

**MODELOS DE SÉRIES TEMPORAIS E REDES
NEURAS ARTIFICIAIS NA PREVISÃO DE VAZÃO**

ANDRÉ LUIZ FRANÇA BATISTA

2009

ANDRÉ LUIZ FRANÇA BATISTA

**MODELOS DE SÉRIES TEMPORAIS E REDES
NEURAS ARTIFICIAIS NA PREVISÃO DE VAZÃO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia de Sistemas, área de concentração em Modelagem de Sistemas Biológicos, para a obtenção do título de Mestre.

Orientadora:

Thelma Sáfydi

LAVRAS
MINAS GERAIS-BRASIL
2009

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Batista, André Luiz França.

Modelos de séries temporais e redes neurais na previsão de
vazão / André Luiz França Batista. – Lavras : UFLA, 2009.
79 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2009.
Orientador: Thelma Sáfadi.
Bibliografia.

1. Modelo SARIMA. 2. Redes neurais artificiais. 3. Vazão
fluvial. 4. Séries temporais. I. Universidade Federal de Lavras. II.
Título.

CDD – 551.4830285
551.4830113

ANDRÉ LUIZ FRANÇA BATISTA

**MODELOS DE SÉRIES TEMPORAIS E REDES
NEURAIS ARTIFICIAIS NA PREVISÃO DE VAZÃO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia de Sistemas, área de concentração em Modelagem de Sistemas Biológicos, para a obtenção do título de Mestre.

APROVADA em 23 de novembro de 2009

Prof. Dr. José Demísio Simões da Silva

INPE

Prof. Dr. Roberto Alves Braga Junior

UFLA

Prof. Dr. Wilian Soares Lacerda

UFLA

Prof.^a Dr.^a Thelma Sáfadi
UFLA
(Orientadora)
LAVRAS
MINAS GERAIS - BRASIL

"Porque o SENHOR dá a sabedoria, e da sua boca vem a inteligência e o entendimento. Ele reserva a verdadeira sabedoria para os retos, é escudo para os que caminham na sinceridade, guarda as veredas do juízo e conserva o caminho dos seus santos." Provérbios 2:6-8

A DEUS, pois a sabedoria é um dom divino, não o simples resultado da capacidade ou esforço humano.

Dedico.

Agradecimentos

"O SENHOR guardará a tua entrada e tua saída, desde agora e para sempre." Salmo 121:8.

Obrigado, SENHOR, por me guardar em todos os momentos da minha pós-graduação, desde o início até agora!

SENHOR, abençoe aqueles que, de uma forma ou de outra, também me ajudaram a concretizar este momento tão especial em minha vida:

João Batista e Suzana, pais sempre presentes, pelo carinho, amor, atenção, educação concedida, zelo, cuidado, orações, e torcida, por mim e pelo Adriano!

Soely França, tia e superamiga, pelo apoio e confiança depositados em mim desde a minha graduação até agora!

Adriano Batista, Danilo Machado e demais colegas de turma, Jackson "Blair" e demais amigos da república, irmãos da ICM, avós, tios e tias... pelo companheirismo e pelas orações!

Thelma Sáfadi, professora, orientadora e amiga, pelos conselhos e orientações, e por me mostrar a importância das séries temporais!

Wiliam, professor e amigo, por me apresentar uma linha de pesquisa apaixonante: redes neurais!

A CAPES, pela concessão da bolsa de estudos.

Aos funcionários do DEG, DEX e DCC pelo auxílio, presteza, e boa vontade.

A ANEEL/CEMIG pelo financiamento do Projeto P & D 176.

Muito obrigado a todos!

SUMÁRIO

RESUMO	i
ABSTRACT	ii
1 INTRODUÇÃO	1
1.1 Objetivos	5
1.2 Organização do trabalho	5
2 CONCEITOS GERAIS SOBRE HIDROLOGIA	6
2.1 Abordagem introdutória	6
2.2 Energia hidráulica	8
2.3 Séries de vazões naturais	10
3 MODELOS DE SÉRIES TEMPORAIS	12
3.1 Introdução	12
3.2 Funções de autocovariância e de autocorrelação	15
3.3 Tendência	16
3.3.1 Teste do Sinal (Cox Stuart)	17
3.4 Sazonalidade	18
3.4.1 Periodograma	19
3.4.2 Teste de Fisher	20
3.5 Modelos de Box e Jenkins	20
3.6 Modelos lineares estacionários	22
3.6.1 Processo linear geral	22
3.7 Modelo autorregressivo e de médias móveis	25
3.7.1 Funções de autocorrelação e autocorrelação parcial	28
3.8 Modelos não estacionários	29
3.8.1 Modelos ARIMA(p, d, q)	30
3.8.2 Modelos ARIMA sazonal - SARIMA	31
3.9 Identificação dos modelos ARIMA	32
3.10 Estimação dos modelos ARIMA	33
3.11 Verificação dos modelos ARIMA	34
3.12 Previsões	35
3.13 Critério para escolha do modelo	37

4	REDES NEURAIAS ARTIFICIAIS.....	40
4.1	Introdução	40
4.2	Breve Histórico da área de RNAs e sua Aplicação em Previsão	41
4.3	Componentes e Treinamento de uma RNA	43
4.3.1	O Neurônio Artificial e as Funções de Ativação	43
4.3.2	Treinamento Supervisionado	46
4.3.3	Treinamento Não Supervisionado	47
4.4	Redes Neurais Multicamadas	48
4.5	Estatística e Redes Neurais Artificiais	51
5	MATERIAIS E MÉTODOS	54
5.1	Base de dados	54
5.2	Séries Temporais	55
5.3	Redes Neurais Artificiais	56
6	RESULTADOS E DISCUSSÃO.....	59
6.1	Modelagem clássica de Séries Temporais.....	59
6.2	Modelagem por Redes Neurais Artificiais	67
6.3	Box & Jenkins x Redes Neurais Artificiais: uma análise comparativa... ..	70
7	CONCLUSÃO	73
	REFERÊNCIAS BIBLIOGRÁFICAS.....	75

RESUMO

BATISTA, André Luiz França. **Análise e previsões de vazões utilizando modelos de séries temporais e redes neurais artificiais**. 2009. 79p. Dissertação (Mestrado em Engenharia de Sistemas) - Universidade Federal de Lavras, Lavras.

*

A previsão do comportamento hidrológico de rios afluentes a reservatórios de usinas hidroelétricas consiste em uma das principais ferramentas para gestão da produção de energia elétrica brasileira. Conhecer os valores futuros da vazão de um rio é de extrema importância para o planejamento dos sistemas hidroelétricos. Diante desse contexto, o presente trabalho investiga duas metodologias distintas para realizar previsão de séries temporais de vazões fluviais: Box & Jenkins e Redes Neurais Artificiais. Os dados utilizados neste trabalho são os valores da vazão média mensal do Rio Grande. O conjunto de dados consiste em 216 observações que abrangem desde Janeiro/1990 a Dezembro/2007. Foram construídos modelos originados da metodologia sugerida por Box & Jenkins e também modelos baseados na técnica de Redes Neurais Artificiais. Tais modelos foram avaliados de acordo com o critério do EQMP e MAPE para que os melhores modelos para a série temporal em estudo fossem selecionados. O modelo estatístico que melhor se adequou ao conjunto de dados foi um SARIMA(0,1,1)(0,1,2)₁₂. O modelo de redes neurais que teve melhor adequação junto ao conjunto de dados foi uma MLP(12,20,1). Os modelos selecionados foram empregados para prever valores futuros da série histórica de vazões naturais do Rio Grande (posto fluviométrico de Madre de Deus de Minas, MG). Foi realizada uma análise comparativa entre ambas as técnicas empregadas no prognóstico da série temporal. Os resultados obtidos na comparação mostram que cada metodologia pode ser ajustada adequadamente ao conjunto de observações em estudo, entretanto cada técnica possui vantagens e desvantagens. A metodologia de Box & Jenkins tem como ponto a seu favor o fato de extrair informações importantes sobre a série temporal, tais como: identificação de ciclos e tendências. Tal extração de informações da série não ocorre na técnica de Redes Neurais Artificiais, o que se torna um ponto negativo para essa técnica. Para a série de vazões do Rio Grande, o ponto positivo da utilização de Redes Neurais foi a obtenção de valores de previsão mais precisos do que os obtidos pelos modelos estatísticos propostos por Box & Jenkins.

* Comitê Orientador: Thelma Sáfiadi - UFLA (Orientadora). Wilian S. Lacerda - UFLA

ABSTRACT

BATISTA, André Luiz França. **River flow analysis and forecasting using time series and artificial neural networks models**. 2009. 79p. Dissertation (Master Program in Systems Engineering) - Federal University of Lavras, Lavras. *

Forecasting the hydrological behavior of inflowing rivers into reservoirs of hydroelectric plants is one of the main tools for managing the production of electric power in Brazil. Knowing the future values of a river's flow is critical when planning hydroelectric systems. Considering such background, this work aims at investigating two different methods to forecast time series of river flows: Box & Jenkins and Artificial Neural Networks. The data used in this work are the values of average monthly flow of Rio Grande (stream gauge station of Madre de Deus de Minas, MG). The data set consists of 216 observations that were done between January/1990 to December/2007. Models originated from the Box & Jenkins method, as well as models based on the Artificial Neural Networks technique, have been constructed. These models were evaluated according to the EQMP and MAPE criteria in order to select the best models for the studied time series. The statistical model that best suited the data set was a SARIMA(0,1,1)(0,1,2)₁₂. The neural networks model that best adjusted to the data set was an MLP(12,20,1). The selected models were used to forecast future values of the historical series of Rio Grande's natural flows. A comparative analysis between both techniques used at the prognostication of time series has been done. The results obtained from this comparison have shown that each method can be adequately adjusted to the set of studied observations; however, each technique has both advantages and disadvantages. The Box & Jenkins method has as an advantage the fact that it extracts important information from the time series, such as identification of cycles and trends. This extraction of information from the series does not occur in the Artificial Neural Networks technique, which is a drawback to this technique. In Rio Grande's flow series, the positive aspect of using Neural Networks was that the obtained prediction values were more accurate than the ones from the statistical models proposed by Box & Jenkins.

*Guidance Committee: Thelma Sáfyadi - UFLA (Major Professor). Wilian S. Lacerda - UFLA

1 INTRODUÇÃO

O parque gerador de energia elétrica no Brasil é predominantemente hidráulico, sendo o setor elétrico responsável pelo gerenciamento da quantidade de água nos grandes reservatórios de acumulação existentes no país. Estes reservatórios, que até a pouco tempo destinavam-se exclusivamente à geração de energia, hoje são alvos de interesses diversos e conflitantes. Nesse contexto, cresce a importância na área de planejamento e controle de sistemas de recursos hídricos.

Um dos principais problemas enfrentados na produção de energia elétrica no Brasil é a variabilidade natural do regime hidrológico, decorrente da precipitação e suas interações com o regime de escoamento.

O planejamento da operação de sistemas de produção de energia elétrica visa a determinar uma política de operação para usinas hidroelétricas e termoeletricas e para os intercâmbios, inclusive internacionais, do sistema que atenda a demanda de energia elétrica de forma econômica e confiável. Em linhas gerais, o objetivo é utilizar da forma mais eficiente possível a geração de base hidráulica para substituir ao máximo a geração térmica, cujo custo do combustível é elevado e crescente com a geração. As usinas hidroelétricas possuem características de produção não-lineares e suas vazões afluentes são variáveis aleatórias, levando assim a um problema estocástico não-linear. Em razão da complexidade do problema, é comum sua decomposição em uma cadeia de planejamento que inclui as etapas de longo, médio e curto prazos, Soares (1987).

A produção energética de um sistema hidroelétrico depende da série de vazões afluentes às diversas usinas do sistema. Por exemplo, quando se adota um aproveitamento hídrico, a série histórica de vazões no local é usualmente utilizada

como dado de entrada para um modelo de simulação e/ou otimização.

As séries históricas de vazões afluentes aos locais de aproveitamento hidroelétricos são obtidas a partir de um processo laborioso baseado em medições e análises como mostra a FIGURA 1.

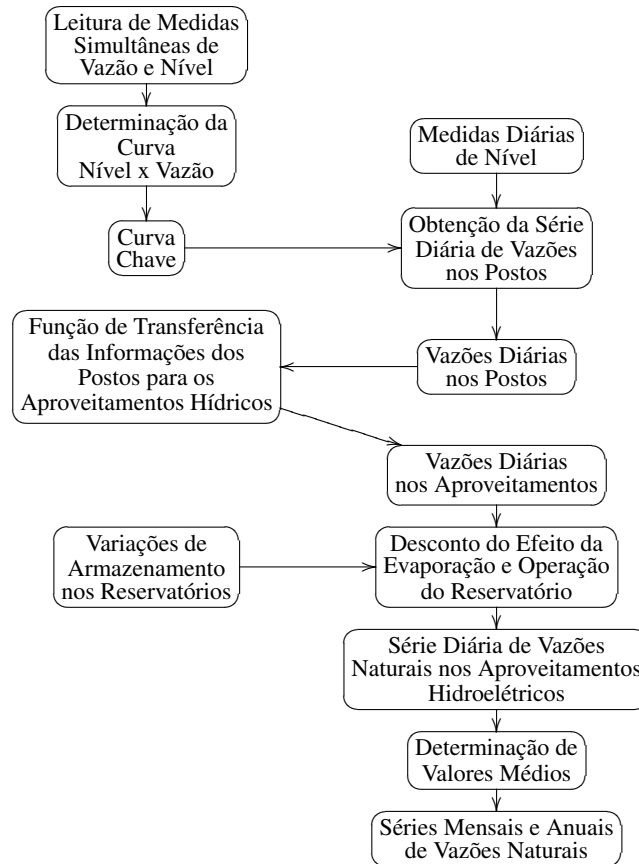


FIGURA 1 Processo de obtenção de séries históricas de vazões
Fonte: Fortunato *et al.* (1990)

As atividades de planejamento e operação de sistemas de energia e recursos hídricos requerem a utilização de muitas variáveis de entrada. No caso do planejamento de sistemas de energia brasileiro, predominantemente hidroelétrico, as previsões de vazões de curto e longo prazos são necessárias para propósitos de

simulação, otimização e tomada de decisões.

Há, atualmente, uma vasta literatura que descreve a previsão de séries temporais em diversas áreas, com exemplos que abrangem desde aplicações em medicina, medindo a irregularidade das batidas do coração; passam pelas pesquisas em hidrologia, prevendo variáveis de recursos hídricos; e chegam até as aplicações em economia, observando o comportamento da volatilidade do mercado financeiro. Muitos modelos são baseados na metodologia Box & Jenkins, a qual assume relações lineares entre as variáveis do sistema, Box *et al.* (1994).

A utilização de modelos de séries temporais em análise e previsão de séries temporais climatológicas é bem aceita pelos pesquisadores. Safadi (2004) utilizou o conceito de séries temporais para descrever a série de vazão de água da represa de Furnas. Nesse trabalho, Safadi (2004) analisou o comportamento da série de vazão de água na barragem de Furnas, empregando análise de séries temporais e estudando o efeito de sazonalidade, tendência e intervenção. Ferraz *et al.* (1999) utilizaram modelos de séries temporais para análise e previsão de séries de precipitação pluviométricas mensais no município de Lavras-MG.

Entretanto, séries temporais, que descrevem uma sequência real de valores, geralmente possuem um comportamento não linear e, nesse caso, as Redes Neurais Artificiais (RNAs) e variantes têm sido largamente adotadas para prever valores futuros com base em valores passados, Gershenfeld & Weigend (1994).

Em anos recentes, redes neurais artificiais tornaram-se extremamente conhecidas para previsão em várias áreas, incluindo finanças, carga elétrica e recursos hídricos. Embora o conceito de neurônios artificiais tenha sido proposto em 1943, por McCulloch & Pitts (1943), pesquisas e aplicações de redes neurais tomaram importância após a publicação em 1986 do algoritmo *back-propagation* por Rumelhart *et al.* (1986).

A utilização de redes neurais para previsão de séries temporais vem sendo proposta por diversos autores, entre eles Lapedes & Farber (1987), Weigend *et*

al. (1990), Zandonade & Souza (1993), Contrell *et al.* (1995), Lachtermacher & Fuller (1995), mostrando a viabilidade de utilizar estes modelos.

A aplicação de redes neurais como preditores de séries temporais climatológicas vem sendo realizadas por alguns pesquisadores, dentre eles Maier & Dandy (1999), Campolo *et al.* (1999), Imrie *et al.* (2000), Ballini *et al.* (2000), Anmala *et al.* (2000), Elshorbagy *et al.* (2000), demonstrando que redes neurais artificiais podem ser usadas com sucesso para previsão de séries temporais de caráter climatológico.

Em vários estudos, redes neurais e técnicas de séries temporais tradicionais vêm sendo comparadas como em Sharda & Patil (1990), Tang *et al.* (1991). A maioria dos estudos realizados utilizam o algoritmo *back-propagation*, ou algumas de suas extensões. Redes neurais com algoritmo *back-propagation* já foram utilizadas com sucesso em diversas áreas, entre as quais sistemas de potência, previsões econômicas e análise de séries temporais.

As comparações de redes neurais com outras abordagens ainda são contraditórias, conforme esclarecem Anmala *et al.* (2000). Alguns estudos concluem que redes neurais artificiais são melhores que os métodos tradicionais, enquanto que outras pesquisas concluem o contrário. As explicações para tais contradições podem ser descritas por diferentes fatores, tais como: a estrutura da rede neural, o tipo de série (estacionária ou não-estacionária) usada nos estudos e a relação do tamanho da rede e o número de entradas das séries temporais.

As redes neurais artificiais, especialmente redes multicamadas com algoritmo *back-propagation*, vêm sendo sugeridas por pesquisadores em inteligência computacional para análise de séries temporais. Isso é decorrente da capacidade das redes neurais para tratar com relações não-lineares de entrada-saída, destacando sua habilidade de aprendizado e capacidade de generalização, associação e busca paralela.

1.1 Objetivos

Objetivo geral

Conduziu-se este trabalho, com o objetivo geral de comparar modelos de séries temporais e redes neurais artificiais para previsão de vazão.

Objetivo específico

Neste trabalho, objetivou-se a vazão média mensal da bacia de drenagem do Rio Grande, com seção de controle no posto fluviométrico de Madre de Deus de Minas. Esse rio é afluente ao reservatório da usina hidrelétrica de Camargos/MG, UHE-Camargos-CEMIG.

1.2 Organização do trabalho

Este capítulo apresentou uma introdução sobre o tema dessa dissertação, descrevendo o contexto, objetivo e as motivações dentro dos quais o trabalho foi desenvolvido.

Antes de modelar, analisar e/ou prever séries de vazões é necessário situá-las de um ponto de vista teórico, que permita identificar o que está em questão para a grande relevância deste trabalho. Nos próximos capítulos serão abordados conceitos básicos sobre hidrologia, geração de energia hidráulica e a necessidade da previsão de vazões. Paralelamente a essa necessidade de prever séries de vazões, serão apresentados os modelos de séries temporais e redes neurais artificiais, e de que forma esses conceitos podem se encontrar.

No Capítulo 2, mostram-se conceitos básicos sobre hidrologia, modelagem hidrológica e geração de energia hidroelétrica. No Capítulo 3, é apresentada uma revisão geral sobre Séries Temporais e o método de modelagem Box & Jenkins. No Capítulo 4, discursa-se sobre Redes Neurais Artificiais, suas teorias e aplicações. No Capítulo 5, apresentam-se as metodologias de pesquisa que foram utilizadas nas investigações deste trabalho. No Capítulo 6, discutem-se os resultados obtidos na investigação realizada. No Capítulo 7, apresentam-se as conclusões referentes às modelagens, análises e comparações realizadas.

2 CONCEITOS GERAIS SOBRE HIDROLOGIA

Neste Capítulo, Conceitos Gerais Sobre Hidrologia, são conceituadas definições básicas sobre recursos hídricos e a importância de conhecer o ciclo hidrológico. Nos itens seguintes, apresenta-se como os recursos hídricos são utilizados na produção de energia elétrica e a relevância das séries de vazões naturais nesse processo de geração de energia elétrica.

2.1 Abordagem introdutória

Antigamente, os recursos hídricos não possuíam a mesma importância que hoje representam, ou seja, podiam ser obtidos à vontade para todas as necessidades de cada época. Portanto, havia pouca preocupação com a capacidade máxima dos rios e das fontes subterrâneas. Com o desenvolvimento da civilização e o constante crescimento da população, a necessidade da utilização e controle dos recursos hídricos tem se tornado cada vez maior.

A compreensão da relação entre os fenômenos que envolvem o ciclo hidrológico permite ao homem simular o comportamento da água nas bacias hidrográficas, possibilitando a previsão de diferentes cenários hidrológicos em decorrência de alterações climáticas e de uso do solo. Surge, então, mediante esse contexto, o conceito de simulação hidrológica. Conceito esse que pode ser entendido como a representação do comportamento do ciclo hidrológico por meio de equações matemáticas, tendo como unidade territorial a bacia hidrográfica, fornecendo as respostas da bacia em termos de vazão, em função das condições de umidade do solo e em resposta à precipitação.

A simulação hidrológica possui várias aplicações, dentre elas possibilita adequada descrição física dos parâmetros em grandes escalas espaciais,

especialmente os relacionados à vegetação, solos e fisiografia, permitindo avaliar o impacto de mudanças climáticas e de uso do solo em bacias hidrográficas.

Outra importante aplicação da simulação hidrológica está associada à previsão de vazões, ou seja, na estimativa do escoamento com antecedência no tempo, visando a minimizar as incertezas do clima, essencial às diversas utilizações dos deflúvios superficiais, podendo ser de curto prazo, quando a antecedência é de poucas horas, ou de longo prazo, que pode chegar a meses, Tucci (2002).

De acordo com Tucci (2005), as limitações de modelos hidrológicos estão associadas à quantidade e à qualidade dos dados hidrológicos, à dificuldade de formulação matemática de alguns fenômenos e à descrição da distribuição espacial das variáveis envolvidas.

A hidrologia é a ciência que estuda a água na terra sob os mais variados aspectos, quais sejam: sua ocorrência, sua circulação, distribuição, suas propriedades físico-químicas e suas relações, e sua relação como meio ambiente, incluindo sua relação com os seres vivos relacionada com toda a água da Terra, Chow (1959).

Tucci (2005) esclarece que a compreensão do ciclo hidrológico é fundamental para o entendimento dos fundamentos que envolvem a simulação hidrológica. Na FIGURA 2, está uma representação dos principais componentes do ciclo hidrológico, que corresponde à dinâmica da água nos diferentes estados físicos e ambientes, tendo como fonte de energia básica a radiação solar e, como unidade territorial, a bacia hidrográfica, a qual define de maneira completa e única a área na qual todo escoamento superficial converge para a sua foz, Tonello *et al.* (2006).

Nos próximos anos, a utilização dos recursos hídricos será, com toda certeza, cada vez mais intensa tornando os dados e o conhecimento sobre os princípios hidrológicos importantíssimos para melhor compreensão e utilização

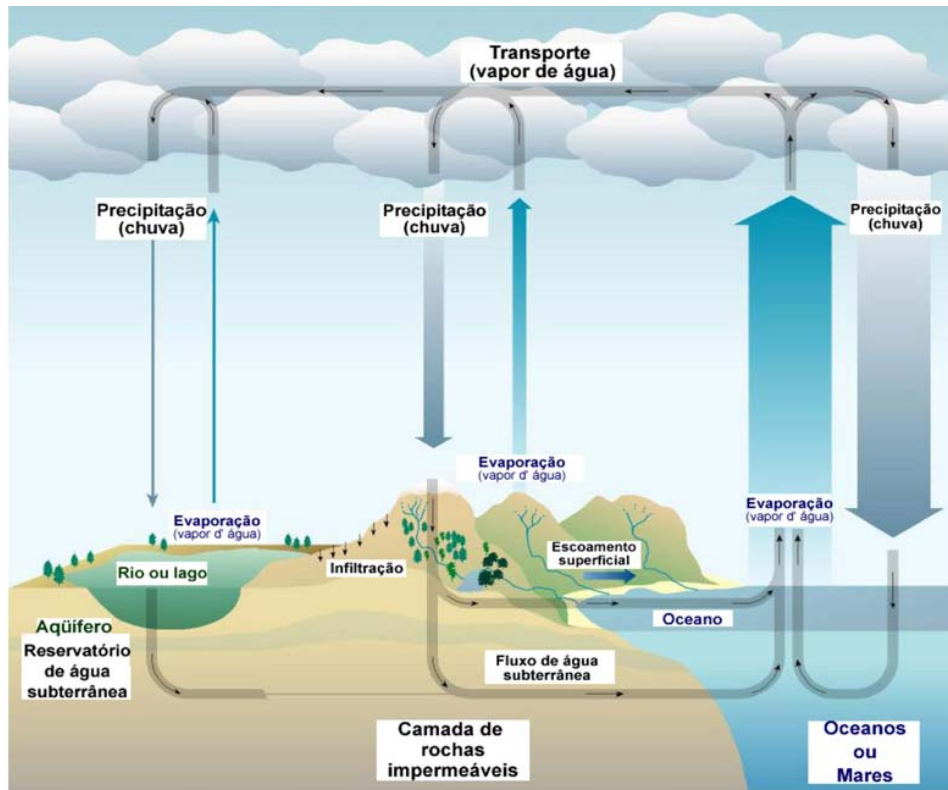


FIGURA 2 Ciclo hidrológico
 Fonte: <http://www.maenatureza.org.br>

desses recursos. A aplicação de recursos hídricos vai desde o uso doméstico, como consumo humano e higiene pessoal. Passa pelo uso rural, utilizada em irrigação e criação de animais, até chegar ao uso industrial, aplicada no funcionamento de máquinas e na produção de energia elétrica.

2.2 Energia hidráulica

A energia hidráulica ou energia hídrica é a energia obtida a partir da energia potencial de uma massa de água. A forma na qual ela se manifesta na natureza é nos fluxos de água, como rios e lagos e pode ser aproveitada por meio de um desnível ou queda d'água. Pode ser convertida na forma de energia mecânica

(rotação de um eixo) por meio de turbinas hidráulicas ou moinhos de água. As turbinas por sua vez podem ser usadas como acionamento de um equipamento industrial, como um compressor, ou de um gerador elétrico, com a finalidade de prover energia elétrica para uma rede de energia (Gil & Sanchez (1997)).

A potência hidráulica máxima que pode ser obtida por meio de um desnível pode ser calculada pelo produto:

$$P = \rho Q H g \quad (2.1)$$

Em unidades do Sistema Internacional de Unidades (SI), em que:

- Potência (P): *Watt(W)*
- Densidade (ρ): *kg/m³*
- Vazão volumétrica (Q): *m³/s*
- Queda (H): *m*
- Aceleração da gravidade (g): *m/s²*

É necessário que haja um fluxo de água para que a energia seja gerada de forma contínua no tempo, por isso, embora se possa usar qualquer reservatório de água, como um lago por exemplo, deve haver um suprimento de água ao lago, caso contrário haverá redução do nível e com o tempo a diminuição da potência gerada, ver Equação (2.1). As represas (barragens) são nada mais que lagos artificiais, construídos num rio, permitindo a geração contínua.

As represas podem ser importantes pois caso a água fosse coletada diretamente de um rio, na medida em que houvesse uma redução da vazão do rio, como em uma época de estiagem, haveria redução da potência gerada. Assim, com a formação de um lago (reservatório da barragem), nas épocas de estiagem pode-se usar a água armazenada e, se este for suficientemente grande, poderá atender a um período de estiagem de vários meses ou mesmo plurianual.

No Brasil, em razão da sua enorme quantidade de rios, a maior parte da energia elétrica disponível é proveniente de grandes usinas hidrelétricas. A energia primária de uma hidrelétrica é a energia potencial gravitacional da água contida numa represa elevada. Antes de se tornar energia elétrica, a energia primária deve ser convertida em energia cinética de rotação, CENTRAIS ELÉTRICAS BRASILEIRAS (2009). O dispositivo que realiza essa transformação é a turbina. Ela consiste basicamente em uma roda dotada de pás, que é posta em rápida rotação ao receber a massa de água. O último elemento dessa cadeia de transformações é o gerador, que converte o movimento rotatório da turbina em energia elétrica.

Um rio não é percorrido pela mesma quantidade de água durante o ano inteiro. Em uma estação chuvosa a quantidade de água aumenta. Para aproveitar ao máximo as possibilidades de fornecimento de energia de um rio, torna-se imprescindível a previsão de sua vazão, a fim de que a usina possa funcionar continuamente com toda a potência instalada.

2.3 Séries de vazões naturais

O planejamento de sistemas hidroelétricos possui um alto grau de complexidade e dificuldade, uma vez que envolve características de produção não lineares e depende de muitas variáveis. As atividades de planejamento e operação de sistemas de energia e recursos hídricos requerem a utilização de muitas variáveis de entrada. Uma das variáveis-chave é a vazão natural. Em hidráulica ou em mecânica dos fluidos, define-se por vazão, o volume por unidade de tempo, que se escoar por meio de determinada seção transversal de um conduto livre (canal, rio ou tubulação com pressão atmosférica) ou de um conduto forçado (tubulação com pressão positiva ou negativa). Isso significa que a vazão é a rapidez com a qual um volume escoar.

Os valores de vazões devem ser previstos com precisão, uma vez que esses valores influenciam significativamente na produção de energia. No caso do

planejamento de sistemas de energia brasileiro, predominantemente hidroelétrico, as previsões de vazões de curto e longo prazos são necessárias para propósitos de simulação, otimização e tomada de decisões.

A análise e a previsão de séries de vazões são de fundamental importância no planejamento da operação de sistemas de recursos hídricos. Uma das grandes dificuldades na previsão das séries de vazões é a presença da sazonalidade decorrente dos períodos de cheia e seca do ano.

A observação sistemática dos valores de vazão gera uma série histórica, ou série temporal, que pode ser usada para análise do fenômeno observado e também prever valores futuros da série. Conforme Tucci (2002), uma série histórica estacionária de vazões é caracterizada pela continuidade dos valores dos parâmetros estatísticos ao longo do tempo. Os modelos estocásticos foram, por um longo tempo, a alternativa mais comum aos modelos determinísticos ou hidrológicos na análise e previsão de vazões, baseados principalmente, na metodologia de Box & Jenkins conforme Box *et al.* (1994).

Neste capítulo, foram apresentados conceitos básicos sobre hidrologia e sobre a importância que os recursos hídricos têm na vida do homem, principalmente na produção de energia elétrica. Destacou-se ainda a relevância das séries de vazões naturais no processo de geração de energia elétrica, uma vez que a principal fonte de energia elétrica no Brasil é hidráulica. Nos capítulos seguintes serão apresentados modelos de Séries Temporais e Rede Neurais Artificiais, respectivamente. Esses modelos possuem muitas aplicações, dentre elas encontram-se a análise e a previsão de séries de vazão.

3 MODELOS DE SÉRIES TEMPORAIS

Existem duas formas de estudar séries temporais: análise e previsão. Uma análise da série temporal é um método para tentar entender a série temporal, de forma a entender a estrutura que gerou a série e também identificar ciclos e tendências. Uma previsão a partir da série temporal procura construir um modelo matemático a partir do qual seja possível prever valores futuros da série.

Os modelos clássicos para análise e previsão de séries temporais são ferramentas bastante conhecidas na literatura. Tais modelos podem ser divididos em dois grandes grupos: métodos paramétricos e não-paramétricos. Em se tratando do contexto de modelagem paramétrica, as metodologias propostas por Box *et al.* (1994) têm sido largamente utilizadas para previsões nas mais diversas áreas, como, por exemplo: economia, finanças, meteorologia e hidrologia.

As séries de vazões apresentam características estatísticas uma vez que não podem ser determinadas por uma equação matemática exata, sendo, portanto, processos estocásticos. Assim, os modelos abordados neste capítulo apresentam características estocásticas e são baseados na metodologia de Box & Jenkins, Box & Pierce (1970).

3.1 Introdução

Uma série temporal é qualquer conjunto de observações ordenadas no tempo. Uma série temporal pode ser determinada por alguma função matemática, tal como:

$$Z_t = \cos(2\pi ft),$$

esta série é conhecida como determinística. Se esses valores podem ser descritos somente em termos de uma distribuição de probabilidades, esta série é conhecida

como não-determinística ou estocástica.

A análise da série temporal pode ser feita no domínio do tempo ou no domínio de frequências, sendo os modelos propostos, paramétricos e não-paramétricos, respectivamente, Morettin & Tolo (2006). Em ambos os domínios esses modelos são processos estocásticos, isto é, processos controlados por leis probabilísticas.

O objetivo da análise em séries temporais é construir modelos para a série, com propósitos determinados, tais como descrever apenas o comportamento da série, investigar o mecanismo gerador da série temporal, fazer previsão de valores futuros e procurar periodicidades relevantes nos dados.

Um modelo clássico para séries temporais supõe que a série $\{Z_1, \dots, Z_n\}$ pode ser escrita como:

$$Z_t = T_t + S_t + a_t \quad t=1,2,\dots,n \quad (3.1)$$

em que Z_t é a série temporal, T_t é a tendência, S_t é a sazonalidade e a_t é um componente aleatório.

Segundo Morettin & Tolo (2006), a tendência pode ser entendida como aumento ou diminuição gradual das observações ao longo do período. A sazonalidade mostra flutuações ocorridas em períodos menores ou iguais a doze meses e a componente aleatória mostra oscilações aleatórias irregulares. A suposição usual é a de que a_t seja uma série puramente aleatória ou um ruído branco independente, isto é, tenha média igual a zero e variância constante.

O modelo (3.1) é dito aditivo, pois, a componente sazonal é independente das outras componentes. Se as amplitudes sazonais variam de acordo com a tendência, o modelo mais adequado é o multiplicativo,

$$Z_t = T_t S_t a_t \quad t=1,2,\dots,n. \quad (3.2)$$

Nesse caso, a variância não é constante, sendo necessária uma transformação de

dados. Tomando-se o logaritmo de (3.2), tem-se:

$$Z_t^* = \log T_t + \log S_t + \log a_t \quad t=1,2,\dots,n. \quad (3.3)$$

em que Z_t^* é o $\log Z_t$.

Estudos realizados comprovaram que transformações não melhoram a qualidade da previsão. Granger & Newbold (1976) mostraram que as previsões dos antilogaritmos dos dados transformados são estimadores viesados, portanto deveriam ser ajustados. Mostraram também que a heterocedasticidade não afeta a adequação da previsão, pois ela não implica em estimadores viesados.

Para ajustar os modelos de Box e Jenkins a uma série temporal Z_t , é necessário que ela seja estacionária, isto é, que ela se desenvolva aleatoriamente no tempo em torno de uma média constante. No entanto, a maioria das séries apresenta componentes como tendência e ou sazonalidade, que são uma forma de não-estacionariedade. A fim de tornar uma série estacionária, um procedimento comum é tomar diferenças sucessivas da série original. A primeira diferença é definida por

$$\Delta Z_t = Z_t - Z_{t-1} \quad t=1,2,\dots,n. \quad (3.4)$$

A segunda diferença é

$$\Delta^2 Z_t = Z_t - 2Z_{t-1} + Z_{t-2} \quad t=1,2,\dots,n. \quad (3.5)$$

Generalizando, a n-ésima diferença é

$$\Delta^n Z_t = \Delta\{\Delta^{n-1} Z_t\} \quad t=1,2,\dots,n. \quad (3.6)$$

Normalmente, são necessárias uma ou duas diferenças para que a série se torne estacionária. Retiradas as componentes T_t e S_t , o que sobra é a componente aleatória a_t .

É importante salientar que o primeiro passo na análise de uma série temporal é a construção do gráfico, que mostra características importantes, como

tendência, sazonalidade, variabilidade e *outliers*.

3.2 Funções de autocovariância e de autocorrelação

Segundo Morettin & Toloi (2006), um processo é dito estocástico quando é definido por uma família $Z = \{Z_t, t \in T\}$, tal que para cada t , Z_t é uma variável aleatória e T é um conjunto arbitrário. Ainda, esse processo é estritamente estacionário se todas as distribuições finito-dimensionais F permanecem as mesmas sob translação no tempo, ou seja,

$$F(z_1, \dots, z_n, t_1 + k, \dots, t_n + k) = F(z_1, \dots, z_n, t_1, \dots, t_n) \quad (3.7)$$

para qualquer $t_1, \dots, t_n, k \in T$.

A análise dos gráficos das funções de autocorrelação é um passo importante para um bom ajuste da série. De acordo com Box *et al.* (1994), no que se segue é apresentado o procedimento para obter a função de autocorrelação.

A autocovariância no *lag* k , γ_k , é definida como a covariância entre Z_t e Z_{t+k} , em que k é o número de intervalos de tempo defasados, que é a esperança do produto do desvio de cada ponto em relação à média do processo estocástico, ou seja:

$$\gamma_k = cov[Z_t, Z_{t+k}] = E[(Z_t - \mu)(Z_{t+k} - \mu)] \quad \forall t, k \in \mathbb{N}. \quad (3.8)$$

A função de autocovariância (facv) de um processo estocástico satisfaz às seguintes propriedades:

- i. $\gamma_0 > 0$
- ii. $\gamma_{-k} = \gamma_k$
- iii. $|\gamma_k| \leq \gamma_0$

iv. γ_k é positiva definida, no sentido que

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma_{k_i - k_j} \quad \forall a_1, \dots, a_n, k_1, \dots, k_n \in \mathbb{N}. \quad (3.9)$$

A autocorrelação no lag k , ρ_k de um processo estacionário é definida por

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad k \in Z, \quad (3.10)$$

A função de autocorrelação (FAC) satisfaz às seguintes propriedades:

- i. $\rho_0 = 1$
- ii. $\rho_{-k} = \rho_k$
- iii. $|\rho_k| \leq \rho_0 = 1$
- iv. ρ_k é positiva definida.

A estimativa da autocovariância γ_k é

$$c_k = \frac{1}{n} \sum_{t=1}^{N-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z}) \quad k = 1, \dots, N-1; \quad (3.11)$$

em que N é o número de observações, $\bar{Z} = \frac{1}{N} \sum_{i=1}^N Z_i$ é a média amostral.

A estimativa da autocorrelação ρ_k é dada por

$$r_k = \frac{c_k}{c_0}. \quad (3.12)$$

3.3 Tendência

Supondo-se a ausência de sazonalidade, tem-se o modelo

$$Z_t = T_t + a_t \quad t=1,2,\dots,n \quad (3.13)$$

em que a_t é um ruído branco. Existem vários métodos para estimar tendência, dentre eles, ajuste polinomial e suavização (filtro), descritos em Morettin & Tolo (2006).

Estimando-se a tendência \hat{T}_t , a série pode ser escrita livre de tendência

$$Y_t = Z_t - \hat{T}_t \quad t=1,2,\dots,n. \quad (3.14)$$

Um outro procedimento utilizado para eliminar a tendência de uma série é tomar a diferença, como foi definido em (3.4) e (3.5).

Pela análise visual, pode-se avaliar se uma série temporal possui tendência. Para confirmar esta avaliação prévia, dois procedimentos são utilizados: análise do gráfico da autocorrelação e ou um teste para tendência.

Dentre os testes para tendência, descritos em Morettin & Tolo (2006), têm-se: teste do sinal (Cox-Stuart), teste baseado no coeficiente de relação de Spearman e teste F para análise de variância.

O Teste do Sinal (Cox Stuart), será apresentado a seguir.

3.3.1 Teste do Sinal (Cox Stuart)

O teste do Sinal é utilizado para verificar a existência de tendência em uma série temporal. O procedimento para aplicar o teste é o seguinte:

1. agrupam-se as observações em pares $(Z_1, Z_{c+1}), (Z_2, Z_{c+2}), (Z_{N-c}, Z_N)$, em que $c = N/2$ se N números de observações for par e $c = (N + 1)/2$ se N for ímpar;
2. para cada par (Z_i, Z_{c+i}) , associa-se um sinal. Este sinal é positivo (+) se $Z_i < Z_{c+i}$ e negativo (-) se $Z_i > Z_{c+i}$. Descarta-se a comparação se $Z_i = Z_{c+i}$;
3. testa-se a hipótese de nulidade.

H_0 : Não existe tendência.

A estatística do teste a ser utilizada é T, sendo que T é igual ao número de sinais positivos atribuídos no final da comparação entre Z_i e Z_{c+i} ;

4. Calcula-se o valor de t , que é dado por:

$$t = \frac{1}{2}(n + \omega_{\alpha/2}\sqrt{n}) \quad (3.15)$$

em que n é o número de vezes que não houve empate na comparação entre Z_i e Z_{c+i} e $\omega_{\alpha/2}$ é o valor tabelado ao nível de significância α . Caso $n \leq 20$, o valor tabelado de $\omega_{\alpha/2}$ é dado pela distribuição binomial com $p = 0,5$ e, caso $n > 20$, utiliza-se a distribuição normal;

5. se $T \leq t$ ou $T \geq n - t$, rejeita H_0 , isto é, a série possui tendência.

3.4 Sazonalidade

A componente sazonal capta características da série que ocorrem regularmente dentro do período de um ano, isto é, os fenômenos sazonais ocorrem regularmente em período de no máximo doze meses. O que se observa em séries sazonais de período 12, é que ocorrem as relações:

- i. entre meses sucessivos, em um ano particular;
- ii. entre observações para o mesmo mês, em anos consecutivos.

De modo geral, pode-se afirmar que as séries sazonais apresentam alta correlação em *lags* sazonais. Os procedimentos mais comuns para se estimar a sazonalidade são o método de regressão (sazonalidade determinística) e o método de médias móveis (sazonalidade estocástica), descritos em Morettin & Tolo (2006).

Estimada a sazonalidade \hat{S}_t , a série pode ser escrita livre da componente sazonal. Se o modelo da série for aditivo, tem-se

$$Z_t^{SA} = Z_t - \hat{S}_t \quad t=1,2,\dots,N \quad (3.16)$$

e se for multiplicativo, tem-se

$$Z_t^{SA} = Z_t / \hat{S}_t \quad t=1,2,\dots,N. \quad (3.17)$$

Como já foi comentado, a componente sazonal também pode ser constatada pela análise visual do gráfico da série. Para se confirmar a existência desta aplica-se um teste para sazonalidade.

Dentre os testes para sazonalidade tem-se: teste de Fridman, teste de Kruskal-Wallis, teste F para Análise de Variância descritos em Morettin & Tolo (2006) e teste de Fisher, encontrado em Priestley (1989) e descrito resumidamente na seção 3.4.2 .

Toda periodicidade acima do período de 12 meses é considerada ciclo. Retirando-se o efeito do ciclo, a série perde muitas observações, reduzindo seu tamanho, o que prejudica e dificulta o ajuste do modelo.

3.4.1 Periodograma

Uma outra forma de se detectar a existência da sazonalidade é por meio da análise do gráfico do periodograma. Segundo Morettin & Tolo (2006), a análise espectral é fundamental em áreas onde o interesse básico é a procura de periodicidade dos dados. Essa análise é feita no domínio da frequência, portanto, é baseada em modelos não paramétricos.

Ferraz *et al.* (1999) comentam que o periodograma foi originalmente usado para detectar e estimar a amplitude do componente senóide de frequência. Priestley (1989) define o periodograma do processo estacionário (a_t) como:

$$I_p(f_i) = \frac{2}{n} \left[\left(\sum_{t=1}^n a_t \cos \frac{2\pi i}{n} t \right)^2 + \left(\sum_{t=1}^n a_t \sin \frac{2\pi i}{n} t \right)^2 \right] \quad (3.18)$$

com $0 < f_i < \frac{1}{2}$ e $t = 1, 2, \dots, n$ em que $I(f_i)$ indica a intensidade da frequência f_i . A periodicidade pode ser verificada observando-se a existência de picos na frequência $f_i = \frac{i}{n}$, o que indica uma periodicidade de período $\frac{1}{f_i}$.

No gráfico do periodograma, a frequência f_i é representada no eixo das ordenadas e a intensidade da frequência $I(f_i)$ no das abscissas. Geralmente, o pico de maior intensidade é o componente periódico. Caso exista mais de um pico,

aplica-se o teste de Fisher, para verificar se o pico é um componente periódico genuíno.

3.4.2 Teste de Fisher

O teste de Fisher é utilizado para confirmar a existência de sazonalidade em uma série temporal. Segundo Priestley (1989), esse teste utiliza os valores do periodograma e detecta grandes periodicidades.

O procedimento para aplicar o teste é o seguinte:

1. traça-se o periodograma, utilizando um pacote estatístico (R, Statistica e outros) ou por meio da frequência padrão dada por (3.18);
2. toma-se a maior periodicidade encontrada no periodograma, $\max(I_p)$.

Calcula-se a estatística:

$$g = \frac{\max I_p}{\sum_{p=1}^{N/2} I_p}; \quad (3.19)$$

3. calcula-se a estatística do Teste de Fisher, z_α , dada por:

$$z_\alpha = \left(\frac{\alpha}{n}\right)^{1/(n-1)} + 1 \quad (3.20)$$

em que $n = \frac{N}{2}$ e α é o nível de significância;

4. testa-se a hipótese de nulidade.

H_0 : Não existe periodicidade.

Se $g > z_\alpha$, rejeita-se H_0 , isto é, existe sazonalidade no período analisado.

3.5 Modelos de Box e Jenkins

Box & Jenkins (1976) apresentaram uma metodologia bastante utilizada na análise de modelos paramétricos. Tal metodologia consiste em ajustar modelos autorregressivos, integrados, de médias móveis, ARIMA(p,d,q) a um conjunto de dados e é baseada em um ciclo iterativo. Esse ciclo consiste em: especificação,

identificação, estimação e diagnóstico. Caso o modelo não seja adequado, o ciclo é repetido, voltando-se à fase de identificação.

Segundo Morettin & Tolo (2006), a fase crítica desse ciclo iterativo é a identificação, pois é possível que vários pesquisadores identifiquem modelos diferentes para a mesma série temporal. E ainda, esse mesmo autor, salienta que as vantagens desses modelos são: que eles são parcimoniosos, isto é, contêm um número pequeno de parâmetros; e que as previsões são bastante precisas, comparando-as com os demais métodos de previsão, tais como os métodos citados por Wheelwright & Makridakis (1998): regressão simples, regressão múltipla, suavização exponencial e decomposição de séries temporais.

Uma notação bastante utilizada nesses modelos é a de operadores. Esses operadores são:

- i. Operador de translação para o passado, denotado por B e definido por

$$BZ_t = Z_{t-1} \quad (3.21)$$

e

$$B^m Z_t = Z_{t-m}. \quad (3.22)$$

- ii. Operador de translação para o futuro, denotado por F e definido por

$$FZ_t = Z_{t+1} \quad (3.23)$$

e

$$F^m Z_t = Z_{t+m}. \quad (3.24)$$

- iii. Operador diferença, denotado por ΔZ_t e definido por

$$\Delta Z_t = Z_t - Z_{t-1} = (1 - B)Z_t \implies \Delta = 1 - B \quad (3.25)$$

e de (3.4) tem-se:

$$\Delta^n Z_t = (1 - B)^n Z_t \implies \Delta^n = (1 - B)^n. \quad (3.26)$$

iv. O operador soma é denotado por S e definido por

$$\begin{aligned} SZ_t &= \sum_{j=0}^{\infty} Z_{t-j} \\ &= (1 + B + B^2 + \dots)Z_t \\ &= (1 - B)^{-1}Z_t \implies S = \Delta^{-1}. \end{aligned} \quad (3.27)$$

3.6 Modelos lineares estacionários

Box *et al.* (1994) afirmam que séries temporais geradas por agregação linear de choques aleatórios são geralmente um modelo linear estocástico. Na FIGURA 3, mostra-se um filtro linear com entrada a_t , saída Z_t e função de transferência ψ_t .

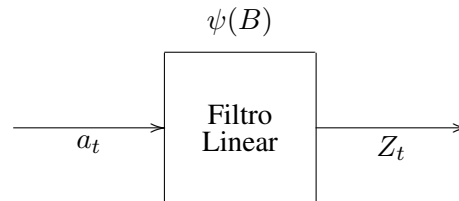


FIGURA 3 Série Temporal como saída de um filtro linear

3.6.1 Processo linear geral

Segundo Morettin & Tolo (2006), esse modelo supõe que a série temporal seja gerada por meio de um filtro linear, (FIGURA 3), cuja entrada é um ruído branco.

Formalmente, tem-se:

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \mu + \psi(B)a_t \quad (3.28)$$

em que

$$\psi(B) = 1 + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots \quad (3.29)$$

é denominada função de transferência do filtro e μ é o parâmetro que determina o nível da série. Chamando $\tilde{Z}_t = Z_t - \mu$, tem-se

$$\tilde{Z}_t = \psi(B)a_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} \quad (3.30)$$

com $\psi_0 = 1$.

O processo linear \tilde{Z}_t , dado por (3.30), é escrito como uma soma ponderada de ruídos branco do presente e do passado. Esse processo consiste em uma sequência de variáveis aleatórias não-correlacionadas com média zero e variância constante, isto é, $E[a_t] = 0$ e $Var[a_t] = \sigma_a^2$. Nessas condições, a função de autocovariância das variáveis aleatórias a_t é dada por:

$$\gamma_k = E[a_t a_{t+k}] = \begin{cases} \sigma_a^2 & \text{se } k = 0, \\ 0 & \text{se } k \neq 0 \end{cases} \quad (3.31)$$

Se a sequência de pesos $\{\psi_j, j \geq 1\}$ for finita ou infinita convergente, isto é, $\sum_{j=0}^{\infty} \psi_j \leq \infty$, o filtro é estável ou somável e \tilde{Z}_t é estacionária. Neste caso, μ é a média do processo. Caso contrário, \tilde{Z}_t é não-estacionária e μ indica o nível da série.

Pode-se escrever \tilde{Z}_t de forma alternativa, como uma soma ponderada de

valores passados de \tilde{Z}_t mais um ruído branco a_t :

$$\begin{aligned}\tilde{Z}_t &= \pi_1 \tilde{Z}_{t-1} + \pi_2 \tilde{Z}_{t-2} + \dots + a_t \\ &= \sum_{j=1}^{\infty} \pi_j \tilde{Z}_{t-j} + a_t \\ a_t &= \tilde{Z}_t - \pi_1 \tilde{Z}_{t-1} - \pi_2 \tilde{Z}_{t-2} - \dots\end{aligned}\quad (3.32)$$

Escrevendo (3.32) em função do operador de translação para o passado, definido em (3.21) e (3.22), segue-se que

$$(1 - \pi_1 B - \pi_2 B^2 - \dots) \tilde{Z}_t = a_t \quad (3.33)$$

então

$$\pi(B) \tilde{Z}_t = a_t \quad (3.34)$$

em que $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$. Multiplicando-se ambos os lados por $\psi(B)$, tem-se

$$\psi(B) \pi(B) \tilde{Z}_t = \psi(B) a_t = \tilde{Z}_t \quad (3.35)$$

Portanto $\psi(B) \pi(B) = 1$, tal que

$$\psi(B) = \pi^{-1}(B). \quad (3.36)$$

A relação (3.36) pode ser usada para determinar os coeficientes (π_1, π_2, \dots) a partir de (ψ_1, ψ_2, \dots) e vice-versa.

Um processo linear será estacionário se a série $\psi(B)$ convergir para $|B| \leq 1$ e será invertível se $\pi(B)$ convergir para $|B| \leq 1$.

A função de autocovariância do processo linear \tilde{Z}_t , dado por (3.30), é

$$\begin{aligned}
 \gamma_k &= E[\tilde{Z}_t \tilde{Z}_{t+k}] \\
 &= E\left[\sum_{j=0}^{\infty} \psi_j a_{t-j} \sum_{h=0}^{\infty} \psi_h a_{t+k-h}\right] \\
 &= E\left[\sum_{j=0}^{\infty} \sum_{h=0}^{\infty} \psi_j \psi_h a_{t-j} a_{t+k-h}\right] \\
 &= \sum_{j=0}^{\infty} \sum_{h=0}^{\infty} \psi_j \psi_h E[a_{t-j} a_{t+k-h}]
 \end{aligned}$$

utilizando (3.31) tem-se que:

$$E[a_{t-j} a_{t+k-h}] = \begin{cases} \sigma_a^2 & \text{se } t-j = t+k-h \implies h = j+k \\ 0 & \text{se } t-j \neq t+k-h \end{cases}$$

Portanto,

$$\gamma_k = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}. \quad (3.37)$$

Os modelos autorregressivos, médias móveis e autorregressivos de médias móveis, AR(p), MA(q) e ARMA(p,q), respectivamente, são casos particulares do processo linear geral.

3.7 Modelo autorregressivo e de médias móveis

O modelo autorregressivo e de médias móveis de ordem (p,q) , ARMA(p,q), é a combinação dos processos autorregressivos de ordem p, AR(p), e de médias móveis de ordem q, MA(q).

Se $q = 0$, tem-se o modelo ARMA(p,0), isto é, autorregressivo de ordem p, AR(p). Esse modelo pode ser escrito como uma soma ponderada finita de valores passados de Z_t mais um ruído branco a_t , ou seja,

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \dots + \phi_p Z_{t-p} + a_t = \sum_{j=1}^p \phi_j \tilde{Z}_{t-j} + a_t \quad (3.38)$$

Isolando a_t , tem-se

$$\begin{aligned} a_t &= \tilde{Z}_t - \phi_1 \tilde{Z}_{t-1} - \dots - \phi_p Z_{t-p} \\ &= \tilde{Z}_t - \phi_1 B \tilde{Z}_t - \dots - \phi_p B^p \tilde{Z}_t \\ &= (1 - \phi_1 B - \dots - \phi_p B^p) \tilde{Z}_t. \end{aligned}$$

em que $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ é o polinômio autorregressivo de ordem p . Portanto,

$$\phi(B) \tilde{Z}_t = a_t \quad (3.39)$$

ou

$$\tilde{Z}_t = \phi^{-1}(B) a_t. \quad (3.40)$$

Como $\phi(B)$ é finito, o processo $AR(p)$ é invertível para qualquer p e será estacionário se as raízes do polinômio, $\phi(B) = 0$, estiverem fora do círculo unitário.

Box *et al.* (1994) consideraram que o modelo autorregressivo, $AR(p)$, é um filtro linear cuja função de transferência é $\phi^{-1}(B)$, a entrada é a_t e a saída é \tilde{Z}_t .

Se $p = 0$, tem-se o modelo $ARMA(0, q)$, isto é, de médias móveis de ordem q , $MA(q)$. Este modelo pode ser escrito como uma soma ponderada finita de ruídos branco do presente e do passado, ou seja,

$$\tilde{Z}_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (3.41)$$

com $\theta_0 = 1$. Pode-se escrever (3.41) da seguinte forma:

$$\begin{aligned}\tilde{Z}_t &= a_t - \theta_1 B a_t - \dots - \theta_q B^q a_t \\ &= (1 - \theta_1 B - \dots - \theta_q B^q) a_t\end{aligned}$$

em que $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ é o polinômio de médias móveis de ordem q . Portanto,

$$\tilde{Z}_t = \theta(B) a_t. \quad (3.42)$$

Como $\theta(B)$ é finito, o processo $MA(q)$ é estacionário para qualquer q e será invertível se as raízes do polinômio, $\theta(B) = 0$, estiverem fora do círculo unitário.

Segundo Box *et al.* (1994), o modelo médias móveis, $MA(q)$, é um filtro linear cuja função de transferência é $\theta(B)$, a entrada é a_t e a saída é Z_t .

Se $p \neq 0$ e $q \neq 0$, o modelo ARMA (p, q) , pode ser escrito como

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \dots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (3.43)$$

ou

$$\phi(B) \tilde{Z}_t = \theta(B) a_t \quad (3.44)$$

em que $\phi(B)$ e $\theta(B)$ são os operadores autorregressivo e de médias móveis, respectivamente.

Uma outra forma de se escrever (3.44) é dada por

$$\tilde{Z}_t = \phi^{-1}(B) \theta(B) a_t. \quad (3.45)$$

Considerando (3.44), tem-se que o processo ARMA (p, q) é um processo linear cuja função de transferência é $\phi^{-1}(B) \theta(B)$.

As condições de estacionariedade e invertibilidade são, respectivamente, que as raízes dos polinômios, $\phi(B) = 0$ e $\theta(B) = 0$, estejam fora do círculo unitário.

3.7.1 Funções de autocorrelação e autocorrelação parcial

Segundo Morettin & Toloí (2006), mutiplicando ambos os membros de (3.43) por \tilde{Z}_{t-k} e tomando-se a esperança desses, obtém-se a função de autocovariância (FAC) de lag k , ou seja,

$$\begin{aligned}\gamma_k &= E[\tilde{Z}_t \tilde{Z}_{t-k}] \\ &= E[(\phi_1 \tilde{Z}_{t-1} + \dots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}) \tilde{Z}_{t-k}] \\ &= \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} + \gamma_{za}(k) - \theta_1 \gamma_{za}(k-1) - \dots - \theta_q \gamma_{za}(k-q)\end{aligned}\quad (3.46)$$

em que γ_{za} é a covariância cruzada entre \tilde{Z}_t e a_t , expressa por $\gamma_{za} = E[\tilde{Z}_{t-k} a_t]$.

Como \tilde{Z}_{t-k} depende apenas de choques a_t ocorridos até o instante $t-k$, obtém-se

$$\gamma_{za} \begin{cases} = 0 & \text{se } k > 0, \\ \neq 0 & \text{se } k \leq 0 \end{cases}.$$

Logo, para $k > q$,

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} \quad (3.47)$$

dividindo (3.47) por γ_0 , obtém-se a função de autocorrelação:

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p}. \quad (3.48)$$

De (3.47), pode-se deduzir que as autocorrelações de lags $k = 1, 2, \dots, q$ são afetadas diretamente pelos parâmetros de médias móveis e para $k > q$ se comportam como uma mistura de polinômios, exponenciais e senóides amortecidas.

Segundo Box *et al.* (1994), pode-se verificar que, para $q < p$, a FAC consiste em uma mistura de exponenciais e senóides amortecidas, entretanto, se $q \geq p$, os primeiros $p - q - 1$ valores de $\rho_0, \rho_1, \dots, \rho_{q-p}$ não seguirão este padrão.

Box & Jenkins (1976) propõem a utilização da função de autocorrelação

parcial (FACP) como um critério para facilitar a identificação do modelo.

O coeficiente de correlação parcial é utilizado para medir o grau de associação entre as observações Z_t e Z_{t-k} , quando os efeitos das defasagens até $k - 1$ são fixadas.

Para se obter os coeficientes de autocorrelação parcial, utilizam-se as equações de Yule-Walker, expressas na forma matricial por

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix}^{-1} = \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} \quad (3.49)$$

em que ϕ_{kj} é o j -ésimo coeficiente de um modelo AR(k) e ϕ_{kk} é o último coeficiente. Substituindo-se os coeficientes ρ_j por suas estimativas r_j em (3.49), têm-se as estimativas dos coeficientes ϕ_{kj} .

Para Morettin & Tolo (2006), as funções de autocorrelação parcial dos processos AR(p), MA(q) e ARMA(p, q) possuem as seguintes características que são úteis na identificação dos modelos:

1. para um processo autorregressivo AR(p) a FACP é tal que: $\phi_{kk} \neq 0$, para $k < p$ e $\phi_{kk} = 0$, para $k > p$;
2. num processo de médias móveis MA(q) a FACP se comporta de maneira similar à FAC de um processo AR(p), sendo determinada por exponenciais e/ou senóides amortecidas;
3. num processo ARMA (p, q) a FACP se comporta como a FACP de um processo MA(q) puro.

3.8 Modelos não estacionários

Segundo Morettin & Tolo (2006), muitas séries encontradas na prática são não estacionárias, ou seja, não se desenvolvem em torno de uma média constante,

como por exemplo, séries econômicas e financeiras que apresentam, em geral, tendências. Pode-se também ter uma forma de não-estacionariedade explosiva, como o crescimento de uma colônia de bactérias.

As séries não-estacionárias cujo comportamento é não explosivo, e que, tomando um número finito de diferenças, tornam-se estacionárias, são chamadas de séries não-estacionárias homogêneas.

3.8.1 Modelos ARIMA(p, d, q)

O modelo autorregressivo, integrado, de médias móveis, ARIMA(p,d,q) é o caso mais geral dos modelos de Box e Jenkins.

Quando a série não apresenta a componente sazonal, mas apresenta a componente tendência, ou estes componentes são não-estacionários homogêneos, ela pode ser representada pelo modelo ARIMA.

Box *et al.* (1994) consideraram o modelo

$$\varphi(B)\tilde{Z}_t = \theta(B)a_t. \quad (3.50)$$

em que $\varphi(B)$ é um operador autorregressivo não-estacionário de ordem, $p + d$, com d raízes sobre o círculo unitário e p raízes fora, ou seja

$$\varphi(B)\tilde{Z}_t = \phi(B)\Delta^d\tilde{Z}_t = \phi(B)(1 - B)^d\tilde{Z}_t \quad (3.51)$$

logo

$$\phi(B)\Delta^d\tilde{Z}_t = \theta(B)a_t \quad (3.52)$$

em que $\Delta^d\tilde{Z}_t$ é estacionária. Tomando

$$W_t = \Delta^d\tilde{Z}_t \quad (3.53)$$

e substituindo em (3.52), tem-se:

$$\phi(B)W_t = \theta(B)a_t. \quad (3.54)$$

Portanto, o modelo (3.52), pode ser representado por um modelo ARMA(p,q). Uma alternativa para escrever (3.53), é

$$\tilde{Z}_t = \frac{W_t}{\Delta^d} = W_t \Delta^{-d} = W_t S^d \quad (3.55)$$

em que S^d é um operador soma. A equação (3.55), implica que o processo estacionário (3.52) pode ser obtido por uma soma (integral), daí o modelo autorregressivo, integrado, de médias móveis.

3.8.2 Modelos ARIMA sazonal - SARIMA

Quando uma série temporal apresenta um componente periódico em um período máximo de 12 meses (sazonalidade), é necessário acrescentar uma componente sazonal no modelo. Podem-se ter dois tipos de modelos sazonais: os determinísticos e os estocásticos.

Segundo Morettin & Toloi (2006), uma série sazonal é estocástica quando apresenta correlações significativas nos *lags* sazonais, isto é, múltiplos de período s , e é determinística, quando torna-se estacionária após tomar D diferenças da série. Muitas vezes, a série pode apresentar essas duas características.

Box & Jenkins (1976) propõem o modelo sazonal multiplicativo, denotado por SARIMA $(p, d, q)(P, D, Q)_s$, representado por

$$\phi(B)\Phi(B^s)\Delta^d\Delta_s^D Z_t = \theta(B)\Theta(B^s)a_t, \quad (3.56)$$

em que

- $\phi(B) = 1 - \phi_1 B^1 - \dots - \phi_p B^p$ é o polinômio autorregressivo de ordem p ;
- $\Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps}$ é o polinômio autorregressivo sazonal de ordem P ;
- $\Delta^d = (1 - B)^d$ é o operador diferença e d é o número de diferenças necessárias para retirar a tendência da série;

- $\Delta_s^D = (1 - B^s)^D$ é o operador diferença generalizado, quando duas observações estão distantes entre si de s intervalos de tempos que apresentam alguma semelhança, e D é o número de diferenças de lags s necessárias para retirar a sazonalidade da série;
- $\theta(B) = 1 - \theta_1 B^1 - \dots - \theta_q B^q$ é o polinômio médias móveis de ordem q ;
- $\Theta(B^s) = 1 - \Theta_s B^s - \dots - \Theta_Q B^{Qs}$ é o polinômio médias móveis sazonal de ordem Q .

Quando a série possui apenas sazonalidade estocástica, ajusta-se o modelo SARIMA com $D = 0$, e se possui apenas sazonalidade determinística, ajusta-se o modelo SARIMA, com $P = Q = 0$.

3.9 Identificação dos modelos ARIMA

Segundo Morettin & Tolo (2006), a identificação do modelo ARIMA é a fase mais crítica do ciclo iterativo do método de Box e Jenkins. Essa escolha é baseada, principalmente, com base nas autocorrelações e autocorrelações parciais estimadas. Geralmente, essa técnica é eficiente, mas, é importante salientar que pode ser identificado mais de um modelo para a mesma série.

O objetivo da identificação é determinar os valores de p, d e q do modelo ARIMA(p,d,q) e também P, D e Q , quando a série possuir a componente sazonal (SARIMA).

O procedimento de identificação consiste em:

1. verificar se a série é estacionária, por meio da análise do gráfico da série original e sua função de autocorrelação;
2. diferenciar a série até que se torne estacionária por meio do operador diferença. O número d de diferenças necessárias é alcançado quando a FAC amostral da série $W_t = \Delta^d Z_t$ decai rapidamente para zero;

3. identificar os valores de p e q , por meio da análise das autocorrelações e autocorrelações parciais, e também P e Q , se a série apresentar sazonalidade estocástica.

3.10 Estimação dos modelos ARIMA

Segundo Morettin & Tolo (2006), tendo-se identificado o modelo provisório para a série, o passo seguinte é estimar os parâmetros. Considerando o modelo ARIMA(p,d,q), conforme (3.54), estima-se o vetor de parâmetros $\xi = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_a^2)$, utilizando-se um procedimento iterativo de estimação não-linear de mínimos quadrados. Os $p + q + 1$ parâmetros do modelo são analisados sob a suposição que $a_t \sim N(0, \sigma_a^2)$.

Um dos métodos utilizados para estimar os parâmetros de ξ é o da máxima verossimilhança. Dada a série $Z_t = (Z_1, \dots, Z_N)$, em que N é o número de observações, os estimadores de máxima verossimilhança (EMV) de ξ são os valores que maximizam a função de verossimilhança $L(\xi|Z_1, \dots, Z_N)$ ou $l = \log L$.

Segundo Box *et al.* (1994), o logaritmo da função de verossimilhança não-condicional é dado por

$$l(\xi) \simeq -n \log \sigma_a - \frac{S(\eta)}{2\sigma_a^2} \quad (3.57)$$

em que

$$S(\eta) = S(\phi, \theta) = \sum_{t=-\infty}^n [a_t(\eta, W)]^2 \quad (3.58)$$

é a soma de quadrados não-condicional com $[a_t(\eta, W)] = E(a_t|\eta, W)$ e $W = W_t = \Delta^d Z_t$ é a série diferenciada com $n = N - d$ observações, sendo d o número de diferenças necessárias para obter estacionariedade.

Segue-se que os estimadores de mínimos quadrados (EMQ), obtidos minimizando-se (3.58), serão boas aproximações para os EMV.

3.11 Verificação dos modelos ARIMA

Após a fase de estimação, o próximo passo é a verificação da adequação do modelo ajustado. O principal procedimento utilizado nessa fase é a análise dos resíduos.

Teste da autocorrelação residual

Para que o modelo definido por (3.54) seja adequado, os resíduos devem ser independentes e normalmente distribuídos (ruído branco). Portanto, suas estimativas

$$\hat{a}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)W_t \quad (3.59)$$

deverão estar próximas de a_t e, aproximadamente, não correlacionadas, isto é, a estimativa da autocorrelação \hat{r}_k é, aproximadamente, igual a zero.

Logo, sob a suposição que o modelo ajustado é adequado, tem-se que

$$\hat{r}_k = \frac{\sum_{t=k+1}^n \hat{a}_t \hat{a}_{t-k}}{\sum_{t=1}^n \hat{a}_t^2} \sim N\left(0, \frac{1}{N}\right). \quad (3.60)$$

Pode-se dizer que o modelo é adequado quando \hat{r}_k estiver dentro dos limites de $\pm \frac{2}{\sqrt{N}}$.

Segundo Morettin & Tolo (2006), a comparação de \hat{r}_k com estes limites fornece uma indicação geral de uma possível quebra de comportamento de ruído branco em a_t , com a condição de que seja lembrado que, para pequenos valores de k , esses limites subestimarão a significância de qualquer discrepância.

Teste de Box Pierce

Box & Pierce (1970) sugeriram um teste para estimação da autocorrelação dos resíduos, baseado nas primeiras k autocorrelações de \hat{a}_t , isto é, nos valores de $r_k(\hat{a})$ para $k = \{1, 2, \dots, K\}$.

Se o modelo ajustado para um processo ARIMA(p,d,q) é apropriado, então

$$Q = n(n+2) \sum_{k=1}^K \frac{\hat{r}_k^2}{n-k} \sim \chi_{k-p-q}^2. \quad (3.61)$$

A hipótese de ruído branco é rejeitada para $Q > \chi_{k-p-q}^2$.

No modelo SARIMA, consideram-se os coeficientes sazonais nos graus de liberdade da distribuição qui-quadrado, tendo, então, $(k - p - q - P - Q)$ graus de liberdade. Neste estudo, serão tomadas as primeiras 40 \hat{r}_k .

3.12 Previsões

Uma das principais razões para decompor uma série temporal em seus componentes é prever valores futuros.

Morettin & Tolo (2006) salientam que a previsão não constitui um fim em si, mas apenas um meio de fornecer informações para uma consequente tomada de decisões, visando a determinar um objetivo.

A previsão Z_{t+h} denotada por $\hat{Z}_t(h)$ é a esperança matemática condicional de Z_{t+h} , dado $Z_t, Z_{t-1} \dots$, isto é,

$$\hat{Z}_t(h) = [Z_{t+h}] = E[Z_{t+h} | Z_t, Z_{t-1}, \dots]. \quad (3.62)$$

Previsões com modelos ARIMA

O modelo ARIMA(p, d, q) pode ser escrito na forma da equação de diferenças

$$Z_{t+h} = \varphi_1 Z_{t+h-1} + \dots + \varphi_{p+d} Z_{t+h-p-d} - \theta_1 a_{t+h-1} - \dots - \theta_q a_{t+h-q} + a_{t+h}, \quad (3.63)$$

tomando a esperança condicional de (3.63), tem-se a equação de previsão

$$\begin{aligned}\hat{Z}_t(h) = & \varphi_1 E[Z_{t+h-1}|Z_{t-1}, Z_{t-2}, \dots] + \dots + \varphi_{p+d} E[Z_{t+h-p-d}|Z_{t-p-d}, \dots] \\ & - \theta_1 E[a_{t+h-1}|a_{t-1}, a_{t-2}, \dots] - \dots - \theta_q E[a_{t+h-q}|a_{t-q}, \dots] \\ & + E[a_{t+h}|a_t, a_{t-1}, \dots],\end{aligned}\quad (3.64)$$

logo

$$\begin{aligned}\hat{Z}_t(h) = & \varphi_1 [Z_{t+h-1}] + \dots + \varphi_{p+d} [Z_{t+h-p-d}] \\ & - \theta_1 [a_{t+h-1}] - \dots - \theta_q [a_{t+h-q}] + [a_{t+h}],\end{aligned}\quad (3.65)$$

para $h \geq 1$. Utilizam-se os seguintes fatos:

$$\begin{aligned}[Z_{t+k}] &= \hat{Z}_t(k), k > 0, \\ [Z_{t+k}] &= Z_{t+k}, k \leq 0, \\ [a_{t+k}] &= 0, k > 0 \\ [a_{t+k}] &= a_{t+k}, k \leq 0.\end{aligned}\quad (3.66)$$

Considerando-se a equação de previsão como uma função de h , com origem t fixa, tem-se

$$\hat{Z}_t(h) = \sum_{i=1}^{p+d} \varphi_i \hat{Z}_{t+h-i}, h > q \quad (3.67)$$

ou

$$\varphi(B) \hat{Z}_t(h) = (1 - B)^d \hat{Z}_t(h) = 0, h > q \quad (3.68)$$

com $\varphi(B)$ operando em h .

Morettin & Toloi (2006) mostram que para $h > q - p - d$, a função $\hat{Z}_t(h)$ consistirá numa mistura de polinômios exponenciais e senóides amortecidas, com sua forma exata determinada pelas raízes G_i^{-1} do operador $\varphi(B) = 0$.

O erro da previsão é definido por

$$e_t(h) = Z_{t+h} - \hat{Z}_t(h) \quad (3.69)$$

e a sua variância por

$$V(h) = (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{h-1}^2)\sigma_a^2. \quad (3.70)$$

Intervalo de confiança

Para obter o intervalo de confiança para Z_{t+h} é necessário fazer as seguintes suposições para os erros:

1. $E[a_t] = 0$;
2. $E[a_t a_s] = 0, t \neq s$;
3. $Var[a_t] = \sigma_a^2, \forall t$;
4. $a_t \sim N(0, \sigma_a^2)$, para cada t .

Dessa forma, dados os valores passados e presentes da série Z_t, Z_{t-1}, \dots , a distribuição condicional da série é $N(\hat{Z}_t(h), V(h))$.

Tem-se que

$$U = \frac{Z_{t+h} - \hat{Z}_t(h)}{[V(h)]^{\frac{1}{2}}} \sim N(0, 1). \quad (3.71)$$

Pode-se obter o valor de u_α , fixando-se o coeficiente α , tal que o intervalo de confiança para Z_{t+h} com $P(-u_\alpha < U < u_\alpha) = \alpha$ é dado por

$$\hat{Z}_t(h) - u_\alpha[V(h)]^{\frac{1}{2}} \leq Z_{t+h} \leq \hat{Z}_t(h) + u_\alpha[V(h)]^{\frac{1}{2}}. \quad (3.72)$$

3.13 Critério para escolha do modelo

A seguir serão apresentados os critérios para escolha do modelo:

Critério de Akaike

Akaike (1973) sugere escolher como melhor modelo, aquele cujas ordens p e q minimizam o critério de Akaike (AIC) normalizado. Na comparação de

diversos modelos, com N fixo, o critério de Akaike pode ser expresso por:

$$AIC = N \log \hat{\sigma}_a^2 + 2(k + l + 1) \quad (3.73)$$

em que $\hat{\sigma}_a^2$ é o estimador de máxima verossimilhança de σ_a^2 , $0 \leq k \leq p$ e $0 \leq l \leq q$. Onde k e l correspondem ao número de parâmetros que minimizam o critério, independente do modelo ser completo ou incompleto. Entre os vários modelos propostos, será escolhido como o melhor modelo aquele que tiver o menor valor de AIC.

Critério Bayesiano

O critério de informação bayesiano (BIC), proposto por Schwarz (1978) tem sido amplamente utilizado para seleção de modelos de séries temporais e regressão linear. O BIC reduz ao máximo o fator de verossimilhança, porque o número de parâmetros é igual ao número de parâmetros do modelo de interesse.

$$BIC = -2 \cdot \ln p(x|k) \approx \text{BIC} = -2 \cdot \ln L + k \ln(n) \quad (3.74)$$

em que x são os dados observados; n é o número de observações em x ; k é o número de parâmetros a serem estimados; $p(x|k)$ é o parâmetro da verossimilhança dado pelo número de parâmetros; e L é o valor maximizado da função de verossimilhança para o modelo estimado. O melhor modelo é aquele que apresentar o menor valor para o critério de informação bayesiano.

Critério do erro quadrático médio de previsão

Segundo Morettin & Tolo (2006), quando o critério de decisão for adotar o modelo que fornece as melhores previsões, o critério do erro quadrático médio de previsão (EQMP) é uma boa medida de comparação entre dois modelos.

As estimativas EQMP são dadas pela média dos quadrados das diferenças entre valores observados e valores previstos. Logo, o EQMP com origem em t é

dado por

$$EQMP_t = \frac{1}{n} \sum_{h=1}^n [Z_{t+h} - \hat{Z}_t(h)]^2. \quad (3.75)$$

O modelo escolhido é o que possuir menor EQMP.

Mean Absolute Percentage Error

De acordo com Morettin & Tolo (2006), uma das medidas de erro utilizadas para analisar os erros gerados pelas previsões encontradas é o MAPE que é dado por:

$$MAPE = \frac{1}{n} \cdot \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (3.76)$$

onde A_t é o valor real e F_t é o valor previsto. A diferença entre A_t e F_t é dividido por A_t . O valor absoluto desse cálculo é somado para cada ponto previsto no tempo e então dividido novamente pelo número de observações previstas. O modelo escolhido é aquele que possuir menor MAPE.

Nesse capítulo foram apresentados alguns conceitos dos modelos clássicos para análise e previsão de série temporais. Inicialmente, apresentaram-se algumas definições básicas de séries temporais a fim de auxiliar o entendimento sobre os modelos existentes. Baseando-se nesses conceitos, modelos lineares estacionários foram descritos, seguindo a metodologia Box & Jenkins. Foram descritos também os modelos não estacionários ARIMA (p, d, q) e ARIMA sazonal (SARIMA).

O próximo capítulo apresentará conceitos importantes sobre Redes Neurais Artificiais que são sistemas computacionais baseados no mecanismo de funcionamento do cérebro humano e, de que forma esses sistemas podem ser utilizados na análise e previsão de séries temporais.

4 REDES NEURAIIS ARTIFICIAIS

Redes Neurais Artificiais (RNAs) são poderosas ferramentas para tratar uma grande diversidade de problemas. Haykin (1994) diz que o modelo *perceptron* multicamadas vem sendo frequentemente aplicado na resolução de uma variedade de problemas de classificação e reconhecimento de padrões, processamento de sinais, aproximação de funções e previsão de séries temporais.

O algoritmo fundamental para a aprendizagem da rede multicamadas é o algoritmo *back-propagation*, que é baseado no método do gradiente descendente. A derivação do algoritmo de retropropagação e considerações sobre as virtudes e limitações das redes *perceptron* multicamadas aplicadas ao problema de previsão de vazões também serão comentadas nesta seção.

4.1 Introdução

As RNAs constituem sistemas de processamento distribuído paralelo que, por serem baseados no funcionamento do sistema nervoso biológico, possuem propensão natural de armazenar conhecimento experimental, Haykin (1994). Assim como o cérebro humano é capaz de transmitir e/ou armazenar informações por meio de impulsos nervosos entre os neurônios, as RNAs possuem conexões inter-neurais, representadas por pesos sinápticos, com função de armazenar e processar o conhecimento.

Do ponto de vista estrutural, as RNAs podem ser classificadas como redes neurais estáticas ou como redes neurais dinâmicas, segundo Haykin (1994). A principal diferença funcional entre as duas estruturas é a presença ou não de conexões que realimentem os neurônios presentes na arquitetura da rede. Quando não existe dependência das saídas atuais com seus valores passados, as redes são

estáticas. Entre essas estruturas, se destacam as rede neurais multicamadas (*Multi Layer Perceptron* - MLP) por serem amplamente utilizadas na literatura.

Por outro lado, as redes recorrentes são capazes de representar uma grande variedade de comportamentos dinâmicos, uma vez que a presença de realimentação de informação permite a criação de representações internas e dispositivos de memória capazes de processar e armazenar informações temporais e sequenciais, conforme descreve Zuben (1996).

4.2 Breve Histórico da área de RNAs e sua Aplicação em Previsão

O primeiro modelo matemático para um neurônio biológico foi proposto por McCulloch & Pitts (1943) e considera a saída do neurônio como uma resposta binária a um estímulo fornecido. Conforme Haykin (1994), apesar de apresentar grande poder computacional, faltavam-lhe características adaptativas e de aprendizagem, imprescindíveis em modelos de sistemas inteligentes.

Após a publicação do trabalho de McCulloch e Pitts, muitos foram os avanços na área de redes neurais durante as décadas subsequentes por pesquisadores de áreas como a neurofisiologia com Hebb (1949) e a engenharia com Rosenblatt (1958), introduzindo e aperfeiçoando paradigmas de aprendizagem. Em meados da década de 60, Hu (1964) sugere o primeiro modelo de RNA para previsão de séries temporais, utilizado em dados meteorológicos.

Apesar do aparente sucesso dos modelos desenvolvidos, em 1969, Minsky (1969) mostrou as limitações da rede *perceptron* com uma única camada para problemas não linearmente separáveis, o que levou a um declínio nas investigações de metodologias relacionadas à área de redes neurais.

Entretanto, Zuben (1996) afirma que alguns trabalhos importantes foram publicados durante a década de 60, destacando-se os mapas auto-organizáveis de Kohonen (1982). No início da década de 80, o interesse na teoria das RNAs foi reativado após a publicação dos estudos de Hopfield (1982) baseados no princípio físico de armazenamento de informação em configurações

dinamicamente estáveis, sendo este um dos primeiros modelos a introduzir dinâmica em RNAs.

Em 1986, por meio da publicação do livro *Parallel Distributed Processing* de Rumelhart *et al.* (1986), o algoritmo de retropropagação (*backpropagation*), previamente proposto por Werbos (1974), foi divulgado e, atualmente, é o algoritmo mais popular para o treinamento de RNAs estáticas multicamadas. A partir disso, pesquisadores de diversas áreas aumentaram seu interesse em aplicações de RNAs, particularmente no campo de previsão de séries temporais, foco deste trabalho. Nesse contexto, Weigend *et al.* (1990), conduziram um trabalho sobre estruturas neurais para previsão de séries temporais reais.

Segundo Zhang *et al.* (1998), além de considerarem as não linearidades presentes entre as entradas e as saídas, as RNAs possuem características muito interessantes:

- Não requerem, *a priori*, informações sobre o modelo, pois aprendem a partir de exemplos e, assim, capturam relações funcionais entre os dados mesmo se tais relações sejam dificilmente descritíveis. Para que uma RNA possa aprender e garantir um bom desempenho no resultado, o conjunto de dados disponíveis deve ser suficientemente diverso. Nesse sentido, esses modelos podem ser tratados como um método estatístico multivariado não-paramétrico e não-linear.
- Têm a capacidade de generalizar. Após a fase de aprendizagem, quando já possuem algum conhecimento sobre as características do conjunto de dados, as RNAs conseguem inferir na população restante mesmo se os dados amostrais contêm ruído nas informações.
- As MLP são aproximadores universais, isto é, podem aproximar uma função contínua, definida sob um domínio compacto, com a precisão desejada. Exemplos: funções seno e cosseno, funções polinomiais, dentre outras.

Diversos estudos, comparando o desempenho dos modelos de RNAs e das clássicas abordagens propostas por Box *et al.* (1994) para previsão de séries temporais vêm sendo realizados. Entre outros trabalhos, destacam-se Tang *et al.* (1991), Tang & Fishwick (1993), Gershenfeld & Weigend (1994).

A seguir, os parâmetros que compõem a estrutura de uma RNA e métodos de aprendizagem são descritos detalhadamente.

4.3 Componentes e Treinamento de uma RNA

O funcionamento e o desempenho de uma RNA dependem diretamente da escolha dos componentes adotados em sua estrutura e dos métodos adotados em seu processo de aprendizagem.

4.3.1 O Neurônio Artificial e as Funções de Ativação

Inicialmente, é necessário compreender a unidade básica de processamento de uma RNA, denominada neurônio artificial. Essa unidade processa suas entradas por meio da soma ponderada e de uma função de ativação. A função de ativação é o primeiro componente a ser escolhido na estrutura de um neurônio para uma dada aplicação.

Na FIGURA 4, mostra-se o modelo genérico de um neurônio artificial onde $\varphi(\cdot)$ é a função de ativação do neurônio.

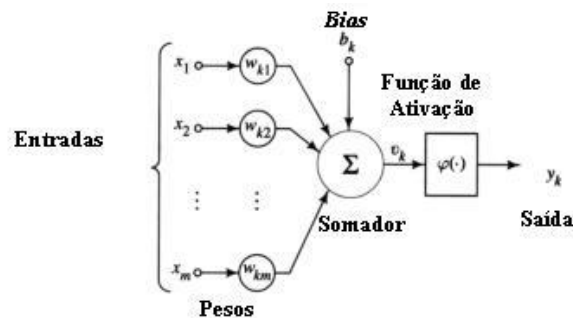


FIGURA 4 Modelo de neurônio artificial

Matematicamente, a saída do neurônio é dada por:

$$y_k = f\left(\sum_{i=1}^m x_i w_{ki}\right) = f(\mu) \quad (4.1)$$

sendo $x_i, i = 1, \dots, m$, as entradas e w_i os pesos sinápticos correspondentes.

Para o modelo de neurônio proposto por McCulloch & Pitts (1943), a função de ativação $f(\mu)$ é uma função degrau da seguinte forma:

$$f(\mu) = \begin{cases} 1 & \text{se } \mu > b_k, \\ 0 & \text{se } \mu \leq b_k. \end{cases} \quad (4.2)$$

A função de ativação pode ser definida principalmente dentro de quatro tipos mais comuns de funções: degrau (*threshold*), linear, logística ou tangente hiperbólica, conforme descreveu Haykin (1994). Na FIGURA 5, apresentam-se os gráficos das principais funções de ativação:

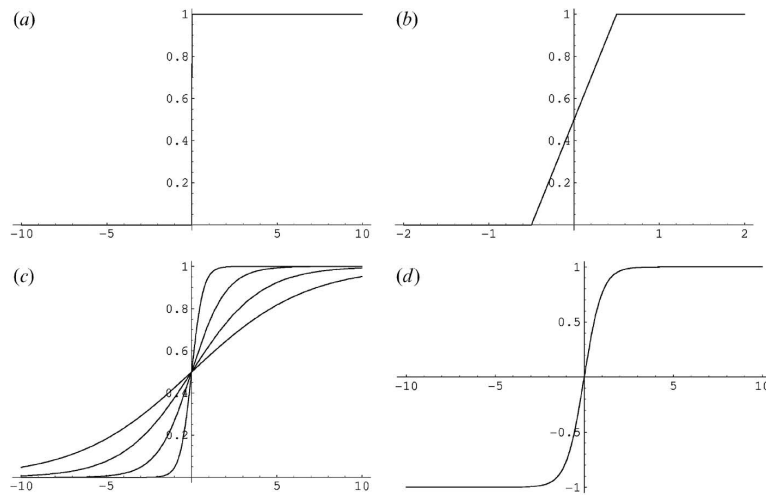


FIGURA 5 Funções de ativação (a) degrau (*threshold*), (b) linear, (c) logística (*sigmoid*) e (d) tangente hiperbólica

A função de ativação linear é descrita da seguinte forma:

$$f(\mu) = \begin{cases} 0, & \text{se } \mu \leq \alpha, \\ \frac{z-\alpha}{\beta-\alpha}, & \text{se } \alpha < \mu \leq \beta, \\ 1, & \text{se } \mu > \beta. \end{cases} \quad (4.3)$$

com, α e β previamente definidos. Já as funções logística e tangente hiperbólica são representadas, respectivamente, como:

$$f(z) = \frac{1}{1 + e^{-\gamma z}} \quad (4.4)$$

$$f(z) = \frac{e^{\gamma z} - e^{-\gamma z}}{e^{\gamma z} + e^{-\gamma z}} \quad (4.5)$$

sendo, γ o ponto de inflexão de uma função sigmoial. Nota-se que, as funções sigmoiais (4.4) e (4.5) possuem suas imagens em intervalos diferentes, isto é, para a função logística $f(z) \in [0, 1]$, enquanto que, para a tangente hiperbólica $f(z) \in [-1, 1]$.

A organização dos neurônios de uma RNA é feita em camadas, sendo a primeira, denominada camada de entrada e a última, camada de saída, conforme FIGURA 6. O número de camadas intermediárias e o número de neurônios em cada uma delas são escolhidos empiricamente de acordo com a maior ou menor necessidade de se processar as entradas da rede, isto é, os números de camadas e de neurônios aumentam, caso o conjunto de dados exija um processamento mais refinado para capturar características não lineares.

Os neurônios recebem as informações das entradas e propagam-nas por meio dos pesos $w_i, i = 1, \dots, m$, entre as camadas, procedimento que simula o processamento sináptico. Haykin (1994) afirma que os pesos sinápticos são parâmetros adaptáveis ajustados via processo de treinamento. Esse processo pode ser classificado, principalmente, como treinamento supervisionado e treinamento não supervisionado. Ambos serão descritos com mais detalhes a seguir.

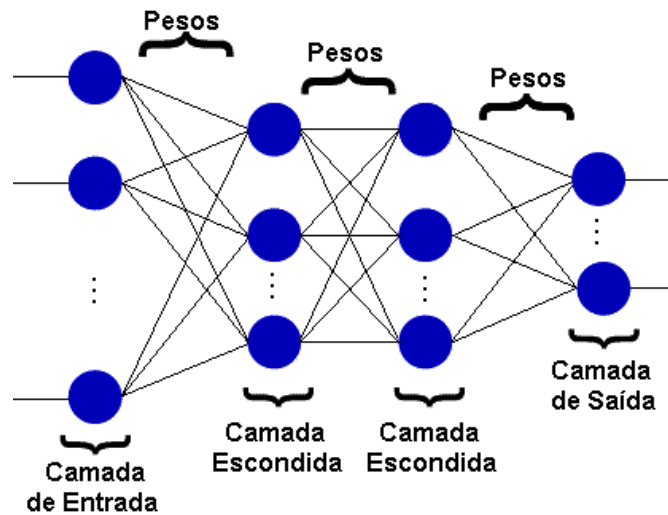


FIGURA 6 Esquema de uma rede de multicamadas
 Fonte: <http://www.lncc.br/labinfo/tutorialRN/>

4.3.2 Treinamento Supervisionado

Haykin (1994) esclarece que treinamento supervisionado caracteriza-se pela necessidade de conhecimento ou informação sobre um sistema na forma de padrões entrada-saída. Aplicando um algoritmo de treinamento supervisionado, uma RNA adquire conhecimento ou informação relevante sobre um problema de forma análoga àquela utilizada pelo ser humano e outros animais, ou seja, a partir de exemplos e de experiência. Zuben (1996) complementa dizendo que essa experiência é baseada apenas nos exemplos fornecidos pelo conjunto de treinamento. Caso apareça uma situação nova e particular, a RNA pode apresentar generalização na modelagem, diminuindo o desempenho.

Um processo de treinamento supervisionado segue, basicamente, o seguinte esquema. Considerando um conjunto de treinamento composto por padrões entrada-saída desejada; para uma dada entrada, o objetivo é ajustar os parâmetros da rede por meio da comparação entre a saída da rede e a saída desejada

correspondente. O ajuste é feito com base no sinal de erro, calculado pela diferença entre a saída desejada (conjunto de treinamento) e a saída da rede.

Segundo Rumelhart *et al.* (1986), o algoritmo de treinamento mais difundido e utilizado na literatura é o de retropropagação (*backpropagation*). Esse método consiste em duas fases. Na primeira, conhecida como fase de propagação direta (*forward*), as entradas são apresentadas e propagadas por meio da rede, camada a camada, calculando a saída de cada neurônio. Durante essa fase os pesos são fixos e a saída calculada é comparada com a saída desejada, resultando em um erro para cada unidade de saída. Na segunda fase, o erro calculado é propagado da camada de saída para a camada de entrada, fase de propagação reversa (*backward*), e os pesos são ajustados de acordo com a regra de correção do erro, originando o termo retropropagação do erro. A segunda fase pode ser formulada como um problema de otimização não linear irrestrita, tendo os pesos sinápticos como variáveis de decisão e o objetivo é minimizar a somatória dos erros quadráticos de todos os neurônios de saída. No algoritmo de retropropagação tradicional, para resolver o problema de otimização, adota-se o método do gradiente descendente, que considera apenas condições de primeira ordem a partir da função objetivo.

Jacobs (1988) conclui esclarecendo que muitas metodologias de treinamento foram desenvolvidas com base no algoritmo de retropropagação a fim de aumentar o desempenho ou a velocidade de convergência.

4.3.3 Treinamento Não Supervisionado

Bishop (1997) relata que nos métodos não supervisionados o treinamento se processa sem que as saídas corretas sejam fornecidas. Esses métodos capacitam as redes neurais a extrair correlações presentes nos dados de entrada. Após a aprendizagem, a rede neural torna-se capaz de classificar suas entradas de acordo com as correlações aprendidas.

O método utilizado durante essa etapa é baseado em métodos competitivos utilizados pelas redes de Kohonen, ou seja, na aprendizagem, após receberem um

sinal de entrada, os neurônios competem entre si e vence aquele que apresentar a menor distância entre os correspondentes pesos sinápticos e as entradas.

4.4 Redes Neurais Multicamadas

A arquitetura da rede MLP, mostrada na FIGURA 7, é composta por neurônios conectados, seguindo uma formação em camadas. Essas conexões são responsáveis por propagar as entradas $x_i, i = 1, \dots, M$. As unidades de entrada, que compõem essa camada, têm o objetivo de difundir o sinal inicial sem nenhuma modificação para a segunda camada. Na FIGURA 7, o j -ésimo neurônio da camada $k, k = 1, \dots, K$ é representado por p_j^k e está conectado ao neurônio p_i^{k-1} da camada anterior com o peso sináptico w_{ji}^k correspondente. O sinal de entrada é processado ao longo da rede e, após percorrer todas as suas camadas, produz uma resposta gerada pelos neurônios da camada de saída. Essa resposta é representada pelo sinal $y_j, j = 1, \dots, M_K$.

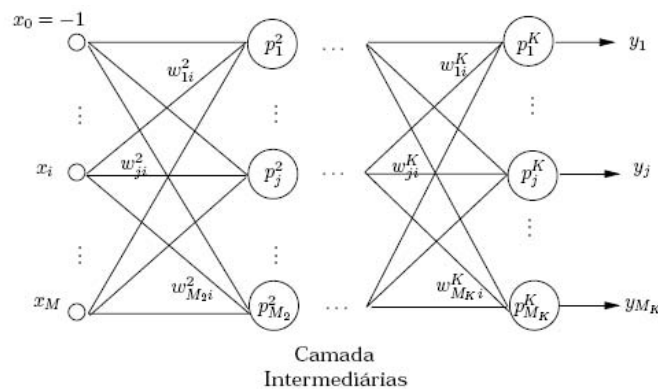


FIGURA 7 Arquitetura de uma MLP

Assim, Haykin (1994) apresenta três características distintas na arquitetura de uma MLP:

- Variações quanto ao número de camadas e neurônios intermediários da rede;
- Tipo de conexão determinada pelas sinapses (estática ou inibitória);

- Função de ativação.

Como foi dito na Seção 4.3.2, o método mais popular para o treinamento supervisionado é o de retropropagação do erro. A extração do conhecimento é realizada a partir de um conjunto de N padrões entrada-saída $(x^1, d^1), \dots, (x^l, d^l), \dots, (x^N, d^N)$, com $x^l = (x_1^l \dots x_i^l \dots x_M^l)$ e $d^l = (d_1^l \dots d_j^l \dots d_{M_K}^l)$. Na fase de propagação direta (*forward*), o sinal de entrada x^l é propagado camada a camada e a respectiva saída $y^l = (y_1^l \dots y_j^l \dots y_{M_K}^l)$ é comparada à saída desejada d^l . Assim, para cada neurônio p_j^K , $j = 1, \dots, M_K$ da camada de saída, na t -ésima apresentação de um padrão (x^l, d^l) , define-se um erro associado:

$$e_j^l(t) = d_j^l(t) - y_j^l(t) \quad (4.6)$$

O valor instantâneo do erro quadrático para o neurônio de saída j é definido como sendo $\frac{1}{2}(e_j^l(t))^2$. A soma dos erros quadráticos é obtida para todos os neurônios da camada de saída, isto é:

$$\varepsilon^l(t) = \frac{1}{2} \sum_{j=1}^{M_K} (e_j^l(t))^2 \quad (4.7)$$

Considerando o número total N de pares entrada-saída contidos no conjunto de treinamento, o erro quadrático médio é obtido pela soma dos valores instantâneos dos erros quadráticos correspondentes a cada padrão, normalizado com relação ao número de padrões N :

$$E = \frac{1}{N} \sum_{l=1}^N \varepsilon^l \quad (4.8)$$

O objetivo do processo de aprendizagem é ajustar os parâmetros livres (pesos) da rede para minimizar E . Para a derivação do algoritmo de retropropagação, é considerado um método de treinamento em que os pesos são ajustados padrão-a-padrão, ou seja, o ajuste dos pesos é realizado a partir do erro

calculado para cada entrada apresentada à rede, conforme a Equação (4.7).

Para a resolução do problema de otimização do erro, Equação (4.8), adota-se o algoritmo de aproximação estocástica. Essa aproximação é aplicada em problemas em que o ambiente é desconhecido, como, por exemplo, no caso do treinamento de retropropagação, em que a função a ser minimizada é definida por amostragem e não tem uma equação conhecida. Dessa forma, os pesos seguem uma trajetória aleatória ao minimizar uma estimativa instantânea da função E , Equação (4.8). Haykin (1994) define então que o vetor gradiente é "aleatório" e sua precisão aumenta à medida que o número de iterações t tende ao infinito. A correção do erro é calculada para cada camada k , composta por M_k neurônios da seguinte forma:

$$\Delta w_{ji}^k(t) = -\alpha \frac{\partial E(t)}{\partial w_{ji}^k(t)} \quad (4.9)$$

sendo α a taxa de aprendizagem.

Para aumentar a velocidade de aprendizagem sem aumentar a taxa de aprendizagem e tornar a rede instável, utiliza-se um método simples que modifica a regra (4.9), adicionando um termo *momentum*, como proposto por Rumelhart *et al.* (1986):

$$\Delta w_{ji}^k(t) = -\alpha \frac{\partial E(t)}{\partial w_{ji}^k(t)} + \beta \Delta w_{ji}^k(t-1) \quad (4.10)$$

ou seja, uma generalização da Equação (4.9) para $\beta \neq 0$. O termo *momentum* relaciona as alterações dos pesos sinápticos na iteração t com as alterações realizadas na iteração $t-1$. A utilização desta constante dificulta a mudança de tendência na atualização dos pesos, podendo ainda ser usada para reduzir a possibilidade de ocorrer mínimos locais conforme descreve Haykin (1994).

Explicitando a função $E(t)$ e aplicando sucessivamente a regra da cadeia, calcula-se a derivada parcial $\partial E(t)/\partial w_{ji}^k$. Dessa forma, o ajuste dos pesos

sinápticos é dado por:

$$w_{ji}^k(t+1) = w_{ji}^k(t) + \Delta w_{ji}^k(t) \quad (4.11)$$

Haykin (1994) exemplifica que a aprendizagem termina, caso a rede tenha alcançado o índice de desempenho estabelecido, isto é, $\|E\| \leq \delta_{max}$, sendo $\|\cdot\|$ a norma Euclidiana e δ_{max} um limiar suficientemente pequeno. Alternativamente, a aprendizagem termina caso o número de iterações ultrapasse um valor máximo especificado.

As abordagens de inteligência artificial, como modelagem de redes neurais, têm sido adotadas com entusiasmo por pesquisadores, e profissionais, em recursos hídricos. Ao fazerem uso dessa nova abordagem, alguns pesquisadores têm encontrado dificuldades associadas aos seus conhecimentos prévios sobre a estatística tradicional. A próxima seção tentará encurtar um pouco essa distância que existe entre as definições das áreas de estudo que serão utilizadas nesse trabalho: Estatística e Redes Neurais.

4.5 Estatística e Redes Neurais Artificiais

Ao contrário do que possa parecer à primeira vista, as redes neurais artificiais têm vários pontos de contato com os modelos estatísticos e econométricos tradicionais. Essas ligações acabam perdendo o sentido em razão do uso de jargões técnicos distintos.

Quando a interação entre as comunidades de pesquisadores em estatística e pesquisadores em redes neurais era muito pequena, e os modelos de cada uma eram desenvolvidos de maneira independente, isso desencadeou uma série de diferenças nas abordagens de modelagem, apesar dos modelos possuírem semelhanças. As RNAs têm suas origens nas pesquisas em Inteligência Artificial (IA) e foram desenvolvidas por engenheiros e cientistas da computação. Isso resultou na diferença entre as terminologias usadas nas modelagens estatística e de redes neurais, o que tem causado alguma confusão. Sarle (1994) criou um glossário

de termos frequentemente utilizados em RNAs associando com suas equivalências em estatística. Esse glossário está sumarizado na TABELA (1).

TABELA 1 Dicionário Redes Neurais x Econometria

REDES NEURAIAS	ESTATÍSTICA
Pesos	Parâmetros
Conjunto de treinamento	Amostra
Entradas	Variáveis independentes
Saídas	Valores preditos
Retropropagação	Aproximação estocástica
Treinamento ou aprendizado	Estimação
Sinal de entrada	Valor das variáveis exógenas
Sinal de saída	Valor estimado
Alvo	Valor da variável endógena
Padrões	Observações

Na TABELA (1), tenta-se fornecer um dicionário reduzido com o objetivo de facilitar a comunicação entre economistas/econometristas e conexionistas. O conexionismo é uma das grandes linhas de pesquisa da Inteligência Artificial (IA) e tem por objetivo investigar a possibilidade de simulação de comportamentos inteligentes por meio de modelos baseados na estrutura e funcionamento do cérebro humano.

Um dos maiores mal-entendidos que surgem nas diferenças de terminologia é que muitos pesquisadores que aplicam RNAs em problemas de recursos hídricos afirmam que RNAs podem "aprender com exemplos" e que isso é uma das maiores vantagens das RNAs sobre os outros métodos. Entretanto, a fase de "aprendizado" ou "treinamento" de uma RNA não é diferente da fase de estimação de parâmetros nos modelos estatísticos convencionais. A razão para usar essa terminologia na área de IA é para distinguir abordagens baseadas em regras, como sistemas especialistas, daquelas que "aprendem" com exemplos empíricos.

As redes neurais artificiais são largamente utilizadas na literatura para a solução de uma grande variedade de problemas. Com o avanço dos métodos de aprendizagem e das arquiteturas, as RNAs têm se mostrado ferramentas bastante promissoras, em particular, para o problema de previsão de séries temporais.

Foi vista, neste capítulo, uma introdução básica aos conceitos de Redes Neurais Artificiais, onde foram apresentadas as suas partes e seus componentes. Foi esclarecido o processo de treinamento, conceituando os treinamentos supervisionados e não-supervisionados. Finalmente, foram esclarecidos alguns pontos de divergências entre as terminologias de Redes Neurais e de Estatística.

5 MATERIAIS E MÉTODOS

5.1 Base de dados

A base de dados de vazões utilizada neste trabalho está disponível no Operador Nacional do Sistema Elétrico (ONS), OPERADOR NACIONAL DO SISTEMA ELÉTRICO (2009), que é uma entidade de direito privado, sem fins lucrativos, criada em 26 de agosto de 1998, responsável pela coordenação e controle da operação das instalações de geração e transmissão de energia elétrica no Sistema Interligado Nacional (SIN), sob a fiscalização e regulação da Agência Nacional de Energia Elétrica (Aneel).

Essa base de dados consiste em séries históricas de vazão do rio Grande, com dados coletados no posto fluviométrico de Madre de Deus de Minas. O Operador Nacional do Sistema Elétrico (ONS), OPERADOR NACIONAL DO SISTEMA ELÉTRICO (2009), mantém os dados referentes ao Rio Grande atualizados até a data de 31/12/2007. A série histórica utilizada nessa pesquisa abrange o período entre 1990 a 2007. As principais informações sobre o Posto fluviométrico de Madre de Deus de Minas estão descritas na TABELA (2).

TABELA 2 Posto fluviométrico de Madre de Deus de Minas

Identificador	61012000
Operadora	IGAM
Rio	Grande
Área dren. (km^2)	2026

Para o ajuste dos modelos de séries temporais foi utilizada a série de vazões médias mensais do rio Grande, com dados coletados no posto fluviométrico de Madre de Deus de Minas. Essa série contém 216 observações, sendo que as 24

últimas observações (referentes ao período de janeiro de 2006 a dezembro de 2007) foram reservadas para serem comparadas com as previsões.

Foram realizadas análises e previsões utilizando modelos clássicos de séries temporais e modelos de redes neurais artificiais, para previsão um passo à frente para a série objeto de estudo do trabalho. Ambas as metodologias estão descritas nas seções a seguir. O desempenho dos modelos foram comparados e os resultados mostraram quais modelos apresentam melhor, ou pior, desempenho.

5.2 Séries Temporais

Na análise temporal, utilizando métodos paramétricos, os modelos de Box & Jenkins têm recebido muita atenção nas últimas décadas. Tais métodos consistem em ajustar modelos autoregressivos, médias móveis, modelos mistos autorregressivos médias móveis e os modelos integrados autorregressivos médias móveis.

O ajuste do modelo autorregressivo é realizado por meio da análise das funções de autocorrelação e autocorrelação parcial e a escolha da ordem mais adequada do modelo é feita a partir dessas funções.

A metodologia que será utilizada para ajuste dos modelos de séries temporais é a seguinte:

1. Analisar o gráfico da série original e sua função de autocorrelação, verificando os indícios da existência de tendência e sazonalidade;
2. Confirmar, por meio de testes, a existência de tendência e sazonalidade;
3. Tomar o número de diferenças necessárias para retirar a tendência e a sazonalidade determinística da série;
4. Analisar as funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da série diferenciada (livre de tendência e ou sazonalidade determinística). As correlações significativas da FAC indicam a ordem do

modelo de médias móveis (MA) e as da FACP indicam a ordem do modelo autorregressivo (AR). Caso haja correlações significativas nos *lags* múltiplos de s (período sazonal), deve-se ajustar a parte sazonal do modelo;

5. Ajustar o modelo e verificar se suas estimativas são significativas;
6. Testar a partir da FAC e FACP dos resíduos do modelo ajustado se o resíduo é um ruído branco (teste da autocorrelação residual) e confirmar pelo teste de Box & Pierce;
7. Fazer as previsões;
8. Aplicar o critério de AIC e BIC para escolha do melhor modelo que se ajusta aos dados;
9. Escrever o modelo matemático com os parâmetros estimados.

Os métodos de Box & Jenkins baseiam-se na proposição de que o valor atual da série temporal é a combinação de p valores precedentes e q impactos aleatórios anteriores, mais o impacto atual. Os p valores antecedentes formam o componente autorregressivo e os q impactos prévios formam o componente de média móvel da série. A modelagem de uma série temporal objetiva, então, a determinação dos valores de p e q , seguida da estimação dos respectivos coeficientes da combinação linear.

As análises estatísticas foram realizadas utilizando-se as funções dos pacotes *stats* e *tseries* do programa R.2.7.2, R DEVELOPMENT CORE TEAM (2009).

5.3 Redes Neurais Artificiais

Considerável atenção tem sido dedicada a métodos alternativos para o estudo de séries com padrões não-lineares, destacando-se a utilização de Redes Neurais Artificiais. O emprego das arquiteturas MLP (*multi-layer perceptron*)

e RBF (*radial basis function*) trouxe resultados muito positivos no campo da predição de valores futuros em séries temporais, em virtude do caráter essencialmente não-linear dessas estruturas.

Para o emprego de uma rede neural artificial como preditor de um passo à frente, é necessário definir quais valores passados da série serão utilizados na definição da entrada da rede neural. Feito isso, o problema de síntese do preditor se transforma em um problema de treinamento supervisionado, quando, o que se deseja é obter um mapeamento multidimensional não-linear de entrada-saída. A sequência de passos abaixo mostra a metodologia usada para ajuste dos modelos de redes neurais.

1. Obter a série temporal, ou seja, os valores históricos da variável a ser predita um ou vários passos à frente. Se necessário, normalizar os dados (média zero e variância unitária), evitando que o intervalo de excursão dos valores seja qualquer;
2. Definir quais valores passados da série serão considerados na predição. Separar esses valores em 3 conjuntos: conjunto de treinamento, conjunto de validação e conjunto de teste;
3. Treinar a rede neural com o conjunto de treinamento e parar o treinamento de acordo com critérios de parada. O treinamento pode ser interrompido quando: for atingido o valor mínimo do erro quadrático médio para os dados de validação; ou for alcançado um número pré-determinado de épocas de treinamento; ou ocorrer uma combinação dos fatores anteriores;
4. Executar o preditor para fazer previsões utilizando como entrada o conjunto separado para treinamento;
5. Aplicar o critério do EQM para escolha do melhor modelo que se ajusta aos dados;

6. Avaliar o preditor com os dados de teste.

As implementações dos modelos de redes neurais foram feitas utilizando as funções do pacote *AMORE* do programa R.2.7.2, R DEVELOPMENT CORE TEAM (2009).

6 RESULTADOS E DISCUSSÃO

6.1 Modelagem clássica de Séries Temporais

O gráfico da série temporal de vazões médias mensais do Rio Grande está apresentado na FIGURA 8. A partir da inspeção visual preliminar podemos extrair algumas informações sobre o conjunto de observações estudado.

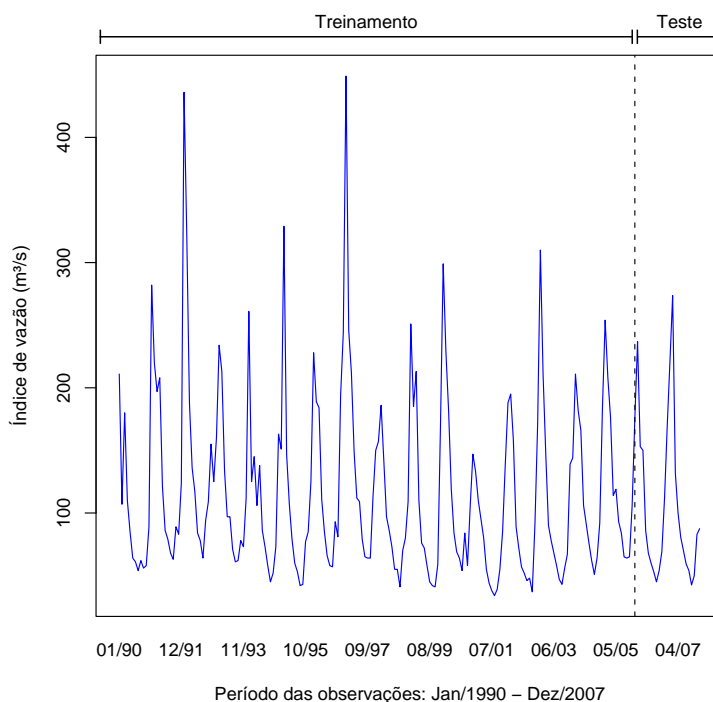


FIGURA 8 Série temporal das vazões médias mensais do Rio Grande entre Janeiro/1990 e Dezembro/2007

Pela análise do gráfico da FIGURA 8, verifica-se que aparentemente a série

possui a componente sazonalidade e, aparentemente, não possui tendência, mas é necessário confirmar tais informações através de testes matemático-estatísticos.

Foram utilizadas as funções de autocorrelação e autocorrelação parcial da série temporal em estudo. Os gráficos das funções de autocorrelação e autocorrelação parcial estão na FIGURA 9.

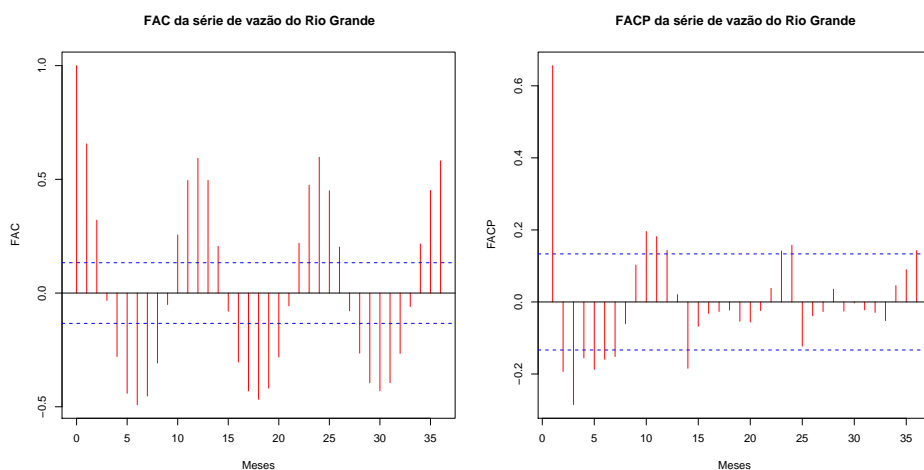


FIGURA 9 Funções de autocorrelação e autocorrelação parcial da série temporal das vazões médias mensais do Rio Grande

O comportamento senoidal da FAC indica que a série possui a componente sazonalidade. A FACP não decai rapidamente para zero, indicando que a série não é estacionária.

Foi construído um periodograma para verificação da existência de sazonalidade no conjunto de dados. O periodograma da série de vazões está apresentado na FIGURA 10.

Pode-se inferir da FIGURA 10 que a série possui uma periodicidade em $p = 12$. Isso confirma a existência da componente sazonalidade.

O teste de Box-Stuart (teste do sinal) foi utilizado para verificação da existência da componente tendência na série. A série possui 216 observações,

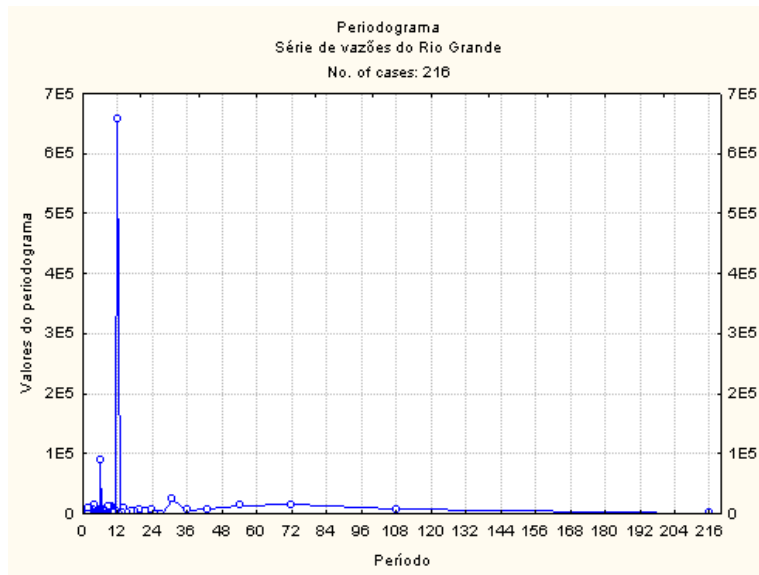


FIGURA 10 Periodograma da série temporal das vazões médias mensais do Rio Grande

logo: $c = 108$ e $n = 108$. A quantidade de sinais positivos T é igual a 43, e o valor de t é dado por: $t = 0,5 \cdot (108 + 1,96 \cdot \sqrt{108}) = 63,6372$. Portanto, $T = 43 \leq t = 63,6372$, mostra que a série apresenta a componente tendência.

Mediante as análises preliminares realizadas, confirmou-se que a série possui tendência e sazonalidade. Logo, torna-se necessário realizar diferenças na série afim de torná-la sem tendência e sem sazonalidade. Foram aplicadas duas diferenças na série, sendo a primeira de $lag = 1$ para eliminação da tendência. E a segunda de $lag = 12$ para remoção da sazonalidade. O esboço da série estacionária está ilustrado na FIGURA 11.

Pode-se verificar na FIGURA 11 que a série diferenciada é estacionária. Na FIGURA 12, mostram-se as funções de autocorrelação e autocorrelação parcial da série estacionária.

A interpretação visual da FIGURA 12 revela um lag significativo em 12, indicando um ajuste sazonal com periodicidade $p = 12$.

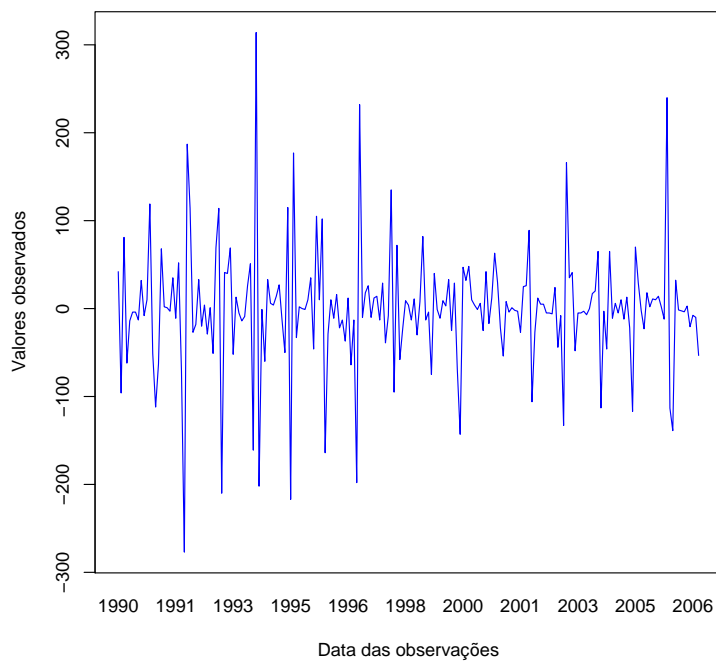


FIGURA 11 Série de vazões do Rio Grande sem tendência e sem sazonalidade

Baseados na metodologia de Box & Jenkins, foram modelados três modelos SARIMA sendo o primeiro um $SARIMA(0,1,1)(0,1,1)_{12}$, o segundo um $SARIMA(0,1,1)(0,1,2)_{12}$ e o último um $SARIMA(1,1,0)(2,1,0)_{12}$. Esses modelos foram escolhidos após realizadas as primeiras análises da série utilizada.

Na TABELA 3, apresenta-se a estimativa dos parâmetros dos modelos de Box & Jenkins propostos, bem como os respectivos erros padrão.

Após a estimação dos parâmetros foram aplicados os testes de autocorrelação residual e o teste de Box-Pierce (Seção 3.11) para verificar se os resíduos dos modelos propostos caracterizavam ruído branco. A aplicação do teste

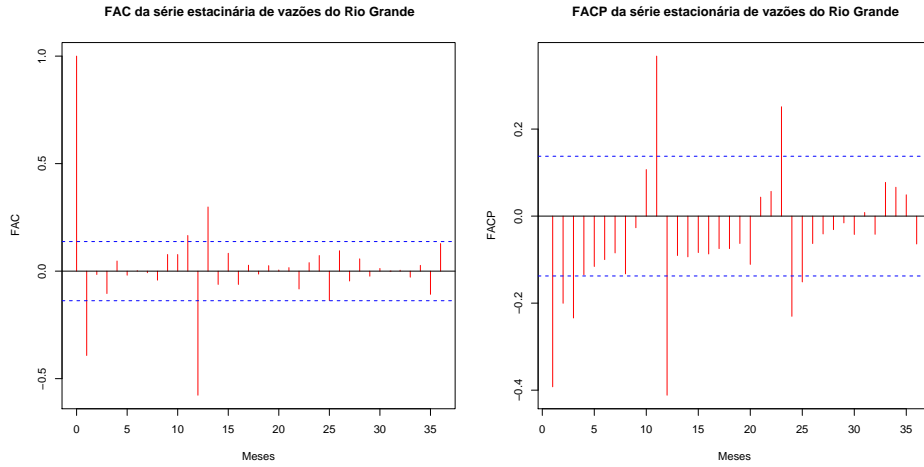


FIGURA 12 Funções de autocorrelação e autocorrelação parcial da série estacionária

TABELA 3 Estimativas dos parâmetros dos modelos e respectivos erros padrão

Modelo	Parâmetros	Estimativas	Erros padrão
SARIMA(0,1,1)(0,1,1) ₁₂	θ_1	0,65740	0,06441
	Θ_1	0,78482	0,03641
SARIMA(0,1,1)(0,1,2) ₁₂	θ_1	0,63647	0,06593
	Θ_1	1,02030	0,07101
	Θ_2	-0,29320	0,07307
SARIMA(1,1,0)(2,1,0) ₁₂	ϕ_1	-0,43860	0,06430
	Φ_1	-0,89620	0,06134
	Φ_2	-0,54570	0,06154

de Box-Pierce forneceu os seguinte resultados:

$$SARIMA(0, 1, 1)(0, 1, 1)_{12} \longrightarrow Q_{40} = 33.91843 < \chi_{38,0.05}^2 = 53.384 \quad (6.1)$$

$$SARIMA(0, 1, 1)(0, 1, 2)_{12} \longrightarrow Q_{40} = 28.49069 < \chi_{37,0.05}^2 = 52.192 \quad (6.2)$$

$$SARIMA(1, 1, 0)(2, 1, 0)_{12} \longrightarrow Q_{40} = 38.72510 < \chi_{37,0.05}^2 = 52.192 \quad (6.3)$$

De acordo com (6.1), (6.2) e (6.3), confirma-se que a_t é ruído branco em todos os três modelos SARIMA. Essa afirmação corrobora com a análise dos gráficos das funções de autocorrelação e autocorrelação parcial dos resíduos dos modelos. Tais gráficos estão na FIGURA 13.

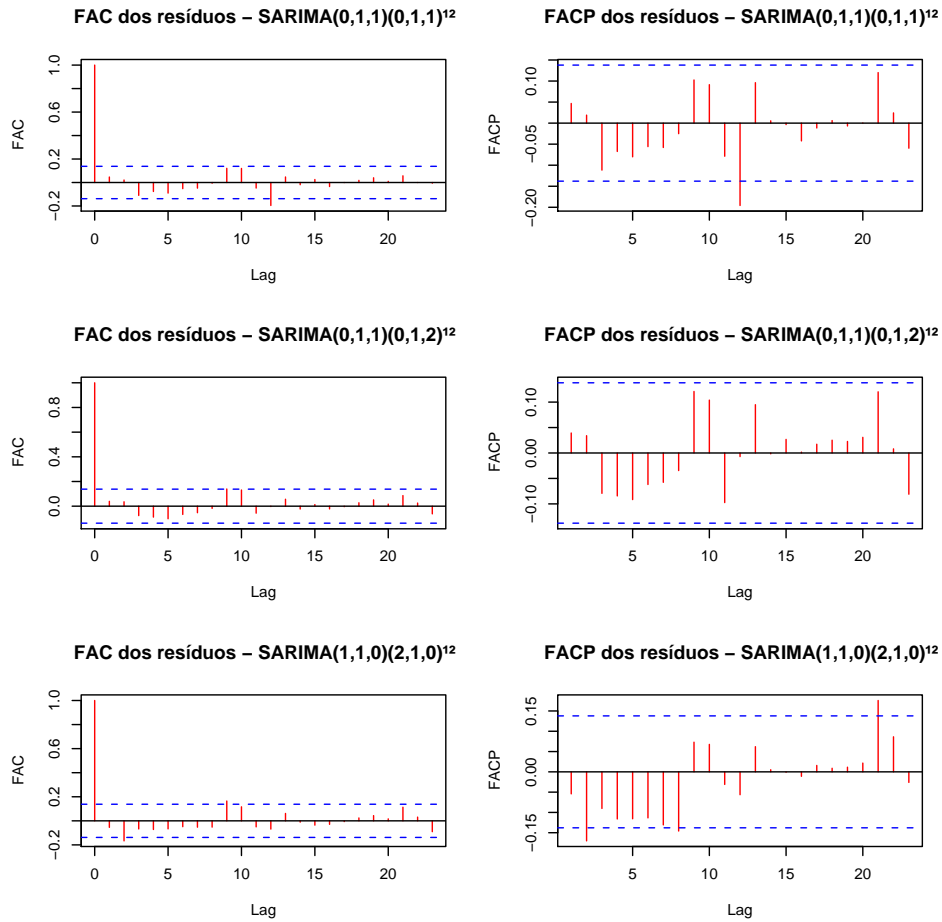


FIGURA 13 Funções de autocorrelação e autocorrelação parcial dos resíduos do modelos SARIMA

As correlações da FAC dos modelos (Figuras 13) são não significativas, indicando que os resíduos são ruído branco.

Para seleção do melhor modelo foram empregados os critérios de informação de Akaike (AIC) e de informação Bayesiana (BIC). Os resultados da aplicação de tais critérios estão na TABELA 4.

TABELA 4 Resultados dos Critérios de Informação Akaike (AIC) e Bayesiana (BIC)

Modelo	AIC	BIC
SARIMA(0,1,1)(0,1,1) ₁₂	7,535115	7,566368
SARIMA(0,1,1)(0,1,2) ₁₂	7,490599	7,553104
SARIMA(1,1,0)(2,1,0) ₁₂	7,549857	7,581109

O modelo SARIMA(0,1,1)(0,1,2)₁₂ apresentou o menor AIC e o menor BIC. Após realizadas as previsões é necessário aplicar algumas medidas de erro sobre os erros de previsão encontrados em cada modelo SARIMA. As medidas de erro aplicadas foram *Mean Absolute Percentage Error* (MAPE) e o Erro Quadrático Médio de Previsão (EQMP). Na TABELA 5, mostram-se os resultados obtidos na utilização de cada medida de erro em cada modelo.

TABELA 5 Medidas de erro de previsão dos modelos SARIMA

Modelo	MAPE	EQMP
SARIMA(0,1,1)(0,1,1) ₁₂	1,10200	1597,799
SARIMA(0,1,1)(0,1,2) ₁₂	1,15357	1497,856
SARIMA(1,1,0)(2,1,0) ₁₂	1,25366	1901,093

O modelo SARIMA(0,1,1)(0,1,1)₁₂ foi o modelo que obteve o menor valor para o MAPE. O modelo SARIMA(0,1,1)(0,1,2)₁₂ obteve o menor valor para o EQMP, enquanto que o modelo SARIMA(1,1,0)(2,1,0)₁₂ foi o modelo que encontrou os maiores valores tanto para o MAPE quanto para o EQMP.

Na FIGURA 14, mostra-se uma comparação gráfica das previsões encontradas pelos três modelos SARIMA.

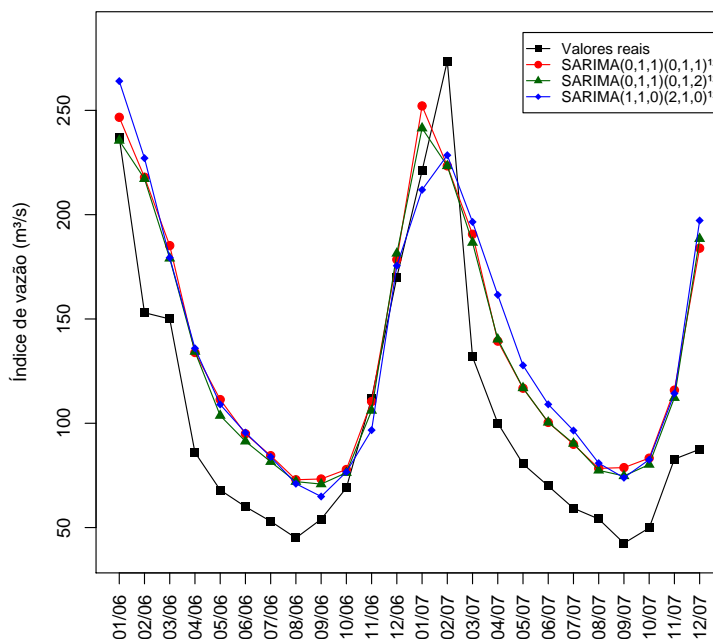


FIGURA 14 Previsões encontradas pelos modelos SARIMA para o período de Janeiro/2006 a Dezembro/2007

Conforme pode-se inferir do gráfico da FIGURA 14, o modelo que aparentemente melhor se ajustou ao conjunto de dados e, conseqüentemente, encontrou melhores previsões foi o modelo SARIMA(0,1,1)(0,1,2)₁₂.

De acordo com os critérios AIC e BIC, expostos na TABELA 4; com a análise dos erros das previsões obtidas por cada modelo apresentados na TABELA 5; e com a ilustração da FIGURA 14, o melhor modelo para o conjunto de dados de vazões médias mensais do Rio Grande é o SARIMA(0,1,1)(0,1,2)₁₂. Tal modelo tem a seguinte fórmula:

$$Z_t = \frac{(1 - \theta_1 B) \cdot (1 - \Theta_1 B^{12} - \Theta_2 B^{24})}{(1 - B) \cdot (1 - B^{12})} \cdot a_t \quad (6.4)$$

Ou seja, utilizando as estimativas encontradas (TABELA 3), o modelo pode ser escrito da seguinte forma:

$$Z_t = \frac{(1 - 0,63647B) \cdot (1 - 1,02030B^{12} + 0,29320B^{24})}{(1 - B) \cdot (1 - B^{12})} \cdot a_t \quad (6.5)$$

6.2 Modelagem por Redes Neurais Artificiais

A modelagem por Redes Neurais Artificiais consistiu em modelar três arquiteturas distintas para que fossem avaliadas entre si e selecionado o modelo que apresentasse melhores resultados de previsão. As redes utilizadas neste trabalho são do tipo *perceptron* de múltiplas camadas, ou MLP (*multi-layer perceptron*).

Em razão da sazonalidade de 12 meses encontrada na série de vazões médias mensais do Rio Grande, foi estabelecido o tamanho da camada de entrada em 12 neurônios, onde cada neurônio da camada de entrada recebe o valor referente a cada mês do ano.

As arquiteturas modeladas possuem as seguintes configurações: MLP(12,10,1), MLP(12,20,1), e MLP(12,30,1), onde a notação MLP(A,B,C) representa redes *perceptron* em múltiplas camadas com: A, número de neurônios na camada de entrada; B, quantidade de neurônios na camada intermediária; C, neurônio presente na camada de saída.

Para execução do treinamento das redes foi utilizado o algoritmo de retropropagação de erros, ou *back-propagation*. A função de ativação utilizada foi a *sigmoid logistic* apresentada na Equação (6.6).

$$S(t) = \frac{1}{1 + e^{-t}} \quad (6.6)$$

O conjunto de observações em estudo foi separado em duas partes: Conjunto de treinamento, composto por 192 observações; Conjunto de validação e teste, composto por 24 observações.

O treinamento das redes neurais modeladas consistiu em apresentar o conjunto de treinamento à rede para que os pesos dos neurônios pudessem ser ajustados conforme o comportamento da série temporal. Os critérios de parada utilizados foram a tolerância do erro definida em 10^{-3} e o número máximo de épocas definido em 10^5 .

Cada arquitetura teve o seu treinamento repetido por dez vezes, afim de que pudesse ter um valor médio do valor do erro quadrático médio (EQM) seguido dos seus desvios-padrões. A média do EQM, bem como os seus desvios-padrões, e os demais parâmetros dos treinamentos executados em cada uma das arquiteturas estão explicitados na TABELA 6. O parâmetro η corresponde a taxa de aprendizado, e o parâmetro γ corresponde a taxa da constante *momentum*.

TABELA 6 Parâmetros utilizados nos treinamentos das redes neurais com dez repetições

Arquitetura	η	γ	EQM	Desvios padrões	Épocas	Tempo médio
MLP(12,10,1)	10^{-3}	0,99	$6,85 \times 10^{-4}$	$2,29 \times 10^{-4}$	10^5	≈ 1 min
	10^{-3}	0,50	$6,91 \times 10^{-4}$	$1,92 \times 10^{-4}$		
	10^{-2}	0,99	$7,87 \times 10^{-4}$	$2,06 \times 10^{-4}$		
	10^{-2}	0,50	$7,65 \times 10^{-4}$	$1,22 \times 10^{-4}$		
MLP(12,20,1)	10^{-3}	0,99	$7,37 \times 10^{-4}$	$2,35 \times 10^{-4}$	10^5	≈ 2 min
	10^{-3}	0,50	$7,45 \times 10^{-4}$	$1,77 \times 10^{-4}$		
	10^{-2}	0,99	$7,55 \times 10^{-4}$	$2,14 \times 10^{-4}$		
	10^{-2}	0,50	$7,96 \times 10^{-4}$	$2,29 \times 10^{-4}$		
MLP(12,30,1)	10^{-3}	0,99	$6,70 \times 10^{-4}$	$3,06 \times 10^{-4}$	10^5	≈ 3 min
	10^{-3}	0,50	$6,84 \times 10^{-4}$	$2,38 \times 10^{-4}$		
	10^{-2}	0,99	$7,31 \times 10^{-4}$	$1,04 \times 10^{-4}$		
	10^{-2}	0,50	$7,59 \times 10^{-4}$	$2,15 \times 10^{-4}$		

Em destaque, na TABELA 6, tem-se os menores valores do EQM de treinamento em cada arquitetura. As configurações com os parâmetros $\eta = 10^{-3}$ e $\gamma = 0,99$ obtiveram melhor desempenho.

Após a execução das redes como preditores, torna-se necessária a aplicação de algumas medidas de erro sobre os erros de previsão encontrados em cada rede neural. As medidas de erro aplicadas foram *Mean Absolute Percentage Error* (MAPE) e o Erro Quadrático Médio de Previsão (EQMP). Na TABELA 7, mostram-se os resultados obtidos na utilização de cada medida de erro em cada modelo.

TABELA 7 Medidas de erro de previsão dos modelos de redes neurais artificiais

Modelo	MAPE	EQMP
MLP(12,10,1)	0,81229	968,862
MLP(12,20,1)	0,59199	779,590
MLP(12,30,1)	0,90429	873,752

Todas as arquiteturas implementadas encontraram bons resultados para previsão da série temporal de vazões médias mensais do Rio Grande. Porém, o modelo MLP(12,20,1) encontrou os menores valores para MAPE e EQMP. Sendo assim, a arquitetura MLP(12,20,1) foi o modelo escolhido para realizar as previsões da série temporal objeto de estudo desta pesquisa. Para fins de inspeção visual, a FIGURA 15 ilustra uma comparação gráfica das previsões encontradas pelas redes neurais modeladas.

Por meio dos resultados encontrados pelas redes neurais pode-se perceber que essa metodologia conseguiu bom desempenho como previsor de séries climatológicas. Com a rede devidamente modelada e os dados estudados representando bem o comportamento da série, as redes neurais se mostram eficientes para tratar problemas não-lineares como previsão de séries temporais.

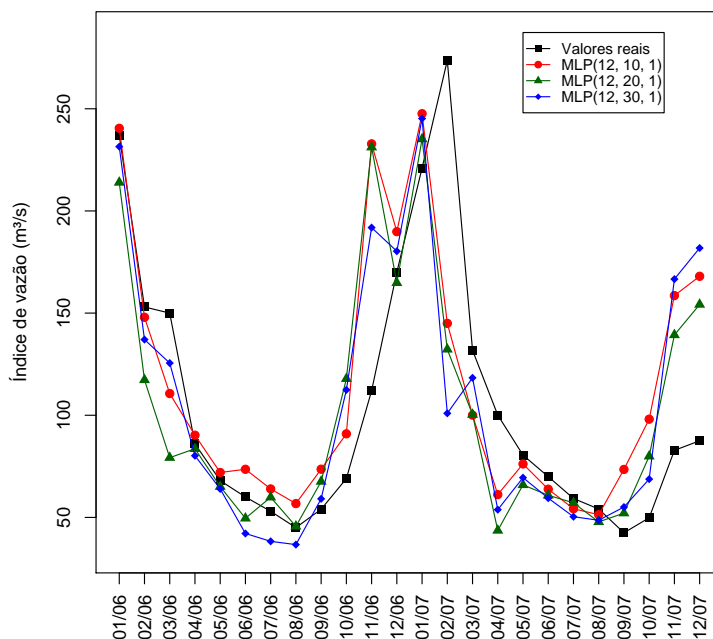


FIGURA 15 Previsões encontradas pelas redes neurais modeladas para o período entre Janeiro/2006 a Dezembro/2007

6.3 Box & Jenkins x Redes Neurais Artificiais: uma análise comparativa

Conforme descrito na Seção 1.1, neste trabalho, objetivou-se comparar duas metodologias distintas para análise e previsão de séries temporais climatológicas. Após as análises, modelagens e previsões obtidas por cada uma das técnicas pesquisadas neste trabalho, são apresentadas as comparações entre a metodologia de Box & Jenkins e a técnica de Redes Neurais Artificiais.

A modelagem baseada na metodologia de Box & Jenkins mostrou-se adequada para modelar o conjunto de dados utilizados nesse trabalho, encontrando bons resultados de previsão. A análise de séries temporais sugerida por essa técnica permite identificar padrões de comportamento das observações, tais como

tendências e sazonalidades. Por meio dessa análise, foi possível identificar a presença das componentes sazonalidade e tendência no conjunto de dados estudado neste trabalho.

A aplicação de Redes Neurais Artificiais na previsão de séries temporais neurais mostrou-se como uma ferramenta relevante para previsão de índices de vazão naturais, podendo ser estendida para outras séries climatológicas. Com essa metodologia, não é possível fazer uma análise mais precisa da série temporal, objetivando obter mais informações sobre os dados. Porém, os resultados das previsões demonstraram que essa técnica pode ser usada como instrumento de apoio a decisão de sistemas hidrelétricos geradores de energia, com propósitos de simulação e/ou otimização da produção elétrica nacional.

Ao aplicar as medidas de erro sobre os erros de previsão encontrados em cada modelagem, encontramos os resultados expostos na TABELA 8.

TABELA 8 Medidas de erro de previsão dos modelos escolhidos para realizar previsões da série de vazões do Rio Grande

Modelo	MAPE	EQMP
SARIMA(0,1,1)(0,1,2) ₁₂	1,15357	1497,856
MLP(12,20,1)	0,59199	779,590

Conforme pode ser observado na TABELA 8, o desempenho das redes neurais como previsor de séries temporais foi superior no que diz respeito ao conjunto de dados utilizado nesta pesquisa. Para uma melhor visualização do resultado, foi gerado um gráfico com os resultados obtidos pelas duas metodologias. Na FIGURA 16, ilustra-se, graficamente, a comparação entre as previsões fornecidas pelo modelo SARIMA(0,1,1)(0,1,2)₁₂ e pelo modelo MLP(12,20,1).

Na FIGURA 16, ilustra-se que ambas as metodologias conseguiram resultados bem próximos dos valores reais nos meses separados para previsão.

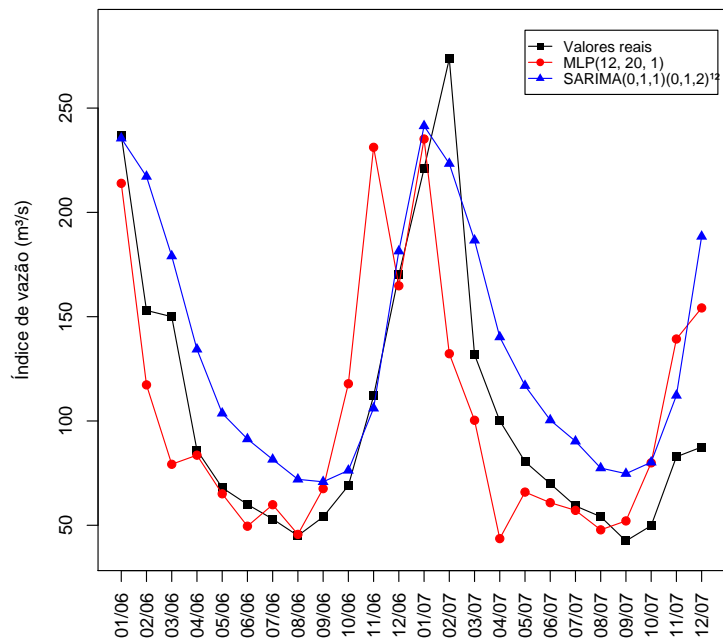


FIGURA 16 Previsões encontradas por Box & Jenkins, e Redes Neurais Artificiais, para o período entre Janeiro/2006 e Dezembro/2007

7 CONCLUSÃO

As conclusões obtidas neste trabalho de dissertação são referentes aos dados históricos da série de vazão do Rio Grande que é afluente ao reservatório da usina hidrelétrica de Camargos/MG, UHE-Camargos-CEMIG. Os modelos de Box & Jenkins e Redes Neurais Artificiais selecionados e apresentados a seguir foram modelados a partir do conjunto de dados citado anteriormente.

Em relação a modelagem por meio da técnica sugerida por Box & Jenkins concluiu-se que o melhor modelo é o SARIMA(0,1,1)(0,1,2)₁₂. Os resultados das análises revelaram que a metodologia citada anteriormente fornece informações importantes sobre o padrão comportamental de séries temporais. Informações tais que auxiliam na elaboração de modelos mais eficientes referentes a série temporal em questão.

Referente à modelagem por Redes Neurais Artificiais concluiu-se que a melhor arquitetura para a rede é a MLP(12,20,1). Os resultados obtidos neste trabalho demonstram a capacidade das redes neurais artificiais tipo *multi-layer perceptron* em realizar com sucesso a associação de padrões e generalizar para padrões desconhecidos. Tal capacidade pode ser observada inclusive com poucos dados de treinamento, desde que representativos do espaço amostral e a rede tenha sido apropriadamente modelada.

Ao realizar a comparação entre o modelo SARIMA(0,1,1)(0,1,2)₁₂ e a arquitetura MLP(12,20,1) foi identificado que esta se adequou melhor aos dados do que aquela. As medidas de erro utilizadas na comparação mostram que redes neurais artificiais do tipo MLP encontraram melhores resultados de previsão do que o modelo proposto por Box & Jenkins.

Entretanto, somente por meio de testes matemáticos utilizados na

modelagem estatística puderam ser confirmadas matematicamente a presença das componentes sazonalidade e tendência no conjunto de dados investigados.

Propostas de continuidade podem incluir:

- Aplicação de outros modelos de redes neurais artificiais, como por exemplo: redes de Função de Base Radial (*Radial Basis Function*, RBF), redes recorrentes, redes de Elman, redes de Kohonen, e/ou redes neurais construtivas;
- Análises abordando outras variáveis disponíveis, tais como: precipitação e evapotranspiração. O objetivo de tais análises seria verificar qual é a influência dessas variáveis na taxa de vazão.
- Analisar os dados de vazões de outras bacias hidrográficas. O objetivo seria verificar se o modelo de Box & Jenkins consegue melhor ajuste do que as redes neurais MLP.
- Realizar análises utilizando um conjunto de observações de vazões que abrange um período de tempo maior e mais recente.
- Implementar as arquiteturas MLP investigadas nesse trabalho em outra linguagem, ou ambiente, de programação. O objetivo seria reduzir o tempo de treinamento possibilitando, assim, a entrada de um conjunto maior de dados.

REFERÊNCIAS BIBLIOGRÁFICAS

- AKAIKE, H. Maximum likelihood identification of gaussian autoregressive moving average models. **Biometrika**, London, v. 21, p. 243–247, 1973.
- ANMALA, J.; ZHANG, B.; GOVINDARAJU, R. S. Comparison of artificial neural networks and empirical approaches for predicting watershed runoff. **Journal of Water Resources Planning and Management**, New York, v. 126, n. 3, p. 156–166, 2000.
- BALLINI, R.; FIGUEIREDO, M.; SOARES, S.; ANDRADE, M.; GOMIDE, F. A seasonal steamflow forecasting model using neurofuzzy network. **Information, Uncertainty and Fusion**, Massachusetts, v. 1, p. 257–276, 2000.
- BISHOP, C. **Neural Networks for Pattern Recognition**. 2. ed. Oxford: Oxford University, 1997. 482 p.
- BOX, G.; JENKINS, G. **Time series analysis, forecasting and control**. San Francisco: Holden-Day, 1976. 575 p.
- BOX, G.; JENKINS, G.; REINSEL, G. **Time series analysis: forecasting and control**. 3. ed. New Jersey: Prentice Hall, 1994. 598 p.
- BOX, G.; PIERCE, D. Distribution of residual auto-correlations in autoregressive-integrated moving average time series models. **Journal of the American Statistical Association**, Alexandria, v. 65, n. 332, p. 1509–1529, 1970.
- CAMPOLO, M.; ANDREUSSI, P.; SOLDATI, A. River flow forecasting with a neural network model. **Water Resources Research**, Washington, v. 35, n. 34, p. 1191–1198, 1999.
- CENTRAIS ELÉTRICAS BRASILEIRAS. **Potencial hidrelétrico brasileiro**. Portal de Minas e Energia do Governo Federal, 2009. Disponível em: <<http://www.eletrobras.gov.br>>. Acesso em: 05 set 2009.
- CHOW, V. **Handbook of applied hydrology: a compendium of water resources technology**. New York: McGraw-Hill, 1959. 174 p.

- CONTRELL, M.; GIRARD, Y.; MANGUEAS, M.; MULLER, C. Neural modelling for time series: A statistical stepwise method for weight elimination. **IEEE Transaction on Neural Networks**, New York, v. 6, p. 1355–1364, 1995.
- ELSHORBAGY, A.; SIMONOVIC, S. P.; PANU, U. S. Performance evaluation of artificial neural networks for runoff prediction. **Journal of Hydrologic Engineering**, Reston, v. 5, n. 4, p. 424–427, 2000.
- FERRAZ, M. I. F.; SAFADI, T.; LAGE, G. Uso de modelos de séries temporais na previsão de séries de precipitação pluviiais mensais no município de Lavras-MG. **Revista Brasileira de Agrometeorologia**, Santa Maria, v. 7, n. 2, p. 259–267, 1999.
- FORTUNATO, L. A. M.; NETO, T. A. A.; ALBUQUERQUE, J. C. R.; PEREIRA, M. V. F. **Introdução ao Planejamento da Expansão e Operação de Sistemas de Produção de Energia Elétrica**. Niterói: Ed. Universitária UFF, 1990.
- GERSHENFELD, N.; WEIGEND, A. **The future of time series: learning and understanding, time series prediction: forecasting the future and understanding the past**. Massachusetts: Addison-Wesley, 1994.
- GIL, M.; SANCHEZ, C. **Hydraulic Energy**. 2. ed. Madri: Espuma, 1997. 350 p.
- GRANGER, C.; NEWBOLD, J. Forecasting transformed series. **Journal of the Royal Statistical Society**, London, v. 38, n. 2, p. 189–203, 1976.
- HAYKIN, S. **Neural Networks: a comprehensive foundation**. 2. ed. New York: Macmilan College, 1994. 696 p.
- HEBB, D. **The organization of behavior**. New York: J. Wiley, 1949. 378 p.
- HOPFIELD, J. Neural networks and physical systems with emergent collective computational abilities. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 79, p. 2554–2558, 1982.
- HU, M. **Application to the adaline system to weather forecasting**. 51 p. Tese (Doutorado em Engenharia Elétrica) — Stanford Electronic Laboratories, Stanford, 1964.
- IMRIE, C. E.; DURUCAN, S.; KORRE, A. River flow prediction using artificial neural networks: generalisation beyond the calibration range. **Journal of Hydrology**, Amsterdam, v. 233, p. 138–153, 2000.

- JACOBS, R. Increased rates of convergence through learning rate adaptation. **Neural Networks**, New York, v. 1, p. 295–307, 1988.
- KOHONEN, T. Self-organized formation of topologically correct feature maps. **Biologica Cybernetics: communication and control in organisms and automata**, New York, v. 43, p. 59–69, 1982.
- LACHTERMACHER, G.; FULLER, J. Backpropagation in time series forecasting. **Journal of Forecasting**, Chichester, v. 14, p. 381–393, 1995.
- LAPEDES, A.; FARBER, R. **Nonlinear signal processing using neural networks: prediction and system modeling**. Los Alamos: Los Alamos National Laboratory, 1987.
- MAIER, H. R.; DANDY, G. C. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. **Environmental Modelling and Software**, Oxford, v. 15, p. 101–124, 1999.
- MCCULLOCH, W.; PITTS, W. A logical calculus of ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, New York, v. 5, p. 115–133, 1943.
- MINSKY, M. **Perceptrons**. Cambridge: MIT, 1969.
- MORETTIN, P.; TOLOI, C. **Análise de séries temporais**. 2. ed. São Paulo: E. Blücher, 2006. 564 p.
- OPERADOR NACIONAL DO SISTEMA ELÉTRICO. **Séries históricas de vazões**. 2009. Disponível em: <<http://www.ons.org.br>>. Acesso em: 15 nov 2009.
- PRIESTLEY, M. **Spectral analysis and time series. Volume 1: Univariate series. Volume 2: Multivariate series, prediction and control**. New York: Academic, 1989. 890 p.
- R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. Vienna: [s.n.], 2009. Disponível em: <<http://www.R-project.org>>. Acesso em: 05 set 2009.
- ROSENBLATT, F. A probabilistic model for information storage and organization in the brain. **Psychological Review**, Washington, v. 65, p. 386–408, 1958.

- RUMELHART, D.; HINTON, G.; WILLIAMS, R. Learning representations by backpropagation errors. **Nature**, London, v. 323, p. 533–536, 1986.
- SAFADI, T. Uso de séries temporais na análise de vazão de água na represa de Furnas. **Ciência e agrotecnologia**, Lavras, v. 28, n. 1, p. 142–148, 2004.
- SARLE, W. Neural networks and statistical models. In: ANNUAL SAS USERS GROUP INTERNATIONAL CONFERENCE, 19. **Proceedings...**, Cary: SAS Institute, p. 1538–1550, 1994.
- SCHWARZ, G. E. Estimating the dimension of a model. **Annals of Statistics**, Haywards, v. 6, p. 461–464, 1978.
- SHARDA, R.; PATIL, R. Neural networks as forecasting experts: an empirical test. **International Joint Conference on Neural Networks**, Washington: IEEE, v. 1, p. 441–494, 1990.
- SOARES, S. Planejamento da operação de sistemas hidrotérmicos. **Revista Brasileira de Automática**, Campinas, v. 1, p. 122–123, 1987.
- TANG, Z.; ALMEIDA, C.; FISHWICK, P. Time series forecasting using neural networks vs Box & Jenkins methodology. **Simulation Council**, San Diego, v. 57, n. 5, p. 303–310, 1991.
- TANG, Z.; FISHWICK, P. Feedforward neural nets as models for time series forecasting. **ORSA Journal on Computing**, Baltimore, v. 5, p. 374–384, 1993.
- TONELLO, K. C.; DIAS, H. C. T. D.; SOUZA, A. L. d.; RIBEIRO, C. A. A. S.; LEITE, F. P. Morphometric characteristics of Cachoeira das Pombas watershed, Guanhanes - MG. **Revista Árvore**, Viçosa, v. 30, n. 5, p. 849–857, 2006.
- TUCCI, C. E. M. **Regionalização de Vazões**. Porto Alegre: UFRGS/ANEEL/ANA/ABRH, 2002. 250 p.
- TUCCI, C. E. M. **Modelos Hidrológicos**. 2. ed. Porto Alegre: UFRGS, 2005. 678 p.
- WEIGEND, A.; RUBERMAN, B.; RUMELHART, D. Predicting the future: a connectionist approach. **International Journal of Neural Systems**, Singapore, v. 1, p. 193–209, 1990.
- WERBOS, P. **Beyond regression: New tools for prediction and analysis in the behavioral sciences**. 235 p. Tese (Doutorado em Matemática Aplicada) — Harvard University, Cambridge, 1974.

WHEELWRIGHT, S.; MAKRIDAKIS, S. **Forecasting methods for management**. New York: J. Wiley, 1998. 392 p.

ZANDONADE, E.; SOUZA, R. C. Forecasting via neural networks: a comparative study. **Pesquisa operacional**, Rio de Janeiro, v. 14, n. 2, p. 15–27, 1993.

ZHANG, Q.; PATUWO, B.; HU, M. Forecasting with artificial neural networks: the state of the art. **International Journal of Forecasting**, Amsterdam, v. 14, p. 35–62, 1998.

ZUBEN, F. V. **Modelos paramétricos e não paramétricos de redes neurais artificiais e aplicações**. 244 p. Tese (Doutorado em Engenharia Elétrica) — Universidade Estadual de Campinas, Campinas, 1996.