

# 1 Introdução

Recentemente, muitos modelos estatísticos para análise e previsão de séries temporais têm sido propostos na literatura. Os modelos mais famosos e que ganharam mais destaque nas pesquisas científicas pertencem ao grupo dos modelos estatísticos lineares. A razão dessa popularidade vem do fato destes modelos tratarem dados homocedásticos, estacionários e Gaussianos (Box, Jenkins e Reisel, 1994). Os modelos lineares possuem características vantajosas e importantes, tais como, fácil interpretação, cálculos de intervalo de confiança, resultados assintóticos, entre outras.

Muito embora essas características possuam vantagens, a natureza é intrinsecamente não-linear e muitos fenômenos podem não ser capturados pelos modelos lineares. Com essa motivação, nos últimos anos, muitos modelos estatísticos não-lineares vêm sendo propostos. Alguns deles são da classe de modelos STAR (*Smooth Transition Autoregressive*), com aplicações em séries temporais econométricas.

Tong (1978) propôs o modelo TAR (*Threshold Autoregressive*), o qual tem como idéia central a mudança dos parâmetros lineares do modelo auto-regressivo, de acordo com o valor assumido por uma variável de transição (neste caso, uma variável indicadora). Este modelo atribui um modelo linear diferente para distintas regiões onde se encontram os valores dessa variável. Caso a variável de transição seja uma defasagem da variável endógena, o modelo é, então, denominado SETAR (*Self-Exciting Threshold Autoregressive*).

Uma generalização do modelo SETAR com dois regimes, incorporando uma transição suave entre eles, foi proposta por Chan e Tong (1986). Este modelo foi denominado modelo STAR (*Smooth Threshold Autoregressive*). Para uma revisão consulte Teräsvirta (1994). Outras extensões dos modelos TAR têm sido propostas na literatura, podendo citar o modelo MRSTAR (*Multiple Regimes Smooth Transition Autoregressive*), proposto por Dijk e Franses (1999) e o

modelo TV-STAR (*Time-Varying Smooth Transition Autoregressive*), proposto por Lundbergh e Teräsvirta (2000).

Muitos assuntos em diferentes áreas do conhecimento requerem diversas relações para serem especificados. Por isso, tornaram-se necessárias técnicas para lidar com aspectos não-lineares de sistemas. A maioria delas refere-se ao modelo VAR não-linear, denominado STVAR (*Smooth Transition Vector Autoregressive*). O modelo STVAR é a versão multivariada do modelo STAR, assim como ocorre com os modelos lineares VAR (*Vector Autoregressive*) e AR (*Autoregressive*). Muitas de suas aplicações foram feitas no campo da Macroeconomia.

Os modelos STAR e STVAR podem ser vistos como modelos lineares autoregressivos, univariado e multivariado, respectivamente, onde os seus coeficientes são determinados pela posição do vetor de variáveis explicativas dentro do denominado espaço de transição.

É crescente, também, o uso de metodologias estruturadas por árvores de decisão, tendo em vista uma metodologia alternativa aos tradicionais métodos de classificação e modelos de regressão. O algoritmo CART (*Classification and Regression Tree*) (Breiman et al., 1984) é considerado o principal marco na utilização de metodologias estruturadas por árvores, cuja filosofia desta formulação é utilizar modelos mais simples para sub-amostras dos dados, dividindo de forma conveniente o problema em partes, através do particionamento recursivo do espaço de variáveis de transição.

As árvores de classificação são úteis quando a variável dependente é categórica, enquanto que as árvores de regressão devem ser utilizadas na solução de problemas com variável dependente contínua.

O principal atrativo da metodologia CART é a interpretabilidade proporcionada pela estrutura de árvore de decisão obtida no modelo final que, também pode ser lido, como um conjunto de sentenças lógicas a respeito das variáveis explanatórias.

O foco desta dissertação é uma nova formulação de modelo não-linear multivariado, o qual combina o modelo não-linear STVAR com a metodologia CART. O modelo resultante é denominado STVAR-Tree e sua versão univariada é o modelo STAR-Tree, proposto por Rosa, Veiga e Medeiros (2008).

A modelagem STVAR-Tree tem como base o conceito de múltiplos regimes, os quais são definidos por uma árvore binária. Desta forma, temos um modelo não-linear paramétrico através de uma árvore de decisão. No STVAR-Tree, os coeficientes do modelo são determinados através da combinação de diferentes modelos auto-regressivos multivariados, que podem apresentar variáveis exógenas no conjunto das variáveis explicativas.

A especificação do modelo é feita através de testes de hipóteses do tipo LM (*Lagrange Multiplier*). Assim, o crescimento da árvore é condicionado à existência de não-linearidade nas séries. Para cada divisão, são estimados os parâmetros lineares e não-lineares do modelo. No momento em que o teste LM rejeita a hipótese de não-linearidade (e divisão de todos os nós) na profundidade em que a árvore se encontra, o procedimento de crescimento da árvore é finalizado e tem-se, portanto, o modelo final estimado.

Como forma de avaliação do modelo STVAR-Tree proposto nesta dissertação, foram feitos diversos experimentos de Monte Carlo.

Após as simulações, aplicou-se o modelo STVAR-Tree às séries brasileiras de Vazão de Rios e às séries de Preço Spot de energia elétrica do mercado brasileiro.

### **1.1. Vazão de Rios**

No Brasil, o sistema gerador de energia elétrica tem como base as vazões de rios que afluem naturalmente no país. Por serem originadas das precipitações, as vazões fluviais possuem muitas irregularidades, tornando-as inconstantes e difíceis de prever. Por causa dessas irregularidades, o Brasil conta com usinas termoeletricas de complementação e com reservatórios de acumulação, com o propósito de regularizar os regimes fluviais. Então, quando as vazões fluviais são escassas (período de seca), a água que foi armazenada durante o período de grandes afluências naturais (período de cheia) é utilizada.

O Planejamento da Operação do Sistema Interligado (SIN), executado atualmente no Operador Nacional do Sistema Elétrico (ONS), permite o melhor aproveitamento das vazões naturais, evitando o desperdício de água e gastos excessivos com combustíveis nas usinas termoeletricas. Devido a sua

complexidade, este planejamento é feito por etapas com base nos modelos desenvolvidos no Centro de Pesquisas de Energia Elétrica (CEPEL). Em cada etapa, os modelos utilizados possuem diferentes horizontes de planejamento, discretização do tempo, e graus de detalhamento em suas representações (Maceira *et al.*, 2002).

Potencialmente, o sistema brasileiro de geração de energia elétrica pode beneficiar-se em larga escala com o contínuo aprimoramento dos modelos de previsão hidrológica.

A geração de energia elétrica futura é influenciada pelas afluências hidrológicas futuras, quando as previsões e incertezas envolvidas devem ser controladas no planejamento. Isto porque a qualidade das previsões hidrológicas feitas pelos modelos afeta substancialmente o desempenho do sistema. Com isso, os benefícios e a confiança aumentam e os custos diminuem.

O CEPEL utiliza o modelo PREVIVAZM (Costa *et al.*, 2003) para fornecer previsões mensais de afluências de vazão de rios para um horizonte de até 12 meses. Este modelo é uma ferramenta para estudos especiais de verificação de condições de atendimento da demanda energética para o horizonte anual.

Outro modelo é utilizado pelo CEPEL, o modelo PREVIVAZ, para obter as previsões de afluências semanais, num horizonte de até seis semanas. Essas previsões são utilizadas no primeiro mês do Programa Mensal de Operação (PMO), um planejamento de curto prazo. Este modelo é, portanto, executado ao final de cada mês para a elaboração do PMO do mês seguinte e, durante o mês em curso, é executado todas as semanas para a realização das revisões do PMO.

Os modelos PREVIVAZ e o PREVIVAZM para obtenção de previsões de vazões semanais e mensais, respectivamente, utilizam diversas alternativas de classe de modelo, ordem dos modelos, agrupamento da estrutura de auto-correlação e métodos de estimação dos parâmetros, acoplado a diferentes pré-transformações das séries históricas, tipo Box-Cox ou logarítmicas (Box e Cox, 1964), ou sem transformação, e a diferentes formas de estimação de parâmetros dos modelos. Os algoritmos de previsão destes modelos são testados por um esquema robusto de validação cruzada e o algoritmo selecionado é aquele de menor erro quadrático médio de previsão.

As formulações lineares de previsão de séries temporais foram classificadas em modelos estacionários e modelos periódicos. Na classe de modelos

estacionários estão a média e os modelos auto-regressivos de média-móvel, ARMA(p,q). Já na classe dos modelos periódicos estão os modelos periódicos auto-regressivos de média-móvel, PARMA(p,q), e o caso particular PAR(p), os quais se caracterizam por apresentar uma equação de regressão para cada período.

## **1.2. Preço Spot de energia elétrica**

No mercado spot de energia elétrica, toda a energia excedente dos contratos bilaterais é comprada e vendida no Mercado Atacadista de Energia Elétrica. Estes contratos formalizam a compra e venda de energia estabelecendo preços, prazos e montantes de suprimento em intervalos temporais determinados pelos agentes.

O preço spot, também chamado de Preço de Liquidação das Diferenças (PLD), é definido com base nos Custos Marginais de Operação (CMO), obtidos por meio de uma cadeia de programas computacionais. O CMO indica quanto custa a produção de uma unidade de energia adicional à última unidade consumida pelo mercado.

O Centro de Pesquisas de Energia Elétrica (CEPEL), criado por iniciativa do Ministério de Minas e Energia de empresas do setor de energia elétrica, utiliza o software NEWAVE para planejar a operação de subsistemas hidrotérmicos interligados em longo prazo. Este programa determina, mensalmente, metas de geração para cada usina, de maneira a atender a demanda, minimizando o valor esperado do custo total de operação ao longo do período de planejamento. Em resumo, o NEWAVE é responsável por gerar o CMO, valor base para a determinação do preço spot de energia elétrica no Brasil. Outro software utilizado neste processo é o DECOMP, responsável por gerar valores para o preço, semanalmente.

A Câmara Comercializadora de Energia Elétrica (CCEE) é o órgão responsável pela determinação e publicação semanal do PLD, calculado individualmente para cada um dos quatro sub-mercados existentes no Brasil (Sudeste/Centro, Sul, Nordeste e Norte). Ao final do mês, após o registro e validação de todos os contratos, a CCEE utiliza o PLD para liquidar as operações de compra e venda de energia.

O preço spot é único e sua definição depende das condições de oferta e procura do mercado. Como o preço não é, de fato, de mercado e sim fornecido por modelos computacionais, grandes variações semanais podem ocorrer nos seus valores. Além disso, fatores externos também podem influenciar o preço spot, como o nível de armazenamento dos reservatórios das usinas hidrelétricas, a evolução prevista da demanda de energia e a disponibilidade atual e futura de usinas e linhas de transmissão de energia elétrica.

Outro problema é quando os próprios modelos matemáticos (da cadeia de modelos do CEPEL) são abandonados. Um exemplo disso ocorreu no período do racionamento de 2001/2002 quando alguns valores do PLD semanal foram regulados pelo governo.

Além disso, alterações nos planos de reparo e manutenção de unidades térmicas, limites na transmissão de energia entre sub-mercados, e outros motivos, têm feito com que a volatilidade das séries de preço aumente.

No caso do setor elétrico brasileiro, o preço da energia é função da natureza da indústria de eletricidade, isto é, da disponibilidade de água nos reservatórios e no nível de precipitação pluviométrico. A volatilidade está relacionada à dinâmica das afluições. Outro problema do PLD é o fato de que não leva em conta a reação da demanda, sendo a apenas a hidrologia - presente e a previsão futura – a formadora do preço.

### **1.3. Organização da dissertação**

A dissertação está organizada da seguinte forma: o Capítulo 2 apresenta uma revisão dos modelos lineares multivariados. Já no Capítulo 3, os modelos não-lineares são revisados, reunindo os modelos univariados e multivariados, além da metodologia CART. No Capítulo 4 consta a teoria sobre os modelos STVAR-tree, PAR(p) e Neuro-Fuzzy. As aplicações estão presentes no Capítulo 5. E a dissertação conclui no Capítulo 6.