

1

Introdução

Diferentes áreas de pesquisas observam fenômenos da natureza, cujos estudos necessitam de ferramentas matemáticas para manipular os valores dessas observações, a fim de tornar possível a interpretação desses fenômenos.

Muitas das vezes, os fenômenos naturais apresentam características aleatórias que se modificam ao longo do tempo ou do espaço, apresentando uma dependência onde o acontecimento no instante atual depende do que aconteceu no passado [7]. A solução de problemas que envolvem componentes aleatórias necessita de uma matemática adequada, que permita modelar também as imprevisibilidades.

Um ramo da teoria da probabilidade, onde se define um conjunto diversificado de modelos que permitem realizar o estudo dos problemas com componentes aleatórias é denominado processo estocástico [47].

Ao observar um fenômeno com características aleatórias por um período de tempo, consegue-se obter uma trajetória ou realização do processo observado. Ao realizar a mesma observação em um período de tempo diferente é possível obter uma outra trajetória, diferente da primeira. Um processo estocástico corresponde ao conjunto de todas as possíveis trajetórias que podem ser observadas desse fenômeno. Cada trajetória observada é chamada de série temporal. Portanto, uma série temporal é considerada uma realização de um processo estocástico [53].

Uma série temporal consiste de um conjunto de observações ordenadas no tempo, que apresentam dependência serial (observações consecutivas são dependentes umas das outras) [62]. Na prática, apenas uma realização do processo estocástico, que corresponde à série dos dados observados, denominada série histórica, é o que se tem disponível. A série histórica é apenas uma trajetória dentre muitas que poderiam ter sido observadas.

Assim, a cada instante de tempo, o processo estocástico é considerado uma variável aleatória. O valor observado em um instante i qualquer (valor da série histórica no instante i) consiste de um valor “amostrado” da distribuição de probabilidade associada à variável aleatória do processo estocástico.

Portanto, um processo estocástico é totalmente descrito pelo conjunto de todas as séries temporais que o compõem ou pela distribuição de probabilidade conjunta de todas as variáveis aleatórias envolvidas [53]. Como na prática essas informações não estão disponíveis, é preciso fazer um estudo da série histórica. Através da análise e modelagem de séries temporais, realizada na série histórica, é possível investigar o mecanismo gerador da série temporal, ajustar um modelo e a partir dele gerar séries sintéticas, que representam as séries temporais que podem ser "amostradas" pelo processo estocástico correspondente ao fenômeno observado. Um modelo é, conseqüentemente, uma simplificação da realidade, onde se considera apenas os aspectos mais relevantes, contendo informações essenciais, ao problema objeto de interesse [35].

O estudo de séries temporais é um campo de pesquisa relevante em diversas áreas de conhecimento. A análise de séries temporais é parte integrante, muitas vezes fundamental, de um processo de tomada de decisão, visando a objetivos específicos [59]. Normalmente as séries temporais são analisadas a partir de seus principais comportamentos, tais como: tendência, ciclo, sazonalidade e variações aleatórias.

Ao longo dos anos, muitas ferramentas para modelagem de séries temporais têm sido desenvolvidas e estudadas, abrangendo dois enfoques [55]:

1. A construção de um modelo conceitual para o fenômeno gerador da série, através do uso de uma teoria sobre o domínio de aplicação;
2. A construção de um modelo paramétrico, que consiste em uma função com parâmetros livres ajustados a partir dos dados disponíveis (obtidos da série histórica).

Os modelos conceituais têm a desvantagem de necessitar de um conhecimento aprofundado do domínio de aplicação para a sua construção, além de ter aplicação limitada ao domínio abordado. Já nos modelos paramétricos, os parâmetros dos modelos gerados não têm uma interpretação direta dentro do domínio do problema, uma vez que a série é modelada através de parâmetros ajustados a partir dos dados históricos.

Através da modelagem paramétrica, entretanto, é possível criar e ajustar um modelo estatístico que capture toda a estrutura de dependência existente na série histórica. Essa dependência é fornecida pelas relações entre as variáveis "causais" (explicativas, independentes ou preditoras) com as variáveis "efeitos" (respostas, dependentes ou previstas), com as primeiras podendo também corresponder a valores atrasados da variável dependente modelada. A relação de dependência entre essas variáveis pode ser do tipo linear ou não-linear.

1.1 Motivação

Os modelos paramétricos que ganharam mais atenção da comunidade científica são os modelos estatísticos lineares de Box & Jenkins [9], que têm sido aplicados satisfatoriamente em uma diversidade de problemas reais e seus princípios servem até hoje como base para outros modelos [55].

Porém, muitos problemas reais apresentam características complexas, tais como não-linearidades e comportamento caótico, para os quais uma aproximação linear pode fornecer como resultado um modelo pouco eficiente, de aplicabilidade limitada ou inadequada [35].

Além disso, algumas séries temporais exibem uma estrutura de auto-correlação que depende não somente do intervalo de tempo entre as observações, mas também do período observado. Um exemplo são as séries hidrológicas sazonais mensais. Essas séries têm como característica o comportamento periódico das suas propriedades probabilísticas, como por exemplo, a média, a variância, a assimetria e a estrutura de auto-correlação [64, 39]. A análise deste tipo de séries pode ser feita pelo uso de formulações auto-regressivas cujos parâmetros apresentam um comportamento periódico, as quais denomina-se modelos auto-regressivos periódicos. Na classe de modelos periódicos, o que mais se destaca é o modelo PAR(p) [37], que ajusta para cada período da série um modelo auto-regressivo [9]. Mas como o modelo auto-regressivo é um modelo estatístico linear, como foi dito, a sua aplicação em séries temporais de comportamento caótico, pode não capturar as reais características da série histórica analisada, gerando resultados inadequados ou inconsistentes.

Isso motivou a elaboração de um novo modelo de processo estocástico genérico, intrinsecamente não-linear, que possa ser aplicado em uma grande gama de problemas de fenômenos não-lineares de comportamento estocástico e/ou com características periódicas de suas propriedades.

Entre as metodologias que possibilitam o tratamento apropriado de problemas complexos e não-lineares destacam-se as metodologias da inteligência computacional [15], mais especificamente as redes neurais artificiais (RNAs), que têm sido aplicadas a diversos estudos com resultados promissores.

As redes neurais artificiais (RNAs) são modelos não-lineares, simples de entendimento e implementação que recentemente vêm sendo aplicadas em diversas pesquisas em análise de séries temporais [3], [33], [67]. A aplicação de RNAs em análise de séries temporais é interessante devido à habilidade de aprendizado e capacidade de generalização, associação e computação paralela

que as RNAs apresentam. Estas qualidades as tornam capazes de identificar e assimilar as características das séries históricas, tais como sazonalidade, periodicidade e tendência, na maioria das vezes implícitas devido a presença de ruídos. Uma outra vantagem na sua utilização é que as RNAs não necessitam de conhecimentos a priori do ambiente da série ou de uma complexa formulação teórica, o que normalmente é imprescindível em alguns métodos estatísticos não-lineares [4]. Isto é, não é preciso pressupor um comportamento para a série quando se utiliza RNAs, como se faz em alguns métodos estatísticos, pois a própria rede irá aprender o comportamento da série. É claro que qualquer informação sobre a natureza da série temporal é útil para melhorar o ajuste do modelo da RNA que irá tratá-la, mas não é uma condição necessária. Outro ponto importante é o relacionamento direto entre os modelos de redes neurais com modelos estatísticos [44]. Muitos modelos de RNAs podem ser considerados como extensões de modelos estatísticos clássicos, ou ainda como uma forma de implementar modelos estatísticos não-lineares [55].

Entretanto, as RNAs não são indicadas para simular realizações estocásticas. Uma vez ajustados os pesos sinápticos no treinamento, ao se apresentar uma dada seqüência de entradas sempre se obterá a mesma seqüência de respostas de uma RNA. Porém, em muitos problemas existe a necessidade de simular cenários estocásticos. Através desses cenários é possível representar por semelhança, na totalidade ou em parte, as propriedades e comportamentos de um sistema real em uma escala menor, permitido a sua manipulação para que se faça um estudo detalhado do sistema.

Isso motivou, nesse trabalho, a elaboração de um novo modelo de processo estocástico utilizando RNAs na sua construção. Através das RNAs desse processo estocástico, que são modelos não-lineares, é possível capturar o comportamento de séries históricas de comportamento estocástico e também de séries que apresentam comportamento periódico de suas propriedades. A utilização desse modelo na geração de séries temporais sintéticas igualmente prováveis à série histórica, pode ser aplicado na solução de problemas que envolvem fenômenos climatológicos (ex: vazão, temperatura, precipitação...) e econômicos (ex: preços de ações), entre outros.

O estudo de séries temporais climatológicas, por exemplo, é de muita utilidade nos setores de energia no Brasil [16], pois, o sistema de produção e transmissão de energia elétrica brasileiro é um sistema de grande porte, onde, atualmente, 93,3% da energia gerada provém de recursos hídricos [50]. Essas séries temporais apresentam alta complexidade devido aos processos físicos envolvidos, ao comportamento não-linear e à alta variabilidade de fatores,

como exemplo, a precipitação de chuva [8]. Uma forma tradicional de tratar estas séries temporais é através de modelos estocásticos que utilizam as séries históricas e suas relações estatísticas temporais do passado para estimar o futuro.

Um exemplo são as séries de afluências mensais. Uma das dificuldades na modelagem dessas séries é a sua natureza não estacionária, devido aos períodos de cheia e seca do ano. Por este motivo, é de fundamental importância o estudo e o desenvolvimento de modelos capazes de capturar a natureza não-linear, associada à sazonalidade e elevado nível de incerteza, apresentadas nas vazões dos rios brasileiros [35].

Isso motivou o estudo de caso dessa tese, que consiste em aplicar o novo modelo proposto, o processo estocástico utilizando RNAs, para tratar as incertezas das séries de afluências mensais, dentro do contexto do planejamento da operação energética do sistema hidrotérmico brasileiro. O objetivo desse planejamento é minimizar o valor esperado do custo total da operação energética ao longo do período de planejamento, sendo este custo dependente das vazões dos rios brasileiros. Como não se conhecem as vazões afluentes futuras que ocorrerão aos reservatórios das usinas hidrelétricas, durante as decisões operativas, este problema é inerentemente estocástico [54, 38].

1.2 Objetivo

O objetivo dessa tese é desenvolver um novo modelo de processo estocástico baseado em redes neurais artificiais (RNAs), denominado Processo Estocástico Neural (PEN), com os seguintes requisitos:

- modelo genérico que possa ser aplicado em uma grande gama de problemas envolvendo fenômenos não-lineares de comportamento estocástico, como, por exemplo, fenômenos relacionados ao clima, temperatura, precipitação, vazão, entre outros;
- modelo intrinsecamente não-linear, devido ao uso de RNAs na sua formação;
- modelo auto-regressivo, pois as entradas das RNAs são formadas por valores temporalmente anteriores da série modelada;
- capacidade de modelar ou representar dados periódicos ou com correlações periódicas;
- capacidade de gerar séries sintéticas, abordando qualquer período de tempo, na quantidade necessária para o estudo da série temporal em foco.

Portanto, o Processo Escotástico Neural é classificado como um modelo estocástico periódico não-linear auto-regressivo genérico.

Tem-se também como objetivo desse trabalho aplicar o modelo PEN para tratar as incertezas das séries de energias afluentes mensais, dentro do contexto do planejamento da operação energética do sistema hidrotérmico brasileiro. Portanto, o caso de estudo desse trabalho consiste em:

- ajustar o modelo PEN com as séries históricas de energias afluentes mensais;
- utilizar o PEN ajustado para gerar um conjunto de cenários hidrológicos, a serem usados no planejamento da operação energética de médio prazo, para os próximos 5 anos;
- utilizar os cenários gerados pelo PEN no planejamento de médio prazo da operação do sistema hidrotérmico brasileiro, realizado através do modelo de otimização monocritério da operação de sistemas hidrotérmicos utilizando algoritmos genéticos, desenvolvido em [60].

1.3

Descrição do Trabalho

Este trabalho se inicia com uma breve introdução sobre processos estocásticos (PE) [7] e sobre a metodologia de redes neurais artificiais (RNAs) [28], fazendo um levantamento bibliográfico dos métodos e modelos que abrangem o tema da aplicação de RNAs na análise e modelagem de séries temporais [1], [3], [4], [8], [10] [20], [33], [55] e [67].

O estudo sobre PE e RNA compreendeu um resumo dos principais conceitos para a formulação do modelo proposto nesse trabalho: um modelo estocástico periódico não-linear auto-regressivo genérico, denominado Processo Estocástico Neural (PEN).

Em relação à modelagem de RNAs, verificou-se na literatura que muitas aplicações em análise de séries temporais utilizavam redes neurais *feedforward Multilayer Perceptron* (MLP), consistindo em geral de uma única camada escondida, treinadas com o algoritmo de aprendizado supervisionado de retropropagação do erro (*backpropagation* [28]), ou variações desse algoritmo.

Para que essas RNAs MLP se comportem como um modelo de processamento temporal, é preciso que elas apresentem habilidades de memória, a qual é simulada através da técnica de "janelamento" [24]. Essa técnica consiste em introduzir memória, fornecendo aos neurônios valores de entrada atuais e valores temporalmente anteriores. Por isso o PEN é classificado como um mo-

delo auto-regressivo, uma vez que as variáveis independentes ou entradas do modelo utilizam valores atrasados da variável dependente modelada.

No estudo de caso desse trabalho, decidiu-se utilizar RNAs MLP, com uma única camada escondida, treinadas com o algoritmo de aprendizado supervisionado *Levenberg- Marquardt* (LM) [25], uma variação do *backpropagation*. Isso porque esse tipo de RNA com esse algoritmo de treinamento é muito aplicada no estudo de séries temporais.

Os parâmetros do modelo correspondem aos pesos sinápticos das RNAs. O ajuste desses parâmetros ocorre na fase de aprendizagem das RNAs, onde um conjunto de exemplos é apresentado à rede, que extrai automaticamente as características necessárias para representar a informação fornecida e posteriormente gerar as respostas do problema analisado [68]. Nesse trabalho, os parâmetros do modelo são ajustados através de exemplos retirados diretamente da série histórica original.

Escolheu-se como estudo de caso a modelagem de séries de afluências mensais, dentro do contexto do planejamento da operação energética do sistema hidrotérmico brasileiro. As séries de afluências mensais são séries sazonais, isto é, são séries que apresentam correlações periódicas de 12 meses. O objetivo é utilizar o PEN para gerar cenário hidrológicos para serem usados no planejamento de médio prazo da operação energética, para os próximos 5 anos. No planejamento de médio prazo ocorre a agregação dos reservatórios das usinas em reservatórios equivalentes de energia em cada região demarcada por um subsistema do Sistema Interligado Nacional (SIN). As energias afluentes que chegam nesses reservatórios equivalentes são chamadas de Energia Natural Afluente (ENA).

O primeiro passo na aplicação do modelo proposto nesse estudo de caso é o ajuste do PEN às séries históricas de ENA. A estrutura do PEN é composta de 12 redes neurais artificiais (RNAs), uma para cada mês do ano. Dentro do PEN ocorre um encadeamento entre as suas RNAs, isto é, uma realização estocástica fornecida por uma rede de um certo mês faz parte da janela temporal de entrada da rede do mês seguinte. O valor de ENA de um mês processado pelo PEN é composto por duas partes: uma parte determinada pelos parâmetros, que corresponde à saída da RNA deste mês, e outra que corresponde a um valor aleatório. Esse valor aleatório é oriundo da distribuição de probabilidade dos resíduos da validação da RNA desse mês. Então a estrutura do PEN é formada por 12 componentes estocásticas (CE), onde cada CE contém a RNA de um mês e a distribuição de probabilidade teóricas dos resíduos dessa RNA. Foram testadas várias arquiteturas para as

RNAs, considerando modificações no número de variáveis de entrada e no número de neurônios na camada escondida. Verificou-se o desempenho dessas redes através de duas métricas principais: o erro médio absoluto percentual (MAPE) e a média dos resíduos oriundos da diferença entre a saída da rede e a resposta desejada.

Uma vez escolhida a configuração de RNA e a distribuição de probabilidade de cada mês, o PEN pode ser utilizado para gerar cenários. No caso de estudo dessa tese são gerados cenários de ENA. Com o propósito de avaliar o desempenho desses cenários de ENA gerados pelo PEN, foram aplicados alguns testes de hipótese relevantes, com o intuito de verificar a aderência do conjunto de séries sintéticas com a série histórica.

No contexto do planejamento da operação energética do sistema hidrotérmico brasileiro a geração de cenários é muito importante pois o único cenário disponível na prática, que é o registro de fluências observadas (chamado de série histórica), é insuficiente para compor uma amostra de tamanho necessário para estimar índices de risco aceitáveis [54, 53]. Através dos cenários é possível reproduzir as características básicas da série histórica, extraindo melhor as informações dessa série temporal, permitindo a avaliação de riscos e incertezas pertinentes a um sistema hidroelétrico.

Os cenários sintéticos foram gerados pelo PEN para serem aplicados no modelo de otimização monocritério da operação de sistemas hidrotérmicos utilizando algoritmos genéticos, desenvolvido em [60]. O objetivo principal do planejamento da operação energética do sistema hidrotérmico brasileiro é obter, para cada etapa, as metas de geração de cada usina hidrelétrica e termoelétrica do Sistema Interligado Nacional de forma a atender a demanda e minimizar o valor esperado do custo de operação ao longo do período de planejamento, que no médio prazo consiste em um horizonte de 5 anos.

Para uma comparação dos resultados implementou-se uma versão do modelo auto-regressivo periódico (PAR(p)) [37], atualmente utilizado no setor elétrico para gerar cenários de ENA, e utilizou-se os mesmos testes de avaliação dos cenários do PEN nos cenários do PAR(p) para fazer comparações.

1.4

Contribuição do Trabalho

As principais contribuições dessa tese são:

- o desenvolvimento de um modelo genérico de processo estocástico baseado em rede neurais artificiais, denominado Processo Estocástico Neural (PEN);

- o PEN pode ser aplicado na modelagem de séries temporais de fenômenos de comportamento estocástico e em séries temporais que apresentam características periódicas;
- o modelo utiliza a capacidade de aprendizado, através de exemplos, das redes neurais artificiais (RNAs) para realizar a estimação do processo estocástico (identificando as características da série histórica desses fenômenos), e a capacidade de generalização para poder simular séries sintéticas igualmente prováveis à série histórica;
- o aprendizado das RNAs é feito diretamente com a série histórica, cuja única transformação necessária é a normalização dos dados dentro de um intervalo pré-estabelecido. Portanto, não é necessário pressupor nenhum comportamento da série histórica para utilizá-la no PEN, todo o comportamento da série é capturado pelas RNAs, que utilizam valores temporalmente anteriores da série na entrada, chamados de termos auto-regressivos da série.

Outras contribuições desse trabalho são:

- a utilização do modelo PEN no tratamento de incerteza de afluências mensais. O modelo PEN é ajustado com as séries históricas de Energias Naturais Afluentes (ENAs) mensais para gerar séries sintéticas igualmente prováveis;
- a análise comparativa do desempenho dos cenários gerados pelo modelo através de testes estatísticos de aderência e através do resultado do planejamento da operação energética de um modelo baseado em algoritmos genéticos desenvolvido em [60].

1.5

Estrutura da Tese

Esta tese está organizado da seguinte forma:

- O Capítulo 2 apresenta os fundamentos teóricos necessários à compreensão dos conceitos utilizados neste trabalho. É realizado um breve resumo sobre a teoria de processos estocásticos e de redes neurais artificiais, conceitos fundamentais do modelo proposto nessa tese. Também apresenta uma breve descrição sobre métodos de amostragem aleatória, a serem utilizados pelas distribuições de probabilidades que compõem o modelo proposto. Os testes de hipóteses utilizados na avaliação do desempenho do modelo também são descritos nesse capítulo;

- O Capítulo 3 apresenta o novo modelo Processo Estocástico Neural (PEN). O capítulo descreve detalhadamente como é realizada a modelagem genérica do PEN para uma série qualquer de interesse. O ajuste do modelo consiste em: determinar a sua estrutura, através da escolha das configurações das RNAs; estimar os seus parâmetros, através do aprendizado das RNAs; e validar os resíduos gerados, determinando as distribuições de probabilidade teóricas que farão parte do modelo;
- O Capítulo 4 apresenta a modelagem completa do Processo Estocástico Neural para tratar especificamente séries de afluências mensais. O objetivo é gerar cenários sintéticos visando o planejamento de médio prazo da operação energética do sistema hidrotérmico brasileiro;
- O Capítulo 5 exhibe os resultados obtidos nos estudos de caso realizados nesse trabalho. Primeiro são apresentados os resultados para se determinar a estrutura do Processo Estocástico Neural. Em seguida, apresenta as avaliações dos cenários gerados pelo PEN comparando com os resultados das avaliações dos cenários gerados pelo PAR(p). Por último apresenta o custo estimado da operação energética calculado pelo modelo baseado em algoritmos genéticos desenvolvido em [60] com os cenários gerados pelo PEN e pelo PAR(p);
- O Capítulo 6 finaliza esta tese apresentando as conclusões gerais e as propostas para trabalhos futuros.