatística, UFPR - Disponível em: http://leg.est.ufpr.br/˜ehlers/notas, 2007.

Emiliano, P. A., E. P. Veiga, M. J. F. Vivanco, and F. S. Menezes. "Critérios de Informação de Akaike Versus Bayesiano: Análise Comparativa." *Simpósio Nacional de Probabilidade e Estatística - 19º SINAPE.* São Pedro-SP, 2010.

EPE, (Empresa de Pesquisa Energética). *Balanço energético nacional 2013 (ano base 2012).* Ministério de Minas e Energia, 2013.

Erdogmus, D., and J.C. Principe. "From linear adaptive filtering to nonlinear information processing - The design and analysis of information processing systems." *IEEE Signal Processing Magazine*, 2006: 14-33.

Feng, G., Z. Qian, and N. Dai. "Reversible water marking via extreme learning machine prediction." *Neurocomputing*, 2012: 62-68.

Ferreira, A. M. "Testes não-paramétricos." *Escola Superior Agrária Castelo - Disponível em http://docentes.esa.ipcb.pt/mede/apontamentos/testes_nao_parametricos.pdf.* 2010.

Ferreira, C. C. "Previsão de Vazões Naturais Diárias Afluentes ao Reservatório da UHE Tucuruí Utilizando Técnica de Redes Neurais Artificiais." *Dissertação de Mestrado - Escola de Engenharia Elétrica, Mecânica e de Computação - UFG.* Goiânia, Goiás, 2012.

Fortunato, L. A. M., T. A. A. Neto, J. C. R. Albuquerque, and C. Ferreira. *Introdução ao Planejamento da Expansão e Operação de Sistemas de Produção de Energia Elétrica.* Niteroi-RJ: Universitária, 1990.

Francelin, R., R. Ballini, and M. G. Andrade. "Back-propagation and Box & Jenkins approaches to streamflow forecasting." *Latin-Iberian-American Congress on Operations Research and Systems Engineering - CLAIO, Simpósio Brasileiro de Pesquisa Operacional - SBPO.* Rio de Janeiro, 1996. 1307-1312.

Friedman, M. "The use of ranks to avoid the assumption of normality implicit in the analysis of variance." *Journal of the American Statistical Association*, 1937: 675–701.

Guilhon, L. G. F. "Modelo Heurístico de Previsão de Vazões Naturais Médias Semanais Aplicado à Usina de Foz do Areia, ." *Tese de Doutorado, COPPE - UFRJ.* 2002.

Guilhon, L.G.F, V. F. Rocha, and J. C. Moreira. "Comparação de Métodos de Previsão de Vazões Naturais Afluentes a Aproveitamentos Hidroelétricos." *Revista Brasileira de Recursos Hídricos*, 2007: 13-20.

Guyon, I., and A. Elisseeff. "An introduction to variable and feature selection." *Journal of Machine Learning Research*, 2003: 1157–1182.

Feature Selection." *Journal of Machine Learning Research*, 2003: 1157-1182.

Haber, R., and H. Unbehauen. "Structure identification of nonlinear dynamic systems – a survey on input." *Automática*, 1990: 651-677.

Haykin, S. *Adaptive Filter Theory.* Prentice Hall, 1997.

—. *Neural Networks and Learning Machines.* 3rd. Prentice Hall, 2008.

Haykin, S., and B. Van Veen. *Sinais e Sistemas.* Bookman, 2001.

Hebb, D. O. *The organization of behavior.* Wiley & Sons, 1949.

Hestenes, M. *Conjugate Directions Methods in Optimization.* New York: Springer Verlag, 1980.

Hippel, K. W., and A. I. McLeod. *Time Series Modelling of Water Resources and Environmental Systems.* Amsterdã - Holanda: Elsevier Science B. V., 1994.

Hippert, H.S., C.E. Pedreira, and R.C. Souza. "Neural Networks for Short-Term. Load Forecasting: A Review and Evaluation." IEEE Transaction on Power Systems 2001: 44-55.

Hollander, M., and D.A. Wolfe. *Nonparametric Statistical Methods.* 2a. John Wiley & Sons, Inc., 1999.

Hopfield, J. J. "Neural networks and physical systems with emergent collective computational proprietes." *Proceedings of the National Academy of Sciences of the USA*, 1982: 2554-2558.

Huang, G.-B., Q.-Y. Zhu , and C.-K. Siew. "Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks." *Proc. of Intern. Joint Conf. on Neural Networks*, 2004: 985-990.

Huang, G.-B., H. Zhou, X. Ding, and R. Zhang. "Extreme Learning Machines for Regression and Multiclass Classification." *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 2012: 513-529.

Huang, G.-B., L. Chen, and C.-K. Siew. "Universal approximation using incremental constructive feedforward networks with random hidden nodes." *IEEE Trans. on Neural Networks*, 2006: 879- 892.

Huang, G.-B., Q.-Y. Zhu , and C.-K. Siew. "Extreme Learning Machine: Theory and Applications." *Neurocomputing*, 2006: 489-501.

Hyvärinen, A., J. Karhunen, and E. Oja. *Independent Component Analysis.* New York: John Wiley & Sons, 2001.

Jaeger, H. "The Echo State Approach to Analyzing and Training Recurrent Neural Networks." *Bremen: German National Research Center for Information Technology.* Vol. Tech. Rep. GMD Report 148. 2001.

Jaeger, H. "Short Term Memory in Echo State Networks." *Bremem: German National Research Center for Information Technology*, Tech. Rep. GMD Report 152, 2002a.

Jaeger, H. "Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the “echo state network” approach." Bremem: German National Research Center for Information Technology, Tech. Rep. GMD Report 159, 2002b.

Kentel, E. "Estimation of river flow by artificial neural networks and identification of input vectors susceptible to producing unreliable flow estimates." *Journal of Hydrology*, 2009: 481–488.

Kligerman, A. S. "Operação Ótima de Subsistemas Hidrotérmicos Interligados Utilziando Programação Dinâmica Dual Estocástica." *Dissertação de Mestrado, FEEC - UNICAMP.* 1992.

Kohavi, R., and G. John. "Wrappers for Feature Subset Selection." *Artificial Intelligence*, 1997: 273-324.

Kohzadi, N., M. S. Boyd, B. Kermanshahi, and I. Kaastra. "A comparison of artificial neural network and time series models for forecasting commodity prices." *Neurocomputing*, 1996: 169-181.

Kulaif, A. C. P., and F.J. Von Zuben. "Improved Regularization in Extreme Learning Machines." *1st BRICS Countries Congress (BRICS-CCI) and 11th Brazilian Congresson Computational Intelligence (CBIC).* 2013.

Lachtermacher, G., and J. D. Fuller. "Backpropagation in time-series forecasting." *Journal of Forecasting*, 1995: 381-393.

Lin, X., Z. Yang, and Y. Song. "Short-term stock price prediction based on echo state networks." *Expert Systems with Applications*, 2009: 7313-7317.

Luenberger, D. G. *Linear and nonlinear programming.* Springer, 2003.

Lukoševičius, M., and H. Jaeger. "Reservoir computing approaches to recurrent neural network training." *Computer Science Review*, 2009: 127-149.

Luna, I. "Análise de Séries Temporais e Modelagem Baseada em Regras Nebulosas." *Tese de Doutorado, FEEC-UNICAMP.* 2007.

Luna, I., and R. Ballini. "Top-Down Strategies Based on Adaptive Fuzzy Rule-Based Systems for Daily Time Series Forecasting." *Intern. Journal of Forecasting*, 2011: 708-724.

Luna, I., R Ballini, and S. Soares. "Técnica de Identificação de Modelos Lineares e Não Lineares de Séries Temporais." *Controle e Automação*, 2006: 245–256.

Maass, W., T. Natschlager, and H. Markram. "Real-time computing without stable states : A new framework for neural computation based on perturbations." *Neural Computation*, 2002: 2531-2560.

Maceira, M. E. P. "Operação Ótima de Reservatórios com Previsão de Afluências." *Dissertação de Mestrado - COPPE, UFRJ.* 1989.

Maceira, M. E. P., and J. M. Damázio. "The use of PAR (p) model in the stochastic dual dynamic programming optimization scheme used in the operation planning of the Brazilian hydropower system." *8th International Conference on Probabilistic Methods Applied to Power Systems.* Iowa State University, Ames, Iowa., 2004.

Maceira, M. E. P., F. S. Costa, J. M. Damázio, M. Denício, and L. G. Guilhon. " Modelo Estocástico de Previsão de Vazões Mensais - PREVIVAZM." *Anais do XV Simpósio Brasileiro de Recursos Hídricos.* Curitiba - PR, 2003.

Maceira, M. E. P., L. A. Terry, F. S. Costa, J. M. G. Damázio, and A. C. Melo. "Chain of Optimization Models for Setting the Energy DIspach and Price in the Brazilian System." *14th Power Systems Computation Conference (PSCC).* Sevilha - Espanha, 2002.

Magalhães, M. H. "Redes Neurais, Metodologias de Agrupamento e Combinação de Previsores Aplicados à Previsão de Vazões Naturais." *Dissertação de Mestrado - FEEC - UNICAMP.* 2004.

Marcato, A. L. M. "Representação Hídrica de Sistemas Equivalentes e Individualizados para o Planejamento da Operação de Médio Prazo de Sistemas de Potência de Grande Porte." *Tese de Doutorado - PUC-RJ.* Rio de Janeiro, RJ, 2002.

Marinho. "Previsão de Vazões Afluentes Vários Passos à Frente Via Agregação de Vazões para o Planejamento Energético da Operação de Sistemas Hidrotérmicos de Potência." *Tese de Doutorado - FEEC - UNICAMP.* 2005.

Mason, J. C., A. Teme'me, and R. K Price. "A neural network model of rainfall runoff using radial basis functions." *Journal of Hydraulic Research*, 1996: 537–548.

Mathews, V. L., and G. L. Sicuranza. *Polynomial Signal Processing.* New York: John Wiley & Sons, 2001.

McCulloch, W., and W. Pitts. "A Logical Calculus of Ideas Immanent in Nervous Activity." *Bulletin of Mathematical Biophysics*, 1943: 115–133.

McLeod, A. I. "Diagnostic Checking of Periodic Autoregression Models with Applications." *Journal of Time Series Analysis*, 1994: 221-233.

Miche, Y., A. Sorjamaa, P. Bas, O. Simula, C. Jutten, and Lendasse A. "OP-ELM: Optimally pruned extreme learning machine." *IEEE Transactions on Neural Networks*, 2010: 158–162.

Minsky, M. L., and S. A. Papert. *Perceptrons.* MIT Press, 1969.

Moller, M.F. "A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning." *Neural Networks*, 1990: 525-533.

Moon, Y. I., B. Rajagopalan, and U. Lall. "Estimation of mutual information using kernel density estimators." *Physical Review E*, 1995: 2318–2321.

Morettin, P. A. *Séries Temporais.* São Paulo: Atual, 1986.

Morettin, P. A., and C. M. C. Toloi. *Análise de Séries Temporais.* São Paulo: Egard Blucher, 2006.

—. *Previsão de Séries Temporais.* São Paulo: Atual, 1987.

Muller, I. I., C. M. Kruger, and E. Kaviski. "Análise da Estacionariedade de Séries Hidrológicas na Bacia Incremental de Itaipu." *Revista Brasileira de Recursos Hídricos*, 1998: 51-71.

Nelson, M., T. Hill, T. Remus, and M. O'Connor. "Time series forecasting using NNs: Should the data be deseasonalized." *Journal of Forecasting, 1999, 18, .*, 1999: 359–367.

Oliveira, F. L. C., and R. C. Souza. " A new Approach to Identify the Structural Order of PAR (p) Models." *Pesquisa Operaciona*, 2011: 487-498.

ONS, Operador Nacional do Sistema Elétrico. *http://www.ons.org.br/operacao/vazoes_naturais.aspx.* 2012.

ONS, Operador Nacional do Sistema Elétrico. "Previsão de Vazões Diárias no Reservatório de Três Maras Usando Técnicas de redes Neurais." 2009.

Oppenheim, A. V., A. S. Willsky, and S. H. Nawab. *Signals and Systems.* 2a. Prentice Hall, 1997.

Ozturk, M. C., D. Xu, and J. C. Principe. "Analysis and Design of Echo State Networks." *Neural Computation*, 2007: 111–138.

Pearlmutter, B.A. "Fast Exact Calculation by the Hessian." *Neural Computation*, 1994: 147-160.

Pereira, M. V. F. "Optimal Stochastic Operations Scheduling of Large Hydroelectric Systems." *Electrical Power & Energy Systems*, 1989: 161-169.

—. "Optimal Scheduling of Hydrothermal Systems - An Overview." *IFAC Symposium on Planning and Operation of Electric Energy Systems.* Rio de Janeiro, Brasil, 1985.

Pereira, M. V. F., G. C. Oliveira, C. C. G. Costa, and J. Kelman. "Stocastic Streamflow Models for Hudroelectric Systems." *Water Resources Research*, 1984: 379-390.

Quenouille, M. H. "Aproximate Testes of Correlation in Time Series." *Journal Royal Statistical Society*, 1949: 64-68.

Rasmussen, R. F., J. D. Salas, L. Fagherazzi, J. C. Rassam, and R. Bobee. "Estimation and validation of contemporaneous PARMA models for streaflow simulation." *Water Resource Research*, 1996: 3151-3160.

Rosenblatt, F. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain." *Psychological Review*, 1958: 386–408.

Rumelhart, D., G. Hinton, and R. Williams. "Learning Representations by Backpropagation Errors." *Nature*, 1986: 533–536.

Sacchi, R. "Política de operação preditiva estabilizada via termo inercial utilizando "analytic signal", "dynamic modelling" e sistemas inteligentes na previsão de vazões afluentes em sistemas hidrotérmicos de potência." *Tese de Doutorado - USP.* São Carlos, 2009.

Sacchi, R., M. C. Ozturk, J. C. Príncipe , and A. A. F. M. Carneiro. "Water Inflow Forecasting Using the Echo State Network: a Brazilian Case Study." *IEEE Proc. of Intern. Joint Conf. on Neural Networks*, 2007: 2403-2408.

Santos, E.P. Dos, and F.J. Von Zuben. "Improved Second-Order Training Algorithms for Globally and Partially Recurrent Neural Networks." *IEEE Proc. of Intern. Joint Conf. on Neural Networks*, 1999: 1501–1506.

Santurio, D. S., and M. H. R. Gomes. "Estudo comparativo entre o método bayesiano e o método clássico para modelação de séries temporais hidrológicas." *XIX SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS.* Maceió, AL., 2011.

Schafer, A. M., and H. G. Zimmermann. "Recurrent Neural Networks are Universal Approximators." *Int. J. of Neur.Syst*, 2007: 253-263.

Schwarz, G. E. "Estimating the dimension of a model." *Annals of Statistics*, 1978: 461–464.

Sharda, R., and R. B. Patil. "Conectionist approach to time series prediction: An empirical test." *Journal of Intelligent Manufacturiong*, 1992: 317-23.

Sharma, A. "Seasonal to internannual rainfall probabilisticforecasts for improvedwater supply management: Part 1 – A strategy for system predictor identification." *Journal of Hydrology*, 2000: 232–239.

Sheng, C., J. Zhao, Y Liu, and W. Wang. "Prediction for noisy nonlinear time series by echo state network based on dual estimation." *Neurocomputing*, 2012: 186-195.

Showkati, H., A.H. Hejazi, and S. Elyasi. "Short Term Load Forecasting using Echo State Networks." *International Joint Conference on Neural Networks (IJCNN).* 2010. 1-5.

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.* 2009.

Siqueira, H. V., R. Attux, and C. Lyra Filho. "Exploração de Alternativas Lineares para Previsão de Séries de Vazões." *Mecánica Computacional*, 2010: 9629-9644.

Siqueira, H. V., L. Boccato, R. Attux, and C. Lyra Filho. "Echo State Networks and Extreme Learning Machines: a Comparative Study on Seasonal Streamflow Series Prediction." *Lecture Notes in Computer Science*, 2012a: 491-500.

—. "Echo State Networks for Seasonal Streamflow Series Forecasting." *Lecture Notes in Computer Science*, 2012b: 226-236.

—. "Echo State Networks in Seasonal Streamflow Series Prediction." *Learning and Nonlinear Models*, 2012c: 181-191.

—. "Previsão de séries de vazões com redes neurais de estados de eco." *10th Brazilian Congress on Computational Intelligence.* Fortaleza-CE, Brasil, 2011. 1-7.

Snedecor, G. W., and W. G. Cochran. *Statistical Methods.* 7ª. Iowa State University Press, 1980.

Soares, M. P. "Otimização Multicritério da Operação de Sistemas Hidrotérmicos Utilizando Algoritmos Genéticos." *Dissertação de Mestrado, COPPE-UFRJ, Brasil.* 2008.

Soares, S. "Planejamento da Operação de Sistemas Hidrotérmicos." *Controle e Automação*, 1987: 122-123.

Sorjamaa, A., J. Hao, N. Reyhani, Y. Ji, and A. Lendasse. " Methodology for long-term prediction of time series." *Neurocomputing*, 2007: 2861–2869.

Souza, B. B. "Avaliação do Impacto da Representação Explícita de Bacias Hidrográficas Através do Acoplamento Hidráulico no Planejamento da Operação Energética de Médio Prazo." *Dissertação de Mestrado, COPPE-UFRJ, Brasil.* 2008.

Souza, R. C., A. L. M. Marcato, B. H. Dias, and I. C. Silva Júnior. "A Pesquisa Operacional e o Planejamento de Sistemas Energéticos." *Minicurso - 42º SBPO*. Bento Gonçalves, RS, 2010.

Souza, R. C., and F. L. C. Oliveira. "Uma nova abordagem para identificação das ordens "p" em modelos auto autoregressivos PAR (p)." *International Research Reports - DEE - PUC RJ*. 2009.

Srinivasan, D., A. C. Liew, and C. S. Chang. "Neural network short term load forecaster." *Electric Power Systems Research*, 1994: 227-234.

Stedinger, J. R. "Report on the Evaluation of CEPEL's PAR Models." *Techinical Report, School of Civil and Environmental Engineering - Cornell University*. Ithaca - New York, 2001.

Tang, X., and M. Han. "Partial Lanczos extreme learning machine for." *Neurocomputing*, 2009: 3066–3076.

Tang, Z., and P. A. Fishwick. "Feed-forward neural nets as models for time series forecasting." *ORSA Journal of Computing*, 1993: 374-386.

Tang, Z., C. de Almeida, and P. A. Fishwick. "Time series forecasting using neural networks vs. Box-Jenkins methodology." *Simulation*, 1991: 303-10.

Tong, M. H., A. D. Bickett, E. M. Christiansen, and G. W. Cottrell. "Learning grammatical structure with Echo State Networks." *Neural Networks*, 2007: 424-432.

Turing, A. M. "Intelligent Machinery." In *Cibernetis: Key Papers*, by C. R. Evans and A. D. Robertson, edited by C. R. Evans and A. D. Robertson. Baltimore Md. and Manchester: University Park Press, 1968.

Vecchia, A. V. "Maximum Likehood Estimation fo Periodic Autoregressive-Moving Average Models." *Technometrics*, 1985: 375-384.

Verplancke, T., et al. "A novel time series analysis approach for prediction of dialysis in critically ill patients using echo-state networks." *Medical Informatics and Decision Making*, 2010.

Verstraeten, D., J. Dambre, X. Dutoit, and B. Schrauwen. "Memory versus non-linearityin reservoirs." *Proceedings of the IEEE WCCI 2010.* 2010. 2669-2676.

Viali, L. "Estatística não-paramétrica." *PUC-RS - Disponível em http://www.mat.pucrs.br/famat/viali/.* 2012.

Villanueva, W. J. P. "Comitê de Máquinas em Predição de Séries Temporais." *Dissertação de Mestrado - FEEC - UNICAMP.* 2006.

Villanueva, W. J. P., E. P. Santos, and F. J. Von Zuben. "Data partition and variable selection for time series prediction using wrappers." *Proceedings of the IEEE International Joint Conference on Neural Networks.* Vancouver - Canadá, 2006. 9490-9497.

—. "Long-term time series prediction using wrappers for variable selection and clustering for data partition." *IEEE Proceedings of the International Joint Conference on Neural Networks.* 2007. 1797.

Von Zuben, F. J. "Modelos Paramétricos e Não-Paramétricos de Redes Neurais Artificiais e Aplicações." *Tese de Doutorado - FEEC - UNICAMP.* 1996.

Von Zuben, F. J., and M. L. A. Netto. "Second-Order Training for Recurrent Neural Networks Without Teacher-Forcing." *Proceedings of IEE International Conference on Neural Networks*, 1995: 795-800.

Wang, G., and F. H. Lochovsky. "Feature selection with conditional mutual information maximin in text categorization." *Conference on Information and Knowledge Management.* 2004. 342-349.

Wang, Y., F. Cao, and Y. Yuan. "A study on effectiveness of extreme learning machine." *Neurocomputing*, 2011: 2483-2490.

Werbos, P. J. "Beyond regression: New tools for prediction and analysis in the behavioral ciences." *Tese de Doutorado - Harvard University.* 1974.

—. "Backpropagation through time: what it does and how to do it." *Proceedings of the IEEE*, 1990: 1550-1560.

Widrow, B., and M. Hoff. "Adaptative Switching Circuits." *IRE WESCON Convention Record.* 1960. 96-104.

Wyffels, F., B. Schrauwen, D. Verstraeten, and D. Stroobandt. "Band-pass reservoir computing." *International Joint Conference on Neural Networks.* 2008. 3204-3209.

Yu, L., and H. Liu. "Efficient feature selection via analysis of relevance and redundancy." *Journal of Machine Learning Research*, 2004: 1205-1224.

Zealand, C. M., D. H. Burn, and S. P. Simonovic. "Short term streamflow forecasting using artificial neural networks." *Journal of Hydrology*, 1999: 32-48.

Zhang, G. P., and M. Qi. "Neural network forecasting for seasonal and trend time series." *European Journal of Operation Research*, 2005: 501-514.

Zheng, G. L., and S. A. Billings. "Radial Basis Function Network Configuration Using Mutual Information and the Orthogonal Least Squares Algorithm." *Neural Networks*, 1995: 1619–1637.