



Marcella Lanzetti Daher de Deus

**Séries Temporais Aplicadas ao Planejamento da
Operação do Sistema Interligado Nacional – SIN**

Dissertação de Mestrado

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da PUC-Rio como requisito parcial para obtenção do grau de Mestre em Engenharia Elétrica. Aprovada pela Comissão Examinadora abaixo assinada.

Orientador: Reinaldo Castro Souza

Rio de Janeiro
Abril de 2008

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.



Marcella Lanzetti Daher de Deus

**Séries Temporais Aplicadas ao Planejamento da
Operação do Sistema Interligado Nacional-SIN**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Reinaldo Castro Souza
Orientador

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Ricardo Tanscheit

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Luiz Guilherme Barbosa Marzano
CEPEL

Prof. José Eugenio Leal
Coordenador Setorial do Centro
Técnico Científico

Rio de Janeiro, 24 de abril de 2008

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e do orientador.

Marcella Lanzetti Daher de Deus

Graduou-se em Engenharia Elétrica na Universidade Federal Fluminense em 2005. Atualmente trabalha como Analista de Gestão de Processos no Operador Nacional do Sistema Elétrico.

Ficha Catalográfica

Deus, Marcella Lanzetti Daher de

Séries temporais aplicadas ao Planejamento da Operação do Sistema Interligado Nacional - SIN / Marcella Lanzetti Daher de Deus ; orientador: Reinaldo Castro Souza. – 2008.

130 f. : il. ; 30 cm

Dissertação (Mestrado em Engenharia elétrica) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008.

Inclui bibliografia

1. Engenharia elétrica - Teses. 2. Séries temporais. 3. Modelos estocásticos. 4. Afluências. 5. Planejamento energético. I. Souza, Reinaldo Castro. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia elétrica. III. Título.

CDD: 621.3

Dedicatória

À minha família querida e amada pelo apoio, carinho e compreensão,
e por nunca terem me deixado desistir.

Agradecimentos

A Deus, por ter me apontado o caminho a ser seguido, por ter colocado as pessoas certas na minha vida, e por sempre me ajudar a alcançar meus objetivos.

A minha amada família que sempre me incentivou e me deu a base necessária para concretizar este projeto.

Em especial, agradeço à minha irmã a Engenheira Flávia Lanzetti Daher de Deus por ter me incentivado e me ajudado durante a confecção deste trabalho.

Aos amigos antigos, agradeço pelo apoio e incentivo durante todo o mestrado.

Aos novos amigos, que me ajudaram durante a tese, em especial os engenheiros Murilo Pereira Soares e Luiz Guilherme Ferreira Guilhon e o Meteorologista Márcio Cataldi, agradeço pela ajuda e comprometimento gratuito, sem os quais não seria possível terminar este trabalho.

Ao Engenheiro Mário Jorge Daher, agradeço pela ajuda em busca de um tema.

Aos colegas de trabalho que aceitaram minhas ausências e sempre me incentivaram.

Ao meu Co-orientador, Francisco José Arteiro de Oliveira, que apesar da escassez de tempo, me ajudou a concretizar este estudo.

Ao meu Orientador pela confiança depositada.

Ao CNPQ e à PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Resumo

Deus, Marcella Lanzetti Daher; Souza, Reinaldo Castro. **Séries Temporais Aplicadas ao Planejamento da Operação do Sistema Interligado Nacional – SIN**. Rio de Janeiro, 2008. 130p. Dissertação de Mestrado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

A vocação natural do Brasil para a hidroeletricidade fez com que o Sistema Interligado Nacional – SIN fosse desenvolvido com forte predominância de geração de origem hidroelétrica. Entretanto, ao se optar por uma base hidroelétrica há de se lidar com as significativas incertezas associadas às afluências futuras aos rios e, por extensão, a todas as bacias hidrográficas do país. Logo, a estrutura de produção de energia hidroelétrica do Brasil foi concebida de forma a minimizar os riscos associados ao comportamento aleatório das afluências. Para contemplar a estocasticidade das afluências no Planejamento da Operação do SIN, o Operador Nacional do Sistema Elétrico – ONS utiliza uma cadeia de modelos dentre os quais estão contidos o modelo de previsões de vazões determinísticas para o curto prazo, e os modelos de geração de cenários de afluências. Estes modelos fornecem insumos para que os modelos de otimização possam estabelecer as Estratégias e Políticas de Operação para o médio e curto prazo, considerando a volatilidade das afluências. Esta dissertação descreve os processos de séries temporais empregados no modelo de previsões determinísticas para o curto prazo e nos modelos de geração de cenários de afluências para o médio e curto prazo. Além disso, é apresentado um estudo de casos do Planejamento da Operação do SIN que avalia o acoplamento feito entre os modelos de otimização de médio e curto prazo através dos cenários hidrológicos de médio e curto prazo. Com esta análise, é possível verificar como o acoplamento entre os modelos de otimização pode impactar as Estratégias e Políticas de Operação para o médio e curto prazo.

Palavras - chave

Séries temporais, previsão de vazões, modelos estocásticos, afluências, planejamento energético.

Abstract

Deus, Marcella Lanzetti Daher; Souza Reinaldo Castro. **Time Series Applied to Operation Planning of the National Interconnected Electric System**. Rio de Janeiro, 2008. 130p. M.Sc. Dissertation - Electrical Engineering Department, Pontifical Catholic University of Rio de Janeiro.

The natural vocation of Brazil for the hydroelectricity made the National Interconnected Electric System – NIS to be developed with strong predominance of hydroelectric origin creation. However, choosing for a hydroelectricity base you have to deal with significant uncertainty associated to the rivers inflows and all hydrographical basins of the country. Therefore, the production structure of Brazilian hydroelectric energy was created to minimize the risks associated to the random behavior of inflows. To contemplate the inflows stochasticity in the operation planning of NIS, the National Operator of the Electrical System - ONS uses a chain of models that contains a model of inflows forecasting for the short term, and a model to generate scenarios of inflows. These models provide inputs for the optimizations model can establish the strategies and policies for the operation of medium and short term, contemplating the volatility of inputs. This dissertation describes the time series processes used in the model of inflows forecasting for the short term and in the models to generate scenarios of inflows for the medium and short term. Moreover, this paper presents a study of cases of Operation Planning of the NIS that analyze the coupling made between the models for optimization of medium and short term through the hydrological scenarios for medium and short term. By this analysis, is possible realize how the coupling between the models of optimization can impact the strategies and policies for the operation of medium and short term.

Key-Words

Time Series, forecasting inflows, stochastic models, inflows, energetic planning.

SUMÁRIO

LISTA DE TABELAS	11
LISTA DE FIGURAS	12
1 INTRODUÇÃO	13
1.1 CONSIDERAÇÕES INICIAIS	13
1.2 OBJETIVO	15
1.3 ESTRUTURA DA DISSERTAÇÃO.....	16
2 CARACTERÍSTICAS DO SISTEMA INTERLIGADO NACIONAL – SIN..	17
2.1 VOCAÇÃO À HIDROELETRICIDADE	17
2.2 DIMENSÃO DO SISTEMA.....	18
2.3 OTIMIZAÇÃO DA UTILIZAÇÃO DOS RECURSOS ENERGÉTICOS ..	25
3 O PROBLEMA DO PLANEJAMENTO DA OPERAÇÃO E A FUNÇÃO DE CUSTO FUTURO	29
3.1 INTRODUÇÃO	29
3.2 SISTEMAS PURAMENTE TÉRMICOS.....	30
3.3 SISTEMAS HIDROTÉRMICOS	32
3.4 FUNÇÃO DE CUSTO FUTURO – MINIMIZAÇÃO DO CUSTO TOTAL DE OPERAÇÃO.....	36
4 SÉRIES TEMPORAIS APLICADAS AO PLANEJAMENTO ENERGÉTICO DA OPERAÇÃO DO SIN	48
4.1 INTRODUÇÃO	48
4.2 SÉRIES TEMPORAIS	49
4.3 PROCESSOS ESTOCÁSTICOS.....	49
4.4 ESTACIONARIEDADE	50
4.5 PROCESSO ESTOCÁSTICO RUÍDO BRANCO.....	51
4.6 PROPRIEDADES ESTATÍSTICAS	52
4.6.1 MÉDIA E VARIÂNCIA.....	52
4.6.2 AUTOCOVARIANÇA E AUTOCORRELAÇÃO.....	52
4.6.3 PROCESSOS ERGÓTICOS.....	53
4.7 MODELOS LINEARES ESTACIONÁRIOS	53
4.7.1 MODELOS AUTO-REGRESSIVOS AR (p).....	53
4.7.2 MODELOS MÉDIAS MÓVEIS MA (q)	59
4.7.3 MODELOS AUTO REGRESSIVOS MÉDIAS MÓVEIS ARMA (p,q).....	63
4.8 MODELOS PERIÓDICOS.....	68
4.8.1 MODELOS AUTO-REGRESSIVOS PERIÓDICOS – PAR (p).....	69
4.8.2 MODELOS AUTO-REGRESSIVOS MÉDIAS-MÓVEIS PERIÓDICOS – PARMA (p,1)	73
4.9 ERRO MÉDIO QUADRÁTICO DE PREVISÃO	77
5 MODELOS UTILIZADOS NO PLANEJAMENTO E PROGRAMAÇÃO DA OPERAÇÃO DO SIN	79
5.1 MODELOS HIDROLÓGICOS – PREVISÃO DE VAZÕES E GERAÇÃO DE CENÁRIOS NO ONS	79
5.1.1 MODELO PREVIVAZH.....	80
5.1.2 MODELO PREVIVAZ.....	81
5.1.3 MODELO PREVIVAZM	87

5.1.4 MODELO GEVAZP.....	88
5.2 MODELOS DE OTIMIZAÇÃO ENERGÉTICA	91
5.2.1 MODELO NEWAVE	94
5.2.2 MODELO DECOMP.....	95
6 ESTUDO DE CASOS – ANÁLISE DO ACOPLAMENTO ENTRE OS MODELOS DE MÉDIO E CURTO PRAZO.....	99
6.1 INTRODUÇÃO	99
6.2 PREMISSAS.....	103
6.3 METODOLOGIA	104
6.4 RESULTADOS.....	105
6.4.1 Caso I – PMO Outubro de 2007	105
6.4.2 Caso II – PMO Janeiro de 2008	111
7 CONCLUSÕES	117
8 REFERÊNCIAS BIBLIOGRÁFICAS.....	119
ANEXOS	122

LISTA DE TABELAS

Tabela 2.1: Empreendimentos em Operação na Matriz Energética Brasileira.....	19
Tabela 4.1: Resumo das Características dos Modelos AR(p), MA (q) e ARMA (p,q).....	65
Tabela 5.1: Modelos Implementados no PREVIVAZM	83
Tabela 5.2: Resumo das Características dos Modelos Utilizados no Planejamento Energético do SIN	88
Tabela 6.1: Número de Cenários do DECOMP por Mês.....	104
Tabela 6.2: Número de Cenários do DECOMP que Extrapolaram o NEWAVE	105
Tabela 6.3: Custo Total de Operação.....	100
Tabela 6.4: Variação Percentual do Custo Total de Operação (%)	108
Tabela 6.5: Custo Marginal de Operação do Subsistema Sudeste	109
Tabela 6.6: Custo Marginal de Operação do Subsistema Nordeste	109
Tabela 6.7: Energia Natural Afluente Prevista e Verificada – Caso I	110
Tabela 6.8 – Atualização dos Dados Diários de Vazão	110
Tabela 6.9: Número de Cenários do DECOMP que Extrapolaram o NEWAVE Caso II.....	112
Tabela 6.10: Custo Total de Operação – Caso II	115
Tabela 6.11: Variação Percentual do Custo Total de Operação (%) – Caso II.....	115
Tabela 6.12: Custo Marginal de Operação do Subsistema Sudeste – Caso II.....	116
Tabela 6.13: Custo Marginal de Operação do Subsistema Nordeste – Caso II.....	116
Tabela 6.14: Energia Natural Afluente Prevista e Verificada – Caso II.....	117

LISTA DE FIGURAS

Figura 2.1: Comparação Brasil – Europa.....	19
Figura 2.2: Energias Naturais Afluentes	20
Figura 2.3: Malha de Interligação do SIN	22
Figura 2.4: Integração das Bacias Hidrográficas	23
Figura 2.5: Principais Usinas em Operação	24
Figura 2.6: Capacidade de Armazenamento	25
Figura 2.7: Importância das Interligações de Bacias.....	27
Figura 3.1: Mecanismo de Despacho de Usinas Térmicas em Função do Preço.....	31
Figura 3.2: Associação Presente e Futuro.....	33
Figura 3.3: Curva do Custo Total	38
Figura 3.4: Hipóteses de Afluências Para o Cálculo do Custo Futuro Médio.....	39
Figura 3.5: Derivada da Função de Custo Futuro.....	40
Figura 3.6: Problema da Otimização.....	40
Figura 3.7: Derivada da Função de Custo Futuro.....	41
Figura 3.8: Enfoques Árvore e Pente.....	42
Figura 3.9: Construção da Função de Custo Futuro.....	44
Figura 3.10: Construção da Função de Custo Futuro para Todas as Hipóteses de Afluência	44
Figura 3.11: Construção da Função de Custo Futuro para o Enfoque Pente	45
Figura 3.12: Uma Função de Custo Futuro por Etapa no Enfoque Pente.....	46
Figura 3.13: Enfoque Árvore X Pente	47
Figura 4.1: Autocorrelação do Ruído Branco.....	51
Figura 4.2: Autocorrelação do Modelo AR (p)	57
Figura 4.3: Função Teórica da Autocorrelação Parcial do Modelo AR (p)	59
Figura 4.4: Função Teórica da Autocorrelação Parcial do Modelo MA (2).....	63
Figura 4.5: Função Teórica da Autocorrelação do Modelo ARMA (p,q).....	66
Figura 4.6: Função Teórica da Autocorrelação Parcial de um Modelo ARMA (1,1).....	67
Figura 5.1: Possíveis Estruturas de Correlação do PREVIVAZ	83
Figura 5.2: Critério de Seleção do Sistema de Modelos PREVIVAZ.....	85
Figura 5.3: Cadeia de Modelos Hidrológicos	87
Figura 5.4: Cadeia de Modelos Hidrológicos	89
Figura 5.5: Interação do GEVAZP com o DECOMP.....	90
Figura 5.6: Interação do GEVAZP com o NEWAVE.....	91
Figura 5.7: Cadeia de Modelos de Otimização	92
Figura 6.1: Simulação Forward – Determinação dos Estados de Armazenamento.....	101
Figura 6.2: Simulação Backward – Cálculo da Função de Custo Futuro	102
Figura 6.3: Desacoplamento Parcial entre NEWAVE e DECOMP na Região SE.....	106
Figura 6.4: Desacoplamento Parcial entre NEWAVE e DECOMP na Região NE	107
Figura 6.5: Variação do Custo Total de Operação	108
Figura 6.6: Desacoplamento Parcial entre NEWAVE e DECOMP na Região SE – Caso II	112
Figura 6.7: Desacoplamento Parcial entre NEWAVE e DECOMP na Região NE – Caso II	113
Figura 6.8: Variação do Custo Total de Operação – Caso II.....	114

1

INTRODUÇÃO

1.1

CONSIDERAÇÕES INICIAIS

O Sistema Interligado Nacional – SIN caracteriza-se por ser um sistema hidrotérmico de grande porte, com significativa predominância de geração de origem hidroelétrica.

Em função desta característica, observa-se uma estreita correlação entre os estoques existentes nos reservatórios das usinas e as afluições a estes para a garantia do atendimento futuro aos seus requisitos de carga. Todavia, as afluições futuras apresentam significativas incertezas, o que confere ao planejamento da operação do SIN um viés estocástico.

Entretanto, há necessidade de se mitigar os riscos decorrentes da relação entre os estoques presentes, as afluições futuras e o não atendimento aos requisitos de carga do SIN.

Neste contexto, a expansão do SIN foi efetuada ao longo dos anos através da inserção de usinas com reservatórios de regularização. Tipicamente, estes reservatórios armazenam água nos períodos úmidos, disponibilizando este recurso durante o período seco. Este fato permite que se mantenha um adequado fluxo de água nos rios, disponibilizando recursos hídricos para a geração das usinas localizadas a jusante.

Além disso, houve a expansão do sistema de transmissão possibilitando a interligação das diversas regiões do país. Este fato foi de extrema importância, uma vez que o SIN dispõe de bacias hidrográficas com uma significativa diversidade hidrológica. Assim sendo, a malha de transmissão passou a funcionar como uma verdadeira “usina virtual”, permitindo a produção de energia elétrica nas bacias onde os recursos hídricos são mais abundantes, reduzindo o custo de geração.

Além da implantação dos reservatórios de regularização, houve a expansão de um parque gerador térmico, para efetuar a complementação dos recursos de geração hidroelétrica.

Desta forma, fez-se necessário o estabelecimento de parâmetros que indiquem o momento de despachar os recursos térmicos, para reduzir o uso dos estoques armazenados nos reservatórios das usinas hidroelétricas.

Para tal, foi desenvolvido o conceito de Valor da Água – VA (Pereira, 1989) para valorar os estoques presentes armazenados nos reservatórios do SIN.

Entretanto esta valoração depende das afluições futuras, uma vez que há um acoplamento temporal entre a decisão presente de se usar os estoques armazenados e o estado futuro de armazenamento do sistema.

As estratégias e Políticas de Operação do SIN são definidas com o auxílio de modelos computacionais de otimização da operação de médio e curto prazo (Maceira et al., 2002) que analisam os estados presente e futuro dos estoques do SIN, levando em conta o acoplamento temporal das decisões tomadas e a estocasticidade das afluições.

Estes modelos têm como função objetivo minimizar o valor esperado do Custo Total de Operação, representado pelo Custo Presente mais o Custo Futuro, no horizonte temporal do Planejamento da Operação Energética.

As decisões adotadas por esta cadeia de modelos envolvem a definição da geração hidrotérmica, e dos intercâmbios de energia entre os subsistemas que compõem o SIN e, eventualmente, de cortes de carga.

O Valor da Água é calculado para cada um dos quatro subsistemas que compõem o SIN, e permite estabelecer onde se encontram os recursos energéticos de menor custo, através da comparação entre os valores da água para os diversos subsistemas e entre estes e o custo de operação das usinas térmicas.

Admitindo-se que para um determinado horizonte de planejamento o valor presente da água seja inferior ao Custo Variável Unitário de Operação – CVU das usinas térmicas, não haverá despacho térmico para complementação energética. Isto significa que o estado presente de armazenamento do sistema, bem como os armazenamentos futuros decorrentes dos cenários futuros de afluições, indicam o atendimento aos requisitos de carga apenas por geração de origem hidroelétrica.

Neste raciocínio, à medida que há o desestoque presente de água nos reservatórios e os cenários futuros de afluições não se apresentem favoráveis, haverá uma elevação no valor presente da água. Assim sendo, haverá o despacho por ordem de mérito de custo de todas as usinas cujo custo de operação seja inferior ao Valor da Água para complementação energética, ou seja, para reduzir o uso dos estoques presentes de água existentes nos reservatórios.

Todavia, ao longo do tempo, tem-se observado uma redução continuada na capacidade de regularização do SIN, dado que houve uma significativa redução na capacidade de regularização das novas usinas. Este fato pode ser analisado por duas

vertentes: a primeira, em função da significativa necessidade de capital para a implantação de usinas com reservatórios de regularização. A segunda, em função dos requisitos crescentes associados ao Meio Ambiente e ao Uso Múltiplo da Água, que podem ter inviabilizado a construção de usinas hidroelétricas com grandes reservatórios.

Estes fatos conduziram a uma perda significativa na capacidade de regularização do SIN, o que se traduz em uma crescente dependência de períodos hidrológicos favoráveis. Neste contexto, acentua-se a importância da definição das previsões de aflúências e dos cenários hidrológicos para a composição das Estratégias e Políticas de Operação.

1.2

OBJETIVO

Tendo o exposto como referência, esta dissertação visa apresentar uma aplicação prática do uso de séries temporais no planejamento energético do Sistema Interligado Nacional.

Para tal, a abordagem consistirá na apresentação dos modelos da cadeia de planejamento energético do SIN, com enfoque na estrutura matemática dos modelos de previsão de vazões e geração de cenários de aflúências, visto que os mesmos baseiam-se em séries temporais.

Neste contexto, haverá a caracterização do acoplamento do modelo de planejamento de médio prazo, responsável pela definição das estratégias de operação a sistemas equivalentes¹, com o de planejamento de curto prazo, que define as políticas de operação a usinas individualizadas. Sumariamente, este acoplamento é definido a partir da Função de Custo Futuro – FCF (Benders, 1962), que fornece o Valor da Água (em valor presente) em função do estado de armazenamento dos reservatórios do SIN. Para efetuar-lo, o modelo de curto prazo utiliza-se de cenários de aflúências gerados para o segundo mês do estudo (mês probabilístico) tendo como referência as aflúências determinísticas² previstas para o primeiro mês (mês determinístico). Assim sendo, ao percorrer os cenários do segundo mês, o modelo de curto prazo efetua o acoplamento com a Função de Custo Futuro – FCF produzida pelo modelo de médio prazo. Entretanto, os estados de armazenamento obtidos pelo modelo de

1 Sistemas equivalentes – A energia é armazenada por subsistema, e não por usinas individualizadas. Para se obter um sistema equivalente, todas as usinas do subsistema são somadas de modo que resulte em um único reservatório agregado.

2 Aflúências Determinísticas – As previsões semanais de aflúências são consideradas determinísticas por que é dado um único valor de probabilidade futura.

curto prazo ao percorrer os cenários de afluências do segundo mês podem se acoplar na FCF, gerada pelo modelo de médio prazo, em uma Região desta não tão bem construída. Este fato pode distorcer o sinal econômico do Valor da Água, com reflexos diretos nas Políticas de Operação de curto prazo. Considerando-se a relevância deste fato, esta dissertação fará uma análise de sua ocorrência em dois casos do Planejamento Mensal elaborados para o SIN, destacando suas conseqüências.

1.3

ESTRUTURA DA DISSERTAÇÃO

No Capítulo 2 é feita uma explanação sobre o Sistema Interligado Nacional e suas características.

O Capítulo 3 faz uma abordagem sobre o planejamento da operação de sistemas elétricos, estabelecendo as diferenças entre sistemas térmicos e hidrotérmicos.

O Capítulo 4 descreve os modelos de séries temporais utilizados na previsão de vazões e geração de cenários hidrológicos, quais sejam: os modelos estacionários AR (p), MA (q), ARMA (p,q) e os modelos Periódicos PAR (p), e PARMA (p,q).

O Capítulo 5 discorre sobre as ferramentas computacionais utilizadas pelo ONS na execução de suas funções de planejamento e otimização da operação, preparando o leitor para um melhor entendimento do estudo de casos.

No Capítulo 6 são mostrados dois estudos de casos de aplicação de séries temporais ao planejamento energético do Sistema Interligado Nacional – SIN.

O Capítulo 7 conclui o estudo deixando clara a importância da previsão para o planejamento energético do SIN.

2

CARACTERÍSTICAS DO SISTEMA INTERLIGADO NACIONAL – SIN

2.1

VOCAÇÃO À HIDROELETRICIDADE

O sistema de produção e transmissão de energia elétrica do Brasil – Sistema Interligado Nacional (SIN) – pode ser classificado como um sistema hidrotérmico de grande porte e com forte predominância de usinas hidroelétricas.

A análise da potência instalada no Brasil indica, de forma evidente, que a expansão do parque gerador ocorreu, principalmente, através da instalação de usinas hidroelétricas, que hoje representam cerca de 71% da potência elétrica instalada no País, cabendo às usinas térmicas (incluindo-se as Centrais Nucleares de Angra dos Reis) cerca de 21%. Ainda, com uma pequena participação, as usinas eólicas representam 0,2% da capacidade instalada no País.

Essa característica do Sistema Interligado Nacional se deve ao fato do Brasil possuir diversas bacias hidrográficas com centenas de rios perenes e caudalosos, que se espalham pelas diversas regiões geográficas do País, topografia favorável à construção de grandes reservatórios de regularização plurianual e regimes de chuvas diferentes entre si ao longo de um mesmo ano – períodos marcadamente chuvosos e secos (sazonalidade anual). Este fato induziu à vocação hidroelétrica do nosso sistema que, no entanto, não tem todo o seu potencial hidroelétrico aproveitado, o qual é da ordem de 260.000 MW.

Atualmente, as usinas hidroelétricas são responsáveis por mais de 90 % da energia produzida anualmente no SIN, percentual este que é variável ao longo dos anos em função das condições hidrológicas e das políticas de otimização eletroenergética verificadas em cada ano.

A predominância das usinas hidroelétricas na capacidade instalada do SIN, independentemente da oportunidade técnica, estratégica, econômica e necessária de novas fontes de geração, complementar ou distribuída ao longo da rede, deverá perdurar por várias

décadas, pois, as expectativas da escala de crescimento do mercado de energia elétrica aliadas à vocação natural à hidroeletricidade conferem ao Setor Elétrico Brasileiro o desafio de, a cada ano, elevar, consideravelmente, sua oferta de energia elétrica.

Entretanto, apesar de seu grande potencial disponível, a expansão do parque hidroelétrico vem encontrando barreiras devido a questões ambientais, conduzindo a uma maior participação das usinas térmicas na matriz energética brasileira, como verificado nos últimos leilões de energia nova.

A Tabela a seguir apresenta a capacidade instalada dos empreendimentos em operação na matriz energética brasileira, incluindo o SIN e sistemas isolados, em 19 de dezembro de 2007.

Tabela 2.1: Empreendimentos em Operação na Matriz Energética Brasileira
Fonte: ANEEL

Tipo	Capacidade Instalada		%	Total		%	
	N° de Usinas	kW		N° de Usinas	kW		
Hidro	665	76.848.711	70,8%	665	76.848.711	70,8%	
Gás	Natural	78	10.193.502	9,4%	108	11.344.480	10,5%
	Processo	30	1.150.978	1,1%			
Petróleo	Óleo Diesel	575	2.918.706	2,7%	597	4.388.600	4,0%
	Óleo Residual	22	1.469.894	1,4%			
Biomassa	Bagaçõ de Cana	238	3.016.625	2,8%	282	4.074.599	3,8%
	Licor Negro	13	794.817	0,7%			
	Madeira	26	224.207	0,2%			
	Biogás	2	20.030	0,0%			
	Casca de Arroz	3	18.920	0,0%			
Nuclear	2	2.007.000	1,8%	2	2.007.000	1,8%	
Carvão Mineral	Carvão Mineral	7	1.415.000	1,3%	7	1.415.000	1,3%
Eólica		16	247.050	0,2%	16	247.050	0,2%
Importação	Paraguai		5.650.000	5,2%	8.170.000	7,5%	
	Argentina		2.250.000	2,1%			
	Venezuela		200.000	0,2%			
	Uruguai		70.000	0,1%			
Total	1.677	108.495.440	100,0%	1.677	108.495.440	100,0%	

2.2

DIMENSÃO DO SISTEMA

Outra característica singular do SIN é sua dimensão continental que lhe confere uma maior complexidade operacional quando comparado aos sistemas de outros países.

Para a visualização desta característica, a Figura 2.1 representa, em escala, a inserção da Rede Básica³ do Sistema Interligado Nacional no continente europeu. Nesta Figura podemos observar que a interligação entre os Subsistemas do SIN, tendo-se como referência as cidades de Porto Alegre no Subsistema Sul e Manaus no Subsistema Norte, corresponde à inserção no continente europeu de uma malha de interligação entre as cidades de Lisboa (Portugal) e Estocolmo (Suécia).

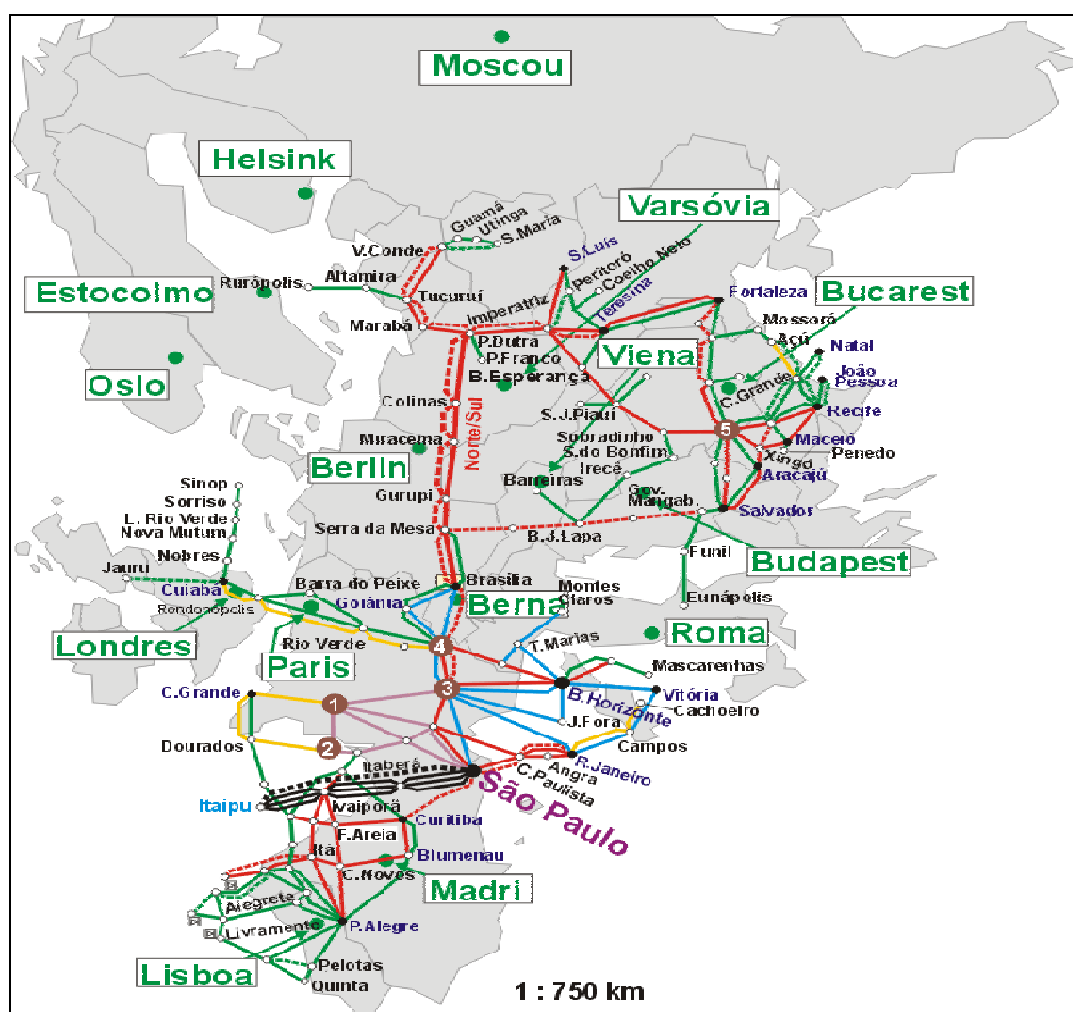


Figura 2.1: Comparação Brasil – Europa
Fonte: ONS

³ A rede básica é integrada pelas linhas de transmissão, barramentos, transformadores de potência e equipamentos com tensão igual ou superior a 230 kV e algumas linhas de 138 kV.

A fim de se aproveitar, de forma ótima, os recursos energéticos existentes no País e a sazonalidade hidrológica própria de cada Região, foi necessário o estabelecimento de uma extensa malha de transmissão que funciona como uma usina virtual, pois possibilita a transferência de excedentes energéticos permitindo a otimização dos estoques armazenados nos reservatórios das usinas hidroelétricas.

A Figura 2.2 permite uma adequada visualização da complementariedade hidrológica entre as Regiões do País que compõem os Subistemas do SIN. Nela estão representadas as Energias Naturais Afluentes, ENAs, em valores médios mensais do histórico de afluências de 1931 a 2006, bem como seu valor médio anual para cada uma das Regiões do País.

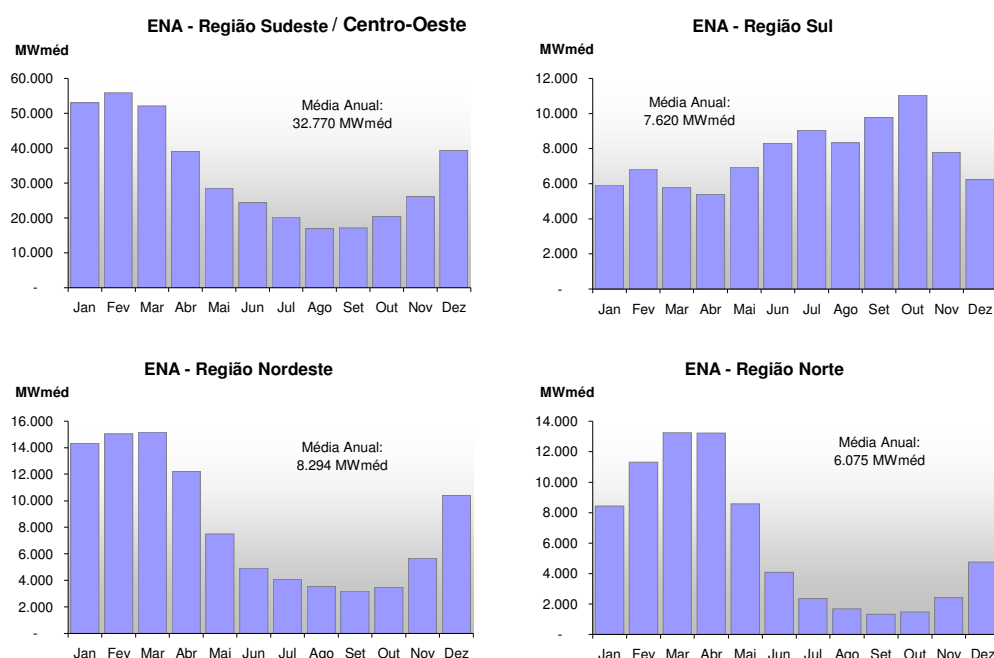


Figura 2.2: Energias Naturais Afluentes
Fonte: ONS

Observa-se, inicialmente, a complementariedade hidrológica existente entre as Regiões Sudeste/Centro-Oeste e Sul. A distribuição das energias naturais afluentes médias mensais mostra que os respectivos períodos secos e úmidos não são coincidentes nessas Regiões.

Na Região Sudeste/Centro-Oeste o período úmido ocorre de dezembro a abril, onde se observam afluências superiores à média anual. Embora a Região Sul não possua período úmido e seco bem delimitados, neste mesmo período as afluências encontram-se abaixo do

valor médio anual, de modo que o Sul pode importar os excedentes energéticos da Região Sudeste.

De forma semelhante, quando a situação hidrológica da Região Sul está mais favorável, isto é, as ENAs mensais estão superiores à média anual o subsistema Sul pode exportar o seu excedente energético para as Regiões Sudeste/Centro-Oeste, já que estas encontram-se em seus períodos secos.

Com isso, há uma clara complementariedade hidrológica que indica a necessidade da transferência de excedentes energéticos entre essas Regiões, bem como do uso dos estoques que apresentarem menor custo.

A análise das ENAs das Regiões Norte e Nordeste nos permite observar a inexistência da complementariedade hidrológica entre essas Regiões, uma vez que há coincidência entre os seus respectivos períodos secos e úmidos. Entretanto, durante o período úmido as afluições à UHE Tucuruí, na Região Norte, são extremamente elevadas, conduzindo à ocorrência de vertimentos turbináveis⁴ e passíveis de serem exportados.

Desta forma, os excedentes energéticos da Região Norte são importantes para a complementação energética da Região Nordeste, apesar dos regimes hidrológicos das Regiões Norte e Nordeste serem coincidentes.

A fim de permitir a otimização integrada entre as Regiões do SIN, no início da década de 80 foram estabelecidas as interligações entre as Regiões Sudeste/Centro-Oeste e Sul e entre as Regiões Norte e Nordeste e, em 1999 foi concluída a interligação entre o Norte e o Sul do País.

Os quatro subsistemas do País possuem as seguintes características:

- *Subsistema Sudeste/Centro-Oeste*: Grande mercado de demanda no País e importador de outras Regiões durante a maior parte do ano. Possui grande capacidade de armazenamento em múltiplos reservatórios.
- *Subsistema Sul*: Sistema hidrotérmico com grande variabilidade de armazenamento. Os sentidos dos intercâmbios com o Subsistema Sudeste/Centro-Oeste são altamente dependentes da sua hidrologia.

4 Vertimento Turbinável: Quantidade de água que passa pelos vertedouros, quando o reservatório encontra-se cheio, mesmo com capacidade de geração disponível, porém sem demanda suficiente para absorver essa geração.

- *Subsistema Nordeste*: Importador ou exportador de energia conforme condições hidrológicas na Região. Com a conclusão das expansões da malha de interligação com a Região Sudeste/Centro-Oeste, será um potencial exportador no período de ponta.

- *Subsistema Norte*: Eminente exportador de energia. Fornece energia para o período de ponta por pelo menos nove meses do ano.

O SIN é responsável pelo suprimento de energia a 96,6% do território nacional. Os 3,4% restantes constituem-se de sistemas isolados localizados na Região Norte do país que dispõem de sistemas hidrotérmicos e/ou térmicos locais.

A Figura 2.3 mostra a atual malha de interligações do Sistema Interligado Nacional.

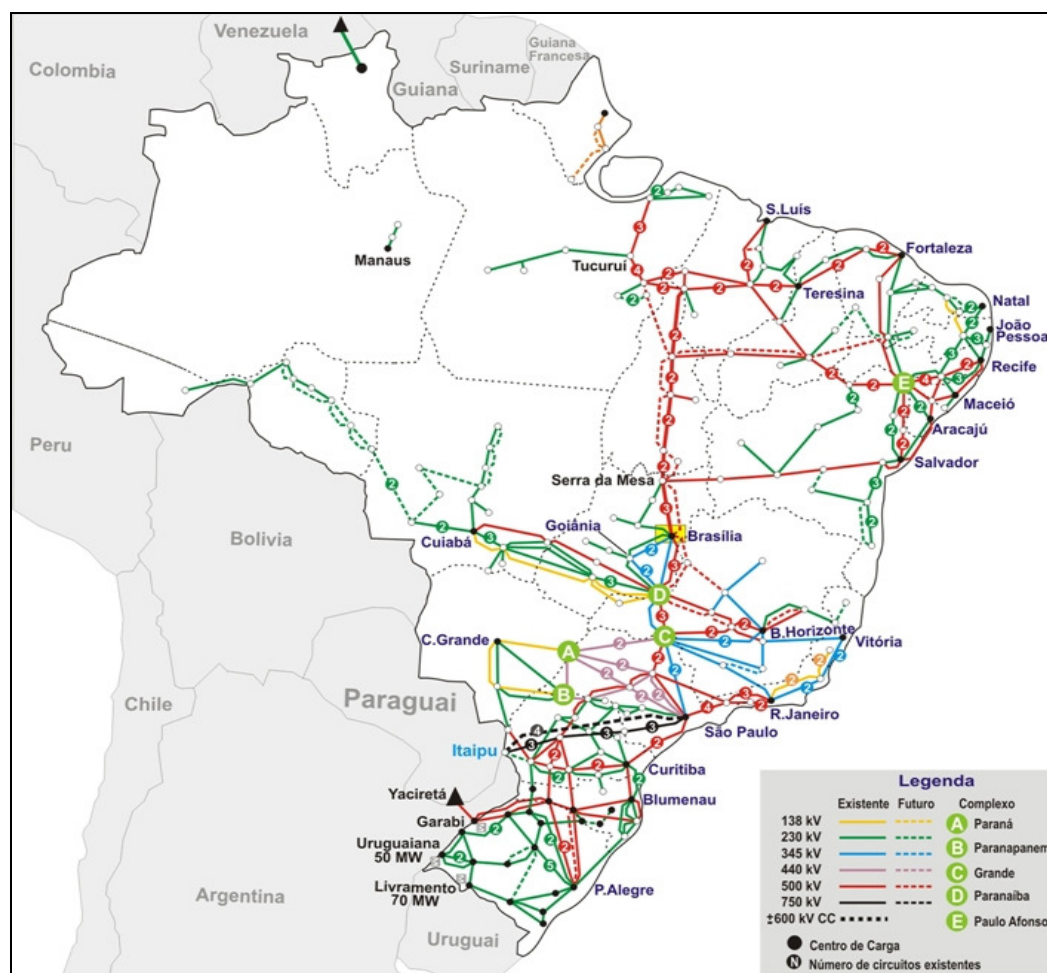


Figura 2.3: Malha de Interligação do SIN
Fonte: ONS

A Figura 2.4 apresenta as interligações entre os Subsistemas e bacias hidrográficas do SIN. São estas interligações que permitem que a complementariedade hidrológica das Regiões seja explorada através do intercâmbio energético. Elas propiciam a “transposição de bacias”, visto que utilizam o excedente energético de uma bacia em outra.

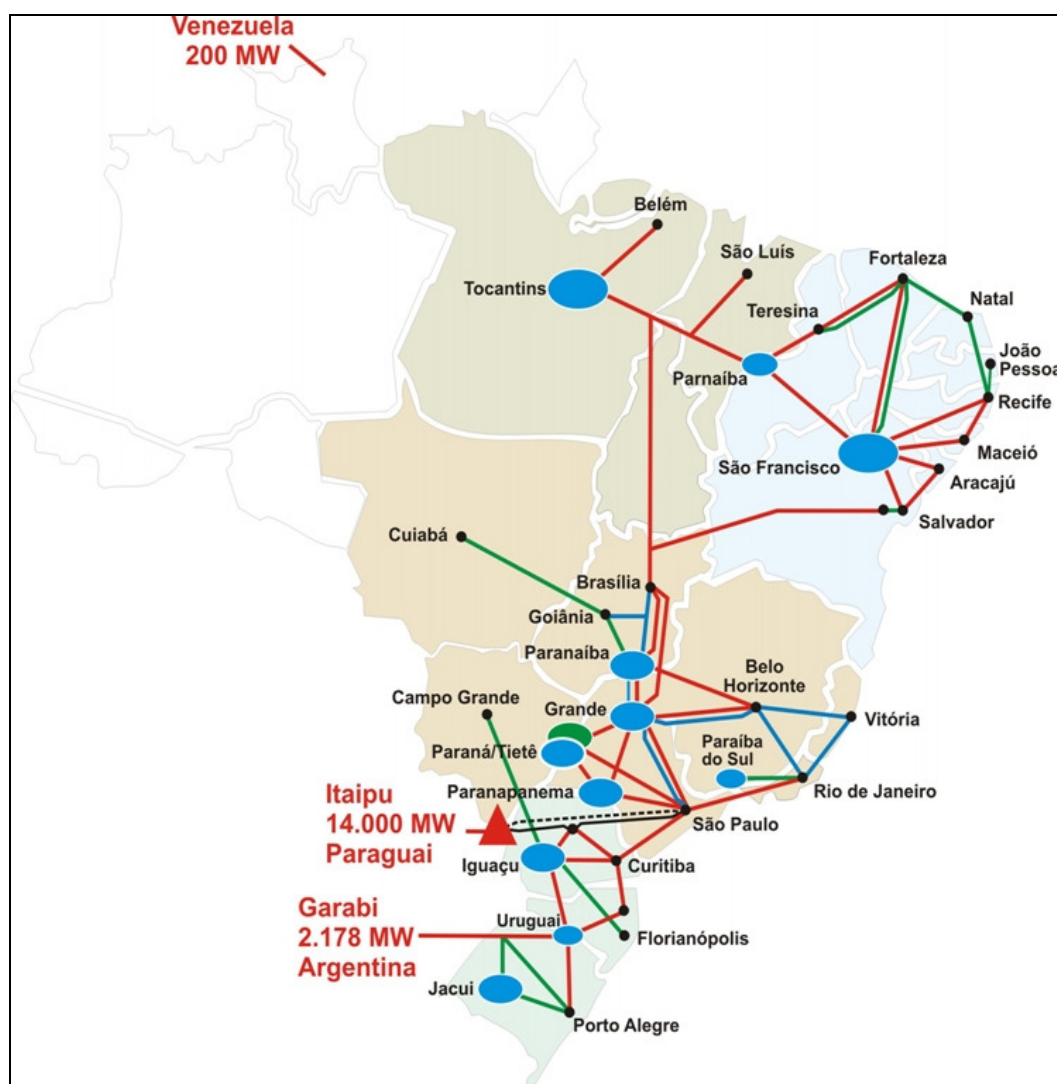


Figura 2.4: Integração das Bacias Hidrográficas
Fonte: ONS



Figura 2.5: Principais Usinas em Operação
Fonte: ONS

A Figura 2.5 apresenta uma visão das usinas com capacidade instalada superior ou igual a 30MW, que são aquelas despachadas de forma centralizada pelo ONS.

Atualmente a capacidade de armazenamento do SIN é de 272.556 MWmês, distribuídos pelas Regiões, conforme mostrado na Figura 2.6. Observa-se a predominância do

subsistema Sudeste/Centro-Oeste, com cerca de 69,7% do armazenamento total, bem como, em destaque, o quadrilátero dos principais reservatórios de regularização da Região Sudeste/Centro-Oeste, localizados nos estados de Minas Gerais e Goiás.

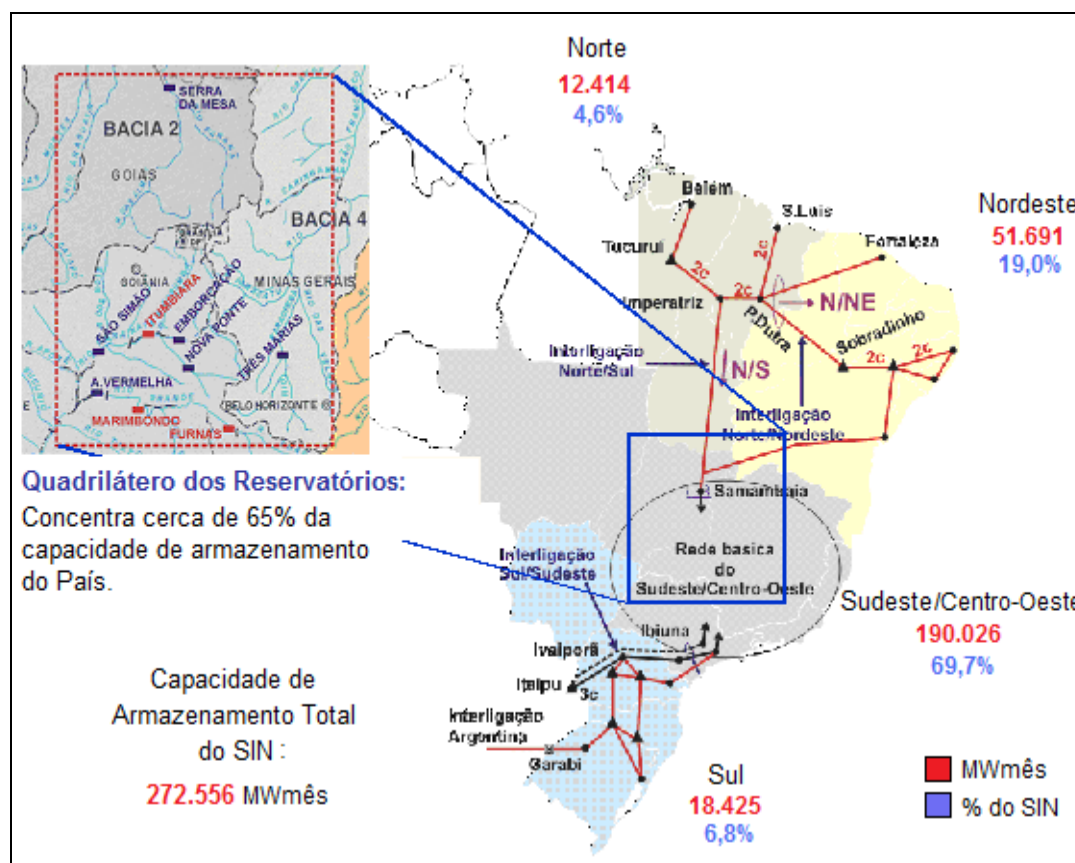


Figura 2.6: Capacidade de Armazenamento
Fonte: ONS

2.3

OTIMIZAÇÃO DA UTILIZAÇÃO DOS RECURSOS ENERGÉTICOS

Desde meados da década de 70, o sistema eletroenergético brasileiro é operado de forma coordenada, visando obter ganhos sinérgicos, minimizando os custos globais de produção de energia elétrica.

A operação de um sistema predominantemente hidroelétrico, com regularização pluri-anual, deve ser integrada, pois quando uma usina com reservatório regulariza uma bacia,

os benefícios são incorporados não só à própria bacia, mas também a todas as usinas que estão na bacia, buscando-se a operação ótima do sistema.

De um modo geral a otimização energética pode ser definida como o resultado do conjunto de ações que visam ao atendimento da carga do SIN ao menor custo, ou seja, evitando vertimentos nos reservatórios das usinas, minimizando a utilização de geração térmica e equalizando, na medida do possível, os custos marginais de operação entre as regiões interligadas. Esta otimização sinaliza a necessidade da utilização de geração térmica em complementação à geração hidroelétrica e da transferência de energia entre regiões ou bacias, bem como indica a adequada produção de energia por bacia, considerando as restrições operativas associadas a cada aproveitamento, de caráter ambiental e de uso múltiplo da água.

O despacho de geração térmica para complementar os recursos energéticos das usinas hidroelétricas é efetuado sempre que o Valor da Água for superior ao custo de operação variável das usinas térmicas. Outra situação em que as usinas térmicas também podem ser despachadas é para atender a requisitos elétricos do SIN; nesta situação, o despacho de geração térmica é efetuado analisando o aumento de confiabilidade proporcionado pelo despacho de geração, e não através da comparação do custo de geração térmica com o Valor da Água.

A transferência de energia entre regiões se dá ordinariamente quando da ocorrência de vertimentos turbináveis em uma região ou existindo desequilíbrio significativo entre as condições de armazenamento e/ou as energias naturais afluentes nessas regiões, conforme pode ser visto na Figura 2.7. Desse modo, é possível minimizar vertimentos, reduzir a utilização de geração térmica, reduzir riscos de racionamento e aumentar os estoques de energia armazenada do SIN, conferindo ao sistema de transmissão o papel de verdadeira usina virtual ou realizador de transposições de bacias.

Observando-se a Figura 2.7, para a situação I percebe-se que devido à inexistência de interligações entre os Subsistemas A e B é impossível transmitir o excedente energético do subsistema A (que está tendo vertimento turbinável já que existe excedente de água, porém não há mercado para consumir esta energia) para atender ao mercado do Subsistema B. Porém, ao adicionar-se uma interligação entre os Subsistemas A e B (situação II) torna-se possível transmitir o excedente energético de A para atender o mercado em B.

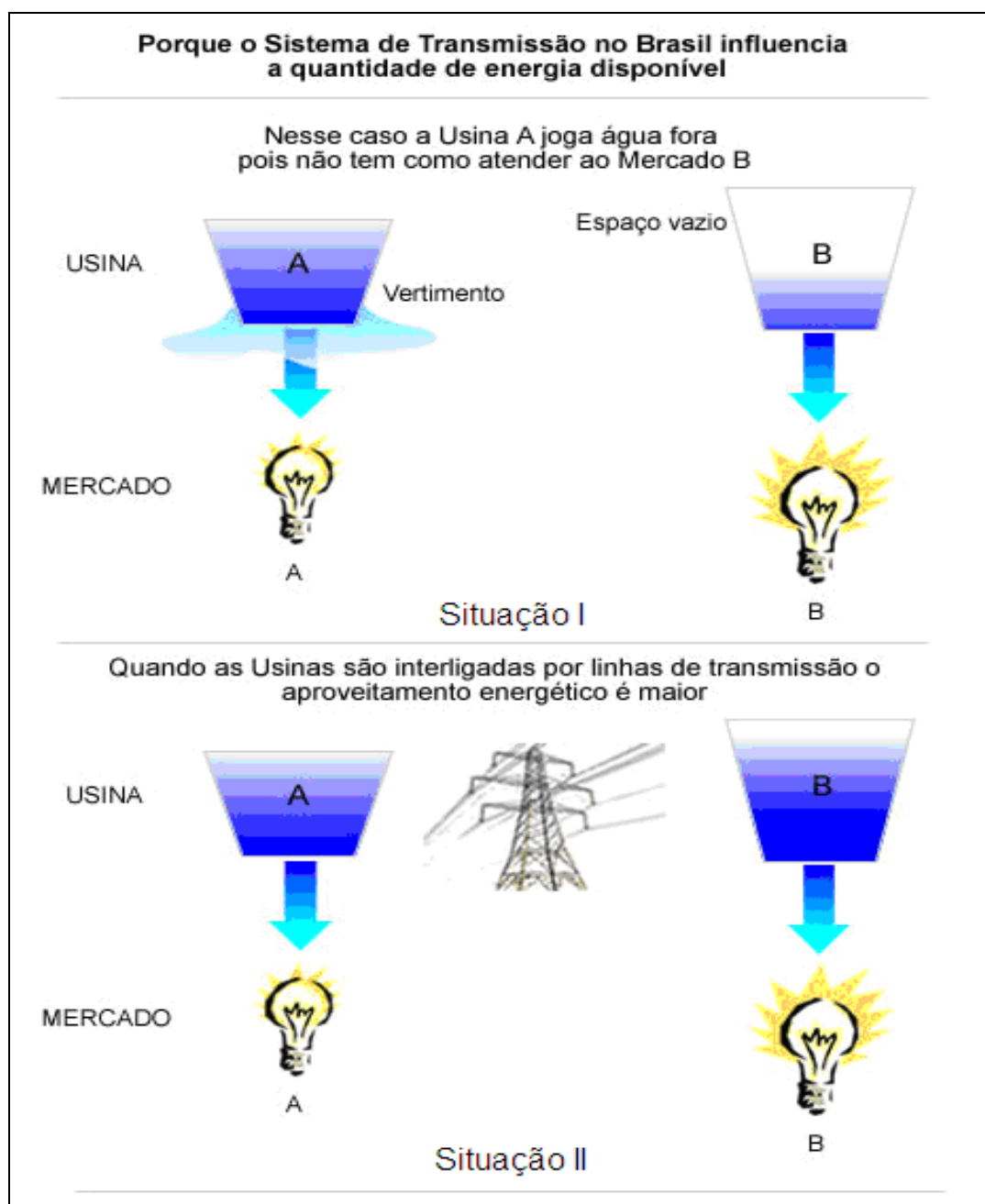


Figura 2.7: Importância das Interligações de Bacias.

A otimização energética do SIN também indica o montante de participação de cada bacia hidrográfica (usinas e reservatórios) no atendimento à carga e é obtida através de cálculos realizados por modelos computacionais, cuja função objetivo consiste em garantir o

atendimento do mercado consumidor de energia elétrica minimizando o custo total de operação, garantindo assim a modicidade tarifária ao consumidor final.

Essa otimização energética leva em consideração o conjunto de restrições operativas associadas a cada bacia, relativas aos condicionantes ambientais e ao uso múltiplo de suas águas e, em especial, aquelas associadas às condições de confiabilidade e segurança operacional do Sistema Interligado.

Face às características singulares do SIN, quais sejam: predominância hidroelétrica, aproveitamentos em cascata com grandes reservatórios de regularização e múltiplos proprietários, diferentes bacias hidrográficas com diversidades hidrológicas, rede de transmissão extensa e geração térmica complementar, os procedimentos necessários para a otimização de seus recursos energéticos apresentam um elevado grau de complexidade, com reflexo direto no processo de planejamento e programação da operação do Sistema.

Estas características demandam a operação centralizada do Sistema Interligado Nacional que é embasada na interdependência operativa entre as usinas, na interconexão dos sistemas elétricos, na integração dos recursos de geração e transmissão e no atendimento ao mercado.

Em resumo, a natureza do Sistema Interligado Nacional faz com que exista um acoplamento espacial e temporal das decisões tomadas na sua operação eletroenergética. O uso, no presente, mais ou menos intensivo dos estoques de água nos diversos reservatórios, frente às incertezas das condições hidrometeorológicas e do consumo, irá afetar a operação futura do Sistema, em termos de garantia de atendimento e de custos ao consumidor final.

Por outro lado, as decisões operativas no presente dependem de como se imagina a configuração futura do Sistema – quantas usinas, quantas linhas de transmissão, qual o mercado a ser atendido, enfim, pode-se usar mais ou menos a “poupança energética dos reservatórios” no presente dependendo da expansão prevista da rede de transmissão e da oferta de energia elétrica.

Essas características, aliadas ao fato de que a expansão da geração e da transmissão demanda longos tempos de maturação, devido à necessidade de vários estudos prévios e exigências ambientais cada vez maiores antes da operação, fazem com que a operação em tempo real do SIN seja precedida de estudos de Planejamento e Programação, de forma a garantir, ao menor custo, a confiabilidade do sistema, a qualidade e a quantidade de energia requerida pelo mercado consumidor.

3

O PROBLEMA DO PLANEJAMENTO DA OPERAÇÃO E A FUNÇÃO DE CUSTO FUTURO

3.1

INTRODUÇÃO

A definição da política ótima de geração de um sistema elétrico tem por objetivo o atendimento aos requisitos de carga, a cada intervalo de tempo do planejamento, ao mínimo custo, preservando a confiabilidade da operação elétrica.

A análise deste conceito pressupõe a distinção entre sistemas puramente térmicos e sistemas hidrotérmicos, uma vez que se faz necessária a definição de parâmetros para se avaliar qual o recurso de geração de menor custo, a cada intervalo de planejamento. A natureza distinta destes tipos de sistemas elétricos conduz a diferentes abordagens para a determinação do recurso de menor custo.

Em uma análise sumária, pode-se observar que um sistema puramente térmico tem um caráter de planejamento determinístico, uma vez que as variáveis associadas à disponibilidade futura de geração, tais como: disponibilidade de combustíveis, cronograma de manutenção, restrições operativas etc., são conhecidas, ou seja, são determinísticas para quaisquer intervalos do planejamento.

No caso de sistemas hidrotérmicos, principalmente naqueles com predominância hidroelétrica, como é o caso do Sistema Interligado Nacional, as incertezas associadas às afluições futuras fazem com que o problema tenha uma característica estocástica. Este fato pressupõe que a cada decisão de uso de um recurso de geração de menor custo haja uma premissa de cenário futuro de afluições e um risco associado.

Assim, há necessidade da definição de parâmetros para se avaliar qual é o recurso de menor custo. Em sistemas puramente térmicos, o custo variável das usinas térmicas é o parâmetro utilizado para a ordenação dos recursos, por ordem de mérito de custo, para o atendimento aos requisitos da carga.

Para a operação de sistemas hidrotérmicos, foi necessário o desenvolvimento de uma metodologia que pudesse valorar a água estocada nos reservatórios, associada às incertezas quanto às afluências futuras.

Com este objetivo foi estabelecida para o SIN uma metodologia para o cálculo de um parâmetro denominado Valor da Água – VA (Pereira, 1989). Este parâmetro determina o custo do desestoque incremental da água armazenada nos reservatórios que compõe os subsistemas do SIN, para atender a um incremento de carga.

O despacho de geração térmica, para complementar os recursos energéticos das usinas hidroelétricas, é efetuado sempre que o Valor da Água for superior ao custo de operação variável das usinas térmicas. Não obstante, as usinas térmicas também podem ser despachadas para atender a requisitos elétricos do SIN; nesta situação, o despacho de geração térmica é efetuado analisando-se o aumento de confiabilidade proporcionado pelo despacho de geração, e não através da comparação do custo de geração térmica com o Valor da Água.

Assim sendo, para o SIN, o processo de definição do recurso de menor custo – processo de otimização energética – conjuga o custo variável das usinas térmicas, basicamente composto pelo custo do combustível, com o Valor da Água das usinas hidroelétricas de cada Subsistema, além de custos associados a eventuais déficits de energia.

É importante notar que o equilíbrio entre o Valor da Água e o Custo da Térmica corresponde à premissa inicial de minimização do Custo Total de Operação, que é dado pela soma do Custo Presente e do Custo Futuro.

Para que se tenha uma estratégia operativa para cada estado de armazenamento e afluências que venham a ocorrer futuramente no sistema, foi criada uma função capaz de valorar a água armazenada nos reservatórios do SIN e trazer este custo para valor presente. Essa função foi chamada de Função de Custo Futuro (Benders, 1962).

3.2

SISTEMAS PURAMENTE TÉRMICOS

Em sistemas de geração compostos somente de unidades térmicas, o custo de cada usina depende basicamente do custo variável de operação, cuja parcela mais significativa é aquela associada ao custo do combustível. Portanto, o problema do planejamento da operação consiste em determinar a combinação de usinas que minimize o custo total de operação (representado pelo custo determinístico de geração das usinas térmicas a carvão, óleo

combustível, nucleares, etc.) necessário para atender a demanda, observando-se os limites de geração e as restrições operativas de cada usina térmica.

A solução clássica deste problema consiste em alocar-se na curva de carga, a cada intervalo de tempo, os geradores térmicos por ordem de mérito de custo, respeitando-se as restrições de cada um deles, bem como aquelas de caráter sistêmico.

A cada intervalo temporal o Custo Marginal de Operação (Pereira e Pinto, 1985) será representado pelo custo variável do último gerador térmico despachado, caso este ainda tenha disponibilidade para suprir o próximo incremento de carga.

Na hipótese do último gerador despachado ter utilizado sua plena disponibilidade, o próximo incremento de carga será atendido pelo gerador de custo imediatamente superior, ainda não despachado. Neste caso, o Custo Marginal de Operação, nesse intervalo temporal, corresponderá ao custo marginal desse gerador, mesmo que este não tenha sido programado para despacho.

A Figura a seguir exemplifica o mecanismo de despacho de usinas térmicas, por ordem de mérito de custo, para atender uma curva de carga prevista, em sistemas puramente térmicos.

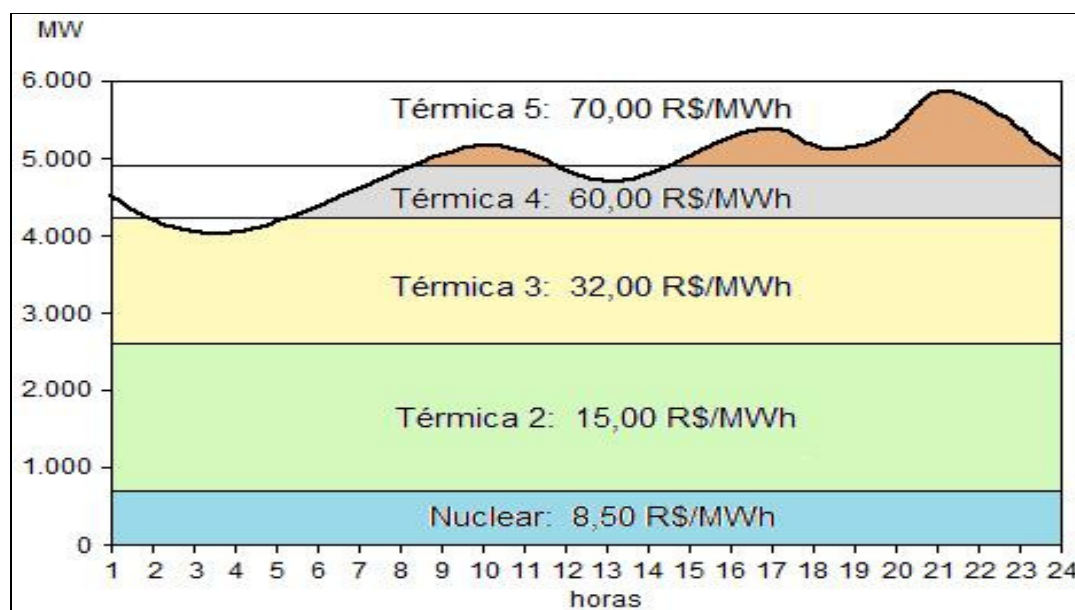


Figura 3.1: Mecanismo de Despacho de Usinas Térmicas em Função do Preço

Com base na Figura 3.1, o mecanismo de despacho térmico por ordem de mérito pode ser entendido da seguinte forma:

Passo 1 – As usinas térmicas são enfileiradas em ordem crescente de custo (Nuclear < térmica 2 < térmica 3 < térmica 4 < térmica 5).

Passo 2 – A primeira térmica da fila (a mais barata) é despachada para atender a carga. Se sua capacidade não for suficiente para atender a carga, então é despachada a próxima térmica. Este procedimento se repete para cada intervalo de tempo da operação.

Passo 3 – A última unidade despachada indica o Custo Marginal de Operação do sistema. De acordo com a Figura, para $t = 4$ a usina nuclear e térmicas 1, 2 e 3 atendem a 100% da carga do sistema, então o CMO é dado pelo valor da última térmica despachada, térmica 3, de custo de operação de R\$ 32,00/MWh. Para $t = 10$ é necessário despachar a térmica 5, e neste caso o CMO será R\$ 70,00/MWh.

A operação de sistemas puramente térmicos deve levar em consideração as características básicas do problema, que podem ser resumidas da seguinte forma:

- É desacoplado no tempo, quer dizer, uma decisão de operação hoje não tem efeito no custo de operação do próximo estágio;
- As unidades têm um custo direto de operação, quer dizer, o custo de operação de uma unidade não depende do nível de geração de outras unidades. Além disso, a operação de uma unidade não afeta a capacidade de geração ou a disponibilidade de outra unidade;
- A confiabilidade do fornecimento de energia depende somente da capacidade total de geração disponível e não da estratégia de operação das unidades do sistema.

3.3

SISTEMAS HIDROTÉRMICOS

As incertezas associadas às afluições futuras conduzem a análise de sistemas hidrotérmicos, principalmente àqueles com predominância hidroelétrica, a apresentar características estocásticas.

Desta forma, há necessidade de se buscar a otimização do uso dos recursos energéticos disponíveis, representados majoritariamente pelos estoques existentes nos reservatórios das usinas hidroelétricas, bem como pelos recursos de geração térmica e importação de energia de outros Subsistemas, dentre outros, considerando-se o custo imediato e o custo futuro do atendimento à carga associado à decisão a ser adotada.

Desta forma, em uma análise conceitual, podemos distinguir os cenários associados às decisões e seus respectivos impactos nos custos imediato e futuro, como ilustrado na Figura a seguir.

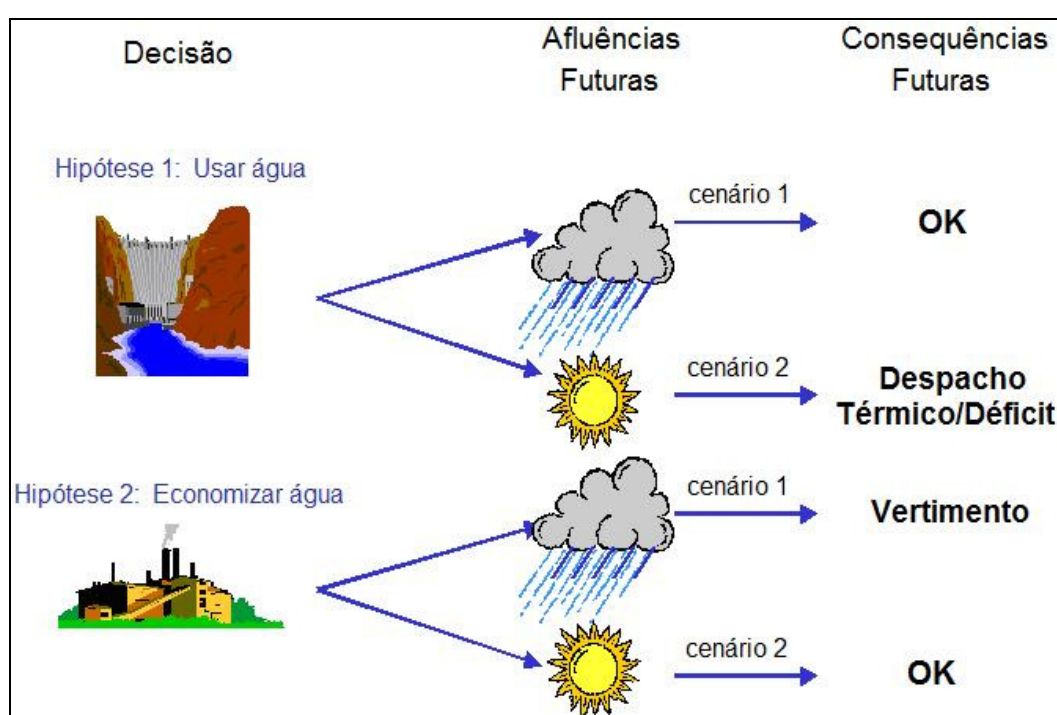


Figura 3.2: Associação Presente e Futuro
Fonte: ONS

Hipótese 1

A decisão consistiu em efetuar o atendimento da carga no presente com o desestoque da água existente nos reservatórios.

Esta decisão conduz a um custo imediato baixo, uma vez que não foi necessário o uso de recursos de geração térmica, que apresentam custos mais elevados para o atendimento da carga.

Entretanto, no futuro, esta decisão terá consequências em função do cenário hidrológico que venha a ocorrer.

Admitindo-se a ocorrência do Cenário 1 – Altas Afluências

Neste cenário, as afluências futuras são favoráveis. Com isso o estoque utilizado é recomposto sem a necessidade de geração térmica para complementação dos recursos energéticos.

Assim sendo, há um custo futuro também baixo, que demonstra o acerto na decisão presente de se atender a carga apenas com recursos de geração hidroelétrica.

Neste caso houve:

$$\begin{array}{l} \text{CUSTO IMEDIATO BAIXO} \\ + \\ \text{CUSTO FUTURO BAIXO} \end{array} = \text{OPERAÇÃO ECONÔMICA}$$

Admitindo-se a ocorrência do Cenário 2 – Baixas Afluências

Neste cenário, as afluências futuras não são favoráveis. Dessa forma, há necessidade de despacho térmico complementar para a reposição dos estoques das usinas hidroelétricas utilizadas no presente para o atendimento à carga. Se o despacho térmico não for suficiente e não houver estoques adequados nas usinas hidroelétricas no futuro, poderá, inclusive, ser necessário implantar um processo de racionamento, ou seja, poderá haver déficit de energia decorrente de uma decisão presente equivocada.

Desta forma, a operação não foi econômica, pois:

$$\begin{array}{l} \text{CUSTO IMEDIATO BAIXO} \\ + \\ \text{CUSTO FUTURO ALTO} \end{array} = \begin{array}{l} \text{CUSTO FUTURO ALTO} \\ \text{E} \\ \text{POSSIBILIDADE DE DÉFICITS} \end{array}$$

Hipótese 2

A decisão consistiu em efetuar o atendimento da carga no presente com o despacho de geração térmica, visando economizar os estoques de água armazenados nas usinas hidroelétricas. Com isso, o custo imediato é alto.

Assim sendo, faz-se necessário analisar as conseqüências dessa decisão admitindo-se a ocorrência dos cenários de afluências apresentados anteriormente.

Admitindo-se a ocorrência do Cenário 1 – Altas Afluências

Neste cenário, as afluências futuras são favoráveis. Com isso, os reservatórios apresentam significativa probabilidade de incorrer em vertimentos, o que caracterizaria um desperdício do gasto efetuado no presente com o uso da geração térmica.

Os estoques das usinas hidroelétricas seriam repostos, inclusive apresentando excedentes energéticos superiores aos requisitos da carga, caracterizando vertimentos turbináveis, ou seja, aqueles em que há capacidade de geração na usina, porém sem carga suficiente para consumir esses excedentes energéticos, conduzindo o custo marginal a valor nulo.

Desta forma, neste cenário teríamos uma operação não econômica, pois:

$$\begin{array}{l} \text{CUSTO IMEDIATO ALTO} \\ + \\ \text{CUSTO FUTURO NULO} \end{array} = \text{VERTIMENTOS (DESPERDÍCIO)}$$

Admitindo-se a ocorrência do Cenário 2 – Baixas Afluências

Neste cenário as afluências futuras não seriam favoráveis. Com isso, a decisão presente de manter o estoque das usinas hidroelétricas e atender a carga total ou parcialmente com despacho de geração térmica mostrou-se adequada.

Embora no presente esta operação tenha custo elevado, a operação futura terá custo baixo, uma vez que a carga poderá ser atendida total ou parcialmente com geração hidroelétrica.

Logo, neste cenário teríamos:

$$\begin{array}{l} \text{CUSTO IMEDIATO ALTO} \\ + \\ \text{CUSTO FUTURO BAIXO} \end{array} = \text{OPERAÇÃO ECONÔMICA}$$

Desta forma, o planejamento da operação de sistemas hidrotérmicos possui características que podem ser assim resumidas:

➤ É acoplado no tempo, ou seja, é necessário avaliar as conseqüências futuras de uma decisão no presente. A solução ótima é um equilíbrio entre o benefício presente do uso da água e o benefício futuro de seu armazenamento, medido em termos da economia esperada dos combustíveis das unidades térmicas;

➤ A este problema dinâmico se agrega o problema da volatilidade das vazões afluentes aos reservatórios, que variam sazonalmente e regionalmente. A incerteza com respeito às vazões, aliada à incerteza com respeito à demanda de energia, faz do planejamento da operação de sistemas hidrotérmicos um problema essencialmente estocástico;

➤ É acoplado no espaço, ou seja, há interdependência na operação de usinas hidroelétricas, pois a quantidade de água liberada em uma usina afeta a operação de outra situada a jusante;

➤ O valor da energia gerada por uma hidroelétrica somente pode ser medido em termos da economia resultante nos custos de geração térmica ou déficits evitados e não diretamente como uma função apenas do estado da usina.

➤ O sistema deve ser operado objetivando minimizar o Custo Total de Operação, que é dado pela soma do Custo Imediato e do Custo Futuro. Para se conhecer o custo futuro, deve ser calculada uma função que dê o valor da água em função do estado futuro de armazenamento do sistema, essa função é conhecida como Função de Custo Futuro.

3.4 FUNÇÃO DE CUSTO FUTURO – MINIMIZAÇÃO DO CUSTO TOTAL DE OPERAÇÃO

A Função de Custo Futuro – FCF traduz a estratégia futura de operação para os modelos de curto prazo da cadeia de otimização (DECOMP (CEPEL, 2004) e DESSEM (Maceira et al., 2003)). Além disso, é através da Função de Custo Futuro, que é feito o encadeamento dos modelos de otimização, possibilitando que as políticas de operação de curto prazo (definidas pelo Modelo DECOMP e, futuramente pelo Modelo DESSEM) estejam em conformidade com as estratégias de operação de médio prazo, definidas pelo Modelo NEWAVE (CEPEL, 2006).

O grande objetivo do planejamento da operação é operar com segurança o Sistema Interligado Nacional, visando atingir o menor Custo Total possível em todo o horizonte de estudo. O Custo Total é formado pela soma dos custos variáveis de todos os recursos utilizados, como exemplo, a geração térmica, o intercâmbio entre regiões e o custo de déficit, que é o custo associado à demanda não suprida.

Para isso, todos os modelos da cadeia realizam a otimização dos recursos disponíveis, visando alcançar o mínimo custo, cabendo a esta estratégia gerenciar as decisões de geração de origem hidroelétrica ou termoelétrica, por usina, as decisões de intercâmbio entre as regiões do SIN e de corte de carga.

O Custo Total constitui-se de duas parcelas: o Custo Imediato referente às decisões tomadas no presente e o Custo Futuro referente às decisões que serão tomadas no futuro. No entanto, sabe-se que devido ao acoplamento temporal das decisões, causado pela predominância da hidroeletricidade, o Custo Futuro é influenciado pelas decisões tomadas no presente.

Conforme dito anteriormente, o Custo Total é composto pela soma do Custo Imediato com o Custo Futuro. O seu valor mínimo pode ser facilmente encontrado através das derivadas das funções de custo futuro e custo imediato, conforme é mostrado na Figura 3.3.

A derivada da curva de custo futuro em relação ao volume armazenado é o Valor da Água, ou seja, quanto mais próximo o volume estiver de zero maior será esse valor e quanto mais próximo estiver de 100% do armazenamento esse custo tenderá a zero. Por outro lado, a derivada do custo imediato em relação ao armazenamento representa o custo de geração térmica, ou de déficit caso não haja disponibilidade térmica para geração. A soma dessas duas inclinações se anula no ponto de menor custo de operação, este fato pode ser observado na Figura 3.3.

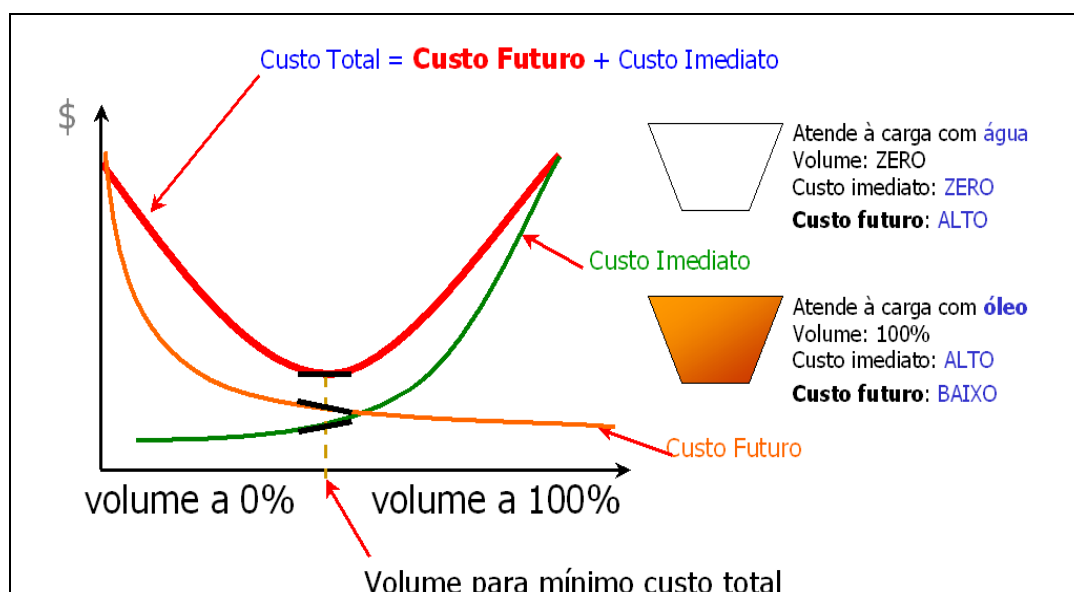


Figura 3.3: Curva do Custo Total
 Fonte: Kligerman, 2005

Portanto, toda a estratégia resume-se, então, em tomar as decisões no presente para que o reservatório atinja, ao final deste mês, o volume indicado pela seta, pois com isso garante-se o menor Custo Total.

A Figura acima expõe com clareza a dependência do custo de geração em relação à quantidade de água armazenada nos reservatórios. Isto implica em estudar o comportamento futuro das afluições, para poder conhecer o Custo Futuro relativo a cada estado de armazenamento que venha a ocorrer.

Sabe-se que a família de modelos Auto-regressivos Periódicos – PAR (p) ajusta-se bem ao comportamento das afluições, cujo caráter periódico está diretamente ligado a sazonalidade do regime hidrológico.

Conhecendo-se o modelo estocástico que representa as afluições pode-se sortear um grande número de hipóteses e acompanhar a evolução do sistema nos próximos cinco anos para cada trajetória correspondente a cada hipótese de afluição e calcular o custo médio de todas as hipóteses. Este é o Custo Futuro Médio.

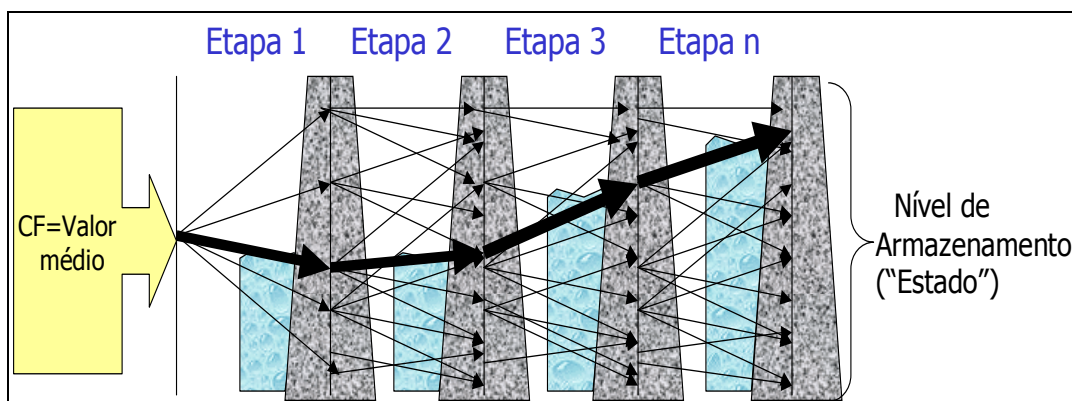


Figura 3.4: Hipóteses de Afluências Para o Cálculo do Custo Futuro Médio
Fonte: Kligerman, 2005

Para acompanhar a evolução do sistema e trazer as informações do futuro para o presente é utilizado o método de Programação Dinâmica Estocástica (Bellman, 1957). Este método além de trazer como resultado o Custo Futuro Médio para o Estado⁵ de partida do estudo, também encontra o Custo Futuro a partir de qualquer outro estado pertencente a qualquer Etapa e a qualquer uma das trajetórias sob as diversas hipóteses de afluências. Além disso, o método garante que o Custo Futuro calculado em cada estado é ótimo, ou seja, é o mínimo possível para as hipóteses consideradas.

Entretanto, o método não é adequado ao sistema brasileiro, pois se forem calculados custos futuros para 100 níveis de armazenamento, para cada reservatório do SIN em cada Etapa, resultaria em um número gigantesco de estados, o que impediria o uso da Programação Dinâmica aplicado a vários reservatórios. Este problema é conhecido na literatura como **Maldição da Dimensionalidade** (Bellman, 1961).

A fim de resolver o problema da maldição da dimensionalidade o método de Programação Dinâmica Dual Estocástica (Pereira e Pinto, 1984) deixa de calcular o custo futuro para todos os Estados de armazenamento, e passa a calcular para apenas alguns Estados. Com isso, passa a ser calculada uma outra grandeza a **Derivada do Custo Futuro**, que é a taxa de variação nas vizinhanças do estado em que foi calculado o custo futuro.

Desta forma, ficou mais fácil calcular o custo futuro em qualquer outro Estado, visto que basta considerar a envoltória superior das retas que representam as derivadas. É possível,

⁵ O Estado do sistema corresponde ao nível de armazenamento do reservatório de uma usina, e cada Etapa corresponde a um mês, uma semana, enfim, um intervalo de tempo.

portanto, calcular o Custo Futuro para o Estado V, vide Figura 3.5, assim como para qualquer outro Estado.

Matematicamente, achar o Custo Futuro em V utilizando a envoltória superior das retas pode ser escrito da seguinte forma:

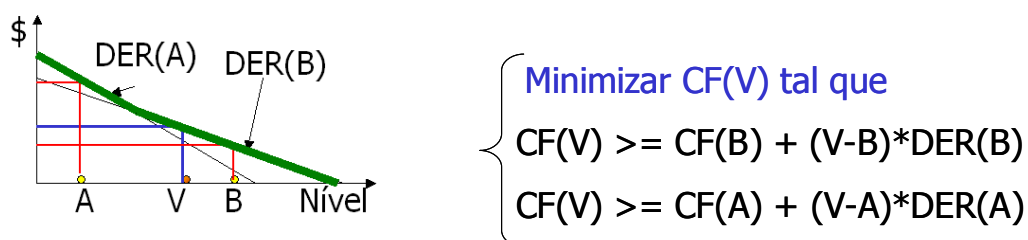


Figura 3.5: Derivada da Função de Custo Futuro
Fonte: Kligerman, 2005

Uma vez calculada a Função de Custo Futuro ela é utilizada para tomar decisões operativas como: o que vale mais, o Valor da Água, dado pela derivada da Função de Custo Futuro, ou o valor do combustível das usinas térmicas que é dado pela derivada da curva de custo imediato.

A Função de Custo Futuro é utilizada para comandar as decisões que determinam o estado de armazenamento a ser atingido em cada etapa. É ela quem executa a otimização da operação de cada uma das trajetórias em cada etapa, atendendo as premissas de minimização do Custo Total e de Atendimento da Carga.

Sendo assim, o problema de minimização do Custo Total é feito para cada etapa, de modo que sempre respeite o Atendimento da Carga, o Balanço Hídrico, as Restrições Hidroelétricas (de uso múltiplo e Ambientais) e as Restrições Elétricas, e o Custo Futuro esteja condicionado à Função de Custo Futuro, conforme mostra a Figura 3.6.



Figura 3.6: Problema da Otimização

Para que cada uma dessas condições seja satisfeita o modelo tem que obedecer as seguintes equações:

☒ Atendimento da Carga:

$$\text{Geração Hidro} + \text{Geração Termo} + \text{Recebimentos} - \text{Fornecimentos} + \text{Déficit} = \text{Carga}$$

☒ Balanço Hídrico:

$$\text{Nível Final} = \text{Nível Inicial} + \text{Afluência} - \text{Geração Hidro} - \text{Vertimento}$$

☒ Função de Custo Futuro:

Como visto anteriormente, a FCF é representada pelas retas que a compõem

$$CF \geq CF(B) + (\text{Nível Final} - B) * \text{Derivada (B)}$$

$$CF \geq CF(A) + (\text{Nível Final} - A) * \text{Derivada (A)}$$

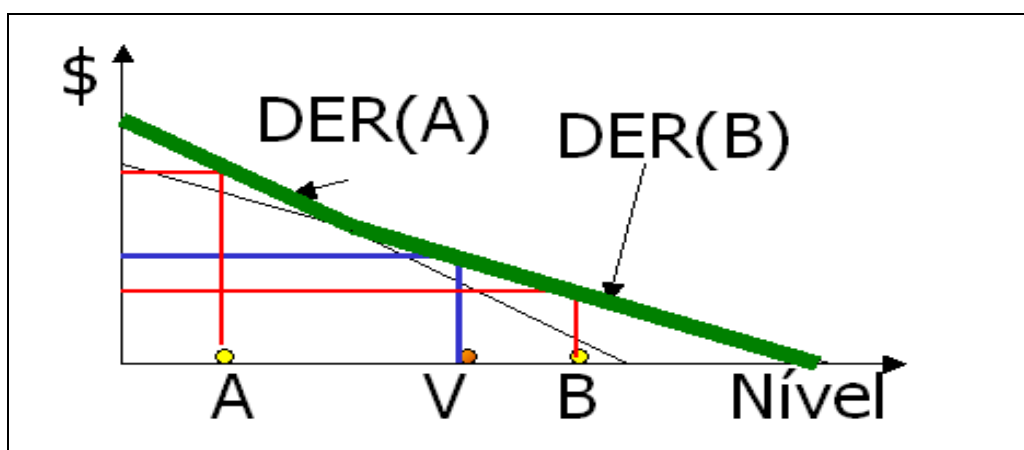


Figura 3.7: Derivada da Função de Custo Futuro
Fonte: Kligerman, 2005

☒ Restrições Hidroelétricas (englobam as restrições de uso múltiplo da água e ambientais):

➤ Controle de Cheias:

$$\text{Nível Final} \leq \text{Nível Máximo para controle de cheias}$$

➤ Defluência Mínima:

$$\text{Geração Hidro} + \text{Vertimento} \geq \text{Defluência Mínima}$$

- Retiradas para Uso Consuntivo⁶:

$$\text{Afluência} = \text{Afluência Bruta} - \text{Retiradas para Uso Consuntivo}$$

⊗ Restrições Elétricas:

- Fluxo Máximo entre Subsistemas:

$$\text{Intercâmbio } (x \rightarrow y) \leq \text{Fluxo Máximo } (x \rightarrow y)$$

- Geração Máxima por usina:

$$\text{Geração Hidro de uma usina} \leq \text{Geração Máxima da usina}$$

- Geração Máxima para um conjunto de usinas:

$$\text{Geração Hidro de um conjunto de usinas} \leq \text{Geração Máx. p/ o conjunto de usinas}$$

Cálculo do Custo Futuro

Como já foi dito antes, o Custo Futuro é determinado somente para poucos Estados, que são determinados através da simulação da operação baseada em seqüências de afluências sorteadas a partir da distribuição estatística.

Sob este ponto de vista, existem dois enfoques utilizados para a determinação do Custo Futuro, que são o enfoque *Árvore*, utilizado pelo DECOMP, onde cada hipótese se ramifica, sucessivamente, em cada etapa, e o enfoque *Pente*, utilizado pelo NEWAVE, onde cada hipótese se ramifica apenas no início do horizonte e dura até o fim do estudo, conforme pode ser visto na Figura 3.8.

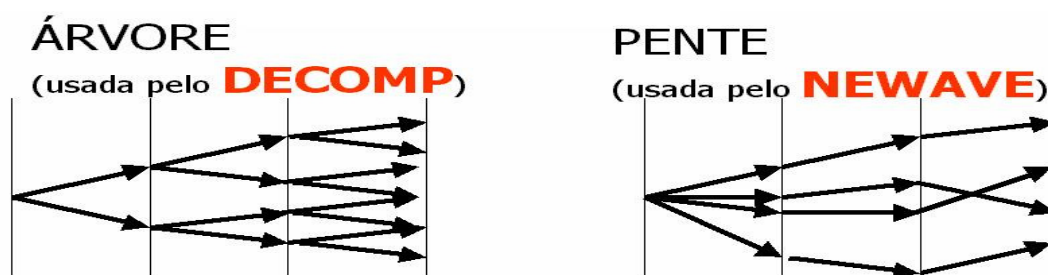


Figura 3.8: Enfoques *Árvore* e *Pente*
Fonte: Kligerman, 2005

⁶ Uso Consuntivo representa a perda de água por diversos usos, como irrigação, abastecimento humano, desvios de água para pecuária e outros fins que diminuem o espelho d'água do lago, o que difere do Uso Múltiplo da Água, que não representa perda de água, mas uma reserva destinada a outros fins que não a geração hidroelétrica como hidrovias, turismo, pesca e lazer.

Cada iteração do processo de otimização dos modelos é dividida em duas fases a fase Forward que vai do início até o final do horizonte calculando o estado de armazenamento dos reservatórios para uma dada aproximação da função de custo futuro, e a fase Backward que, do final do horizonte de estudo até o período inicial, aprimora a função de custo futuro (com a adição de uma nova reta para sua representação) a partir dos estados definidos na fase Forward.

Enfoque ÁRVORE

Sabe-se que, no enfoque Árvore, cada etapa dá origem a diversas hipóteses de afluências. Sendo assim, para cada Estado de Armazenamento em que se deseja calcular o Custo Futuro, tem-se que calcular o Custo Futuro de cada hipótese que parte deste Estado e em seguida fazer a média destes custos. Este custo médio representa o Custo Futuro Estimado para este Estado de armazenamento. Da mesma forma, a Derivada neste Estado também é dada pela média das derivadas fornecidas por cada hipótese que parte deste Estado.

Partindo-se do Custo Futuro obtido com a primeira iteração, é feita uma segunda otimização Forward com as mesmas hipóteses de afluência atingindo-se Estados de Armazenamentos diferentes dos da primeira iteração. Com isso, calcula-se novamente o Custo Futuro e sua Derivada através da otimização Backward, mas para o novo Estado atingido por cada hipótese de afluência.

Cada reta acrescentada à Função de Custo Futuro é chamada “**Corte de Benders**” que são utilizadas na aproximação da verdadeira FCF. Este processo se repete, iterativamente, até que a Função de Custo Futuro se torne uma função bem desenhada. O processo pode ser observado na Figura 3.9.

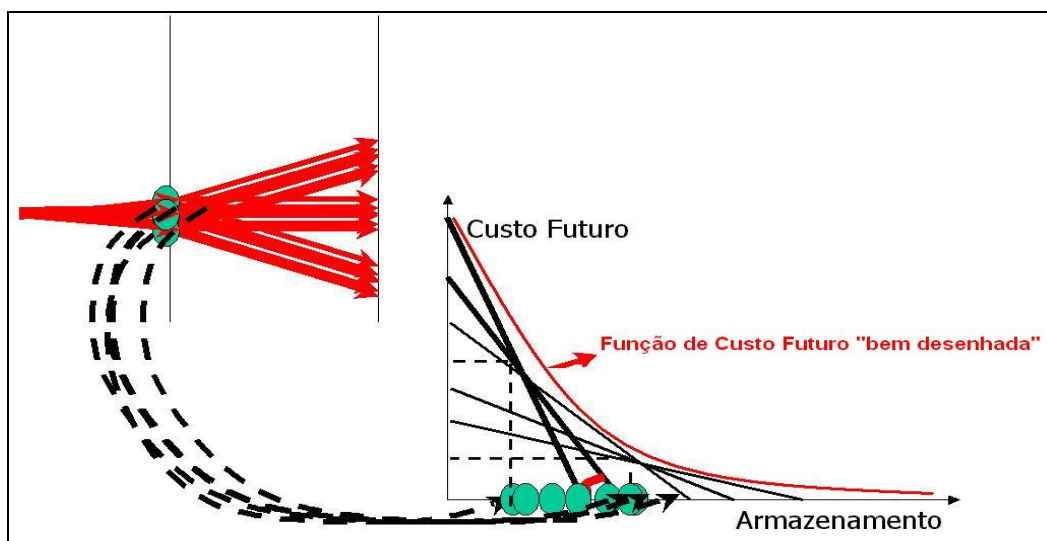


Figura 3.9: Construção da Função de Custo Futuro
Fonte: Kligerman, 2005

Nesta abordagem para o cálculo do Custo Futuro sob o enfoque *Árvore*, foi destacada apenas uma hipótese de afluência e os diferentes Estados atingidos por esta hipótese em cada otimização "forward" conforme pode ser observado na Figura 3.9.

Observando o problema como um todo, no enfoque *ÁRVORE*, é calculada uma Função de Custo Futuro para cada hipótese de afluência da Etapa anterior como pode ser observado na Figura 3.10.

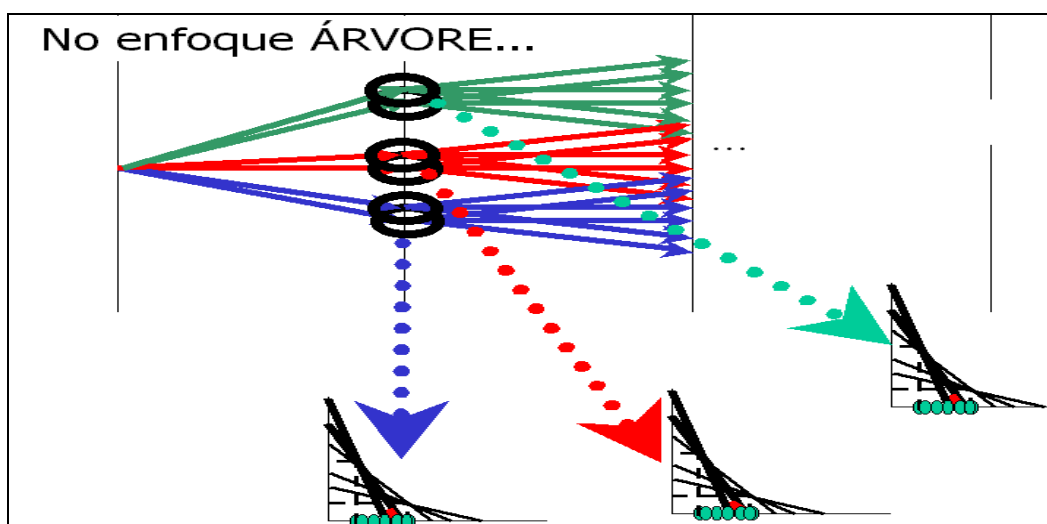


Figura 3.10: Construção da Função de Custo Futuro para Todas as Hipóteses de Afluência
Fonte: Kligerman, 2005

Enfoque PENTE

No enfoque PENTE, sabe-se que as hipóteses de afluições não se ramificam a cada etapa, o que faz com que em cada etapa seja gerada apenas uma Função de Custo Futuro, conforme pode ser visualizado na Figura 3.11. Isso só se tornou possível, quando se passou a considerar além do Armazenamento, as Afluições Passadas como grandeza para a definição do Estado, que antes, no enfoque Árvore, considerava apenas o Armazenamento.

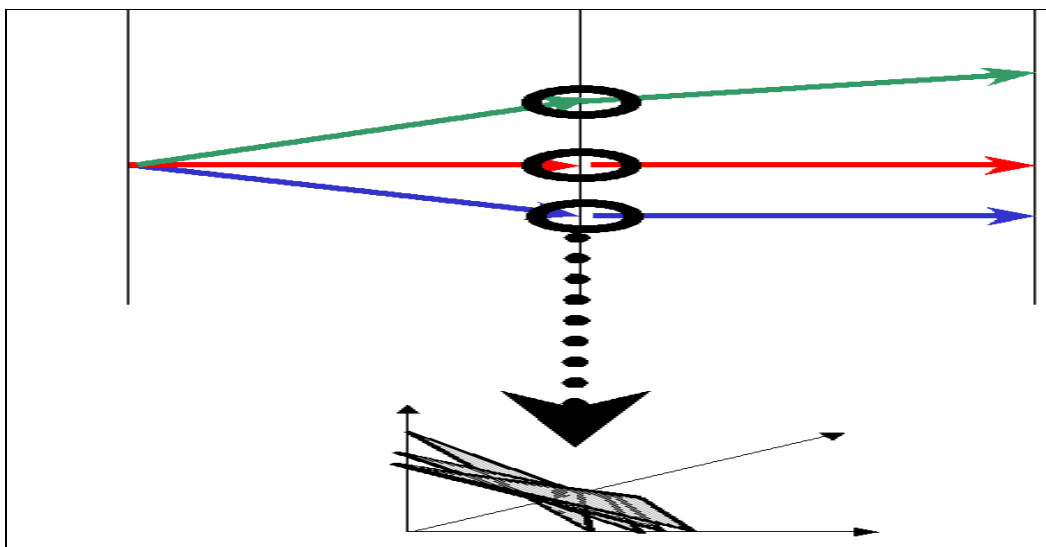


Figura 3.11: Construção da Função de Custo Futuro para o Enfoque Pente
Fonte: Kligerman, 2005

Por conta dessa consideração, a Derivada deixa de ser uma reta, e passa a ser um plano, já que passa a variar em função de duas variáveis, Nível de Armazenamento e Afluência Passada.

Como no enfoque PENTE as hipóteses não se ramificam a cada etapa, não faz sentido calcular o Custo Futuro, visto que não contemplaria a variabilidade do regime hidrológico.

Então, para suprir esta deficiência, para que seja calculado o Custo Futuro em algum Estado, na otimização Backward são sorteadas hipóteses adicionais de afluições partindo do Estado, e é temporariamente desconsiderada a hipótese de afluição utilizada na otimização Forward. Estas hipóteses são denominadas “ABERTURAS”.

Para cada uma destas Aberturas são calculados os Custos Futuros e suas Derivadas, e a média destes Custos e Derivadas resultam no Custo Futuro Estimado.

Cabe ressaltar que como na representação do estado do sistema estão incluídas afluências passadas, todos os estados de uma mesma Etapa, têm seus Custos Futuros e Derivadas representados em uma única Função de Custo Futuro, conforme mostra a Figura 3.12.

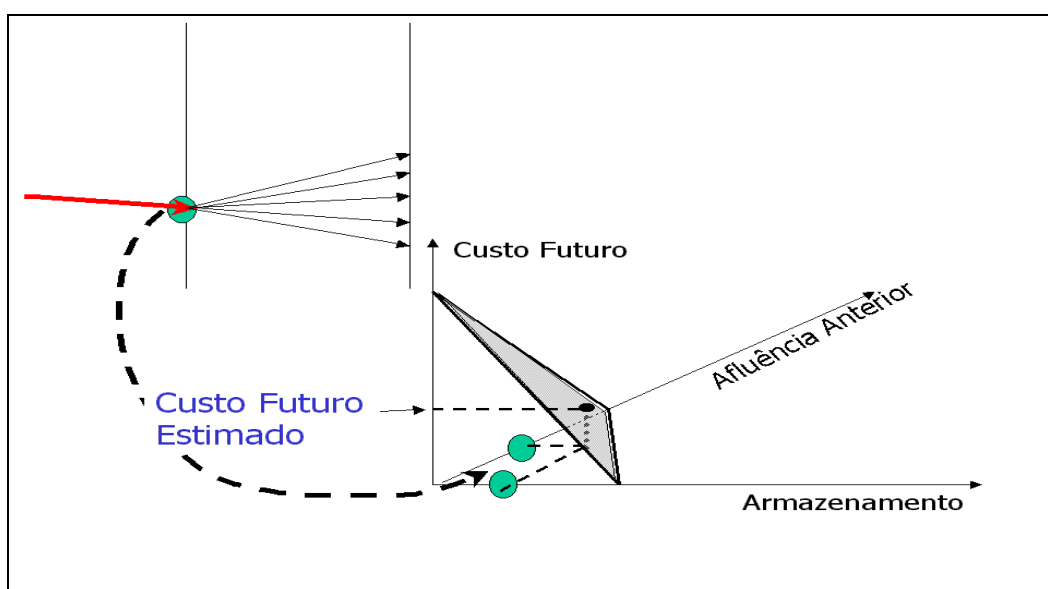


Figura 3.12: Uma Função de Custo Futuro por Etapa no Enfoque Pente
Fonte: Kligerman, 2005

Em resumo, como cada um dos enfoques é utilizado por modelos diferentes, as características de cada enfoque devem se adequar aos modelos que as utilizam. Dessa forma, o DECOMP, que tem um horizonte de curto prazo, pode utilizar o enfoque ÁRVORE e assim trabalhar com quaisquer hipóteses de afluência, mesmo que não sejam oriundas de um modelo auto-regressivo. E o NEWAVE, por trabalhar com um horizonte de médio prazo, é obrigado a utilizar o enfoque PENTE, caso contrário teria uma explosão do número de Estados do enfoque ÁRVORE.

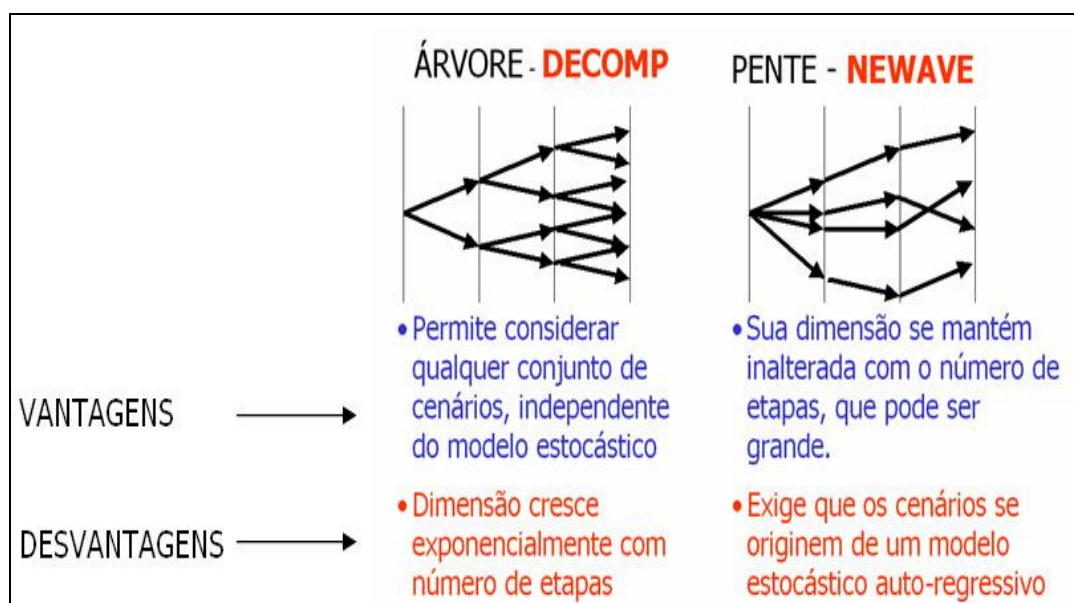


Figura 3.13: Enfoque Árvore X Pente
Fonte: Kligerman, 2005

A fim de se estabelecer o critério de parada para o número de iterações Forward e Backward, acompanha-se a evolução do processo, comparando o custo estimado futuro a partir do início do horizonte com o custo médio simulado ao final do horizonte. Tal critério depende do enfoque utilizado.

Quando o processo converge, no enfoque Árvore, cujo Custo Futuro é calculado considerando as mesmas afluências com as quais é feita a simulação, a tendência é que o Custo Futuro Estimado seja igual ao Custo Médio Simulado ao final do horizonte e no enfoque Pente, cujo Custo Futuro é calculado com aberturas diferentes das afluências utilizadas na simulação, a tendência é do Custo Futuro Estimado permanecer no intervalo de confiança do Custo Médio Simulado.

4

SÉRIES TEMPORAIS APLICADAS AO PLANEJAMENTO ENERGÉTICO DA OPERAÇÃO DO SIN

4.1

INTRODUÇÃO

Grande parte dos fenômenos que ocorrem na natureza possui comportamento aleatório, de modo que é quase impossível afirmar como a natureza vai se comportar no futuro. Entretanto, podem-se obter previsões razoáveis sobre a ocorrência destes fenômenos com o auxílio de ferramentas estatísticas que ajudam a prever comportamentos estocásticos, como o comportamento da vazão natural de um rio.

Modelos estocásticos, também chamados processos estocásticos, são modelos matemáticos capazes de descreverem sistemas que ao longo do tempo são delineados por leis probabilísticas. Estes modelos, em geral modelam observações contínuas medidas em pontos discretos do tempo, por exemplo, séries de dados de negócios, de economia, de engenharia e de ciências naturais que ocorrem sob a forma de séries temporais.

Os processos estocásticos descrevem medidas disponíveis ao longo do tempo, sob a forma de séries temporais e muitas vezes, os dados descritos apresentam uma dependência entre si. O corpo de técnicas disponíveis para a análise de séries de observações dependentes é chamado de Análise de Séries Temporais.

A Análise de Séries Temporais é uma ferramenta que possibilita o entendimento do comportamento de fenômenos naturais. É a partir deste entendimento que cientistas, engenheiros e pesquisadores adquirem embasamento para modelar estes fenômenos e então obter os modelos que representam com veracidade a ocorrência destes na natureza.

A obtenção destes modelos é importante por que:

Eles podem fornecer informações sobre a natureza do sistema que gerou a série temporal;

A partir deles, podem ser obtidas previsões ótimas de valores futuros da série;

Ajudam a entender a relação dinâmica entre variáveis correlacionadas, e com isso entender como a ação de uma variável influencia sobre a outra, por exemplo: temperatura, chuva e vazão afluente.

Podem ser utilizados para obter políticas de controle ótimo, mostrando como uma variável controlada do processo pode ser manipulada para minimizar distúrbios em alguma variável dependente.

4.2

SÉRIES TEMPORAIS

Uma série temporal é um conjunto de dados observados e coletados cronologicamente. Na análise de séries temporais a ordem de ocorrência dos fatos é crucial. Existem séries temporais contínuas que são provenientes de dados coletados em todo instante de tempo ou séries temporais discretas que provêm de dados discretos. As séries temporais, em sua maioria são espaçadas em intervalos constantes de tempo, sejam eles horário, diário, semanal, mensal, anual, etc.

Séries temporais de dados naturais são geralmente discretas, e igualmente espaçadas no tempo. Este tipo de série é o mais utilizado e simplifica muito a matemática que suporta os modelos. Além disso, séries contínuas no tempo podem ser discretizadas em intervalos iguais, como a de temperatura por exemplo. Os dados contínuos podem dar origem a dados horários, calculados pela média horária dos dados contínuos observados.

Existem séries que podem ser perfeitamente modeladas por uma função polinomial, e com isso a observação futura pode ser prevista com exatidão. Este tipo de série segue uma *função determinística*. Quando não se pode prever com exatidão as observações futuras e elas podem ser descritas somente em termos de distribuição de probabilidade, esta série segue um modelo *não determinístico* o que geralmente é um modelo estatístico ou estocástico.

4.3

PROCESSOS ESTOCÁSTICOS

Em se tratando de fenômenos naturais é impossível de se prever deterministicamente o que vai ocorrer no futuro. Meteorologistas nunca afirmam o quanto de chuva vai cair no dia seguinte, eles dizem somente que há probabilidade de chuva. Uma vez ocorrida a chuva,

então a quantidade de chuva é conhecida exatamente. Não obstante, isso continuará acontecendo para as chuvas futuras e a seqüência de toda a precipitação histórica gravada é somente uma realização do ocorrido. Precipitação é um exemplo de um fenômeno estatístico que envolve o tempo e leis probabilísticas. A expressão matemática que descreve a estrutura de probabilidades que foi observada na série temporal é chamada de *Processo Estocástico*. A seqüência de dados históricos coletados é uma realização do processo estocástico que a produziu.

Um processo estocástico muito importante para o planejamento e operação do SIN é o processo que modela as vazões afluentes aos reservatórios do Sistema Interligado Nacional. O ONS possui o histórico das afluições conhecidas e consolidadas no período de 1931 a 2006. Esta seqüência de dados distribuída no tempo é uma série temporal, que representa a série histórica de vazões. A série temporal pode ser modelada por um processo estocástico, que nada mais é do que o conjunto de todas as possíveis séries temporais que podem ser observadas.

A série histórica é a única realização do processo estocástico que está disponível na prática. Portanto, o processo estocástico pode assumir valores aleatórios para cada instante do tempo, o que o transforma em uma variável aleatória. O valor observado em um instante t qualquer da série histórica, nada mais é do que o valor "amostrado" da distribuição de probabilidades associada à variável aleatória do processo estocástico no instante t .

Um processo estocástico é totalmente descrito pelo conjunto de todas as séries temporais que o compõe ou pela distribuição de probabilidades conjunta de todas as variáveis aleatórias envolvidas. Entretanto, não é possível determinar todas as possíveis séries de afluição que o compõem nem as distribuições de probabilidades. Com o objetivo de contornar este problema, ajusta-se um modelo pelo qual acredita-se que a série histórica tenha sido produzida e a partir dele são geradas as *séries sintéticas* que representam as séries temporais que podem ser "amostradas" pelo processo físico que se está observando.

4.4

ESTACIONARIEDADE

Processos estocásticos podem ser classificados como estacionários ou periódicos. Estacionariedade de um processo estocástico pode ser interpretada como uma forma de equilíbrio estatístico. Se ao longo do tempo suas propriedades estocásticas que são a média, o desvio padrão, covariância não sofrerem modificações, ele é dito estacionário. Ou, de uma

forma mais abrangente, significa que a distribuição de probabilidades em um instante t qualquer é válida para qualquer outro instante. A não estacionariedade de um processo estocástico pode ser causada pela intervenção direta do homem, ou da natureza, no processo físico, ou ainda pela presença de ciclos sazonais (características que se repetem dentro de um ano).

Em certas situações, as características estatísticas do processo podem variar em função do tempo. Um exemplo disso é o processo de afluências. Quando analisado anualmente é considerado estacionário, pois possui média única para todo o período e quando analisado numa escala mensal o processo de afluências é considerado não estacionário, já que possui uma média para cada mês.

A modelagem de séries temporais não estacionárias pode ser feita aplicando-se um procedimento de retirar a não estacionariedade da série. Isso pode ser feito subtraindo-se a média e dividindo-se pelo desvio-padrão para retirar a sazonalidade. E em seguida, aplicando-se uma transformação para tornar a série homocedástica.

4.5

PROCESSO ESTOCÁSTICO RÚIDO BRANCO

O processo estocástico ruído branco é o processo cujas variáveis aleatórias componentes são independentes e identicamente distribuídas – “i.i.d.”.

A Figura 4.1 ilustra a autocorrelação deste processo que é dada por:

$$\rho_k = \begin{cases} 1; & k = 0 \\ 0; & k \geq 1 \end{cases}$$

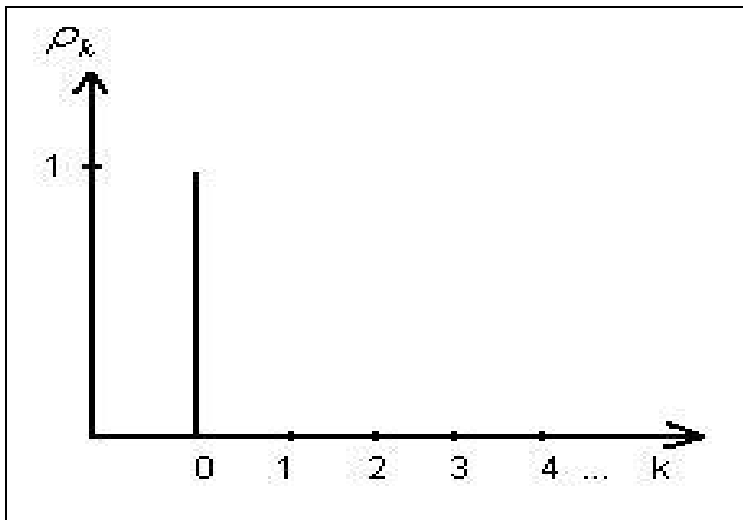


Figura 4.1: Autocorrelação do Ruído Branco

4.6

PROPRIEDADES ESTATÍSTICAS

Seja $[z_1, z_2, \dots, z_N]$ uma série temporal de N valores que foram observados em intervalos iguais de tempo. As propriedades estatísticas a seguir serão definidas para esta série.

4.6.1

MÉDIA E VARIÂNCIA

A média teórica do processo, $\mu = E[Z_t]$, pode ser estimada a partir da amostra realizada pela seguinte equação:

$$\bar{Z} = \frac{1}{N} \sum_{t=1}^N Z_t \quad (5.1)$$

A dispersão de dados do processo espalhados em torno da média é medida pela variância teórica. Esta variância pode ser estimada a partir da série dada pela seguinte equação:

$$\hat{\sigma}_Z^2 = \frac{1}{N} \sum_{t=1}^N (Z_t - \bar{Z})^2 \quad (5.2)$$

4.6.2

AUTOCOVARIÂNCIA E AUTOCORRELAÇÃO

A covariância entre Z_t e o valor Z_{t+k} , separados por k intervalos de tempo é chamada de autocovariância de lag k e é definida pela equação:

$$\gamma_k = \text{cov}[Z_t, Z_{t+k}] = E[(Z_t - \mu)(Z_{t+k} - \mu)] \quad (5.3)$$

A autocovariância lag um mede o grau de dependência linear entre observações contíguas de um processo estocástico. Neste caso a autocovariância lag um indica a dependência linear entre a observação do dado de uma semana com a observação do dado de uma semana imediatamente anterior (se o intervalo de tempo for semanal).

Supondo-se que a estrutura de dependência temporal é estacionária, $\sigma_Z^2 = \gamma_0$ ou seja, é a mesma para o tempo t e para o tempo $t+k$, então, a autocorrelação lag k é dada por:

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad (5.4)$$

4.6.3

PROCESSOS ERGÓTICOS

Um processo estocástico é dito ergótico, se com apenas uma realização do processo é possível caracterizá-lo.

Esta propriedade é importante, pois se deseja caracterizar o processo gerador de uma dada série temporal a partir de uma única realização desta, o que nos leva a pensar que a própria série temporal é a realização de um processo estocástico ergótico.

4.7

MODELOS LINEARES ESTACIONÁRIOS

Nos tópicos seguintes, serão estudados três tipos de modelos lineares estacionários: os modelos Auto-regressivos – AR (p), Médias Móveis – MA (q) e Auto-regressivo – Médias Móveis – ARMA (p, q).

4.7.1

MODELOS AUTO-REGRESSIVOS AR (p)

Modelos Auto-Regressivos descrevem como uma observação depende diretamente de uma ou mais observações feitas em períodos anteriores e do termo de ruído branco. Eles estimam a variável dependente a partir de valores assumidos por ela em tempos anteriores, o que significa dizer que a própria variável é capaz de explicar a sua ocorrência partindo dos dados registrados na série temporal.

Este tipo de modelo é utilizado em vários campos da ciência, economia e até no turismo.

Quando uma observação, Z_t , medida num instante de tempo t depende de valores da série temporal medidos no instante $t-1$ com o ruído branco a_t , este processo é chamado de *Auto-Regressivo de ordem 1* ou AR (1). O Processo **AR (1)** é comumente conhecido como *Processo de Markov* e pode ser escrito sob a forma da seguinte equação:

$$Z_t - \mu = \phi_1 (Z_{t-1} - \mu) + a_t \quad (5.5)$$

Onde,

μ é a média do processo

ϕ_1 é o parâmetro auto-regressivo

a_t é o ruído branco que é independente e identicamente distribuído $(0, \sigma_a^2)$

A seqüência da variável a_t é a parte randômica da série, é o ruído, ou o distúrbio adicionado à série. A premissa mais importante para o ruído branco é que ele é i.i.d. o que infere que os a_t 's são descorrelatados e devem satisfazer a equação:

$$E[a_t a_{t-k}] = \begin{cases} \sigma_a^2, & k = 0 \\ 0, & k \neq 0 \end{cases} \quad (5.6)$$

A equação (5.5) pode ser escrita de outra maneira, utilizando-se o operador de atraso “B” (Backward Shift Operator):

$$BZ_t = Z_{t-1}$$

e onde, k é um inteiro positivo.

$$B^k Z_t = Z_{t-k}$$

Partindo-se da equação (5.5) e do operador B , calcula-se o valor do ruído a_t :

$$Z_t - \mu = \phi_1(BZ_t - \mu) + a_t$$

Ou

$$a_t = Z_t - \mu - \phi_1(BZ_t - \mu)$$

$$a_t = (1 - \phi_1 B)(Z_t - \mu)$$

Onde,

$B\mu = \mu$, desde que a média seja constante para todos os períodos.

$$a_t = \phi(B)(Z_t - \mu) \quad (5.7)$$

Onde $\phi(B) = 1 - \phi_1 B$ é o operador AR de ordem 1.

Partindo-se do caso particular de um AR (1), pode-se chegar ao caso genérico para um modelo AR de ordem p **AR (p)**.

$$Z_t - \mu = \phi_1(Z_{t-1} - \mu) + \phi_2(Z_{t-2} - \mu) + \dots + \phi_p(Z_{t-p} - \mu) + a_t \quad (5.8)$$

De acordo com o caso particular do AR (1), a expressão acima pode ser escrita da seguinte forma:

$$a_t = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(Z_t - \mu)$$

Ou

$$a_t = \phi(B)(Z_t - \mu) \quad (5.9)$$

Onde $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ é o operador AR de ordem p .

Condição de Estacionariedade para o AR (p)

A equação $\phi(B) = 0$ é chamada de *Equação Característica* do processo. É condição suficiente e necessária para que o processo seja estacionário que as raízes desta equação estejam fora do círculo unitário. O círculo unitário é medido em radianos, centrado na origem do plano complexo, onde o eixo do x representa o eixo dos números reais, e o do y representa o eixo dos números imaginários.

A condição de estacionariedade garante que o processo possa ser escrito em termos do ruído branco.

Condição de Invertibilidade para o AR (p)

Todo processo auto-regressivo é invertível, não necessitando de condição alguma para garantir a invertibilidade do processo.

4.7.1.1

Função de Autocorrelação - ACF

A função de autocorrelação mede a dependência linear entre Z_t e Z_{t+k} (ou Z_{t-k} , já que a função é simétrica).

Para estudar a função teórica de autocorrelação de um processo AR (p) estacionário, é necessário multiplicar a equação do modelo, (5.8) por $(Z_{t-k} - \mu)$ para obter:

$$(Z_{t-k} - \mu)(Z_t - \mu) = \phi_1(Z_{t-k} - \mu)(Z_{t-1} - \mu) + \phi_2(Z_{t-k} - \mu)(Z_{t-2} - \mu) + \dots + \phi_p(Z_{t-k} - \mu)(Z_{t-p} - \mu) + (Z_{t-k} - \mu)a_t \quad (5.10)$$

Ao assumir valores esperados na equação (5.10) a equação das diferenças para a função de autocovariância do processo AR (p) é:

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \dots + \phi_p\gamma_{k-p} \quad k > 0 \quad (5.11)$$

A Expressão $E[(Z_{t-k} - \mu)a_t]$ é zero para $k > 0$ por que Z_{t-k} é a única função de distúrbios até o momento, e o ruído a_t não possui correlação com estes distúrbios.

Dividindo-se a equação (5.11) por γ_0 determina-se a expressão da *Função de Autocorrelação Teórica – ACF* do Processo AR (p):

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad k > 0$$

Esta equação pode ser equivalentemente escrita com o auxílio do operador de atraso B ao invés do tempo t . De modo que seria dada pela seguinte forma:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \rho_k = \phi(B) \rho_k = 0 \quad (5.12)$$

Desta forma, a solução geral da equação de diferenças seria:

$$\rho_k = A_1 G_1^k + A_2 G_2^k + \dots + A_p G_p^k \quad (5.13)$$

Onde $G_1^{-1}, G_2^{-1}, \dots, G_p^{-1}$, são raízes distintas da equação característica $\phi(B) = 0$ e os termos A_i 's são constantes.

O gráfico da autocorrelação de um modelo AR (p) é dado por senóides e ou exponenciais que são amortecidas à medida que k cresce, conforme pode ser observado na Figura 4.2.

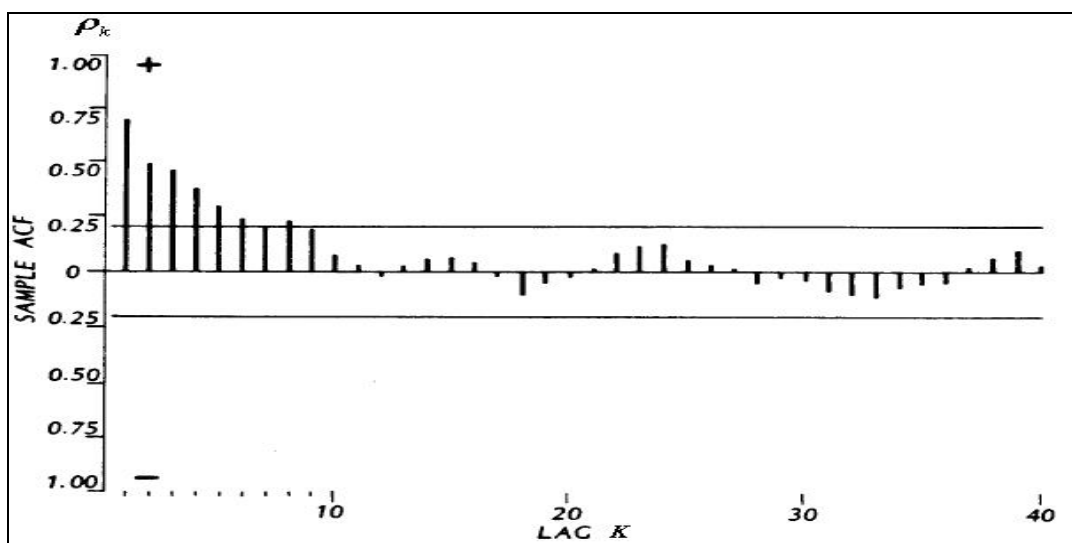


Figura 4.2: Autocorrelação do Modelo AR (p)

Equações de Yule-Walker

Substituindo $k=1, 2, \dots, p$ na equação (5.12) os parâmetros podem ser expressos em termos da ACF teórica. O resultado desse conjunto de equações é chamado de **Equações de Yule-Walker**

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 + \dots + \phi_p \rho_{p-2} \\ &\vdots \quad \vdots \quad \vdots \quad \dots \quad \vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p \end{aligned} \quad (5.14)$$

Escrevendo as equações de Yule-Walker na forma matricial, a solução para os parâmetros em termos das autocorrelações pode ser obtida por:

$$\phi = P_p^{-1} \rho_p \quad (5.15)$$

$$\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \quad \rho_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} \quad P_p = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & 1 \end{bmatrix}$$

Quando $k=0$, a contribuição do termo $E[Z_{t-k} a_t]$ no valor esperado da equação (5.10) é, $E[a_t^2] = \sigma_a^2$, desde que a única parte de Z_t que tem correlação com a_t é o termo mais recente a_t .

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma_a^2 \quad (5.16)$$

Dividindo a equação acima por $\gamma_0 = \sigma_z^2$, a variância é dada pela seguinte expressão:

$$\sigma_z^2 = \frac{\sigma_a^2}{1 - \rho_1 \phi_1 - \rho_2 \phi_2 - \dots - \rho_p \phi_p} \quad (5.17)$$

4.7.1.2

Função de Autocorrelação Parcial- PACF

A Função de Autocorrelação Parcial ϕ_{kk} mede a dependência linear entre Z_t e Z_{t+k} , eliminando os efeitos das variáveis intermediárias $Z_{t+1}, Z_{t+2}, \dots, Z_{t+k-1}$. Devido ao decaimento da função de autocorrelação do modelo AR (p) mostrou-se vantajoso a definição de outra função, que mostrasse nitidamente a ordem do modelo AR. A função de autocorrelação parcial é finita, e mostra um corte brusco após o lag p, o que ajuda a identificar a ordem do modelo.

Ela pode ser obtida a partir das equações de Yule-Walker que podem ser escritas da seguinte forma:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad (5.18)$$

Ou

$$P_k \phi_k = \rho_k$$

Resolvendo-se a equação para $k=1, 2, 3, \dots$, sucessivamente, obtemos $\phi_{11}, \phi_{22}, \phi_{33}, \dots, \phi_{kk}$

O gráfico da função de autocorrelação parcial de um modelo AR (p) é dado por um corte brusco, no lag p, como pode ser observado na Figura 4.3.

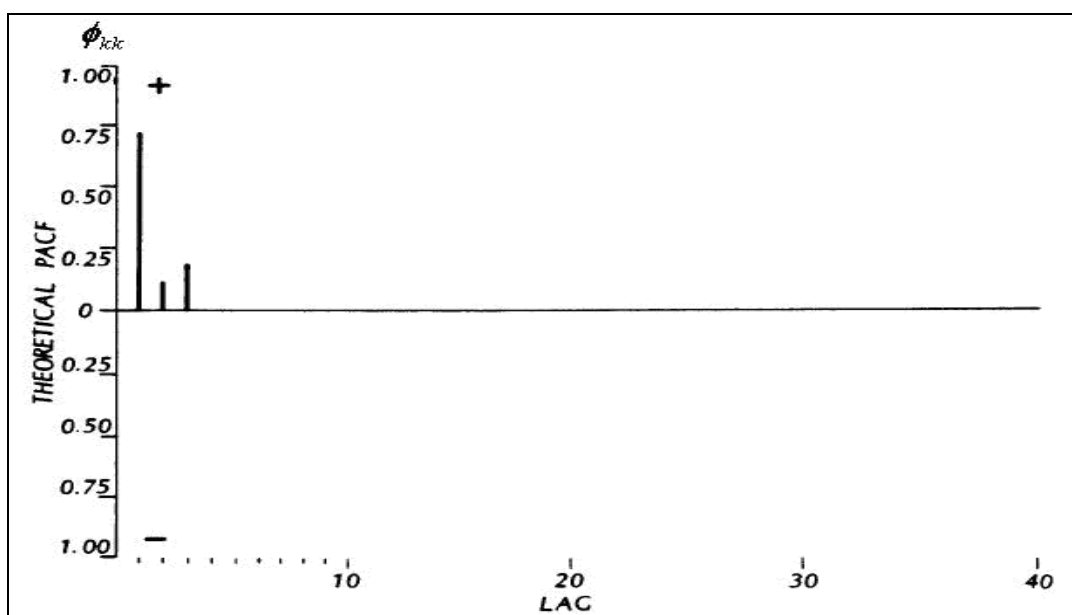


Figura 4.3: Função Teórica da Autocorrelação Parcial do Modelo AR (p)

4.7.2

MODELOS MÉDIAS MÓVEIS MA (q)

Os modelos de médias móveis descrevem quando uma observação depende do último termo do ruído branco assim como de um ou mais termos ruídos branco anteriores.

Quando uma série temporal Z_t depende somente do último ruído branco e mais do ruído corrente, esse processo é chamado de Médias Móveis de ordem 1- **MA (1)**, e pode ser descrito pela seguinte equação:

$$Z_t - \mu = a_t - \theta_1 a_{t-1} \quad (5.19)$$

Onde,

μ é a média do processo

θ_1 é o parâmetro de médias-móveis

a_t é o ruído branco que é independente e identicamente distribuído $(0, \sigma_a^2)$

A equação acima, pode ser reescrita, em função do operador de atraso B :

$$\begin{aligned}
Z_t - \mu &= a_t - \theta_1 B a_t \\
Z_t - \mu &= (1 - \theta_1 B) a_t \\
Z_t - \mu &= \theta(B) a_t
\end{aligned} \tag{5.20}$$

Onde o polinômio $\theta(B) = 1 - \theta_1 B$ é o operador MA (1).

O MA (1) visto acima, pode ser estendido para modelos que possuam q parâmetros de médias móveis. O processo de médias-móveis de ordem q é denotado por **MA (q)** e é dado pela equação:

$$Z_t - \mu = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \tag{5.21}$$

A equação acima, pode ser reescrita, em função do operador de atraso B .

$$\begin{aligned}
Z_t - \mu &= a_t - \theta_1 B a_t - \theta_2 B^2 a_t - \dots - \theta_q B^q a_t \\
Z_t - \mu &= \theta(B) a_t
\end{aligned} \tag{5.22}$$

Onde o polinômio $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ é o operador MA (q).

Condição de Estacionariedade para o MA (q)

Toda série temporal composta por a_t 's é dita estacionária, e como Z_t na equação (5.22) é formado por uma combinação linear de a_t 's, então o processo pode ser escrito em termos do ruído branco o que garante a estacionariedade do processo.

Condição de Invertibilidade para o MA (q)

A equação $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q = 0$ é chamada de *Equação Característica* do processo. É condição suficiente e necessária para que o processo MA (q) seja invertível que as raízes da equação característica estejam fora do círculo unitário.

Quando a condição de invertibilidade é satisfeita para um MA (q), este processo pode ser expresso como um AR puro

4.7.2.1

Função de Autocorrelação – ACF

Utilizando-se a equação (5.3) da função de autocovariância e equação do processo MA (q) (5.21), temos que:

$$\begin{aligned}\gamma_k &= E[(Z_t - \mu)(Z_{t-k} - \mu)] \\ \gamma_k &= E[(a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q})(a_{t-k} - \theta_1 a_{t-k-1} - \theta_2 a_{t-k-2} - \dots - \theta_q a_{t-k-q})] \\ (5.23)\end{aligned}$$

Após a multiplicação e o valor esperado, a função de autocorrelação é dada por:

$$\gamma_k = \begin{cases} (-\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \dots + \theta_{q-k} \theta_q) \sigma_a^2, & k = 1, 2, \dots, q \\ 0, & k > q \end{cases} \quad (5.24)$$

Onde, $\theta_0 = 1$ e $\theta_{-k} = 0$ para $k \geq 1$. Quando $k=0$, na equação (5.23) a variância é:

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_a^2 \quad (5.25)$$

Dividindo-se a autocovariância pela variância, é encontrada a ACF teórica para o processo MA (q).

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \dots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2}, & k = 1, 2, \dots, q \\ 0, & k > q \end{cases} \quad (5.26)$$

A autocorrelação de um modelo MA (q) é finita e apresenta um corte brusco após o lag (q).

4.7.2.2

Função de Autocorrelação Parcial – PACF

Em geral, qualquer MA (q) finito e invertível pode ser expresso como um processo AR infinito. Para processos AR (p) finitos a função teórica de autocorrelação parcial é zero após o lag p. Portanto, para um processo MA (q) ou o equivalente AR (p) infinito, a PACF ϕ_{kk} deve ser amortecida à medida que os lags vão aumentando. Esta característica pode ser observada na Figura 4.4, onde a Função de Autocorrelação Parcial tem a forma de uma senóide amortecida e cujo último lag significativo é o segundo, o que indica o modelo ser um MA (2).

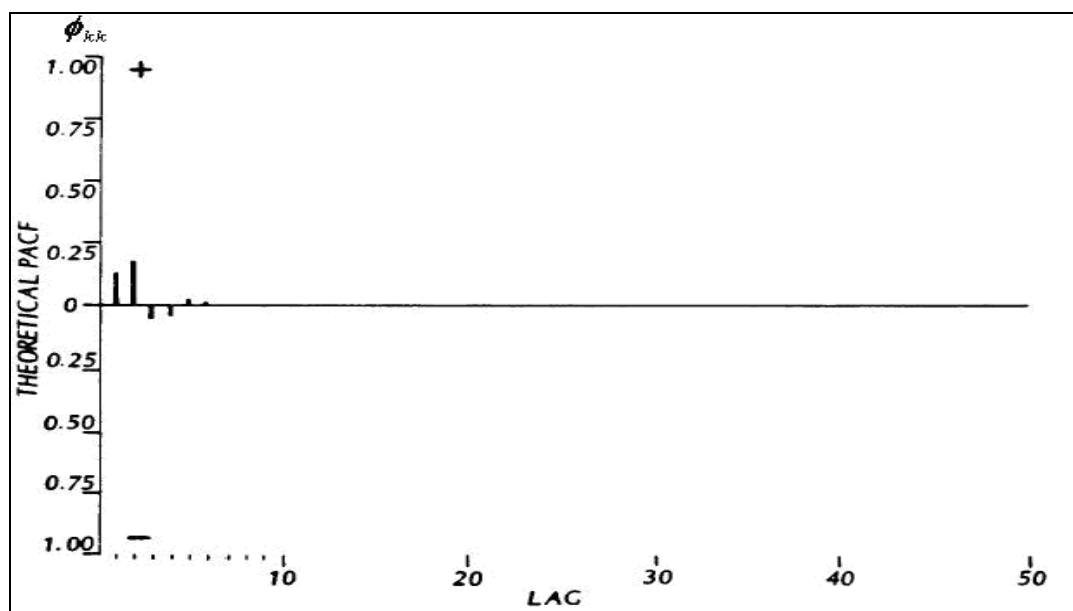


Figura 4.4: Função Teórica da Autocorrelação Parcial do Modelo MA (2)

4.7.3

MODELOS AUTO REGRESSIVOS MÉDIAS MÓVEIS ARMA (p,q)

Um bom modelo deve ser capaz de representar todos os dados de uma série temporal, ou a maior parte deles, com o menor número de parâmetros possível. Partindo-se dessa premissa a família de modelos Auto-Regressivos e de Médias Móveis **ARMA (p,q)** pode representar séries que contenham características tanto de um modelo AR como de um MA, com o menor número de parâmetros possível.

Se um modelo possui um parâmetro AR e um parâmetro MA, então ele é chamado **ARMA (1,1)**, e pode ser descrito pela seguinte equação:

$$(Z_t - \mu) - \phi_1(Z_{t-1} - \mu) = a_t - \theta_1 a_{t-1} \quad (5.27)$$

Utilizando-se o operador B, o modelo ARMA (1,1) pode ser escrito como:

$$(1 - \phi_1 B)(Z_t - \mu) = (1 - \theta_1 B)a_t \quad (5.28)$$

Ou,

$$\phi(B)(Z_t - \mu) = \theta(B)a_t$$

Onde, $\phi(B) = 1 - \phi_1 B$ e $\theta(B) = 1 - \theta_1 B$ são respectivamente os operadores AR e MA.

Em geral, um processo ARMA é formado por p parâmetros da parte auto-regressiva e q parâmetros da parte de médias móveis, formando os processos **ARMA (p,q)**, que são escritos sob a forma:

$$\begin{aligned} (Z_t - \mu) - \phi_1(Z_{t-1} - \mu) - \phi_2(Z_{t-2} - \mu) - \dots - \phi_p(Z_{t-p} - \mu) \\ = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \end{aligned} \quad (5.29)$$

Utilizando-se o operador B, o modelo ARMA (p,q) pode ser escrito como:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(Z_t - \mu) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)a_t$$

Ou, (5.30)

$$\phi(B)(Z_t - \mu) = \theta(B)a_t$$

Onde, $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ é o operador AR de ordem p, e o polinômio $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ é o operador MA de ordem q.

Um processo ARMA (p,q) contém tanto o processo AR puro quanto o MA, de modo que um processo AR (p) pode ser escrito sob a forma ARMA (p,0) assim como um processo MA (q) pode ser escrito sob a forma ARMA (0,q).

Condição de Estacionariedade e Invertibilidade para o ARMA (p,q)

As condições de estacionariedade e invertibilidade discutidas anteriormente para os modelos AR e MA também são válidas para os processos ARMA (p,q). Para um processo ARMA (p,q) ser estacionário as raízes da equação característica $\phi(B) = 0$ devem estar fora do círculo unitário. Similarmente, as raízes do polinômio $\theta(B) = 0$ devem estar fora do círculo unitário, para que o processo seja invertível e possa ser expresso como um AR puro.

4.7.3.1

Função de Autocorrelação – ACF

A função teórica da autocorrelação para um processo ARMA (p,q) é obtida de forma semelhante à ACF do processo AR, multiplicando-se ambos os lados da equação (5.29) por $(Z_{t-k} - \mu)$ e tirando-se o valor esperado.

$$\begin{aligned} \gamma_k - \phi_1 \gamma_{k-1} - \phi_2 \gamma_{k-2} - \dots - \phi_p \gamma_{k-p} = \\ \gamma_{za}(k) - \theta_1 \gamma_{za}(k-1) - \theta_2 \gamma_{za}(k-2) - \dots - \theta_q \gamma_{za}(k-q) \end{aligned} \quad (5.31)$$

Onde $\gamma_k = E[(Z_t - \mu)(Z_{t-k} - \mu)]$ é a função de autocovariância e $\gamma_{za}(k) = E[(Z_{t-k} - \mu)a_t]$ é a covariância entre Z_{t-k} e a_t . Desde que Z_{t-k} seja independente dos ruídos que tenham ocorrido até o tempo $t-k$, temos o seguinte resultado:

$$\begin{aligned} \gamma_{za}(k) = 0, k > 0 \\ \gamma_{za}(k) \neq 0, k \leq 0 \end{aligned} \quad (5.32)$$

Devido aos termos γ_{za} em (5.31) é necessário obter mais uma equação, e esta é obtida multiplicando-se (5.29) por a_{t-k} :

$$\gamma_{za}(-k) - \phi_1 \gamma_{za}(-k+1) - \phi_2 \gamma_{za}(-k+2) - \dots - \phi_p \gamma_{za}(-k+p) = -[\theta_k] \sigma_a^2 \quad (5.33)$$

Onde,

$$[\theta_k] = \begin{cases} \theta_k & , k = 1, 2, \dots, q \\ -1 & , k = 0 \\ 0 & , k < 0 \end{cases}$$

Resolvendo o sistema com as equações (5.31) e (5.33), a autocovariância para o modelo ARMA (p,q) é dada por:

$$\begin{aligned} \gamma_k - \phi_1 \gamma_{k-1} - \phi_2 \gamma_{k-2} - \dots - \phi_p \gamma_{k-p} &= 0 \\ \text{Ou} & \\ \phi(B) \gamma_k &= 0 \end{aligned} \quad (5.34)$$

Dividindo a autocovariância (5.34) por γ_0 , finalmente obtemos a função teórica da autocorrelação ρ_k do modelo ARMA (p,q):

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \rho_k = \phi(B) \rho_k = 0, k > q \quad (5.35)$$

A Figura 4.5 mostra o gráfico de uma função teórica da autocorrelação de um modelo ARMA (p,q).

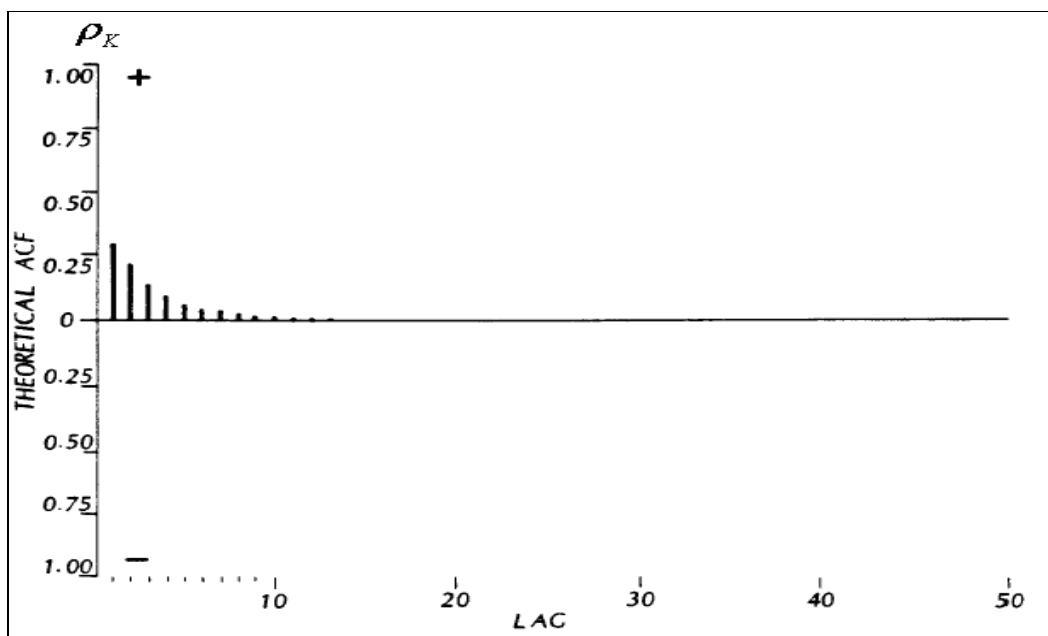


Figura 4.5: Função Teórica da Autocorrelação do Modelo ARMA (p,q)

4.7.3.2 Função de Autocorrelação Parcial – PACF

Como resultado do operador MA em (5.30), o processo ARMA (p,q) pode ser escrito sob a forma de um AR infinito dado por:

$$a_t = \theta(B)^{-1} \phi(B)(Z_t - \mu) \quad (5.36)$$

Onde, $\theta(B)^{-1}$ é uma série infinita em B. Para processos AR (p) finitos a função teórica de autocorrelação parcial é zero após o lag p. Portanto, para um processo MA (q) ou o equivalente AR (p) infinito, a PACF ϕ_{kk} deve, ter formas de senóides e ou exponenciais amortecidas à medida que os lags vão aumentando. A Figura 4.6 mostra o gráfico da autocorrelação parcial de um modelo ARMA (1,1) que é uma exponencial amortecida com o aumento dos lags.

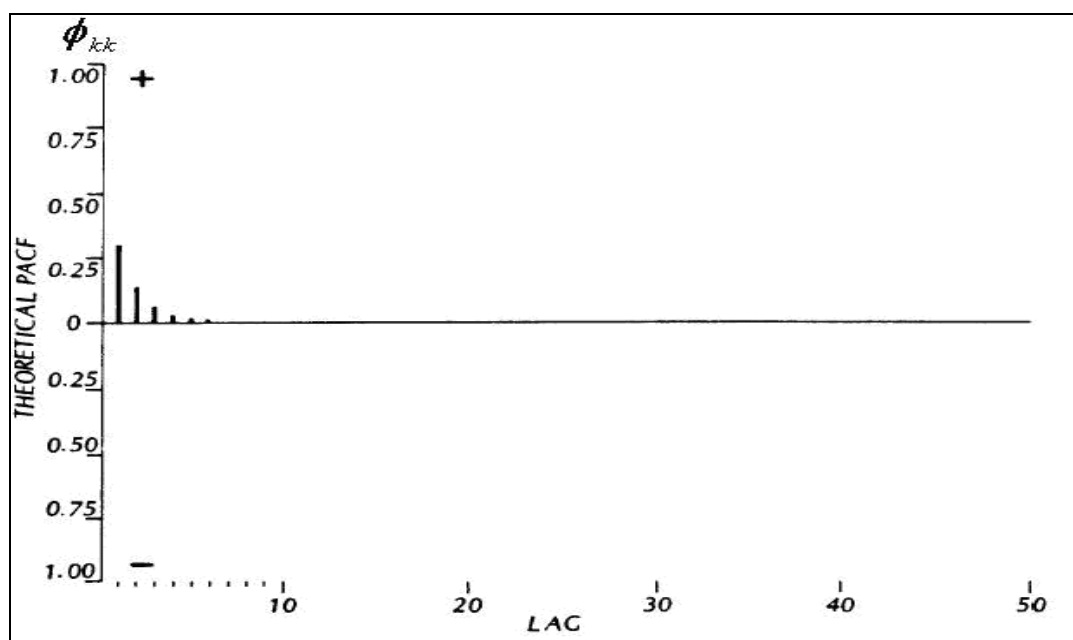


Figura 4.6: Função Teórica da Autocorrelação Parcial de um Modelo ARMA (1,1)

Na tabela seguinte, segue um resumo das principais características dos modelos AR(p), MA (q) e ARMA (p,q).

Tabela 4.1: Resumo das Características dos Modelos AR(p), MA (q) e ARMA (p,q).

	AR (p)	MA (q)	ARMA (p,q)
Modelo	$a_t = \phi(B)(Z_t - \mu)$	$Z_t - \mu = \theta(B)a_t$	$\phi(B)(Z_t - \mu) = \theta(B)a_t$
Condição de Estacionariedade	Raízes de $\phi(B) = 0$ Fora do círculo unitário	Sempre Estacionário	Raízes de $\phi(B) = 0$ Fora do círculo unitário
Condição de Invertibilidade	Sempre Invertível	Raízes de $\theta(B) = 0$ Fora do círculo unitário	Raízes de $\theta(B) = 0$ Fora do círculo unitário
Função de Autocorrelação ρ_k	Infinita – Exponenciais e/ou senóides amortecidas	Finita – corte após o lag “q”	Infinita – Exponenciais e/ou senóides amortecidas após o lag “q-p”
Função de Autocorrelação Parcial ϕ_{kk}	Finita – corte após o lag “p”	Infinita– Exponenciais e/ou senóides amortecidas	Infinita – Exponenciais e/ou senóides amortecidas após o lag “p-q”

4.8

MODELOS PERIÓDICOS

Modelos sazonais hidrológicos, como as vazões naturais de um rio, e outros tipos de séries sazonais apresentam uma estrutura de autocorrelação que depende não só do intervalo de tempo entre as observações, mas também das estações sazonais ao longo do ano. Estes processos, quando analisados em escalas semanal ou mensal, têm como característica o comportamento periódico, refletindo o ciclo das estações do ano. Cada período apresenta um conjunto de suas propriedades probabilísticas, definidas pela média, desvio-padrão e função de autocorrelação.

A média amostral de cada período é dada por:

$$\hat{\mu}_m = \frac{1}{N} \sum_{i=1}^N Z_{(i-1)12+m} \quad \text{para } m = 1, 2, \dots, 12 \text{ meses} \quad (5.37.a)$$

Ou,

$$\hat{\mu}_m = \frac{1}{N} \sum_{i=1}^N Z_{(i-1)52+m} \quad \text{para } m = 1, 2, \dots, 52 \text{ semanas} \quad (5.37.b)$$

O desvio padrão amostral de cada período é dado por:

$$\hat{\sigma}_m = \sqrt{\frac{1}{N} \sum_{i=1}^N Z_{(i-1)12+m} - \hat{\mu}_m} \quad \text{para } m = 1, 2, \dots, 12 \text{ meses} \quad (5.38.a)$$

Ou,

$$\hat{\sigma}_m = \sqrt{\frac{1}{N} \sum_{i=1}^N Z_{(i-1)52+m} - \hat{\mu}_m} \quad \text{para } m = 1, 2, \dots, 52 \text{ semanas} \quad (5.38.b)$$

O valor da auto-covariância amostral do período pode ser obtido da seguinte forma:

$$\hat{\gamma}^m(k) = \frac{1}{N} \sum_{i=1}^N (Z_{(i-1)12+m} - \hat{\mu}_m)(Z_{(i-1)12+m-k} - \hat{\mu}_{m-k}) \quad \text{para } m = 1, 2, \dots, 12 \text{ meses} \quad (5.39.a)$$

Ou,

$$\hat{\gamma}^m(k) = \frac{1}{N} \sum_{i=1}^N (Z_{(i-1)52+m} - \hat{\mu}_m)(Z_{(i-1)52+m-k} - \hat{\mu}_{m-k}) \quad \text{para } m = 1, 2, \dots, 52 \text{ semanas} \quad (5.39.b)$$

O valor da autocorrelação amostral para o mês ou para a semana pode ser obtido da seguinte forma:

$$\hat{\rho}^m(k) = \frac{\hat{\gamma}^m(k)}{\hat{\sigma}_m \hat{\sigma}_{m-k}} \quad \text{para } m \text{ variando de } 1 \text{ a } 12 \text{ meses}, \quad (5.40)$$

ou para m variando de 1 a 52 semanas

Os dois tipos de modelos periódicos mais utilizados na previsão de vazões são os Modelos Periódico Auto-Regressivo – PAR (p) e Auto-Regressivo Média-Móvel Periódico – PARMA (p, q). Quando modelamos uma série sazonal com um modelo PAR (p), um modelo AR é especificado para cada estação do ano. De maneira similar, o modelo PARMA consiste em ter um ARMA separado para cada estação do ano.

4.8.1

MODELOS AUTO-REGRESSIVOS PERIÓDICOS – PAR (p)

O histórico das afluições conhecidas e consolidadas pelo ONS totalizando 75 anos representa a série histórica de vazões do SIN.

O modelo PAR (p) é bem adequado à análise deste tipo de série, em virtude dos seus parâmetros apresentarem comportamento periódico, tal como as séries semanais e mensais de vazões afluentes. Os modelos periódicos consideram para o ajuste de parâmetros somente determinadas semanas de um mês, trimestre ou semestre, agrupando determinados períodos do ano, enquanto que modelos estacionários consideram a série como um todo com todos os meses do ano, no qual meses e trimestres diferentes são tratados do mesmo modo.

Na prática, observa-se que em meses iniciais de período úmido as afluições dependem de 1 ou no máximo 2 meses anteriores. Já em meses iniciais de período seco as afluições dependem de vários meses passados do último período úmido.

O número de termos auto-regressivos do modelo PAR (p) indica a ordem do modelo, que em geral é um vetor, $p = (p_1, p_2, \dots, p_{12})$, onde cada elemento fornece a ordem de cada período.

A essência do modelo PAR (p) é definir um modelo AR para cada estação m do ano. A descrição que se segue, é a formulação matemática do modelo PAR (p), que para séries mensais, $p = (p_1, p_2, \dots, p_{12})$.

$$\left(\frac{(Z_t - \mu_m)}{\sigma_m} \right) = \phi_1^m \left(\frac{(Z_{t-1} - \mu_{m-1})}{\sigma_{m-1}} \right) + \dots + \phi_p^m \left(\frac{(Z_{t-p} - \mu_{m-p})}{\sigma_{m-p}} \right) + a_t \quad (5.41.a)$$

ou

$$\Phi^m(B) \left(\frac{(Z_t - \mu_m)}{\sigma_m} \right) = a_t \quad (5.41.b)$$

Onde:

Z_t é uma série sazonal de período s

s é o número de períodos ($s=12$ para séries mensais)

N é o número de anos

t é o índice do tempo, $t=1, 2, \dots, sN$, função do ano T ($T=1, 2, \dots, N$) e do período m ($m=1, 2, \dots, s$)

μ_m é a média sazonal de período s

σ_m é o desvio-padrão sazonal de período s

ϕ_i^m é o coeficiente AR para a estação m

$\Phi^m(B)$ é o operador auto-regressivo de ordem p_m

$$\Phi^m(B) = (1 - \phi_1^m B - \phi_2^m B^2 - \dots - \phi_p^m B^p) \quad (5.42)$$

B^i aplicado a Z_t resulta em Z_{t-i} ($B^i Z_t = Z_{t-i}$)

p é a ordem de cada operador auto-regressivo

a_t é a série de ruídos independentes com média zero e variância $\sigma_a^{2(m)}$

Condição de Estacionariedade para o PAR (p)

Para que o modelo associado a cada estação m seja estacionário, é necessário que as raízes da equação característica sazonal $\Phi^{(m)}(B) = 0$ estejam fora do círculo unitário. Entretanto, esta não é condição suficiente para que o modelo PAR (p) seja estacionário.

Conforme mostrado por Troutman (1979), o modelo PAR (p) é uma caso particular do PARMA (p,q) para que o modelo PAR (p) seja estacionário, a condição necessária e suficiente é dada pela equação:

$$\sum_{i=0}^{\infty} (\psi_i^{(m)})^2 < \infty \quad m = 1, 2, \dots, s$$

Onde $\psi_i^{(m)}$ é o coeficiente de pesos do modelo PARMA e é dado pela equação:

$$\phi^{(m)}(B)\psi_k^{(m)} = -\theta_k^{(m)}$$

4.8.1.1

Função de Autocorrelação – ACF

Seja $\rho_k^{(m)}$ a correlação entre Z_t e Z_{t-k} , de tal forma que t corresponda ao período m :

$$\rho^{(m)}(k) = E \left[\left(\frac{Z_t - \mu_m}{\sigma_m} \right) \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] \quad (5.43)$$

Multiplicando-se ambos os lados da equação (5.41a) por $\left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right)$ e tomando os valores esperados de cada termo, obtemos para cada período:

$$\begin{aligned} E \left[\left(\frac{Z_t - \mu_m}{\sigma_m} \right) \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] &= \phi_1^m E \left[\left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] + \dots + \\ &+ \phi_p^m E \left[\left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] + E \left[a_t \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] \end{aligned} \quad (5.44)$$

Fazendo $k=1$, a expressão (5.44) resulta em:

$$\rho^m(1) = \phi_1^m + \phi_2^m \rho^{m-1}(1) + \dots + \phi_p^m \rho^{m-1}(p_m - 1) \quad (5.45)$$

Uma vez conhecidos os parâmetros do modelo PAR (p) as funções $\rho_k^{(m)}$ são dadas pela solução da equação (5.44) e podem ser expressas por uma combinação de decaimentos exponenciais e/ou ondas senoidais, o que faz $\rho_k^{(m)}$ tender a zero, à medida que k cresce.

Fixando-se m (como um período qualquer) e variando-se k de 1 a p na equação (5.44) obtemos para cada período o conjunto de equações a seguir, que são conhecidas como equações de Yule-Walker.

$$\begin{bmatrix} 1 & \rho^{m-1}(1) & \rho^{m-1}(2) & \dots & \rho^{m-1}(p-1) \\ \rho^{m-1}(1) & 1 & \rho^{m-2}(1) & \dots & \rho^{m-2}(p-2) \\ \rho^{m-1}(2) & \rho^{m-2}(1) & 1 & \dots & \rho^{m-3}(p-3) \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{m-1}(p-1) & \rho^{m-2}(p-2) & \rho^{m-3}(p-3) & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_1^m \\ \phi_2^m \\ \phi_3^m \\ \dots \\ \phi_p^m \end{bmatrix} = \begin{bmatrix} \rho^m(1) \\ \rho^m(2) \\ \rho^m(3) \\ \dots \\ \rho^m(p) \end{bmatrix} \quad (5.46)$$

Para $k=0$, a expressão (5.44) fica:

$$1 = \phi_1^m \rho^m(1) + \phi_2^m \rho^m(2) + \dots + \phi_p^m \rho^m(p) + E \left[a_t \left(\frac{Z_t - \mu_m}{\sigma_m} \right) \right] \quad (5.47)$$

Multiplicando-se a equação (5.41.a) por a_t e tomando o valor esperado, obtemos:

$$E \left[a_t \left(\frac{Z_t - \mu_m}{\sigma_m} \right) \right] = \sigma_a^{2(m)} \quad (5.48)$$

Substituindo este resultado na equação (5.47), obtemos a seguinte expressão válida para qualquer período m :

$$\sigma_a^{2(m)} = 1 - \phi_1^m \rho^m(1) - \phi_2^m \rho^m(2) - \dots - \phi_p^m \rho^m(p) \quad (5.49)$$

4.8.1.2

Função de Autocorrelação Parcial – PACF

Já que a função de autocorrelação do modelo PAR para um período m vai decaindo com o tempo e não trunca após um lag específico, não determinando a ordem exata do modelo, pode ser útil identificar outras funções que sejam truncadas após determinado lag, e assim ajudem a identificar o modelo. Para conseguir determinar esta função definiu-se a função teórica de autocorrelação parcial – PACF de um modelo PAR (p) semelhante a definição utilizada para o modelo AR (p), de modo que essa função seja bruscamente cortada após o último lag significativo do modelo, e com isso seja determinada a ordem do modelo para o período m .

4.8.2

MODELOS AUTO-REGRESSIVOS MÉDIAS-MÓVEIS PERIÓDICOS – PARMA (p,1)

O modelo PARMA foi criado para definir um modelo ARMA para cada período do ano. A descrição que se segue, é a formulação matemática do modelo PARMA ($p, 1$) que é dada por:

$$\left(\frac{(Z_t - \mu_m)}{\sigma_m} \right) = \phi_1^m \left(\frac{(Z_{t-1} - \mu_{m-1})}{\sigma_{m-1}} \right) + \dots + \phi_p^m \left(\frac{(Z_{t-p} - \mu_{m-p})}{\sigma_{m-p}} \right) - \theta_1^m a_{t-1} + a_t \quad (5.50)$$

Onde:

Z_t é uma série sazonal de período s

s é o número de períodos ($s=12$ para séries mensais)

N é o número de anos

t é o índice do tempo, $t=1, 2, \dots, sN$, função do ano T ($T=1, 2, \dots, N$) e do período m ($m=1, 2, \dots, s$)

μ_m é a média sazonal de período s

σ_m é o desvio-padrão sazonal de período s

$\Phi^m(B)$ é o operador auto-regressivo de ordem p_m

$$\Phi^m(B) = (1 - \phi_1^m B - \phi_2^m B^2 - \dots - \phi_p^m B^{pm}) \quad (5.51)$$

B^i aplicado a Z_t resulta em Z_{t-i} ($B^i Z_t = Z_{t-i}$)

p é a ordem de cada operador auto-regressivo

θ_1^m é coeficiente média-móvel de ordem 1

a_t é a série de ruídos independentes com média zero e variância $\sigma_{a_t}^2$

Multiplicando-se ambos os lados da equação (5.50) por a_t e tomando os valores esperados de cada termo, obtemos para cada período:

$$\begin{aligned} E \left[\left(\frac{(Z_t - \mu_m)}{\sigma_m} \right) a_t \right] &= \phi_1^m E \left[\left(\frac{(Z_{t-1} - \mu_{m-1})}{\sigma_{m-1}} \right) a_t \right] + \dots + \phi_p^m E \left[\left(\frac{(Z_{t-p} - \mu_{m-p})}{\sigma_{m-p}} \right) a_t \right] - \\ &\quad - \theta_1^m E[a_{t-1} a_t] + E[a_t a_t] \end{aligned}$$

$$E \left[\left(\frac{(Z_t - \mu_m)}{\sigma_m} \right) a_t \right] = \sigma_{a_t}^2 \quad (5.52)$$

Multiplicando-se a equação (5.50) por a_{t-1} e tomando os valores esperados, obtemos:

$$\begin{aligned} \mathbb{E} \left[\left(\frac{Z_t - \mu_m}{\sigma_m} \right) a_{t-1} \right] &= \phi_1^m \mathbb{E} \left[\left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) a_{t-1} \right] + \dots + \phi_p^m \mathbb{E} \left[\left(\frac{Z_{t-p} - \mu_{m-p}}{\sigma_{m-p}} \right) a_{t-1} \right] - \\ &\quad - \theta_1^m \mathbb{E}[a_{t-1} a_{t-1}] + \mathbb{E}[a_t a_{t-1}] \end{aligned}$$

$$\mathbb{E} \left[\left(\frac{Z_t - \mu_m}{\sigma_m} \right) a_{t-1} \right] = \phi_1^m \sigma_{a_{t-1}}^2 - \theta_1^m \sigma_{a_{t-1}}^2$$

Ou

$$\mathbb{E} \left[\left(\frac{Z_t - \mu_m}{\sigma_m} \right) a_{t-1} \right] = (\phi_1^m - \theta_1^m) \sigma_{a_{t-1}}^2 \quad (5.53)$$

Multiplicando a equação (5.50) por $\frac{Z_t - \mu_m}{\sigma_m}$ e tomando os valores esperados,

obtemos:

$$\begin{aligned} \mathbb{E} \left[\left(\frac{Z_t - \mu_m}{\sigma_m} \right) \left(\frac{Z_t - \mu_m}{\sigma_m} \right) \right] &= \phi_1^m \mathbb{E} \left[\left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) \left(\frac{Z_t - \mu_m}{\sigma_m} \right) \right] + \dots + \\ &+ \phi_p^m \mathbb{E} \left[\left(\frac{Z_{t-p} - \mu_{m-p}}{\sigma_{m-p}} \right) \left(\frac{Z_t - \mu_m}{\sigma_m} \right) \right] - \theta_1^m \mathbb{E} \left[a_{t-1} \left(\frac{Z_t - \mu_m}{\sigma_m} \right) \right] + \\ &+ \mathbb{E} \left[a_t \left(\frac{Z_t - \mu_m}{\sigma_m} \right) \right] \end{aligned}$$

$$1 = \phi_1^m \rho^m(1) + \dots + \phi_p^m \rho^m(p) + \sigma_{a_t}^2 - \theta_1^m (\phi_1^m - \theta_1^m) \sigma_{a_{t-1}}^2$$

Ou

$$\sigma_{a_t}^2 = 1 - \phi_1^m \rho^m(1) - \dots - \phi_p^m \rho^m(p) + \theta_1^m (\phi_1^m - \theta_1^m) \sigma_{a_{t-1}}^2 \quad (5.54)$$

4.8.2.1

Função de Autocorrelação – ACF

Para obtermos a função de autocorrelação lag um basta multiplicar a equação (5.50) por Z_{t-1} e tomar o valor esperado, obtendo:

$$\begin{aligned} \mathbb{E} \left[\left(\frac{Z_t - \mu_m}{\sigma_m} \right) \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) \right] &= \phi_1^m \mathbb{E} \left[\left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) \right] + \dots + \\ &+ \phi_p^m \mathbb{E} \left[\left(\frac{Z_{t-p} - \mu_{m-p}}{\sigma_{m-p}} \right) \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) \right] - \theta_1^m \mathbb{E} \left[a_{t-1} \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) \right] + \\ &+ \mathbb{E} \left[a_t \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) \right] \end{aligned}$$

$$\rho^m(1) = \phi_1^m + \phi_2^m \rho^{m-1}(1) + \dots + \phi_p^m \rho^{m-1}(p-1) - \theta_1^m \sigma_{a_{t-1}}^2 \quad (5.55)$$

Ou, tirando o valor de θ_1^m , temos:

$$\theta_1^m = \frac{\phi_1^m - \rho^m(1) + \phi_2^m \rho^{m-1}(1) + \dots + \phi_p^m \rho^{m-1}(p-1)}{\sigma_{a_{t-1}}^2} \quad (5.56)$$

Repetindo-se o procedimento acima para Z_{t-2} , vamos obter o coeficiente auto regressivo para o lag 2:

$$\begin{aligned} \mathbb{E} \left[\left(\frac{Z_t - \mu_m}{\sigma_m} \right) \left(\frac{Z_{t-2} - \mu_{m-2}}{\sigma_{m-2}} \right) \right] &= \phi_1^m \mathbb{E} \left[\left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) \left(\frac{Z_{t-2} - \mu_{m-2}}{\sigma_{m-2}} \right) \right] + \dots + \\ &+ \phi_p^m \mathbb{E} \left[\left(\frac{Z_{t-p} - \mu_{m-p}}{\sigma_{m-p}} \right) \left(\frac{Z_{t-2} - \mu_{m-2}}{\sigma_{m-2}} \right) \right] - \theta_1^m \mathbb{E} \left[a_{t-1} \left(\frac{Z_{t-2} - \mu_{m-2}}{\sigma_{m-2}} \right) \right] + \\ &+ \mathbb{E} \left[a_t \left(\frac{Z_{t-2} - \mu_{m-2}}{\sigma_{m-2}} \right) \right] \end{aligned}$$

$$\rho^m(2) = \phi_1^m \rho^{m-1}(1) + \phi_2^m + \dots + \phi_p^m \rho^{m-2}(p-2) \quad (5.57)$$

Em fim, repetindo-se este procedimento até Z_{t-p-1} obteremos o coeficiente auto regressivo para o lag $p+1$:

$$\begin{aligned} & \mathbb{E} \left[\left(\frac{(Z_t - \mu_m)}{\sigma_m} \right) \left(\frac{(Z_{t-p-1} - \mu_{m-p-1})}{\sigma_{m-p-1}} \right) \right] = \phi_1^m \mathbb{E} \left[\left(\frac{(Z_{t-1} - \mu_{m-1})}{\sigma_{m-1}} \right) \left(\frac{(Z_{t-p-1} - \mu_{m-p-1})}{\sigma_{m-p-1}} \right) \right] + \dots + \\ & + \phi_p^m \mathbb{E} \left[\left(\frac{(Z_{t-p} - \mu_{m-p})}{\sigma_{m-p}} \right) \left(\frac{(Z_{t-p-1} - \mu_{m-p-1})}{\sigma_{m-p-1}} \right) \right] - \theta_1^m \mathbb{E} \left[a_{t-1} \left(\frac{(Z_{t-p-1} - \mu_{m-p-1})}{\sigma_{m-p-1}} \right) \right] + \\ & + \mathbb{E} \left[a_t \left(\frac{(Z_{t-p-1} - \mu_{m-p-1})}{\sigma_{m-p-1}} \right) \right] \\ & \rho^m (p+1) = \phi_1^m \rho^{m-1} (p) + \phi_2^m \rho^{m-2} (p-1) + \dots + \phi_p^m \rho^{m-p} (1) \end{aligned} \quad (5.58)$$

Arrumando-se as equações (5.57) e (5.58) em forma matricial, obtemos o seguinte sistema de equações:

$$\begin{bmatrix} \rho^{m-1}(1) & 1 & \dots & \rho^{m-2}(p-2) \\ \rho^{m-1}(2) & \rho^{m-2}(1) & \dots & \rho^{m-3}(p-3) \\ \vdots & \vdots & \dots & \vdots \\ \rho^{m-1}(p) & \rho^{m-2}(p-1) & \dots & \rho^{m-p}(1) \end{bmatrix} \begin{bmatrix} \phi_1^m \\ \phi_2^m \\ \vdots \\ \phi_p^m \end{bmatrix} = \begin{bmatrix} \rho^m(2) \\ \rho^m(3) \\ \vdots \\ \rho^m(p+1) \end{bmatrix} \quad (5.59)$$

Através da solução do sistema de equações acima, são obtidos os parâmetros auto regressivos de cada período.

4.9

ERRO MÉDIO QUADRÁTICO DE PREVISÃO

Seja $\hat{X}_t(k)$ a previsão para X_{t+k} em t , ou seja, com o conhecimento de X_t, X_{t-1}, \dots

Seja $e_t(k)$ o erro desta previsão definido por:

$$e_t(k) = X_{t+k} - \hat{X}_t(k) \quad (5.60)$$

O erro médio quadrático de previsão é dado por:

$$E[(e_t(k))^2] = VAR[e_t(k)] + E^2[e_t(k)] \quad (5.61)$$

Como no instante t , $\hat{X}_t(k)$ já é conhecido, o valor esperado e a variância de $e_t(k)$ dependem apenas de X_{t+k} . Logo, podemos escrever:

$$E[(e_t(k))^2] = VAR[X_{t+k} | x_t, x_{t-1}, \dots] + \left(E[X_{t+k} | x_t, x_{t-1}, \dots] - \hat{X}_t(k) \right)^2 \quad (5.62)$$

Portanto para minimizar o erro médio quadrático de previsão deve-se fazer:

$$\hat{X}_t(k) = E[X_{t+k} | x_t, x_{t-1}, \dots] \quad (5.63)$$

Assim, a previsão do erro quadrático médio mínimo para X_{t+k} é seu valor esperado condicionado às informações disponíveis no instante em que se faz a previsão. Este resultado é geral para qualquer variável aleatória, como descrito em Maceira (1989).

5

MODELOS UTILIZADOS NO PLANEJAMENTO E PROGRAMAÇÃO DA OPERAÇÃO DO SIN

5.1

MODELOS HIDROLÓGICOS – PREVISÃO DE VAZÕES E GERAÇÃO DE CENÁRIOS NO ONS

A previsão de vazões semanais e a geração de cenários hidrológicos são insumos de extrema importância para a operação do Sistema Interligado Nacional. Com base nestas informações, são realizados o Programa Mensal da Operação – PMO e suas revisões.

A principal ferramenta utilizada na previsão é o modelo estocástico univariado, desenvolvido pelo CEPEL, de nome PREVIVAZ (Maceira et al., 1999). Este modelo é utilizado para prever as vazões semanais de quase todas as bacias do SIN, com exceção da bacia do Rio Iguaçu, onde é utilizado um modelo complementar ao PREVIVAZ, o Modelo de Previsão de Classes de Vazão – MPCV (Cataldi et al., 2007), que considera informações de precipitação observada e prevista (Rocha et al., 2007).

As previsões diárias de vazão natural são feitas em sua grande parte pelos diferentes Agentes de Geração e complementadas pelo ONS com auxílio dos modelos PREVIVAZH, desenvolvido pelo CEPEL (2002), e CPINS que faz o Cálculo e Previsão de Vazões Incrementais e Naturais a Sobradinho (Paiva e Acioli, 2007) desenvolvido pela CHESF/ONS. O CPINS é utilizado somente para o trecho da bacia do rio São Francisco incremental a UHE Sobradinho, é um modelo de propagação baseado na rotina de propagação do modelo SSARR (Streamflow Synthesis and Reservoir Regulation) (Rocha, et al., 2007).

Após a fase de previsão diária, o ONS consolida as previsões de vazões médias diárias enviadas pelos diferentes Agentes de Geração e ou processadas pelo próprio ONS,

para o fechamento da vazão natural ⁷ média da semana em curso, utilizada como insumo básico para a obtenção das previsões de vazões semanais para as semanas seguintes.

Uma vez consolidados os dados diários, o ONS dá início ao processamento dos modelos de previsão de vazões semanais (PREVIVAZ e o MPCV) para obter as previsões de vazões semanais até o fechamento do mês em curso para todos os *postos base*⁸. É então, nesta etapa, que as previsões de vazões são calculadas para os demais aproveitamentos hidroelétricos do SIN por regressão, a partir das previsões feitas nos postos-base.

É através destes modelos que o ONS incorpora ao Planejamento da Operação a estocasticidade das chuvas e das vazões, que repercutem diretamente na operação do SIN, pois é com base nas previsões fornecidas que os modelos de otimização eletroenergética são processados.

O Programa Mensal de Operação utiliza o modelo DECOMP desenvolvido pelo CEPEL (2004) que, partindo da previsão de vazões semanais para o mês em curso e dos cenários de aflúncias para o segundo mês, calcula o Custo Marginal de Operação visando à minimização do custo imediato e do custo futuro. Os resultados dessa modelagem determinam a política de geração para cada aproveitamento hidroelétrico e para cada usina térmica para o uso ótimo dos recursos existentes.

A previsão de vazões é de suma importância no Planejamento da Operação do SIN, visto que serve como insumo básico para o cálculo do custo marginal e para a definição dos montantes de geração de todos os aproveitamentos do SIN, objetivando a otimização dos recursos hidrotérmicos e de transmissão existentes. O Custo Marginal de Operação serve de base para a definição dos preços da energia no curto prazo e dos montantes de geração das diferentes usinas, influenciando diretamente no faturamento das empresas de geração.

⁷ Para se obter as vazões naturais aos locais de aproveitamentos é necessário desfazer o efeito de ações antrópicas decorrentes da operação dos reservatórios, assim como incorporar as parcelas de vazões de uso consuntivo e a evaporação líquida (diferença entre a evaporação de lago e a evapotranspiração).

⁸ As previsões de vazões naturais no ONS são elaboradas de forma direta para os locais denominados de *postos base*, ou de forma indireta, a partir de regressões lineares obtidas a partir das previsões realizadas para um posto base próximo. Desta forma, os aproveitamentos que terão suas vazões previstas a partir de um modelo são denominados *postos base*, e aqueles cuja previsão é obtida a partir das equações de regressão são denominados *postos não base*.

5.1.1

MODELO PREVIVAZH

No setor elétrico brasileiro, convencionou-se que a semana operativa tem seu início no sábado e seu término na sexta-feira. Atualmente, o ONS realiza a previsão de vazões para o PMO nas terças-feiras, e as revisões nas quartas-feiras. Desse modo, na data da realização da previsão de vazões do PMO e suas revisões a semana operativa em curso está sempre incompleta.

Como o PREVIVAZ é um sistema de modelos de base semanal, é necessário que todas as semanas do histórico estejam completas, incluindo a semana operativa em curso. Para completar essa última semana, o ONS conta então com as previsões diárias realizadas pelos agentes, ou no caso do não envio dessas previsões, esse Operador executa os modelos de previsões de vazões diárias que são o PREVIVAZH e o CPINS.

O modelo de Previsão de Afluências Diárias – PREVIVAZH, desenvolvido pelo CEPEL (2002), tem como objetivo fornecer previsões diárias de afluências a aproveitamentos hidroelétricos do Sistema Hidroelétrico Brasileiro. Estas previsões são obtidas a partir da desagregação das previsões de afluências semanais fornecidas pelo PREVIVAZ, o que garante a preservação da estrutura de dependência temporal das previsões semanais usadas no Planejamento de Curto Prazo e, incorpora a variabilidade característica do processo natural a nível diário, a partir do tratamento dado às vazões pelo modelo de geração de vazões diárias.

5.1.2

MODELO PREVIVAZ

O planejamento da operação de sistemas hidrotérmicos é uma tarefa complexa, visto que existem várias características do sistema, tais como o acoplamento temporal entre as decisões e a estocasticidade em relação às afluências futuras, que dificultam a tomada de decisão.

A existência de reservatórios de regularização plurianual faz com que qualquer decisão tomada no presente sobre o deplecionamento do reservatório influencie a capacidade de geração do sistema também no futuro. Além disso, devido ao regime aleatório das afluências hidrológicas futuras não é possível garantir a plena capacidade futura de geração hidroelétrica do SIN.

Para contornar este problema, e melhorar o desempenho da operação do sistema aumentando benefícios e reduzindo riscos e custos, o ONS faz uso de previsões hidrológicas e geração de cenários sintéticos de afluências naturais. Os modelos hidrológicos tem sido de grande valia na operação do sistema brasileiro, e cada vez surgem mais técnicas de previsão, como redes neurais e mineração de dados que tendem a melhorar cada vez mais os resultados das previsões, tornando assim o planejamento da operação mais eficiente.

A ferramenta matemática atualmente em uso pelo Operador Nacional do Sistema Elétrico para modelagem estocástica das afluências semanais é o sistema desenvolvido pelo CEPEL, denominado *PREVIVAZ* (Maceira et al.,1999 e CEPEL, 2004). O *PREVIVAZ* é um sistema de modelos estocásticos de previsão de afluências semanais para até seis semanas à frente, que abrange uma combinação de **94 modelagens** resultantes da combinação dos modelos de estrutura de correlação periódica ou estacionária, de métodos de estimação de parâmetros e de diferentes transformações.

Além do *PREVIVAZ*, existem também o *PREVIVAZM* (CEPEL, 2003) e *PREVIVAZH* (CEPEL, 2002) que são modelos de previsão das afluências hidrológicas aos aproveitamentos hidroelétricos com discretização mensal (até 12 meses) e diária (até 14 dias), respectivamente.

A Figura 5.1 ilustra as possíveis estruturas de correlação do *PREVIVAZ*. Somadas a estas possibilidades, existem os métodos de estimação e as transformações que, combinados, resultam nas 94 opções de modelagens do *PREVIVAZ*. Os Métodos de Estimação de Parâmetros que podem ser utilizados são:

- ⊗ Momentos
- ⊗ Regressão em relação à origem das previsões
- ⊗ Regressão simples

As Transformações utilizadas são:

- ⊗ Box & Cox
- ⊗ Box & Cox com ou sem transformação
- ⊗ Logarítmica
- ⊗ Logarítmica com ou sem transformação
- ⊗ Sem transformação

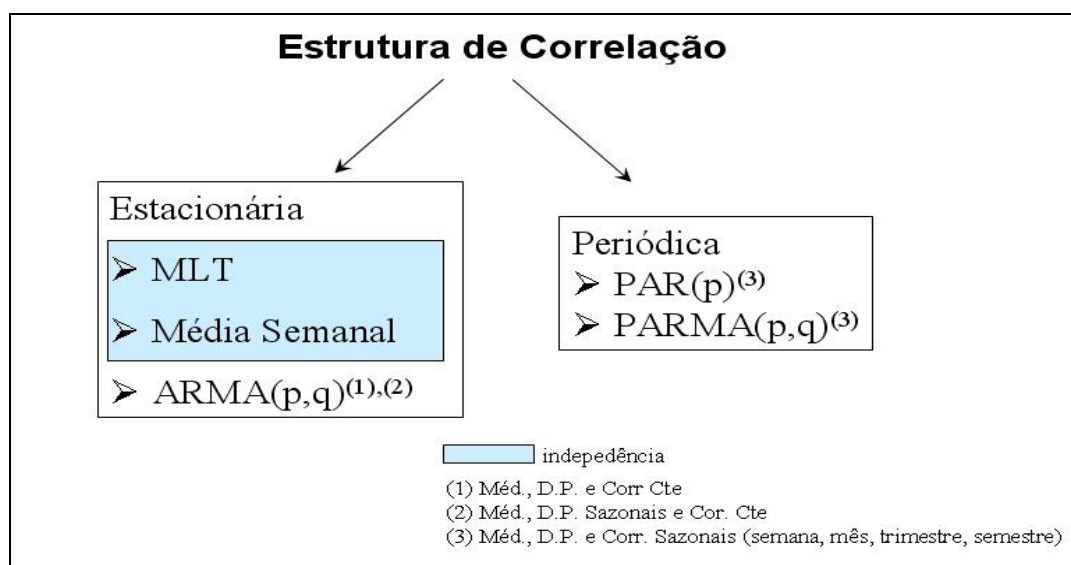


Figura 5.1: Possíveis Estruturas de Correlação do PREVIVAZ
 Fonte: ONS

As previsões semanais das afluências feitas pelo PREVIVAZ são utilizadas como subsídio à Programação do primeiro mês do Programa Mensal da Operação – PMO e suas revisões. Deste modo, ele é executado ao final de cada mês, para elaborar o PMO do mês seguinte e, durante o mês em curso é rodado semanalmente para a realização das revisões do PMO.

Este modelo analisa a série histórica de afluências semanais de cada aproveitamento, e seleciona para cada semana um modelo estocástico dentre as diversas alternativas de modelagem estocástica de vazões semanais.

Para modelar a série deve-se atentar para alguns cuidados, como:

- Representar as principais características da série histórica de vazões.
- Sazonalidade e Estrutura de Correlação Periódica ou Estacionária.
- A modelagem final é uma combinação do modelo estocástico escolhido e do método de estimação adotado e depende das características da série que se deseja preservar.
- Retirar a sazonalidade da série, subtraindo-se a média e dividindo-se pelo desvio padrão.
- Aplicar alguma transformação, se necessário, para tornar a série homocedástica (variância constante).

Os modelos estocásticos do PREVIVAZ variam desde os clássicos auto-regressivos AR (p), ou auto-regressivos e de médias móveis ARMA (p,q) propostos na literatura (Box e Jenkins, 1970) até modelos periódicos mais sofisticados como o modelo PARMA (p,q) (Hipel e Mcleod, 1994), para os quais são admitidas investigações de até quarta ordem.

Os Modelos utilizados são os 47 listados abaixo, que ao se optar entre usar ou não usar alguma transformação, o número de modelos dobra, resultando nas 94 modelagens que seguem.

- ⊗ MLT - (média de todas as semanas)
- ⊗ SAZONAL - (média da semana)
- ⊗ AR(p) - com p de 1 a 4
- ⊗ ARMA(p,1) - com p de 1 a 3
- ⊗ PAR(p) - com p de 1 a 4 - momentos+semanal
- ⊗ PAR(p) - com p de 1 a 4 - momentos+mensal
- ⊗ PAR(p) - com p de 1 a 4 - momentos+trimestral
- ⊗ PAR(p) - com p de 1 a 4 - momentos+semestral
- ⊗ PAR(p) - com p de 1 a 4 - regressão em relação à origem
- ⊗ PARMA(p,1) - com p de 1 a 3 - momentos+semanal
- ⊗ PARMA(p,1) - com p de 1 a 3 - momentos+mensal
- ⊗ PARMA(p,1) - com p de 1 a 3 - momentos+trimestral
- ⊗ PARMA(p,1) - com p de 1 a 3 - momentos+semestral
- ⊗ PARMA(p,1) - com p de 1 a 3 - regressão em relação à origem
- ⊗ PARMA(p,1) - com p de 1 a 3 - regressão simples

A fim de identificar um modelo estatístico que se ajuste ao comportamento estocástico das afluições do SIN o modelo realiza uma análise do histórico das afluições conhecidas e consolidadas para o período com dados de elevada confiabilidade. Desta análise, extraem-se alguns índices como:

- ⊗ Média;
- ⊗ Desvio Padrão
- ⊗ Correlação Temporal (mede a dependência do evento com dados anteriores);
- ⊗ Correlação Espacial (indica o quanto um evento em um local depende do que está ocorrendo em outro local).

OBS: (o PREVIVAZ não usa a correlação espacial)

A escolha entre as 94 modelagens é feita objetivando o uso mais eficiente e robusto da informação contida nos registros recentes de afluições semanais ao aproveitamento analisado. Isto é feito, tendo como base a média dos desvios padrão dos erros de previsão das alternativas de modelagens calculados para a primeira metade da série histórica, tendo-se estimado os parâmetros com a segunda metade e, para a segunda metade da série histórica, tendo-se estimado os parâmetros com a primeira metade. O modelo escolhido será aquele que apresentar a **menor média de desvio padrão do erro de previsão**. Os parâmetros do modelo, são então obtidos utilizando-se toda a série.

Para realizar a previsão das vazões semanais afluentes aos aproveitamentos do SIN, o PREVIVAZ executa o seguinte algoritmo básico para a série de 52 semanas (1 ano):

Passo 1: Realiza-se a transformação da série, se indicada pelo usuário;

Passo 2 : Divide-se a série transformada em duas metades, estimam-se os parâmetros para a primeira metade e verifica-se o erro médio quadrático para a outra metade.

Passo 3 : Repete-se o passo 2, fazendo-se a estimação dos parâmetros com a segunda metade da série e a verificação do erro para a primeira metade, conforme indicado na Figura a seguir.

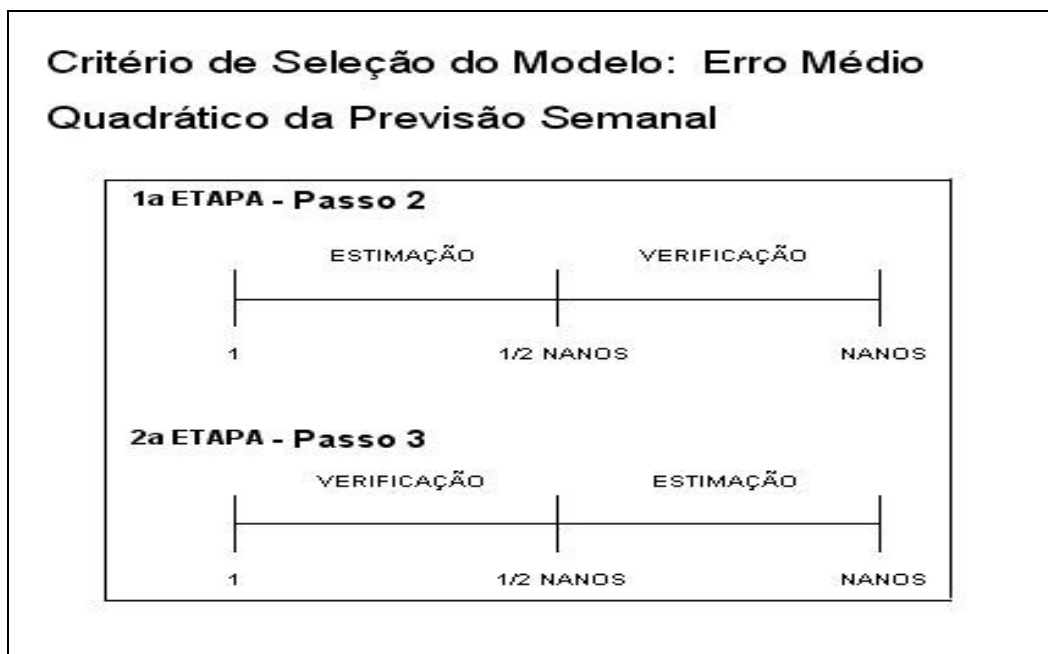


Figura 5.2: Critério de Seleção do Sistema de Modelos PREVIVAZ
Fonte: ONS

O Erro Médio Quadrático de previsão é dado pela seguinte fórmula:

$$\sqrt{\frac{\sum_{i=1}^N (X_{prev}^i - X_{obs}^i)^2}{N}}$$

Onde,

X_{prev}^i - Vazão prevista no instante i

X_{obs}^i - Vazão observada no instante i

N – Número total de semanas da metade do histórico

Passo 4 : Seleciona-se como melhor modelo para cada semana aquele que obtiver menor Erro Total, que é dado pela média dos dois erros médios quadráticos obtidos nos passos 2 e 3.

Passo 5 : A partir do modelo selecionado estimam-se os parâmetros para a série completa, e com este, calcula-se a previsão da semana seguinte.

Passo 6 : Retira-se a transformação dos resultados calculados para obter as vazões desejadas.

Uma vez escolhido o modelo adequado, o PREVIVAZ dá a previsão para cada aproveitamento do SIN com discretização semanal, para o horizonte dentro do primeiro mês (mês do PMO). Partindo da média destas previsões, o modelo gerador de cenários, GEVAZP, (Maceira e Mercio, 1997) fornece cenários de vazões para o segundo mês à frente que servirão para o cálculo do Custo Futuro no curto prazo.

O encadeamento entre os modelos de previsão de vazões e o modelo de geração de cenários hidrológicos pode ser visualizado através da Figura 5.3.

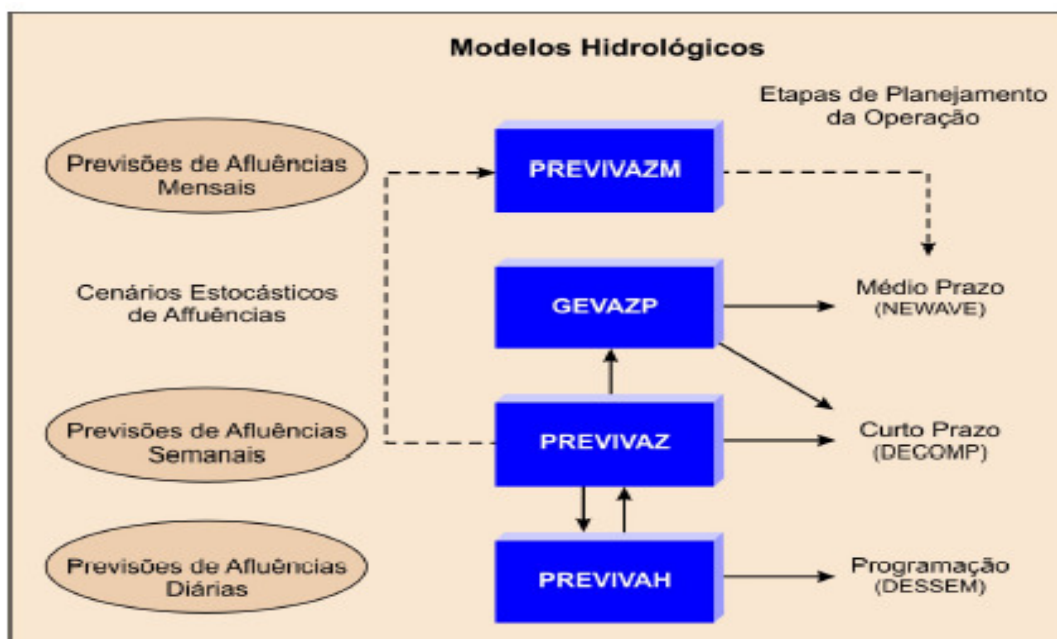


Figura 5.3: Cadeia de Modelos Hidrológicos
Fonte: CEPEL

5.1.3

MODELO PREVIVAZM

O modelo PREVIVAZM desenvolvido pelo CEPEL (2003) tem como objetivo fornecer previsões mensais de afluências a aproveitamentos hidroelétricos do Sistema Hidroelétrico Brasileiro para um horizonte de até 12 meses, sendo uma ferramenta para estudos especiais de verificação de condições de atendimento da demanda energética no horizonte anual.

O PREVIVAZM foi desenvolvido com base no modelo PREVIVAZ e segue a mesma abordagem, adaptando-a ao intervalo mensal e ao horizonte de 12 meses, mantendo-se o procedimento de estimação e escolha do melhor algoritmo de previsão. A Tabela 3 apresenta os algoritmos de previsão implementados no modelo PREVIVAZM. Estes algoritmos podem também ser aplicados às séries históricas de vazões mensais com transformação Box-Cox ou Logarítmica e nas séries sem transformação.

Tabela 5.1: Modelos Implementados no PREVIVAZM

Modelo de Previsão	Característica	Método de Estimação
CONSTANTE	previsão pela média anual	momentos
SAZONAL*	previsão pela média do mês	momentos
AR(p) ($1 \leq p \leq 4$)*	estrutura de correlação estacionária	momentos
ARMA(p,1) ($1 \leq p \leq 3$)*	estrutura de correlação estacionária	momentos
PAR(p)-G1 ($1 \leq p \leq 4$) *	estrutura de correlação sazonal mensal	momentos
PAR(p)-G2 ($1 \leq p \leq 4$) *	estrutura de correlação sazonal trimestral	momentos
PAR(p)-G3 ($1 \leq p \leq 4$) *	estrutura de correlação sazonal semestral	momentos
PARMA(p,1)-G1 ($1 \leq p \leq 4$) *	estrutura de correlação sazonal mensal	momentos
PARMA(p,1)-G2 ($1 \leq p \leq 4$) *	estrutura de correlação sazonal trimestral	momentos
PARMA(p,1)-G3 ($1 \leq p \leq 4$) *	estrutura de correlação sazonal semestral	momentos
PARMA(p,1)-R ($1 \leq p \leq 3$) *	estrutura de correlação sazonal	regressão simples

* Estes modelos são aplicados a séries com transformação Box & Cox, Logarítmica ou nas séries sem transformação.

** G1 => é um modelo de correlação mensal

G2 => é um modelo de correlação trimestral

G3 => é um modelo de correlação semestral

5.1.4

MODELO GEVAZP

A hidrologia é considerada parte determinística e parte estocástica. Para cada semana do primeiro mês do planejamento as aflúências são consideradas conhecidas, utilizando-se as previsões obtidas pelos modelos PREVIVAZ (Maceira et al., 2007) e CPINS (Paiva e Acioli, 2007). Para horizontes de tempo maiores, a incerteza nas aflúências naturais aumenta e desse modo uma única previsão não é suficiente para determinar o processo estocástico que a originou.

Para contornar este problema, ajusta-se um modelo pelo qual acredita-se que a série histórica tenha sido produzida e a partir dele são geradas as *séries sintéticas*. E é neste âmbito que se insere o modelo gerador de cenários GEVAZP (CEPEL, 2002).

O GEVAZP gera cenários para dois escopos, o de curto prazo e o de médio prazo. Partindo das previsões do modelo PREVIVAZ para o primeiro mês do planejamento, o modelo GEVAZP fornece **cenários de vazões** para o segundo mês à frente que são utilizados pelo **DECOMP** (CEPEL, 2004) para calcular o Custo Futuro no curto prazo. Partindo do valor da afluência verificada no mês anterior ao primeiro mês do planejamento, o modelo GEVAZP fornece **cenários de energia afluente** para até cinco anos à frente que são utilizados pelo **NEWAVE** (CEPEL, 2006) para cálculo do Custo Futuro.

Desta forma, são gerados cenários de vazões para o curto prazo pelo GEVAZP Vazões e cenários de energia para o médio prazo pelo GEVAZP Energia.

Os cenários hidrológicos podem ser representados através de uma árvore de afluências, com probabilidades de ocorrência associadas a cada ramo. A Figura 5.4 mostra uma estrutura de cenários para um horizonte de estudo de quatro meses, onde o primeiro mês está dividido em cinco semanas com afluências supostas conhecidas.

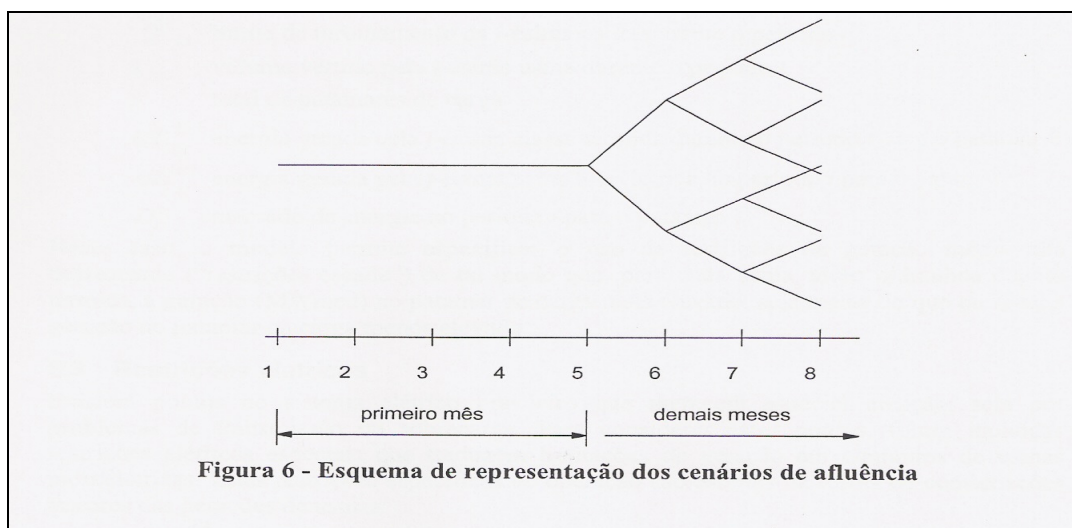


Figura 5.4: Cadeia de Modelos Hidrológicos
Fonte: CEPEL – Manual do DECOMP

O GEVAZP Vazões constrói os cenários de afluências para o segundo mês do planejamento, partindo da média dos valores semanais previstos pelo PREVIVAZ. É através destes cenários, que a estocasticidade das vazões é inserida no planejamento do curto prazo por meio do modelo DECOMP. Desse modo, o DECOMP está apto a prosseguir com a otimização do planejamento energético, considerando a estocasticidade das vazões, porém com a indicação da Função de Custo Futuro.

O GEVAZP Energia constrói os cenários mensais de energia afluyente até cinco anos à frente, que serão utilizados para calcular a Função de Custo Futuro do médio prazo.

Para calcular a Função de Custo Futuro, é necessário um conhecimento prévio do modelo estocástico que representa as afluições. Conhecendo-se este modelo, pode-se sortear um grande número de hipóteses de afluições e acompanhar a evolução do sistema durante os próximos cinco anos para cada trajetória sorteada, correspondente a cada hipótese de afluição, e calcular o custo futuro médio de todas as situações propostas.

O modelo estocástico, utilizado no GEVAZP, capaz de se ajustar ao comportamento das afluições é o modelo PAR (p). Este é o modelo proposto para ser utilizado no Modelo Estratégico de Geração Hidrotérmica a Subsistemas Equivalentes Interligados – NEWAVE (CEPEL, 2006) e no Modelo de Determinação da Coordenação da Operação a Médio Prazo – DECOMP (CEPEL, 2004).

A Figura 5.5 ilustra a seqüência do processo de previsão para o PMO. No primeiro mês (mês da previsão) as previsões são feitas semanalmente pelo PREVIVAZ e a partir da média destes valores o GEVAZP gera os cenários de vazões afluentes para o segundo mês. Estes cenários são utilizados pelo DECOMP para construir a Função de Custo Futuro do curto prazo e com isso, estabelecer a política e a estratégia de operação.

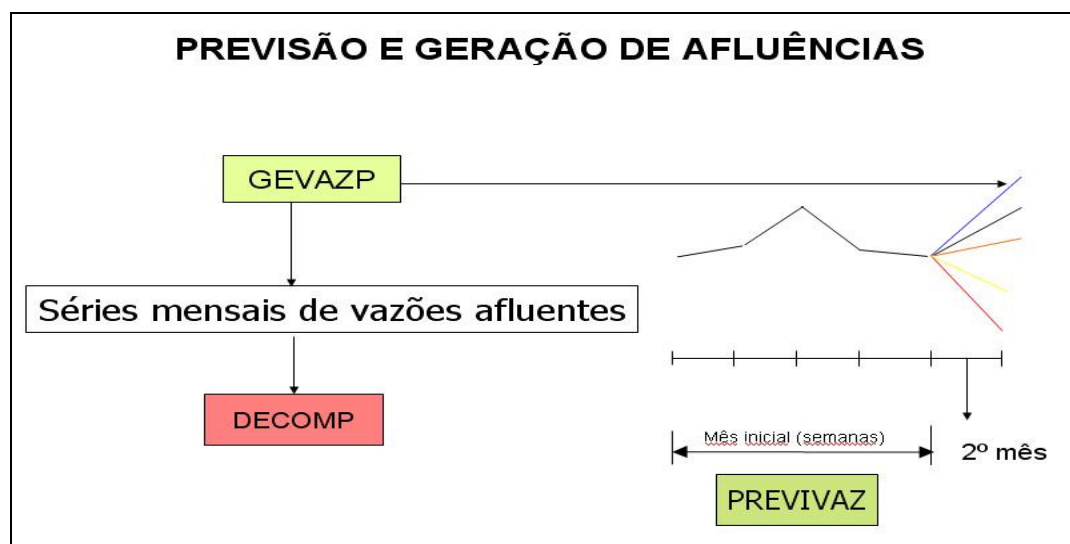


Figura 5.5: Interação do GEVAZP com o DECOMP

A Figura 5.6 ilustra a continuidade do processo anterior, porém no âmbito do médio prazo. O GEVAZP gera, para o NEWAVE, cenários mensais de energia afluyente para um horizonte de cinco anos à frente partindo do valor da afluição verificada no mês anterior. É

com base nestes cenários de afluência que o NEWAVE calcula a estratégia de operação para o médio prazo, e traça a Função de Custo Futuro.

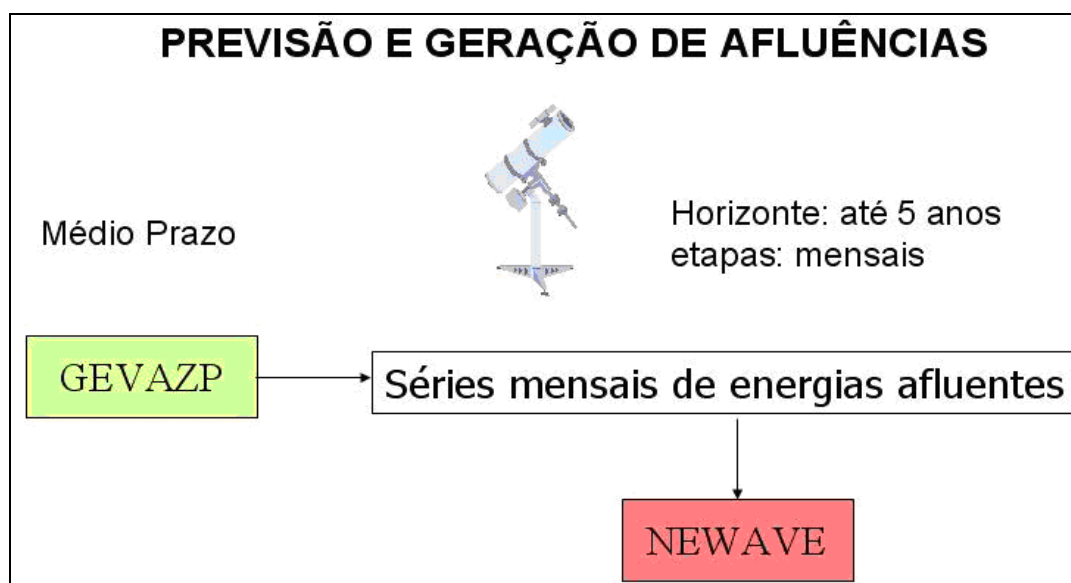


Figura 5.6: Interação do GEVAZP com o NEWAVE

5.2

MODELOS DE OTIMIZAÇÃO ENERGÉTICA

Os objetivos de economia de operação e confiabilidade de atendimento são claramente antagônicos: a máxima utilização da energia hidroelétrica disponível a cada etapa é a política mais econômica, pois minimiza os custos de combustível. Entretanto, esta política é a menos confiável, pois resulta em maiores riscos de déficits futuros. Por sua vez, a máxima confiabilidade de fornecimento é obtida conservando o nível dos reservatórios o mais elevado possível. Contudo, isso significa utilizar mais geração térmica e, portanto, aumentar os custos de operação. O equilíbrio entre os custos de operação e a confiabilidade é obtido através do custo de déficit, que representa o impacto econômico associado à interrupção do fornecimento.

A cada interrupção, existe um custo de déficit associado que varia de acordo com a quantidade de carga não suprida. A ANEEL é responsável pela determinação do custo de déficit, o qual é indispensável para a determinação da política de operação mais adequada para o sistema. Se o custo do déficit é muito baixo, resulta em uma utilização excessiva dos reservatórios e, portanto, em maiores riscos de racionamento no futuro. Se o custo do déficit é muito alto, resulta em uma utilização excessiva dos recursos térmicos do sistema e,

portanto, em custos de operação elevados e maiores probabilidades de ocorrência de vertimentos.

Considerando o exposto adotou-se como solução para o Sistema Interligado Nacional uma cadeia de modelos computacionais com a função objetivo de minimizar o custo total da operação e que levasse em conta o custo do não atendimento, ou o custo de déficit, onde:

$$\text{CUSTO TOTAL} = \text{CUSTO PRESENTE} + \text{CUSTO FUTURO}$$

Em função da dimensão do SIN, o processo de otimização de seus recursos energéticos constitui-se em um problema de grande porte. Desta forma, foi necessário subdividi-lo em etapas temporais associadas a modelos da cadeia de otimização energética que foram desenvolvidos para dar suporte ao processo decisório.

Esta cadeia de modelos tem seu horizonte temporal de análise dividido, variando de 5 anos até 1 semana, conforme pode ser visto na Figura 5.7. É responsável pela definição das estratégias de operação hidrotérmica em horizonte de médio prazo (de 1 mês a 5 anos) e pelas políticas e diretrizes para a operação (horizonte de 1 mês a 1 dia).

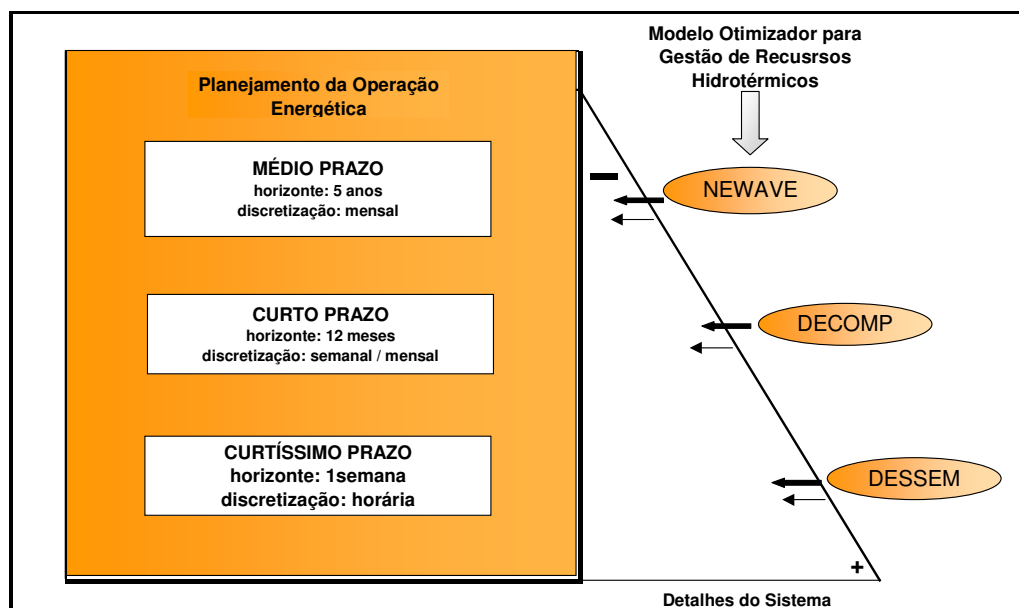


Figura 5.7: Cadeia de Modelos de Otimização
Fonte: ONS

A cadeia de modelos de otimização energética é composta pelos modelos NEWAVE (CEPEL, 2006), DECOMP (CEPEL, 2004) e DESSEM (Maceira et al., 2003), estando o

último em fase de implementação. Estes modelos encontram-se encadeados pela Função de Custo Futuro, de modo que é através desta função que as políticas de operação de curto prazo (definidas pelo modelo DECOMP e, futuramente, também pelo modelo DESSEM) são acopladas às estratégias de operação de médio prazo definidas pelo modelo NEWAVE.

Em cada etapa do planejamento da operação eletroenergética é utilizado um modelo apropriado à fase em questão, mas sempre focado no acoplamento temporal das decisões, visto que a decisão tomada no presente repercutirá no futuro. Os modelos de otimização da cadeia possuem diferentes horizontes de planejamento, discretização do tempo, e graus de detalhamento em suas representações, em particular das afluências hidrológicas futuras, conforme pode ser visto na Tabela 5.2.

Tabela 5.2: Resumo das Características dos Modelos Utilizados no Planejamento Energético do SIN.

Etapa	Médio Prazo	Curto Prazo	Programação da Operação
Modelo Hidrológico	GEVAZP	PREVIVAZ/ GEVAZP	PREVIVAZH
Modelo de Otimização	NEWAVE	DECOMP	DESSEM
Horizonte de planejamento	Até 10 anos	Até 1 ano	Até 14 dias
Discretização temporal	Mensal	Semanal / Mensal	Horária / patamares
Afluências hidrológicas	Estocástico	Determinístico/ Estocástico	Determinístico

Pode-se afirmar que a função objetivo da cadeia de modelos utilizada no planejamento energético pelo ONS, consiste em definir as políticas e diretrizes para a operação para cada período do planejamento, através da minimização do custo total de operação sujeito ao custo de déficit, as restrições operativas, ambientais e de uso múltiplo da água, dentre outras, mantendo-se a confiabilidade da operação elétrica.

5.2.1

MODELO NEWAVE

O Modelo de Despacho Hidrotérmico a Subsistemas Equivalentes – NEWAVE é a ferramenta responsável por elaborar os estudos de planejamento da etapa de médio prazo. Neste modelo o parque hidrotérmico é representado de forma agregada, o que significa que todos os reservatórios de um subsistema são agregados, ficando reduzidos a apenas 4 reservatórios equivalentes, um para cada região do SIN. Depois de agregar os reservatórios é determinada a política de operação mais econômica para os subsistemas equivalentes, levando-se em conta as incertezas das afluições futuras, os patamares de demanda e a indisponibilidade dos equipamentos, CEPEL (2006). E por fim, ele simula a operação do sistema ao longo do período de planejamento para diversos cenários de seqüências hidrológicas, fornecidos pelo GEVAZP.

Para a definição das estratégias futuras da operação, o Modelo NEWAVE percorre diversos cenários hidrológicos para o horizonte de 5 anos e verifica o estado do SIN para cada um desses cenários. Em seguida, compõe a **Função de Custo Futuro**, que consiste no estabelecimento de valores da água para cada subsistema associado aos estados de armazenamento, para cada um dos cenários hidrológicos percorridos pelo modelo.

O Modelo de Despacho Hidrotérmico a Subsistemas Equivalentes – NEWAVE é um modelo de planejamento da operação cujo objetivo é a determinação das estratégias de operação hidrotérmica a médio prazo, com representação agregada do parque hidroelétrico e cálculo da política ótima baseado em Programação Dinâmica Dual Estocástica (Pereira e Pinto, 1984). Suas principais características são:

- Múltiplos subsistemas interligados;
- Configuração estática ou dinâmica;
- Modelo equivalente com produtibilidade variável;
- Energias afluentes modeladas por um processo auto-regressivo periódico de ordem p – PAR (p).

A principal aplicação do NEWAVE na cadeia de procedimentos do PMO é a obtenção da Função de Custo Futuro do Sistema, permitindo assim o acoplamento entre os estudos de médio e curto prazo.

O NEWAVE efetua a análise simultânea da operação integrada dos subsistemas, tomando decisões de geração térmica e intercâmbios de forma a minimizar o custo total de

operação. Para a obtenção desta estratégia, são considerados cenários hidrológicos compostos por séries sintéticas de energias afluentes, com base no histórico de vazões. São utilizadas previsões determinísticas da carga em 3 patamares típicos (carga leve, média e pesada), do cronograma de expansão, dos limites de intercâmbio e dos custos variáveis de geração térmica. Os déficits são representados por uma função linear por partes, possibilitando considerar custos variáveis com sua profundidade.

5.2.2

MODELO DECOMP

O modelo DECOMP é utilizado no Programa Mensal da Operação do Sistema Interligado Nacional – PMO, que tem como principal objetivo estabelecer as metas e diretrizes energéticas de curto prazo da operação coordenada do SIN, assegurando a otimização dos recursos de geração disponíveis.

O DECOMP traça as metas de geração para cada usina, de forma individualizada. Entretanto, por receber o valor de custo obtido pelo NEWAVE, que agrega todas as usinas de um subsistema, o modelo de curto prazo acaba calculando um custo marginal para cada um dos subsistemas. Este custo de operação é composto pelos gastos com combustíveis nas unidades térmicas e eventuais penalizações pelo não atendimento da demanda.

O Custo Marginal de Operação – CMO define sempre quanto custa ao sistema atender o próximo MW de demanda em cada subsistema. Se esse aumento de demanda for atendido com utilização de usinas térmicas, o custo corresponde ao custo de geração da última térmica despachada, conforme mostra a Figura 5.8. Na Figura, as térmicas A e B não foram suficientes para atender a toda a carga o que ocasionou o despacho da unidade térmica C, desta forma, o Custo Marginal de Operação do sistema passou a ser o custo de operação da última unidade térmica despachada. Caso o aumento de demanda seja atendido com geração hidroelétrica o custo corresponde ao Valor da Água desse sistema e se for atendido com aumento de envio de energia de outro subsistema, esse custo será o custo do subsistema fornecedor.

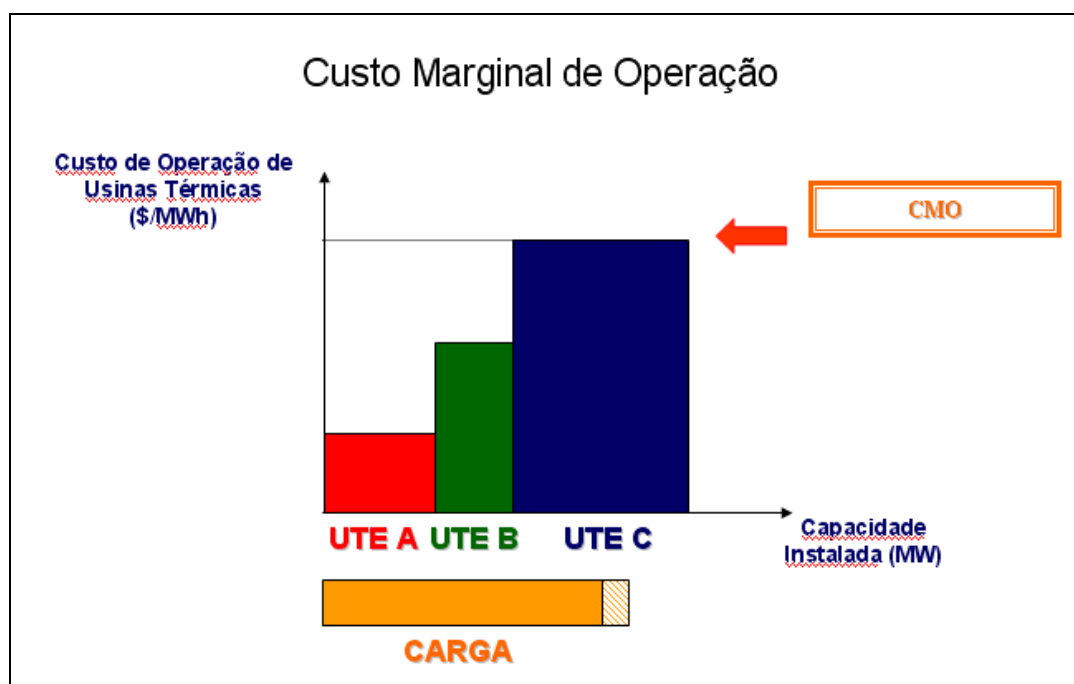


Figura 5.8: Custo Marginal de Operação

Para que os conflitos concernentes ao Uso Múltiplo da Água sejam considerados na operação do SIN, o DECOMP representa as restrições físicas e operativas associadas ao problema, como exemplo a conservação da água, os limites de turbinamento, defluência mínima das usinas, armazenamento dos reservatórios e atendimento à demanda. As incertezas acerca das vazões afluentes aos diversos aproveitamentos do sistema são representadas através de cenários hidrológicos.

No modelo DECOMP estão incorporadas as seguintes características para a operação de um sistema hidrotérmico.

Características gerais de operação do sistema:

- Cenários de afluência;
- Representação de patamares de carga;
- Configuração dinâmica;
- Limites de interligação entre subsistemas;
- Contratos de importação/exportação de energia;
- Representação de curvas de déficit por patamar;
- Restrições elétricas;

- Restrição de transporte Itaipu 50 e 60 Hz e os subsistemas SU/SE;
- Acoplamento com o modelo NEWAVE: cálculo das energias armazenada e afluenta média;
- Representação de subsistemas acoplados hidraulicamente.

Características específicas das usinas hidroelétricas:

- Enchimento de volume morto;
- Cronograma de manutenção;
- Vazão deplecionada mínima;
- Representação de unidades elevatórias;
- Restrições hidroelétricas especiais;
- Restrições de balanço hídrico por patamar para usinas a fio d'água;
- Volume de espera para amortecimento de cheias;
- Produtividade variável: Função de produção energética;
- Evaporação;
- Retiradas de água para outros usos;
- Alteração de dados do cadastro de usinas hidroelétricas;
- Tempo de viagem da vazão defluente dos aproveitamentos;
- Tratamento das bacias especiais;
- Geração de pequenas usinas.

Características específicas das usinas térmicas:

- Geração mínima em usinas térmicas;
- Cronograma de manutenção.

Características do processo de otimização:

- Integração com modelos de planejamento da operação a médio prazo (NEWAVE);
- Penalidades para intercâmbios entre subsistemas e vertimentos em reservatório;
- Revisão da política do mês inicial.

Tendo em vista as características citadas, o DECOMP busca encontrar uma estratégia de operação que forneça as metas de geração para cada usina do sistema, para cada estágio do

período de planejamento, dado o estado de armazenamento em que o sistema se encontra, no início do estágio. Esta estratégia deve minimizar o valor esperado do custo de operação ao longo do período, composto de custo de combustível mais penalizações por eventuais falhas no atendimento.

Desta forma, para se conhecer o custo de operação ao longo de todo o período, é necessário ter uma função que represente o valor esperado do custo de operação da etapa seguinte até o final do horizonte do estudo, e essa função é conhecida como Função de Custo Futuro.

6

ESTUDO DE CASOS – ANÁLISE DO ACOPLAMENTO ENTRE OS MODELOS DE MÉDIO E CURTO PRAZO

Os estudos de caso a seguir representam aplicações concretas da utilização de Séries Temporais no Planejamento da Operação do Sistema Interligado Nacional.

6.1

INTRODUÇÃO

O Programa Mensal da Operação realizado pelo ONS, utiliza a previsão de vazões semanais como um dos principais insumos para estabelecer as metas e diretrizes da política de geração para o mês do planejamento. Esta previsão é elaborada para todas as semanas do mês do PMO a partir da execução dos modelos PREVIVAZ, CPINS e MPCV.

O PREVIVAZ é o modelo utilizado pelo ONS para a geração das previsões semanais para todos os postos-base, com exceção à bacia do Rio Iguaçu que utiliza o Modelo de Previsão de Classes de Vazões – MPCV, e ao trecho da bacia do rio São Francisco incremental a UHE Sobradinho que utiliza o modelo denominado Cálculo e Previsão de Vazões Naturais e Incrementais a Sobradinho - CPINS. O modelo MPCV é um modelo de classes de vazões baseado em Redes Neurais Bayesianas e que, diferente do PREVIVAZ, utiliza as informações de precipitação observada e prevista (Cataldi et.al, 2007). O modelo CPINS é um modelo de propagação de vazões, utilizado para os diversos trechos de rio e entre os reservatórios da bacia do rio São Francisco, baseado na rotina de propagação do modelo SSARR – “Streamflow Synthesis and Reservoir Regulation” (Paiva, e Acioli, 2007).

A previsão inicial do PMO ocorre geralmente na primeira terça-feira que antecede a primeira semana prevista e as revisões são executadas às quartas-feiras. Como as semanas da programação se iniciam num sábado e acabam numa sexta-feira antes do processamento do PREVIVAZ, é necessário completar os dias restantes da semana em curso (quarta, quinta

e sexta-feira). Para tal são utilizadas as previsões diárias feitas em sua grande parte pelos diferentes Agentes de Geração ou pelo modelo PREVIVAZH.

Uma vez que a previsão de vazões semanais foi realizada, é calculada uma média mensal baseada nessas previsões semanais, a partir da qual serão gerados os cenários de afluições pelo GEVAZP.

Com base na previsão de vazões para o mês em curso e dos cenários de afluições para o segundo mês, o DECOMP realiza a otimização visando minimizar o valor esperado do custo total de operação, composto pelo custo imediato mais o custo futuro, e como variável resultante da solução deste problema de otimização tem-se o Custo Marginal de Operação, que representa o impacto no custo total de operação ao se aumentar o requisito de mercado em 1MW.

Para que o DECOMP realize a otimização dos recursos no curto prazo, é necessário que a informação estrutural de médio prazo seja fornecida pelo NEWAVE através da Função de Custo Futuro, possibilitando que a política de operação do médio prazo seja contemplada no curto prazo.

À medida que o mês avança, as previsões semanais feitas para o PMO são substituídas pelos valores verificados. Desta forma, o PREVIVAZ é executado novamente contemplando os dados reais verificados, o que faz com que as previsões seguintes se distanciem das previsões originais feitas no início do mês (revisão 0). Sendo assim, espera-se que na última revisão se esteja o mais próximo da realidade.

O Custo Futuro representa o Valor da Água no futuro. Para a determinação do Custo de Operação de um período futuro T qualquer é necessário conhecer o estado de armazenamento dos reservatórios naquele instante, logo este custo depende das afluições que irão ocorrer nos rios e bacias do SIN.

Conhecendo-se o modelo estocástico que determina as afluições, é possível acompanhar um grande número de hipóteses para estas afluições, e assim determinar o Custo Futuro médio para estas hipóteses.

A Programação Dinâmica Estocástica permite estudar a evolução do sistema e trazer as informações do futuro para o presente. O método encontra o Valor da Água para qualquer Estado de armazenamento pertencente a qualquer Etapa e a qualquer uma das trajetórias sob as diversas hipóteses de afluições. Esse cálculo resulta em uma explosão de valores quando o método é aplicado a mais de dois reservatórios, o que inviabiliza a utilização do método no Planejamento da Operação do SIN. Este problema é conhecido como Maldição da Dimensionalidade.

A técnica utilizada para contornar este problema é a Programação Dinâmica Dual Estocástica. Ela reduz consideravelmente a quantidade de cálculos, uma vez que deixa de calcular o custo futuro para todos os Estados de armazenamento, e passa a calcular para apenas alguns Estados.

Os Estados (combinação de armazenamentos) para os quais a FCF é calculada são determinados pela simulação forward com pode ser visto na Figura 6.1. Logo, ao invés de calcular os Valores da Água para cada combinação de armazenamento, passam a ser calculados valores somente para os estados visitados na iteração FW. Para os estados nos quais a FCF não foi calculada, o Valor da Água é obtido a partir de uma extrapolação/interpolação. Conseqüentemente, há um maior erro de aproximação da FCF nestas regiões.

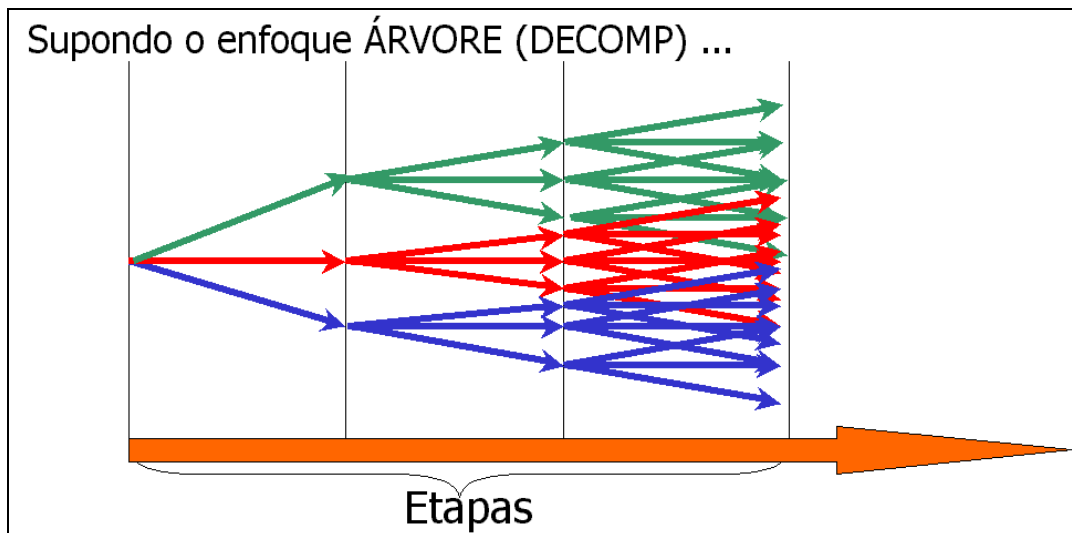


Figura 6.1: Simulação Forward – Determinação dos Estados de Armazenamento

Nos Estados que foram atingidos, em cada Etapa, na otimização "Forward", é então calculado o Custo Futuro, em um processo que parte do final do horizonte e chega ao seu início, no sentido inverso do tempo (otimização "Backward").

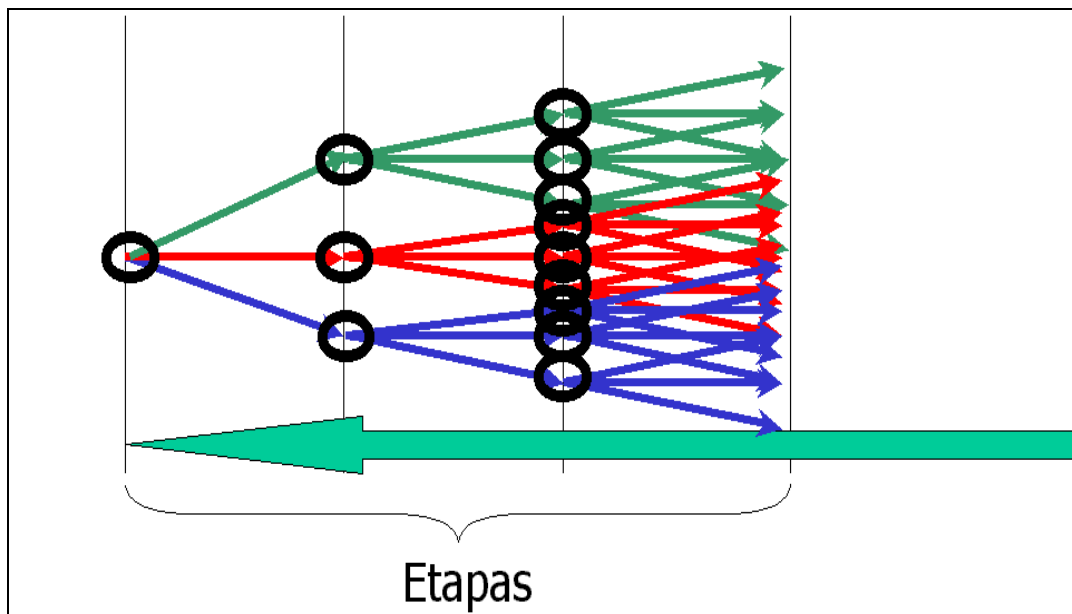


Figura 6.2: Simulação Backward – Cálculo da Função de Custo Futuro

A cada iteração é acrescentado um corte da Função de Custo Futuro de modo que o processo iterativo vai se repetindo até que se atinja uma função bem definida.

Tendo isto exposto, fica clara a necessidade de se ter uma previsão e geração de cenários, no curto e médio prazo, compatíveis entre si e com a realidade.

Ao longo das revisões no curto prazo o DECOMP consulta os cortes da Função de Custo Futuro do NEWAVE que se aproximam das combinações de armazenamentos/cenários encontradas no curto prazo. Com isso, o DECOMP realiza a otimização dos recursos no curto prazo contemplando, através da consulta à Função de Custo Futuro, a informação estrutural de médio prazo fornecida pelo NEWAVE.

Se as combinações de armazenamentos/cenários do DECOMP forem muito distintas dos valores praticados pelo NEWAVE, o DECOMP não vai encontrar a situação hidrológica do médio prazo que represente a situação hidrológica vivenciada no curto prazo de modo que a Função de Custo Futuro pode ser consultada em regiões nas quais não representa adequadamente as condições hidrológicas reais.

Esta ocorrência pode fazer com que a política de operação adotada para o sistema não seja a mais adequada, pois o DECOMP tende a consultar sempre os mesmos cortes da FCF para os cenários de afluências que não estão dentro dos cenários previstos pelo NEWAVE, deixando de visitar outros cortes que poderiam representar melhor a realidade.

Neste estudo de caso, pretende-se avaliar como a informação fornecida pelo PREVIVAZ influencia na geração dos cenários de vazões do GEVAZP (já que estes são traçados com base na média dos valores de vazões semanais fornecidos pelo PREVIVAZ) e, conseqüentemente, no acoplamento entre os modelos de otimização de médio e curto prazo. Essa avaliação se estende ao impacto da atualização das previsões de vazões semanais, que também interfere nos resultados da operação, alterando os custos de operação, armazenamentos, geração térmica e CMO's.

Como medidas de avaliação destes impactos serão analisadas a variação do número de cenários em que houve extrapolação na zona de acoplamento entre os modelos NEWAVE e DECOMP, possivelmente devido à atualização das aflúncias nas revisões semanais, e o impacto sobre o custo total da operação e sobre o CMO.

6.2

PREMISSAS

Este estudo foi realizado com dados reais da operação eletroenergética do SIN. Os casos utilizados no estudo estão disponíveis no site do ONS.

O programa NEWAVE foi executado para determinar a política de geração de médio prazo, e o DECOMP foi executado para determinar a política de operação do curto prazo, gerando os valores do custo total e do CMO para cada revisão.

Para realizar a otimização no horizonte de médio prazo, o NEWAVE utiliza 200 cenários equiprováveis de aflúncia (fornecidos pelo GEVAZP) para a simulação Forward, na qual serão estimados os estados para os quais a FCF será calculada. Na simulação Backward, na qual a FCF é calculada, o NEWAVE utiliza 20 cenários equiprováveis para cada combinação de armazenamentos obtidos na simulação Forward em cada período do estudo, e calcula um corte médio para cada uma dessas combinações. Já o DECOMP realiza as simulações Forward e Backward sobre o mesmo conjunto de cenários, cujo número varia de acordo com cada mês.

O número de cenários do DECOMP foi definido com base em estudos e notas técnicas elaborados pelo ONS em conjunto com os Agentes. A quantidade de cenários varia de mês para mês, conforme pode ser visto na Tabela 6.1. Maiores esclarecimentos sobre o número de cenários do DECOMP podem ser obtidos no Relatório de Validação do Modelo

GEVAZP elaborado em conjunto pelo ONS e pelo Mercado Atacadista de Energia (2002), atualmente chamado de Câmara de Comercialização de Energia.

Tabela 6.1: Número de Cenários do DECOMP por Mês

Mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Nº de cenários	136	116	143	143	193	267	513	353	303	259	228	153

Neste estudo foram analisados dois casos do Planejamento Mensal da Operação do SIN:

- Caso I – PMO de outubro de 2007, com 228 cenários gerados para o segundo mês de planejamento.
- Caso II – PMO de janeiro de 2008, com 116 cenários gerados para o mês de fevereiro.

6.3

METODOLOGIA

Para cada uma das revisões do DECOMP, os cenários de vazões gerados pelo GEVAZP, devidamente convertidos para cenários de energias naturais afluentes, foram comparados com os cenários gerados para a simulação Forward do modelo NEWAVE. Embora os cenários em questão sejam multivariados, ou seja, cada afluência para cada subsistema ocorre conjuntamente com afluências nos outros subsistemas, a análise é realizada de forma marginal, onde as afluências de cada subsistema são avaliadas separadamente.

O NEWAVE é executado uma única vez para cada PMO, para um horizonte de cinco anos à frente, partindo do valor da Energia Natural Afluente verificada no mês anterior. O DECOMP por sua vez, é rodado a cada revisão, e a cada revisão a previsão da afluência semanal da semana que passou é substituída pelo valor da afluência verificada, gerando uma nova média das afluências semanais a partir da qual serão gerados novos cenários para o mês seguinte. Como o segundo mês do planejamento está contido no horizonte do NEWAVE não é necessário mais do que uma execução do NEWAVE para comparar os novos cenários gerados a cada revisão do DECOMP.

Tendo em vista o exposto, os cenários do DECOMP gerados para o segundo mês foram comparados com os cenários do NEWAVE no segundo mês. Assim, para cada cenário utilizado pelo modelo DECOMP que estava abaixo ou acima do conjunto de cenários usados pelo modelo NEWAVE foi considerado como uma extrapolação na consulta à FCF do modelo NEWAVE. Embora não seja considerado neste trabalho, a consulta pelo DECOMP a regiões pouco visitadas pelo modelo NEWAVE, mesmo que internamente à Região delimitada pelo seu menor e maior cenários, também pode acarretar no uso de um Valor da Água distante do real.

6.4

RESULTADOS

6.4.1

Caso I – PMO Outubro de 2007

De acordo com a análise das Figuras 6.3 e 6.4, que representam o acoplamento entre os modelos de médio e curto prazo, nota-se que para cada revisão semanal o número de extrapolações na consulta à FCF variou consideravelmente, o que é resumido na Tabela 6.2.

Em todas as revisões o subsistema Sudeste foi o que obteve o maior número de cenários extrapolados. Como o Sudeste é o maior subsistema do SIN, sua influência nos resultados obtidos pelos modelos de otimização é maior que os demais subsistemas. Dessa forma, a aproximação da FCF nas regiões consultadas pelo DECOMP deve ser a melhor possível, o que nem sempre ocorre quando se consulta essa função em estados pouco detalhados pelo NEWAVE.

Tabela 6.2: Número de Cenários do DECOMP que Extrapolaram o NEWAVE

Outubro 2007	RV0	RV1	RV2	RV3
SUDESTE	11	5	21	32
NORDESTE	0	1	3	8

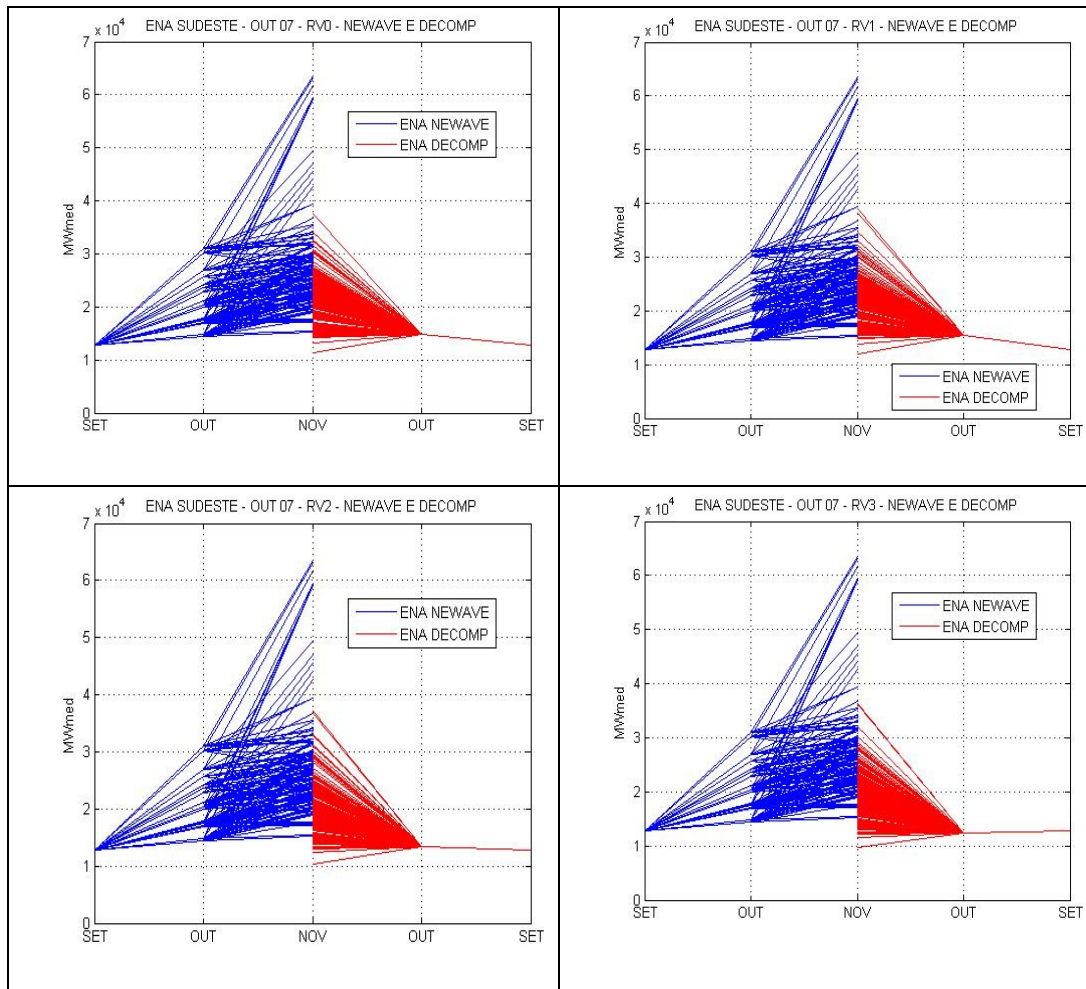


Figura 6.3: Desacoplamento Parcial entre NEWAVE e DECOMP na Região SE

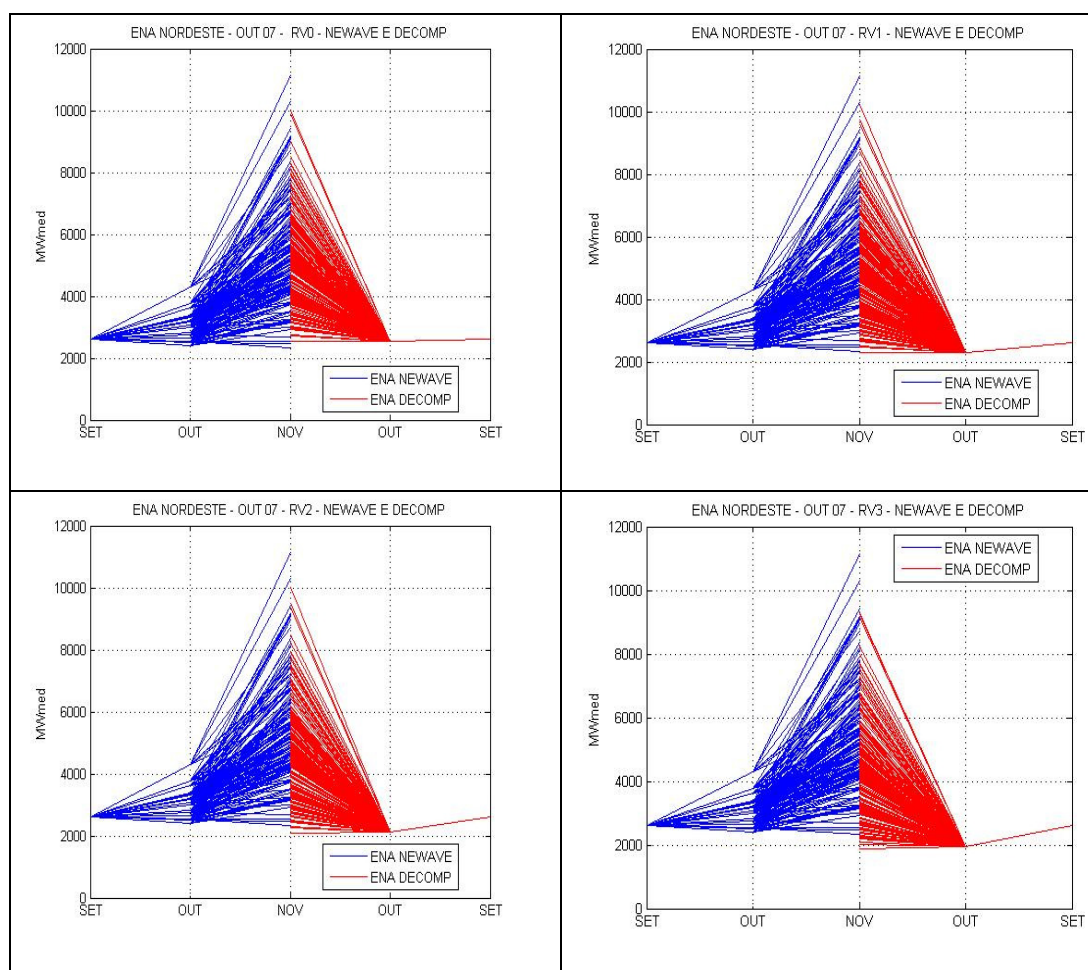


Figura 6.4: Desacoplamento Parcial entre NEWAVE e DECOMP na Região NE

A incorporação dos dados de afluência verificados, no modelo de previsão, faz com que os resultados das previsões das revisões semanais sejam sensivelmente alterados, modificando o acoplamento dos modelos de otimização e também os resultados da operação, como pode ser visto na Tabela 6.3 que mostra como o Custo Total de Operação variou a cada revisão semanal.

Tabela 6.3: Custo Total de Operação

Custo Total de Operação (1.000 \$)

Revisão	Sem 1	Sem 2	Sem 3	Sem 4	Sem 5
0	51003.9	51746.2	48148.7	47356.9	51104.4
1		53291	53796.8	52253.9	50291.6
2			59044.9	52909.5	50621.6
3				61916.5	58600.2

A Figura 6.5 mostra a variação do Custo Total de Operação ocorrida durante o mês do PMO. O Custo Total da Rev.0 foi calculado com a média das previsões de afliências das cinco semanas. O Custo Total da Rev.1 foi calculado com a média do valor verificado de afliência para a semana 1 e as previsões de afliências para as quatro semanas seguintes. O Custo Total da Rev.2 foi calculado com a média dos valores verificados de afliência para as semanas 1 e 2 e as previsões de afliências para as três semanas seguintes. Dessa forma, o processo de Revisões se repete até que a última revisão seja feita com quatro dados de afliência verificada e apenas um valor de afliência prevista. Percebe-se que quanto mais a previsão de afliências se distancia dos dados verificados, maior é a variação do Custo Total de Operação.

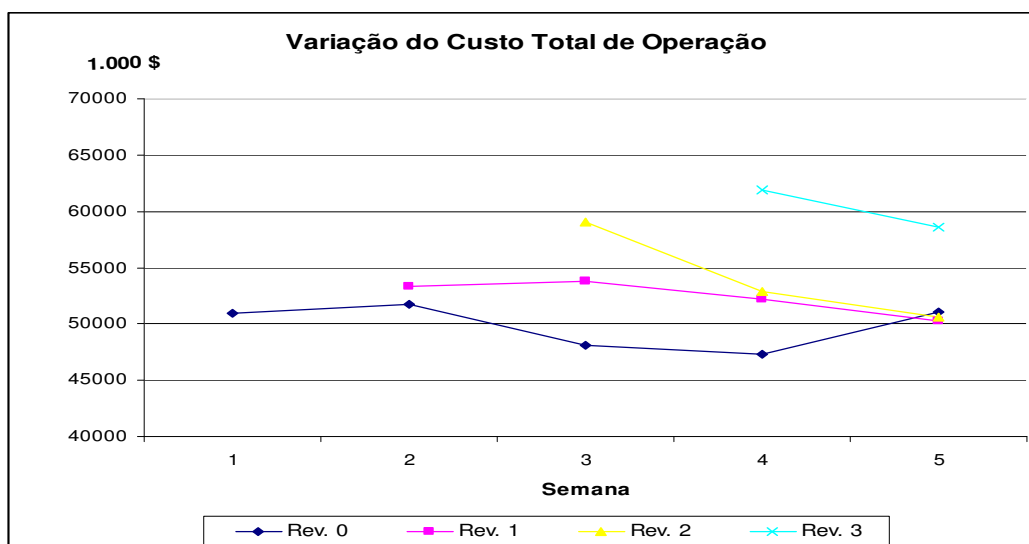


Figura 6.5: Variação do Custo Total de Operação

A Tabela 6.4 mostra quanto o Custo Total de Operação variou de uma revisão para outra, em virtude das revisões do DECOMP.

Tabela 6.4: Variação Percentual do Custo Total de Operação (%)

Revisão	Sem 1	Sem 2	Sem 3	Sem 4	Sem 5
0	-	-	-	-	-
1		2.99	11.73	10.34	-1.59
2			22.63	11.73	-0.94
3				30.74	14.67

Conforme pode ser visto nas Tabelas seguintes, os resultados do Custo Marginal de Operação foram sensivelmente aumentados desde a previsão feita no início do mês, para a previsão feita na última revisão, após ter incorporado os valores de afluições verificados.

Tabela 6.5: Custo Marginal de Operação do Subsistema Sudeste

	Subsistema	Patamar	Sem 1	Sem 2	Sem 3	Sem 4	Sem 5
Rev. 0	Sudeste	Pesada	162,07	161,94	162,12	160,59	158,41
		Médio	160,98	158,37	159,69	158,88	157,06
		Leve	157,82	157,80	158,39	158,82	157,06
	Média Semanal		159,95	158,44	159,48	159,04	157,18
Rev. 1	Sudeste	Pesada		166,62	165,99	164,83	163,49
		Médio		163,70	163,33	163,21	161,85
		Leve		163,70	162,82	163,21	161,85
	Média Semanal		163,96	163,43	163,39	162,00	
Rev. 2	Sudeste	Pesada			200,04	199,22	197,21
		Médio			197,89	197,56	196,02
		Leve			197,89	197,56	196,02
	Média Semanal			198,12	197,74	196,13	
Rev. 3	Sudeste	Pesada				202,87	203,46
		Médio				202,87	201,79
		Leve				201,37	201,79
	Média Semanal				202,33	201,94	

Tabela 6.6: Custo Marginal de Operação do Subsistema Nordeste

	Subsistema	Patamar	Sem 1	Sem 2	Sem 3	Sem 4	Sem 5
Rev. 0	Nordeste	Pesada	162,07	161,94	162,12	160,59	158,41
		Médio	160,98	158,37	159,69	158,88	157,06
		Leve	157,82	157,80	158,39	158,82	157,06
	Média Semanal		159,95	158,44	159,48	159,04	157,18
Rev. 1	Nordeste	Pesada		166,62	165,99	164,83	163,49
		Médio		154,87	163,33	163,21	161,85
		Leve		154,32	161,88	163,21	161,85
	Média Semanal		155,68	163,10	163,39	162,00	
Rev. 2	Nordeste	Pesada			200,04	199,22	195,04
		Médio			197,89	197,56	195,04
		Leve			196,13	197,56	195,04
	Média Semanal			197,49	197,74	195,04	
Rev. 3	Nordeste	Pesada				202,87	203,38
		Médio				202,10	201,79
		Leve				201,37	201,79
	Média Semanal				201,92	201,93	

A variação da Energia Natural Afluente a cada revisão semanal pode ser vista, na Tabela 6.7, onde os valores em vermelho representam as afluências verificadas na semana de cada revisão e os valores em preto representam os dados previstos.

Tabela 6.7: Energia Natural Afluente Prevista e Verificada – Caso I

ENERGIA NATURAL AFLUENTE (MWmed) - SE

out/07	Sem_01	Sem_02	Sem_03	Sem_04	Sem_05
RV0	12097.9	13607.3	14734.6	15813.0	17778.8
RV1	12626	14532.4	15494.8	16528.6	18458.9
RV2	12388	11396	12670.7	14342.1	16655.0
RV3	12390	10935	10667	12767.1	15168.7

ENERGIA NATURAL AFLUENTE (MWmed) - NE

out/07	Sem_01	Sem_02	Sem_03	Sem_04	Sem_05
RV0	2135.4	2272.3	2479.5	2681.0	3124.4
RV1	2300	1978.3	2059.6	2329.0	2814.4
RV2	2320	2161	1880.0	1834.4	2334.6
RV3	2271	2089	1967	1794	1760

Da observação da Tabela 6.7 percebe-se que os dados verificados sofrem variações entre as revisões semanais. Esta diferença existe em virtude das revisões do PMO ocorrerem no meio da semana operativa, geralmente às quartas-feiras. Este fato torna a semana em que a revisão ocorre uma semana estimada, já que esta possui valores observados de vazão de sábado até terça-feira e previstos de quarta até sexta-feira. Na revisão seguinte, os dados desta semana deixam de ser estimados (uma parte observada e uma parte prevista) e passam a ser somente observados.

Além disso, os próprios dados observados sofrem diferentes níveis de consistência, de acordo também com a semana operativa. O último nível de consistência, onde o dado recebe o carimbo de consolidado, é baseado em médias móveis centradas de cinco dias. Desta forma, como a média móvel é centrada, somente após toda a semana operativa possuir dados observados é que esta consolidação pode ser aplicada, o que também ocasiona pequenas diferenças entre as médias semanais observadas. Uma ilustração da distribuição temporal dos diferentes tipos de dados utilizados nas revisões pode ser encontrada na Tabela 6.8.

Tabela 6.8 – Atualização dos Dados Diários de Vazão

	SEMANA 1	SEMANA 2	SEMANA 3	SEMANA 4	SEMANA 5
	S D S T Q Q S	S D S T Q Q S	S D S T Q Q S	S D S T Q Q S	S D S T Q Q S
REV 0	O O O O P P P	P P P P P P P	P P P P P P P	P P P P P P P	P P P P P P P
REV 1	V V V V O O O	O O O O P P P	P P P P P P P	P P P P P P P	P P P P P P P
REV 2	V V V V V V V	V V V V O O O	O O O O P P P	P P P P P P P	P P P P P P P
REV 3	V V V V V V V	V V V V V V V	V V V V O O O	O O O O P P P	P P P P P P P

O => DADOS OBSERVADOS

P => DADOS PREVISTOS

V => DADOS VERIFICADOS E CONSOLIDADOS

O procedimento de consolidação é necessário porque dados diários possuem muitas imprecisões, decorrentes principalmente de:

- Desvios entre a vazão afluyente informada e o valor calculado por meio do método de balanço hídrico do reservatório;
- Diferenças entre tabelas Cota x Volume utilizadas pelos Agentes e pelo ONS;
- Níveis da água do reservatório fora da faixa situada entre os níveis mínimo e máximo operativo;
- Erros nas vazões defluentes totais;
- Incompatibilidades entre vazão turbinada e energia gerada;
- Repetições de informações do dia anterior.

Maiores informações sobre o procedimento de consolidação dos dados de vazão observada podem ser obtidas na Nota Técnica elaborada pelo ONS (2005) sobre os Processos de Consistência e de Consolidação de dados Hidráulicos, Hidrológicos e Hidroenergéticos.

6.4.2

Caso II – PMO Janeiro de 2008

De acordo com a análise feita para o Caso I, os itens a seguir apresentam a mesma análise aplicada ao caso II.

A Tabela 6.9 mostra a quantidade de cenários de Energia Natural Afluyente do DECOMP que extrapolaram os cenários do NEWAVE, e as Figuras 6.6 e 6.7 ilustram a extrapolação dos cenários para as regiões Sudeste e Nordeste.

Tabela 6.9: Número de Cenários do DECOMP que Extrapolaram o NEWAVE – Caso II

Janeiro 2008	RV0	RV1	RV2	RV3
SUDESTE	2	4	9	8
NORDESTE	0	11	21	9

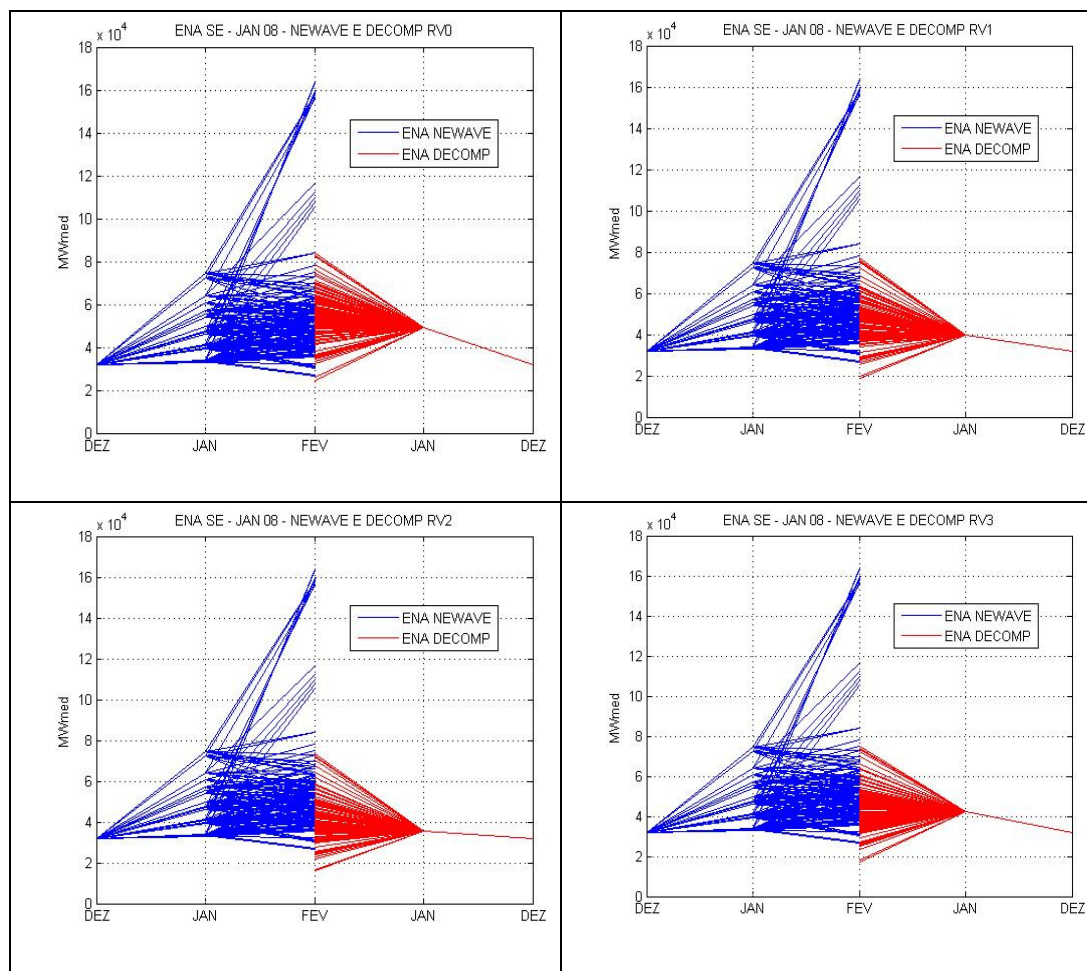


Figura 6.6: Desacoplamento Parcial entre NEWAVE e DECOMP na Região SE – Caso II

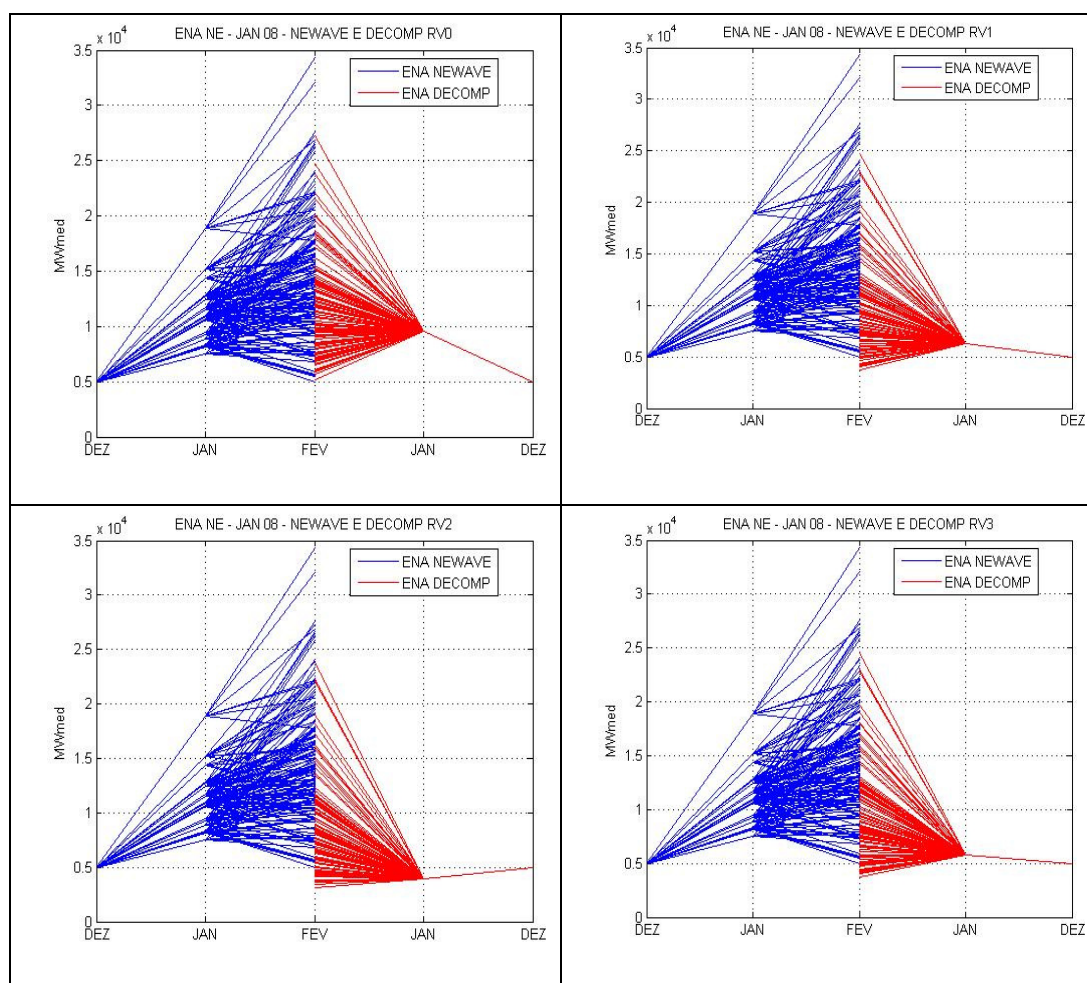


Figura 6.7: Desacoplamento Parcial entre NEWAVE e DECOMP na Região NE – Caso II

Cabe ressaltar que Janeiro é um mês de afluências favoráveis. Entretanto, em janeiro de 2008 o período úmido atrasou, o que fez com que o modelo fizesse previsões bem distantes das afluências verificadas. Percebe-se neste caso, que como a “chuva atrasou”, o valor de afluência verificada foi menor do que o valor previsto, e com isso as previsões para as semanas seguintes também foram mais baixas. Esta diferença reflete-se diretamente sobre o Custo Total de Operação, conforme pode ser observado na Tabela 6.10. À medida que as previsões de afluência foram sendo “mais pessimistas” o modelo passou a enxergar menos água disponível para geração hidroelétrica e como consequência aumentou a geração termoelétrica, aumentando assim o Custo Total de Operação.

Tabela 6.10: Custo Total de Operação – Caso II

Custo Total de Operação (1.000 \$)

Revisão	Sem 1	Sem 2	Sem 3	Sem 4	Sem 5
0	82508.08	82508.09	81409.75	79925.11	81594.33
1		109574.00	103384.17	110766.16	105058.86
2			110633.90	114819.41	115207.30
3				104528.89	119606.84

A Figura 6.8 mostra a variação do Custo Total de Operação ocorrida a cada revisão, durante o mês do PMO.

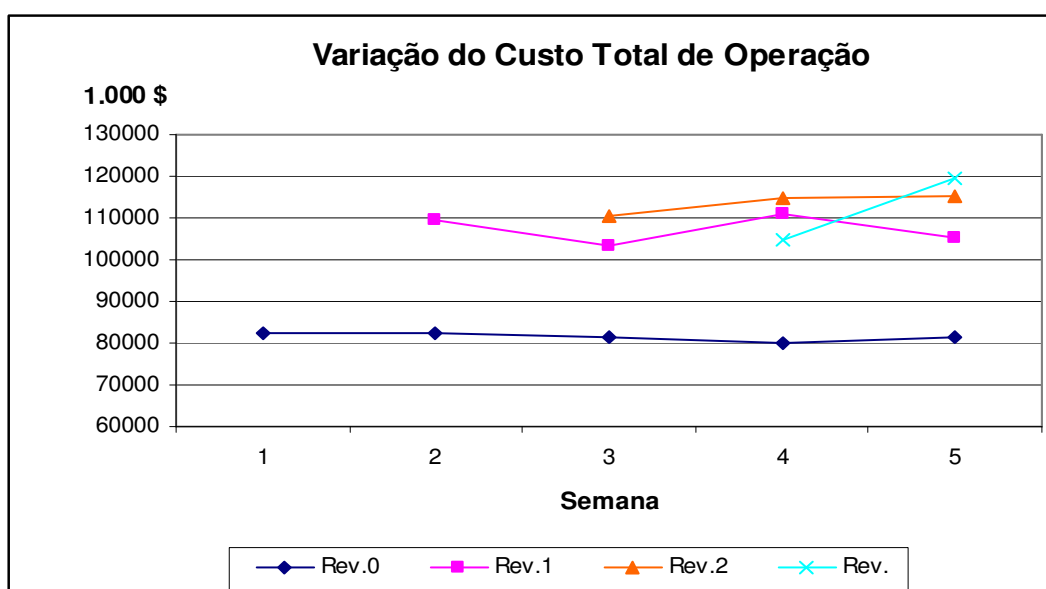


Figura 6.8: Variação do Custo Total de Operação – Caso II

A Tabela 6.11 mostra quanto o Custo Total de Operação variou de uma revisão para outra, em virtude das revisões do DECOMP.

Tabela 6.11: Variação Percentual do Custo Total de Operação (%) – Caso II

Revisão	Sem 1	Sem 2	Sem 3	Sem 4	Sem 5
0	-	-	-	-	-
1		32,80	26,99	38,59	28,76
2			35,90	43,66	41,20
3				30,78	46,59

Conforme pode ser visto nas Tabelas 6.12 e 6.13, o Custo Marginal de Operação teve grande variação desde a previsão feita no início do mês, para a previsão feita na última revisão, após ter incorporado os valores de afluições verificados. E neste caso, como os valores de afluições verificados foram menores que os previstos, devido ao atraso da chuva, a geração que o sistema encontrou disponível para atender a carga, foi a geração térmica, o que ocasionou um aumento considerável do CMO.

Tabela 6.12: Custo Marginal de Operação do Subsistema Sudeste – Caso II

	Subsistema	Patamar	Sem 1	Sem 2	Sem 3	Sem 4	Sem 5
Rev. 0	Sudeste	Pesada	248,79	252,22	252,03	250,29	249,70
		Médio	248,79	252,22	252,03	249,79	248,78
		Leve	248,79	252,22	252,03	249,79	248,69
	Média Semanal		248,79	252,22	252,03	249,84	248,84
Rev. 1	Sudeste	Pesada		472,10	470,67	471,77	468,71
		Médio		472,10	470,67	471,77	468,71
		Leve		472,10	470,67	471,77	468,71
	Média Semanal		472,10	470,67	471,77	468,71	
Rev. 2	Sudeste	Pesada			637,23	633,65	633,50
		Médio			637,23	633,65	633,50
		Leve			637,23	633,65	633,50
	Média Semanal			637,23	633,65	633,50	
Rev. 3	Sudeste	Pesada				581,48	580,49
		Médio				581,48	580,49
		Leve				581,48	580,13
	Média Semanal				581,48	580,36	

Tabela 6.13: Custo Marginal de Operação do Subsistema Nordeste – Caso II

	Subsistema	Patamar	Sem 1	Sem 2	Sem 3	Sem 4	Sem 5
Rev. 0	Nordeste	Pesada	250,79	252,22	252,73	250,29	251,17
		Médio	250,79	252,22	252,73	250,29	251,17
		Leve	250,79	252,22	252,73	250,29	251,17
	Média Semanal		250,79	252,22	252,73	250,29	251,17
Rev. 1	Nordeste	Pesada		472,10	470,67	471,77	468,71
		Médio		472,10	470,67	471,77	468,71
		Leve		472,10	470,67	471,77	468,71
	Média Semanal		472,10	470,67	471,77	468,71	
Rev. 2	Nordeste	Pesada			632,39	622,91	625,05
		Médio			632,39	622,91	625,05
		Leve			632,39	622,91	625,05
	Média Semanal			632,39	622,91	625,05	
Rev. 3	Nordeste	Pesada				559,11	548,52
		Médio				559,11	548,52
		Leve				559,11	548,52
	Média Semanal				559,11	548,52	

A variação da Energia Natural Afluente a cada revisão semanal pode ser vista, na Tabela 6.14, onde os valores em vermelho representam as afluências verificadas na semana de cada revisão e os valores em preto representam os dados previstos.

Tabela 6.14: Energia Natural Afluente Prevista e Verificada – Caso II

ENERGIA NATURAL AFLUENTE (MWmed) - SE

jan/08	Sem_01	Sem_02	Sem_03	Sem_04	Sem_05
RV0	45231.5	46784.2	49090.4	50878.0	53424.3
RV1	26892.0	33307.1	38093.4	42526.9	45972.1
RV2	27320.0	24977.0	30948.1	36085.1	39276.1
RV3	27319.0	24974.0	32669.0	40189.4	44080.4

ENERGIA NATURAL AFLUENTE (MWmed) - NE

jan/08	Sem_01	Sem_02	Sem_03	Sem_04	Sem_05
RV0	7144.1	8541.7	10243.1	10884.7	10908.1
RV1	6821	7262.6	5193.6	5894.1	6676.9
RV2	6805.0	7009.0	3867.1	3469.6	4590.6
RV3	6804.0	6949.0	4528.0	4752.9	6868.2

Os dados verificados na Tabela 6.14 sofrem alterações entre as revisões pelos mesmos motivos já explicados para a Tabela 6.7. Maiores informações sobre o procedimento de consolidação dos dados de vazão observada podem ser obtidas na Nota Técnica elaborada pelo ONS (2005) sobre os Processos de Consistência e de Consolidação de dados Hidráulicos, Hidrológicos e Hidroenergéticos.

7

CONCLUSÕES

O estudo demonstra o quão fundamental são os modelos de previsão de vazões para o Planejamento Energético do SIN. Além disso, evidencia a necessidade de se obter previsões mais próximas da realidade para que os riscos decorrentes da dependência das afluições dos reservatórios do SIN sejam mitigados.

Para um sistema hidrotérmico da extensão do SIN, e com regimes hidrológicos distintos para cada Região é muito difícil que um único modelo seja capaz de representar de forma eficiente toda essa gama de afluições.

Como o Modelo de curto prazo efetua simulações a usinas individualizadas, faz-se necessário a obtenção de afluições determinísticas a seus postos em horizonte mensal, subdividido em semanas operativas.

Além disso, a composição dos cenários de Energias Afluentes no horizonte de cinco anos, com as vazões determinísticas do primeiro mês, é de fundamental importância na análise das variações do CMO e na determinação da Política de Operação, como demonstrado no estudo de caso.

Por isso, para o curto prazo é imprescindível um modelo de previsão de afluições com alta eficiência, pois é partindo destas previsões que serão gerados os cenários para o 2º mês (e conseqüentemente os estados visitados no 2º mês). Com boas previsões iniciais pode-se diminuir a ocorrência de grandes variações de previsões ao longo das revisões do PMO. Caso o modelo de previsão cometa muitos erros, os cenários gerados para o 2º mês podem ser muito distintos, e conseqüentemente, as decisões a respeito da operação também serão distintas, impactando CMO, custos, Política de Operação e etc..

Para superar este problema, o ONS vem investindo sistematicamente em modelos de previsão capazes de incorporar informações de chuva, como é o caso da extensão do modelo MPCV para a Bacia do Uruguai e da elaboração de novos modelos para outras bacias, como é o caso do modelo Neuro-Fuzzy para o Iguazu, do modelo MGB para a Bacia

do Paranaíba e do modelo Smap-Mel para a bacia do rio Paraná. Esses modelos foram concebidos com tecnologias bastante distintas (inteligência artificial, modelagem física e modelam estocástica), mas todos têm em comum o fato de considerarem, além das vazões naturais e afluências passadas, a precipitação observada nos últimos dias e prevista para os próximos 10 dias.

Com estas ações, pretende-se aumentar a qualidade das previsões de vazão natural e com isso, fazer com que as previsões feitas no início de um ciclo do Planejamento Mensal da Operação sejam mais próximas das previsões feitas durante as revisões, e desta forma determinar a melhor Política de Operação para o período.

REFERÊNCIAS BIBLIOGRÁFICAS

ACIOLI, G.C.L., GOMES, L.F.C. e MAGALHÃES, J.K.M. (2004): Previsão de vazões diárias ao reservatório de Sobradinho, VII Simpósio de Recursos Hídricos do Nordeste, São Luiz, Brasil.

ARTEIRO, F.J., DAHER, M.J., VIEIRA, M. A., Analysis of the Environment and Multiple Use of Water Impacts in the Energetic Operation of the BRAZILIAN INTERCONNECTED POWER System / Paris, 2005

BELLMAN. Dynamic Programming. Princeton University Press, 1957.

BELLMAN. Adaptive Control Processes: a Guided Tour. Princeton University Press, 1961.

BENDERS, J.F. Partitioning Procedures for Solving Mixed Variables Programming Problems. Numer. Math., v.4, p. 238-252, 1962.

BOX, G.E.P., JENKINS, G.M., "Time Series Analysis - Forecasting and Control", Holden-Day, 1970.

CATALDI, M. ; Freitas, Bruno G. ; da Silva, Simone B. ; Guilhon, Luiz G. F. ; Achão, Carla da C. L. . Aplicação de Técnica de Mineração de Dados como Complemento às Previsões Estocásticas Univariadas de Vazão Natural: Estudo de Caso para a Bacia do Rio IGUAÇU. Revista Brasileira de Recursos Hídricos, v. 12, n.3, p. 83-92, 2007.

CENTRO DE PESQUISA DE ENERGIA ELÉTRICA – CEPEL, Manual de Referência do Modelo DECOMP V 12.0, Rio de Janeiro, out. 2004.

CENTRO DE PESQUISA DE ENERGIA ELÉTRICA – CEPEL, Manual de Referência – Modelo de Previsão de Vazões Semanais Aplicado ao Sistema Hidroelétrico Brasileiro – Modelo PREVIVAZ, Rio de Janeiro, 2004.

CENTRO DE PESQUISA DE ENERGIA ELÉTRICA – CEPEL, Manual de Referência – Modelo de Geração de Séries Sintéticas de Energias e Vazões – MODELO GEVAZP, Rio de Janeiro, jun. 2002.

CENTRO DE PESQUISA DE ENERGIA ELÉTRICA – CEPEL, Manual de Referência do Modelo NEWAVE, Rio de Janeiro, 2006.

CENTRO DE PESQUISA DE ENERGIA ELÉTRICA - CEPEL, Modelo Estocástico para Previsão de Vazões Diárias para a Programação Diária, Manual de Utilização – PREVIVAZH V 2.5, Rio de Janeiro, fev. 2002.

CENTRO DE PESQUISA DE ENERGIA ELÉTRICA - CEPEL, Modelo de Previsão de Vazões Mensais, Manual de Metodologia – PREVIVAZM, Rio de Janeiro, fev. 2003.

FORGIARINI, F.R., GERALDO, L.S. e CRUZ, J.C. (2007): Gestão dos Recursos Hídricos e Cobrança pelo Uso da Água: Visão da Sociedade da Bacia do Rio Santa Maria / RS/ Artigo publicado na RBRH – Revista Brasileira de Recursos Hídricos Volume 12 n.2 Abr/Jun 2007, 123-133. UFSM, Santa Maria/RS.

GUILHON, L.G.F., 2002, **Modelo Heurístico de Previsão De Vazões Naturais Médias Semanais Aplicado à Usina de Foz Do Areia** - Tese de Mestrado, COPPE - UFRJ.

HIPEL, K. W. e MCLEOD, A. I. (1994), “Time Series Modeling of Water Resources and Environmental Systems”. Elsevier Science B.V., Netherlands, 1013 páginas.

MACEIRA, M.E.P., “Operação Ótima de Reservatório com Previsão de Afluências”, Tese de Msc. COPPE/UFRJ, Rio de Janeiro, 1989.

MACEIRA M.E.P., DAMÁZIO J.M., GHIRARDI A.O. & DANTAS H. Periodic ARMA Models Applied to Weekly Streamflow Forecasts. Proceedings of the 1999 IEEE Powertech, Budapest..

MACEIRA, M.E.P., e MERCIO, C.M.V.B. Stochastic Streamflow Model for Hydroelectric Systems, in Anais do V International Conference PMAPS - Probabilistic Methods Applied to Power Systems, Vancouver, Canada, 1997.

MACEIRA, M.E.P., TERRY, L.A., DAMÁZIO, J.M, COSTA, F.S., MELO, A.C.G. Chain of Optimization Models for Setting the Energy Dispatch and Spot Price in the Brazilian. Power System Computation Conference – PSCC 02, Sevilla, Spain, p. 24-28, 2002.

MACEIRA, M.E.P., COSTA, F.S., DAMÁZIO, J.M. Modelos de Previsão Hidrológica Aplicados ao Planejamento da Operação do Sistema Elétrico Brasileiro/ Artigo apresentado no Workshop de Previsão de Vazões do ONS, 2007.

MACEIRA, M.E.P et al. Aplicação de Técnicas de Relaxação Lagrangeana e Programação Dinâmica Dual ao Modelo de Despacho Horário da Operação Energética/ Artigo apresentado no Seminário Nacional de Produção e Transmissão de Energia Elétrica – SNPTEE, Uberlândia/MG, 2003.

OPERADOR NACIONAL DO SISTEMA ELÉTRICO – Nota Técnica NT-076/2005 - Processos de Consistência e de Consolidação de Dados Hidráulicos, Hidrológicos e Hidroenergéticos, Rio de Janeiro, 2005. 24 p.

OPERADOR NACIONAL DO SISTEMA ELÉTRICO e MERCADO ATACADISTA DE ENERGIA ELÉTRICA. Validação do Modelo GEVAZP – Geração de Vazões Padrão Paralelo e Padrão Árvore com um Estágio à Frente, Rio de Janeiro, 2002. 168 p.

PAIVA, L.G.P., ACIOLI, G.C.L., PREVISÃO DE VAZÕES COM O MODELO CPINS – Cálculo e Previsão de Vazões Naturais e Incrementais a Sobradinho (2007) / *Artigo apresentado no Workshop de Previsão de Vazões do ONS.*

PEREIRA, M.V.F. Optimal Stochastic Operations Scheduling of Large Hydroelectric Systems. *Electrical Power & Energy Systems*, v.11, n.3, p. 161-169, jul. 1989.

PEREIRA, M.V.F., PINTO, L.M.V.G. Operation Planning of Large-scale Hydroelectric Systems. *Proc. 8th Power System Computation Cont.*, 1984.

PEREIRA, M.V.F., PINTO, L.M.V.G. Stochastic Optimization of a Multireservoir Hydroelectric System - a Decomposition Approach. *Water Resources*, v. 21, n. 6, 1985.

ROCHA, V.F., GUILHON, L.G.F., CATALDI, M., SILVA, S.B., e PAIVA, L.F.G. A Articulação Entre o ONS e os Agentes de Geração para a Previsão de Vazões Naturais dos Aproveitamentos Hidroelétricos / *Artigo apresentado no Workshop de Previsão de Vazões do ONS, 2007.*

THOMAS, P.T. (2002). Proposta de uma Metodologia de Cobrança pelo Uso da Água vinculada à Escassez. *Dissertação de Mestrado, COPPE/UFRJ, Rio de Janeiro/RJ, 153 p.*

TROUTMAN, B.M. (1979). Some Results in Periodic Autoregression. *Biometrika*, 66(2):219-228

ANEXOS

RBRH – Revista Brasileira de Recursos Hídricos Volume 12 n.3 Jul/Set 2007, 83-92

Aplicação das Técnicas de Mineração de Dados como Complemento às Previsões Estocásticas Univariadas de Vazão Natural: Estudo de Caso para a Bacia do Rio Iguaçu

Marcio Cataldi; Bruno Goulart de Freitas Machado; Simone Borim da Silva;
Luiz Guilherme Ferreira Guilhon
Operador Nacional do Sistema Elétrico (ONS)
cataldi@ons.org.br

Carla da C. Lopes Achão
Planning Engenharia e Consultoria

Recebido: 05/12/06 - revisado: 30/04/07 - aceito: 05/07/07

RESUMO

Este trabalho apresenta os resultados obtidos a partir da aplicação de técnicas de mineração de dados e de Redes Neurais com treinamento bayesiano, para o balizamento da escolha da melhor previsão de vazões naturais do sistema de modelos estocásticos PREVIVAZ. Para aplicação desta técnica, foram utilizadas informações de precipitação observada e prevista, além das vazões naturais verificadas nas últimas semanas que antecederam a previsão. O estudo foi realizado para os aproveitamentos hidrelétricos de Foz do Areia e Salto Santiago na bacia do rio Iguaçu. Os resultados obtidos indicam que a utilização desta ferramenta pode ser uma solução simples e eficaz para a diminuição dos erros de previsão em horizonte semanal de vazão natural nesta bacia.

Palavras-chave: Data Mining; Redes Bayesianas; Modelos Estocásticos; Previsão de vazões.

HISTÓRICO

A partir do I ENCONTRO TÉCNICO SOBRE PREVISÕES DE SÉRIES TEMPORAIS – SCEN, em maio de 1993, algumas empresas do setor elétrico iniciaram o processo de contratação do desenvolvimento do modelo PREVIVAZ junto ao CEPEL. Naquela ocasião, as empresas participantes do GCOI (Grupo Coordenador para a Operação Interligada) financiaram seu desenvolvimento, que ocorreu entre o segundo semestre de 1994, sendo entregue a versão 1.1 em dezembro de 1996. A partir desse momento, o PREVIVAZ foi utilizado e foi se aperfeiçoando até culminarem, em março de 1998, os testes para sua validação com a posterior elaboração do relatório intitulado “Modelo PREVIVAZ - Testes Finais de Validação – Agosto/1998”. Esse relatório foi aprovado pelo GCOI – Resolução RS-G-2946/98 em 06/10/98. Em janeiro de 1999, o modelo foi implantado para todas as bacias do Sistema Interligado Nacional pelo ONS e passou a ser utilizado no Programa Mensal de Operação Eletroenergética (PMO) a partir de fevereiro de 1999. Em 11/04/2000, o ONS apresentou uma comparação entre o modelo PREVIVAZ (modelo estocástico em base mensal utili-

zado até então pela ELETROBRÁS) durante todo o ano de 1999, e o modelo PREVIVAZ (Horizonte Semanal), tornando-se um consenso a decisão de se substituir a previsão semanal, até então obtida pela desagregação da vazão mensal do modelo PREVIVAZ, pela previsão de vazões semanais calculadas pelo modelo PREVIVAZ, a partir do PMO de maio de 2000.

O PREVIVAZ é um modelo de previsão de vazões médias semanais constituído de um conjunto de alternativas de metodologias para previsão de vazões para um horizonte de até seis semanas, em base estocástica, utilizando combinações de modelos estatísticos estacionários ou periódicos, com diferentes métodos de estimação de parâmetros e diferentes tipos de transformações.

APRESENTAÇÃO DO PROBLEMA

As metodologias estocásticas contidas no modelo PREVIVAZ [CEPEL, 2004] contemplam os modelos autoregressivos e de médias móveis, com estrutura estacionária ou periódica, ou seja, os modelos AR(p) e PAR(p), com “p” de até ordem 4, e

Aplicação das Técnicas de Mineração de Dados como Complemento às Previsões Estocásticas Univariadas de Vazão Natural: Estudo de Caso para a Bacia do Rio Iguaçú

PARMA(p,q) e ARMA (p,q), com "p" de até ordem 3 e q de ordem 1. As transformações podem ser logarítmica, Box & Cox ou sem transformação [Guilhon, 2003]. Os métodos de estimação de parâmetros se baseiam no método da máxima verossimilhança e são o método dos momentos, o método de regressão simples e o de regressão em relação à origem das previsões.

O PREVIVAZ divide o histórico em duas metades, estimando, para cada semana, os parâmetros de todos os modelos para a primeira metade, e verificando o erro médio quadrático da previsão com os dados da segunda metade do histórico, conforme a equação (1) a seguir.

$$\sqrt{\frac{\sum_{i=1}^N (X_{prev}^i - X_{obs}^i)^2}{N}} \quad (1)$$

onde,

X_{prev}^i - Vazão prevista no instante i.

X_{obs}^i - Vazão observada no instante i.

N - número total de semanas da metade do histórico considerada.

Em seguida, de forma análoga, o PREVIVAZ estima os parâmetros de todos os modelos para cada semana da segunda metade da série e verifica o erro médio quadrático da previsão com os valores da primeira metade do histórico. Calcula-se então a média dos erros médios quadráticos de cada modelo para as duas metades do histórico, e é feita uma ordenação de modo a escolher aquela metodologia que apresente o menor valor médio do erro médio quadrático.

Após a escolha do melhor modelo, o PREVIVAZ estima novamente os parâmetros, considerando agora todas as semanas do histórico e passa a utilizar este modelo com os novos parâmetros calculados para cada semana específica.

Esta modelagem, no entanto, não incorpora informações de precipitação na bacia, sejam estas de precipitação observada ou prevista, as quais são fundamentais na composição da vazão natural afluyente.

Neste artigo será mostrado que é possível melhorar o desempenho do sistema PREVIVAZ, mantendo sua característica univariada, a partir da utilização de um critério de seleção capaz de balizar a escolha do melhor modelo deste sistema. Esta escolha ocorre dentro de uma faixa de vazões previstas por um sistema auxiliar, que incorpora outras informações, tais como os dados previstos e observados de precipitação na bacia.

Desta forma, buscou-se por meio da utilização das técnicas de mineração de dados (*Data Mining*), uma classificação para as previsões de vazões naturais de dois aproveitamentos localizados na bacia do rio Iguaçú, a saber: UHE Foz do Arcaia e Salto Santiago.

Para tal, utilizou-se o aplicativo WEKA¹ *Data Mining - Waikato Environment for Knowledge Analysis, software* de domínio público e grande portabilidade, implementado em linguagem Java, que possibilita a aplicação de um grande número de tecnologias distintas para o estudo de classificação.

Dentre as técnicas empregadas pelo *software* WEKA para a previsão das faixas de vazão natural, foram testadas aquelas baseadas em Redes Neurais (RN) e Inteligência Artificial (IA), tais como árvores de decisão ID3 e J48; RN do tipo *Multi Layer Perceptron* e *Lazy*, classificadores baseados em regras de associação, além de RN com treinamento bayesiano e algoritmo de automatização de procura baseado na técnica *Hill-Climbing* (subida da encosta ou gradiente). Esta última técnica foi a que apresentou, em todos os experimentos, os melhores resultados. Uma breve descrição do classificador bayesiano disponível no *software* WEKA, utilizado neste estudo, será apresentada a seguir.

TEORIA BAYESIANA

Esse teorema tem como premissa que os itens e relações de interesse são a manifestação de leis de distribuição de probabilidade. É, portanto, uma abordagem essencialmente quantitativa, olhando para o problema como a escolha da melhor hipótese de um espaço de hipóteses, ou seja, aquela que é mais coerente com os dados do problema. [Friedman et al. 1997]

O teorema de Bayes, sendo um dos resultados mais importantes da teoria das probabilidades, é também o princípio fundamental da aprendizagem bayesiana.

Esse teorema pode ser resumido como:

Teorema: Se $\{A_1, A_2, \dots, A_m\}$ é uma partição do espaço de resultados e B um acontecimento qualquer, com $P(B) > 0$, e para cada i $P(A_i) > 0$, então:

$$P(A_i | B) = \frac{P(A_i)P(B | A_i)}{\sum_{i=1}^m P(A_i)P(B | A_i)} \quad (2)$$

¹ <http://www.cs.waikato.ac.nz/ml/weka/>

$i \in \{1, \dots, m\}$

onde:

$P(A)$ é a probabilidade de ocorrência do acontecimento A

$P(A | B)$ é a probabilidade de A condicionada por B , definida por $P(A \cap B) / P(B)$

De uma maneira geral podemos entender a $P(A | B)$ como sendo a probabilidade do evento A ocorrer, tendo em vista que o evento B já ocorreu.

Uma consequência imediata deste teorema pode ser aplicada para dois acontecimentos A e B , tais $P(A) > 0$ e $P(B) > 0$. Neste caso pode-se assumir então que:

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)} \quad (3)$$

Em um treinamento de redes bayesianas, supomos que o nosso conjunto de dados (instâncias de treino) é designado por D , então pelo teorema anterior, temos uma forma de calcular a probabilidade de ocorrência de uma hipótese h , tendo por base os dados do treinamento, através da seguinte relação:

$$P(h | D) = \frac{P(D | h)P(h)}{P(D)} \quad (4)$$

Nesse caso a probabilidade $P(h | D)$ é a denominada “probabilidade à posteriori” de ocorrência de um evento em h , dado que tenha ocorrido um determinado evento dentro do conjunto de dados D . O termo $P(h)$ é a “probabilidade à priori” da hipótese h . A probabilidade “à priori” é a probabilidade de ocorrência não condicionada ao treinamento, e somente ao conjunto de dados que compõe o treinamento. Ela é calculada avaliando-se a probabilidade inicial de ocorrência de cada classe dentro do conjunto de treinamento.

Considerando que temos um espaço de hipóteses possíveis H , então em uma rede bayesiana, pretende-se determinar qual a melhor hipótese “à posteriori”, levando-se em consideração o conjunto de dados observados D . Se interpretarmos a melhor hipótese como sendo a mais provável, atendendo ao conjunto de dados observados D , isto é, a hipótese com melhor valor de “probabilidade à posteriori”, então o valor procurado, deve ser:

$$h_{MAP} = \arg \max P(h | D) \quad (5)$$

onde $h \in H$.

Na equação 5, então, podemos obter o valor de maior probabilidade de ocorrência para um evento “à posteriori” (h_{MAP}), levando-se em consideração o treinamento bayesiano de probabilidades.

O classificador bayesiano irá criar um conjunto de tabelas de probabilidade organizadas em formato árvore, unidas através de diversos nós que formam um conjunto acíclico de busca, conforme apresentado no exemplo da Figura 1. Nesta figura, os nós representam as variáveis de domínio e os arcos as relações de dependência probabilística direta entre as variáveis que as conectam.

A probabilidade de cada tabela seguirá a configuração da rede bayesiana, de modo que o nosso conjunto de variáveis D seja formado por k variáveis, onde $D = \{x_1, \dots, x_k\}$, com $k > 1$. A probabilidade bayesiana (P_{BS}) de ocorrência de cada tabela será então [Bouckaert 2004]:

$$P_{BS} = \{P(d | pa(d)) | d \in D\} \quad (6)$$

onde $pa(d)$ é a probabilidade de cada subconjunto de d que compõe a tabela.

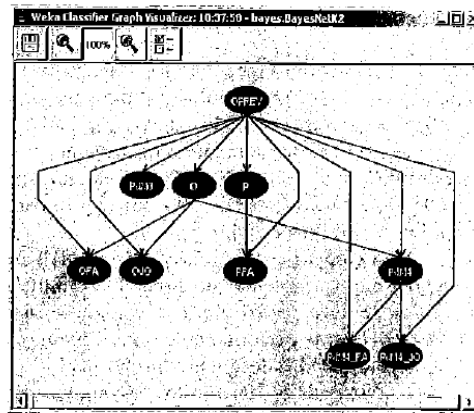


Figura 1 - Exemplo de “árvore de probabilidades bayesianas” para o trecho da bacia do rio Iguaçu a montante da UHE Salto Santiago.

A maior probabilidade bayesiana, ou seja, a classe mais provável de ocorrer, será encontrada através da busca acíclica realizada em todas as tabelas das k variáveis. Essa busca é feita utilizando-se o algoritmo “Hill Climbing”. Maiores detalhes podem ser encontrados em [Buntine 1996].

Aplicação das Técnicas de Mineração de Dados como Complemento às Previsões Estocásticas Univariadas de Vazão Natural: Estudo de Caso para a Bacia do Rio Iguaçu

Vale ressaltar que para a solução da equação 6, o classificador bayesiano necessita de três "hiperparâmetros", sendo que somente um deles, o hiperparâmetro α , é ajustável na versão de classificador bayesiano disponível no Weka. De acordo com o valor atribuído a esse parâmetro, será determinado o peso que cada tabela de probabilidades terá na escolha da classe de maior probabilidade de ocorrência. Maiores detalhes sobre o classificador bayesiano podem ser encontrados em Witten e Frank (2000).

DESCRIÇÃO E JUSTIFICATIVA DOS DADOS

Com o objetivo de obter um desempenho satisfatório nos testes com o *software* WEKA, foram testadas inúmeras configurações, a partir da combinação das variáveis de vazão (observada e prevista) e precipitação (observada e prevista). Cabe notar que as variáveis de precipitação previstas incluídas neste estudo correspondem àquelas geradas pelo modelo ETA [Black, 1994] do Centro de Previsão do Tempo e Estudos Climáticos - CPTEC, referentes aos aproveitamentos da bacia do rio Iguaçu (Foz do Areia e Salto Santiago).

Os dados utilizados compreenderam o período entre 1994 e 2003. A fim de validar a metodologia proposta foram testados dois anos: 2002 e 2003. O período de treinamento correspondente ao teste de 2002 foi de 1994 a 2001. Para se testar o ano de 2003, utilizou-se o período de treinamento compreendido entre 1994 e 2002.

Para compor a precipitação observada para o trecho a montante de Foz do Areia, foram utilizados dados de doze postos pluviométricos, ao passo que para o trecho a montante de Salto Santiago, foram utilizados dados de apenas dois postos.

A distribuição dos postos pluviométricos, dos pontos de grade do modelo ETA e das principais usinas da bacia do rio Iguaçu pode ser observada na Figura 2.

É importante notar que para cada aproveitamento obteve-se o melhor desempenho no *software* de *Data Mining* utilizando uma determinada configuração, isto é, a partir do uso de diferentes conjuntos de variáveis, observando a dependência de cada uma delas em relação às demais variáveis estudadas. A definição da configuração ótima dos aproveitamentos - que possibilitou o melhor desempenho nos testes com *Data Mining* - resultou da apli-

cação de um conjunto de testes, quando foram avaliadas várias combinações dentre as variáveis disponíveis.

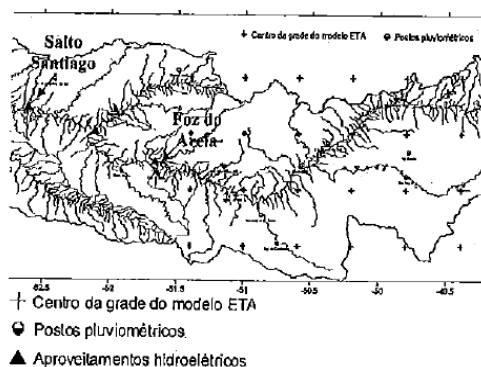


Figura 2 - Topografia da bacia do rio Iguaçu desde a cabeceira até a UHE Salto Santiago.

A escolha das variáveis foi feita através de análises estatísticas padrões nos estudos de *Data Mining*, a saber: matriz correlação, dendogramas e Análise de Componentes Principais (ACP). As faixas de classe destas variáveis foram escolhidas a partir da análise da curva de permanência de cada uma delas, com o objetivo de se obter classes que pudessem caracterizar principalmente períodos de cheia e de recessão, períodos estes onde os erros do modelo PREVIVAZ são maiores, devido principalmente ao fato do modelo não incorporar as informações de precipitação.

Serão apresentadas a seguir algumas figuras utilizadas nas análises realizadas para o trecho da bacia do rio Iguaçu a montante da UHE Foz do Areia. Cabe ressaltar que análises semelhantes foram realizadas para que se pudesse obter a composição final das variáveis apresentadas nas figuras 4 a 7.

A partir dessas análises e da realização de inúmeros testes com o conjunto de treinamento, obteve-se a configuração que possibilitou o melhor desempenho para Foz do Areia. As variáveis descritas na tabela 1.

A precipitação média foi obtida pelo método de Kriging. A metodologia de cálculo, bem como a análise e discussão de seu uso foram abordadas por Cataldi e Machado (2004).

Tabela 1 - Variáveis utilizadas para a UHE Foz do Areia.

Sigla	Significado
Q_1	Vazão natural média observada na semana anterior à semana da previsão (m³/s)
Q	Vazão natural média observada na semana da previsão (m³/s)
QUV	Vazão natural média observada no posto de União da Vitória na semana da previsão (m³/s)
QPUV	Vazão natural média prevista para o posto de União da Vitória na semana seguinte à semana da previsão (m³/s)
Q_PREV ou QMI	Vazão natural média prevista para a semana seguinte à semana da previsão (m³/s)
P**	Precipitação diária acumulada em 7 dias observada na semana da previsão (mm).
PdM4*	Previsão de Precipitação acumulada para os próximos 4 dias a partir da data da previsão (mm)
PdM3*	Previsão de Precipitação acumulada do 4º ao 7º dia a partir da data da previsão (mm)

*Nos testes realizados considerando a previsão "perfeita" de precipitação essas variáveis foram compostas pelos valores de precipitação observada na semana a ser prevista.
 ** Nos testes que utilizaram a previsão de precipitação do modelo ETA, os últimos 3 dias dessa variável foram compostos com a previsão de precipitação, visando completar a semana operativa.

Na Figura 3 pode-se observar a correlação entre as variáveis usadas no estudo. O dendograma, que representa os agrupamentos obtidos, é mostrado na Figura 4. As análises de componentes principais por variável e por período do histórico são apresentadas nas figuras 5 e 6.

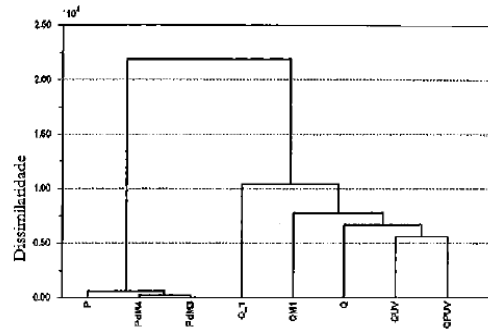


Figura 4 -- Dendograma.

A análise das figuras 5 e 6 nos faz perceber que o conjunto de componentes principais formado somente pelas vazões, em geral, só é capaz de representar todo o conjunto de dados nos casos onde ocorrem pequenas variações entre as vazões naturais semanais.

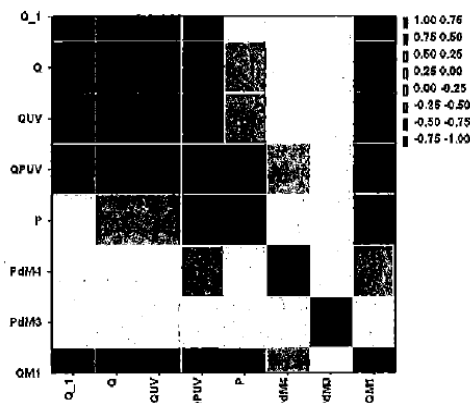


Figura 3 – Matriz correlação.

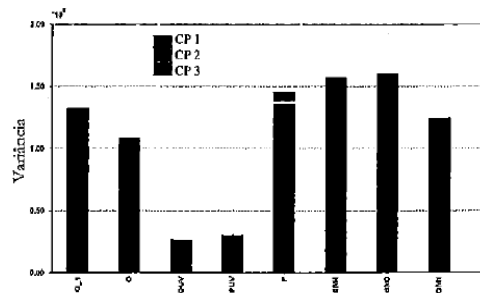


Figura 5 – Análise de Componentes Principais por variável.

As faixas que possibilitaram o melhor desempenho do classificador bayesiano foram obtidas a partir de curvas de distribuição de probabilidades (Figura 7), e estão apresentadas na figura 8.

Aplicação das Técnicas de Mineração de Dados como Complemento às Previsões Estocásticas Univariadas de Vazão Natural: Estudo de Caso para a Bacia do Rio Iguaçu

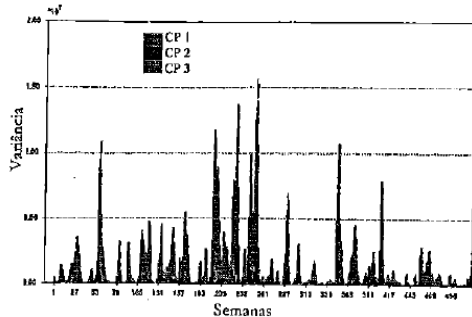


Figura 6 – Distribuição das Componentes Principais ao longo de toda a série histórica.

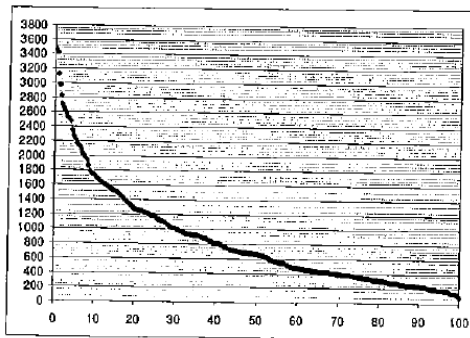


Figura 7 – Curva de permanência em porcentagem das vazões naturais totais a UHE Foz do Areia (m³/s)

Q (m ³ /s)	Pob 4d (mm)	Pob 3d (mm)	Q 1 (m ³ /s)
Q ≤ 160	PdM4 ≤ 5	PdM3 ≤ 5	Q 1 ≤ 160
160 < Q ≤ 300	5 < PdM4 ≤ 20	5 < PdM3 ≤ 15	160 < Q 1 ≤ 300
300 < Q ≤ 470	20 < PdM4 ≤ 40	15 < PdM3 ≤ 30	300 < Q 1 ≤ 470
470 < Q ≤ 650	40 < PdM4 ≤ 60	30 < PdM3 ≤ 50	470 < Q 1 ≤ 650
650 < Q ≤ 950	PdM4 > 60	PdM3 > 50	650 < Q 1 ≤ 950
950 < Q ≤ 1220			950 < Q 1 ≤ 1220
Q > 1220			Q 1 > 1220
P (mm)	PdM4 (mm)	PdM3 (mm)	OPREV (m ³ /s)
P ≤ 5	PdM4 ≤ 5	PdM3 ≤ 5	OPREV ≤ 160
5 < P ≤ 20	5 < PdM4 ≤ 20	5 < PdM3 ≤ 15	160 < OPREV ≤ 300
20 < P ≤ 40	20 < PdM4 ≤ 40	15 < PdM3 ≤ 30	300 < OPREV ≤ 470
40 < P ≤ 60	40 < PdM4 ≤ 60	30 < PdM3 ≤ 50	470 < OPREV ≤ 650
P > 60	PdM4 > 60	PdM3 > 50	650 < OPREV ≤ 950
			950 < OPREV ≤ 1220
			OPREV > 1220

Figura 8 – Faixas de vazões e precipitação utilizadas para a UHE Foz do Areia

Observando a Figura 3 podemos perceber que a previsão de precipitação dos últimos 3 dias da semana a ser prevista (PDM3) foi a variável que apresentou menor correlação com a vazão da semana

a ser prevista (QM1), porém, ela contém informações importantes nos casos onde a variação nos valores de vazão entre as semanas observada e prevista é grande, como pode ser observado na ACP (Figura 5). Observa-se na Figura 4, como era de se esperar, dois grandes grupos formados pelas variáveis analisadas: um formado pelas variáveis de precipitação e outro pelas variáveis de vazão natural.

A ACP apresentada na Figura 5 indica que com os 3 Componentes Principais (CP) encontrados, cerca de 98 % da série poderia ser explicada. A ACP é uma técnica estatística que pode ser utilizada para redução do número de variáveis e para fornecer uma visão estatisticamente privilegiada do conjunto de dados. A ACP consiste em reescrever as variáveis originais em novas variáveis denominadas Componentes Principais - CP, através de uma transformação de coordenadas. Os CP são as novas variáveis geradas através de uma transformação matemática especial realizada sobre as variáveis originais. Cada CP é uma combinação linear de todas as variáveis originais. Nestas combinações, cada variável terá uma importância ou peso diferente.

As variáveis podem guardar entre si correlações que são suprimidas nos CP, ou seja, os componentes principais são ortogonais entre si. Deste modo, cada componente principal traz uma informação estatística diferente das outras. As variáveis originais têm a mesma importância estatística, enquanto que os componentes principais têm importância estatística decrescente, ou seja, os primeiros componentes principais são tão mais importantes que podemos em alguns casos até desprezar os demais.

Vale ressaltar na análise da Figura 5 que o CP (1) é formado pela transformação linear de parte de todas as variáveis do subconjunto de dados, e sozinho é capaz de explicar cerca de 87% dos eventos. Os CP (2) e (3) são formados apenas pela transformação linear dos dados de vazão natural (com defasagem temporal). Podemos observar na Figura 6 que o CP (1) é capaz de explicar os eventos de grande variação entre as semanas (observada e prevista). Já os componentes (2) e (3) juntos não são capazes de identificar essas grandes variações entre as vazões naturais semanais. Essa análise é um indicio de que nessas situações as informações de precipitação, tanto observadas quanto previstas, são de fundamental importância para o conhecimento das vazões futuras em complemento ao conhecimento das vazões passadas.

As outras variáveis e configurações analisadas foram: vazões naturais semanais observadas com defasagem de 2 e 3 semanas, previsão de precipitação para 7 dias agrupados e precipitação observada

acumulada nos últimos 7 dias dividida em dois conjuntos, com 4 e 3 dias. Essas variáveis/configurações foram retiradas do estudo por não apresentarem relevância nas análises estatísticas e/ou por comprometerem o desempenho do classificador bayesiano, sendo que em muitas vezes, o próprio modelo de classificação excluía algumas dessas variáveis/configurações. Esta exclusão se deu pelo fato dessas variáveis/configurações não apresentarem uma relação de causa e efeito significativa, do ponto de vista probabilístico, com a variável a ser prevista.

Um resumo do processo e das tecnologias envolvidas para a criação deste tipo de modelo, que foi batizado como Modelo de Previsão de Classes de Vazão (MPCV), pode ser vislumbrado no fluxograma apresentado na figura 9.

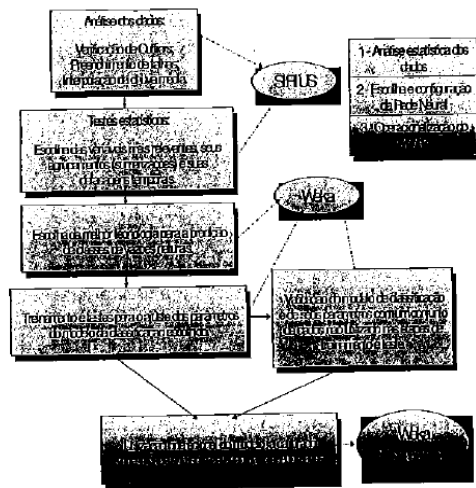


Figura 9 – Fluxograma com as etapas de análise estatística, escolha e configuração da tecnologia de Rede Neural e operacionalização do MPCV.

RESULTADOS OBTIDOS NOS TESTES DO MODELO

Os resultados apresentados a seguir foram obtidos no desenvolvimento do MPCV.

Assim, partindo de informações de precipitação observada e prevista, além das vazões verificadas nas últimas semanas que antecederam a previsão, foram estabelecidas faixas para as variáveis inseridas no classificador bayesiano, de forma a se ter uma classificação associada a cada previsão.

A partir das faixas de vazões semanais previstas pelo classificador bayesiano, interferiu-se na escolha do melhor modelo do PREVIVAZ em todas as semanas em que a sua previsão (aquela realizada pelo modelo melhor classificado pelo sistema PREVIVAZ) se apresentou fora da faixa sugerida. Nestes casos, buscou-se a previsão do modelo melhor posicionado dentro do ranking dos modelos utilizados pelo sistema PREVIVAZ, constantes em seu relatório de resultados, e que estivesse dentro da faixa de vazão determinada pelo modelo de balizamento desenvolvido no software WEKA.

Cabe ressaltar que, para algumas semanas, onde o melhor modelo escolhido pelo PREVIVAZ estava fora da faixa determinada pelo WEKA, o critério de busca descrito acima não obteve sucesso, tendo em vista o fato de as previsões dos modelos do PREVIVAZ estarem fora da faixa determinada. A alternativa encontrada para contornar este problema foi buscar, dentro do ranking dos modelos, aquele cuja previsão mais se aproximava da faixa prevista pelo classificador bayesiano. Esta metodologia foi testada para dois aproveitamentos da bacia do rio Iguaçu, a saber: Foz do Areia e Salto Santiago, considerando-se dados de previsão perfeita e real de precipitação. Conforme elucidado anteriormente, os dados de previsão real de precipitação foram gerados pelo modelo ETA do GPTEC.

Desta forma, para cada aproveitamento e para cada ano de teste (2002 e 2003), foram obtidos dois conjuntos de resultados: um considerando previsão perfeita de precipitação (obtida através da interpolação dos dados observados nos postos pluviométricos) e o outro considerando a previsão real de precipitação, como será apresentado a seguir.

Tabela 3 - Resumo dos erros médios quadráticos das previsões de vazão natural média semanal relativos aos anos de 2002 e 2003.

UHE	Ano	Previsão de Precipitação	Previsão (%)	Previsão com Data Mining (%)
Foz do Areia	2002	Perfeita	50,0	22,0
		Real		27,6
	2003	Perfeita	50,0	35,4
		Real		36,9
Salto Santiago	2002	Perfeita	35,1	24,3
		Real		27,6
	2003	Perfeita	35,1	27,6
		Real		29,4

Aplicação das Técnicas de Mineração de Dados como Complemento às Previsões Estocásticas Univariadas de Vazão Natural: Estudo de Caso para a Bacia do Rio Iguaçu

Nas tabelas 3 e 4 estão disponíveis os principais resultados desse trabalho. Na tabela 3 são apresentadas as comparações dos erros médios quadráticos relativos das previsões de vazão natural média semanal do sistema Previvaz para os anos de 2002 e 2003, com e sem a utilização da metodologia desenvolvida nesse trabalho. Na tabela 4 são apresentadas as comparações somente para as semanas onde foi possível utilizar o critério de seleção proposto. Essas previsões são realizadas pelo Operador Nacional do Sistema Elétrico – ONS uma vez por semana e com uma antecedência de 3 a 4 dias em relação ao início da semana a ser prevista.

Tabela 4 - Resumo dos erros médios quadráticos das previsões de vazão natural média semanal, relativos às semanas em que a aplicação da metodologia interferiu no resultado.

UHE	Ano	Previsão de Precipitação	Previvaz (%)	Previvaz com Data Mining (%)
Foz do Areia	2002	Perfeita	57,3	26,6
		Real	63,3	31,1
	2003	Perfeita	57,3	29,2
		Real	63,3	38,1
Salto Santiago	2002	Perfeita	39,4	20,0
		Real	40,1	20,0
	2003	Perfeita	39,4	25,0
		Real	37,0	24,1

MODELO OPERACIONAL

A aplicação do MPCV de modo operacional na rotina de previsão semanal de vazões naturais do ONS, para a bacia do rio Iguaçu, se deu a partir do ano de 2006 com a autorização da Agência Nacional de Energia Elétrica - ANEEL. Desde então, foram avaliados os valores iniciais previstos pelo modelo PREVIVAZ com o resultado do balizamento sugerido pela técnica de mineração de dados. Ambos os valores previstos foram armazenados e comparados com os valores observados na bacia. Nas figuras 10 e 11, pode-se analisar a evolução da previsão semanal de vazões naturais realizadas com o PREVIVAZ e com o MPCV para as UHE Foz do Areia e Salto Santiago, respectivamente. Essas previsões foram comparadas com os dados de vazão natural total observada nesses trechos até o fechamento da primeira semana de outubro de 2006.

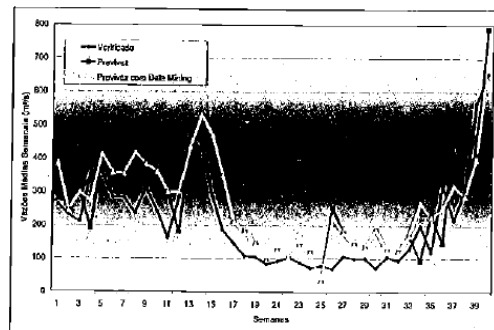


Figura 10 – Acompanhamento da previsão semanal de vazões naturais para a UHE Foz do Areia (m³/s).

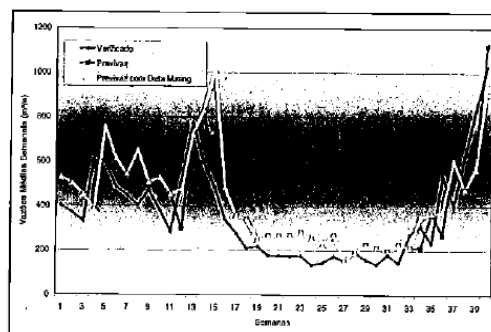


Figura 11 – Acompanhamento da previsão semanal de vazões naturais para a UHE Salto Santiago (m³/s).

De modo geral, verifica-se que para vazões baixas, como as que ocorreram em meados de setembro de 2006 na bacia do rio Iguaçu, o uso do MPCV, na maioria dos casos, não modifica os valores previstos pelo PREVIVAZ.

Em relação ao erro médio obtido entre os valores previstos e observados, nota-se que para o aproveitamento Foz do Areia, no período de 40 semanas, houve 11 ocasiões onde o MPCV indicou uma mudança da faixa de vazões semanais previstas. Deste total, em 8 ocasiões, a alteração resultou numa melhora da previsão e, por conseguinte, na diminuição do erro absoluto entre o valor esperado e o verificado. Em 3 ocasiões a indicação da nova faixa pelo MPCV implicou num afastamento maior da previsão em relação ao valor observado na bacia. Na figura 12 é apresentado a diferença entre os erros médios quadráticos calculados nas ocasiões onde houve alteração do patamar inicial previsto pelo PREVIVAZ. Destaca-se que na 35ª semana, foi obtida

uma melhora significativa de cerca de 80% da previsão quando utilizada a técnica de mineração de dados.

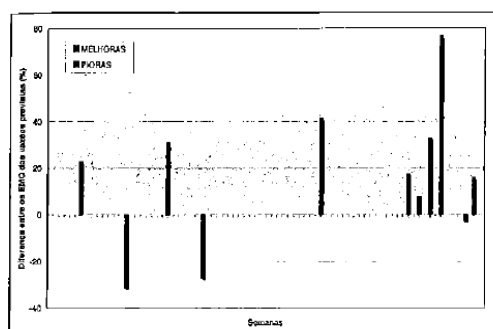


Figura 12 – Diferença entre o erro Médio Quadrático das vazões previstas pelos modelos Previvaz e MPCV para a UHE Foz do Arica (%)

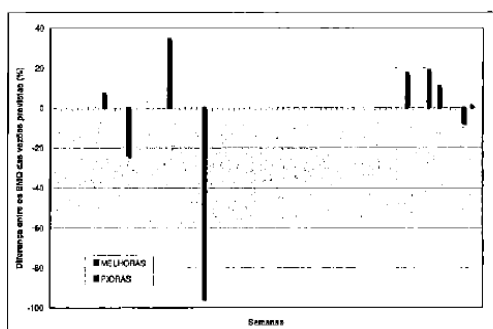


Figura 13 – Diferença entre o erro Médio Quadrático das vazões previstas pelos modelos Previvaz e MPCV para a UHE Salto Santiago (%)

Em relação ao aproveitamento Salto Santiago, como pode ser observado na figura 13, houve um total de 9 mudanças de faixa de vazão prevista pelo MPCV na mesma amostra de 40 semanas do ano de 2006. Deste total, em 6 ocasiões houve aproximação do valor previsto com o verificado e, em 3 ocorrências, houve maior dispersão entre os mesmos. Destaca-se um erro acentuado da previsão na 15ª semana, após a alteração da previsão pelo MPCV. Este tipo de erro na previsão de vazões do MPCV geralmente está associado a grandes desvios entre a precipitação prevista pelo modelo ETA e aquela ocorrida na bacia. Pequenas variações entre os totais de precipitação observada e prevista geral-

mente não implicam em grandes erros nas previsões de vazão do MPCV, já que a previsão de precipitação é inserida no modelo através de faixas de valores, tal como ilustrado na figura 8.

CONCLUSÕES

Este estudo demonstrou que a aplicação das técnicas de *Data Mining* pode se apresentar como uma importante ferramenta para análise de variáveis de interações não lineares, como aquelas que compõem a estrutura dos fenômenos hidrológicos. Dentre as técnicas estudadas, os classificadores bayesianos foram os que apresentaram melhor destreza na predição das classes de vazões naturais, para a maioria dos casos analisados.

Nos anos escolhidos para a validação da metodologia, os resultados obtidos com a interferência do classificador bayesiano melhoraram o índice de acerto das previsões do modelo PREVIVAZ em todas as situações, inclusive naquelas onde foi utilizada a previsão de precipitação do modelo ETA. Destaca-se a previsão para o ano de 2003 no aproveitamento de Foz do Arica, onde os erros foram reduzidos pela metade nas semanas onde o classificador interferiu diretamente no resultado. Cabe ressaltar que a alternativa apresentada nesse estudo é de simples aplicação e possui um custo computacional extremamente baixo.

Em relação à utilização operacional do MPCV, foram observados melhores resultados nas previsões de vazão para UHE Foz do Arica. Esse comportamento pode estar associado a melhor distribuição e cobertura pluviométrica nessa região.

A modelagem estocástica univariada contida no modelo PREVIVAZ, por muitas vezes, dificulta a previsão de mudanças no comportamento das vazões entre uma semana e outra, o que ocasiona um efeito que é conhecido como "efeito sombra". A inserção das variáveis de precipitação como complementação às previsões do PREVIVAZ, se mostrou, ao longo deste estudo, como uma alternativa relativamente eficiente na minimização deste tipo de erro sistemático. Entretanto em muitos casos essa correção não poder ser realizada de forma mais efetiva, devido ao fato de que em algumas semanas, nenhuma das previsões realizadas pelos modelos do sistema PREVIVAZ estar dentro da faixa prevista pelo MPCV.

Por outro lado, nos casos onde a incerteza da previsão de precipitação induziu a previsão do MPCV à faixas de vazões muito distintas dos valores

Aplicação das Técnicas de Mineração de Dados como Complemento às Previsões Estocásticas Univariadas de Vazão Natural: Estudo de Caso para a Bacia do Rio Iguaçu

verificados, os resultados do PREVIVAZ, cuja tendência é se aproximar da média de longo termo, devido a sua natureza estocástica, fizeram com que, especificamente nesses casos, os erros não aumentassem de forma significativa, minimizando o erro associado à inclusão de previsões de precipitação equivocadas no processo de previsão de vazões.

Para dar continuidade a este trabalho esta metodologia está sendo replicada para a bacia do rio Uruguai.

over the weeks that precede the actual forecast target for the Foz do Areia and Salto Santiago hydroelectric plants located in the Iguaçu River Basin. The results obtained indicate that the using these tools can provide a simple and efficient solution to reduce natural inflow forecast errors on a weekly forecast basis for the Iguaçu River Basin

Keywords: Data Mining; Bayesian Networks; Stochastic Models; Inflow Forecasts.

REFERÊNCIAS

- Black T.L., 1994: NMC Notes: The New NMC mesoscale Eta model: Description and forecast examples. *Weather and Forecasting*, 9, 256-278.
- Bouckaert, R. B., "Bayesian Network Classifiers in WEKA", *Internal Notes*, 2004
- Bunline, W.L. "A guide to the literature on learning probabilistic networks from data", *IEEE Transactions on Knowledge and Data Engineering*, 8:195-210, 1996.
- Cataldi, M., Machado, C.O., "Avaliação da previsão de precipitação utilizando a técnica de Downscale do modelo ETA e suas aplicações no setor elétrico", *XIII Congresso de Meteorologia*, 2004.
- CEPEL, "Modelo de Previsão de Vazões Semanais Aplicado ao Sistema Hidroelétrico Brasileiro – Modelo Previvaz", Manual de Referência, 2004.
- Friedman, N., Geiger, D., Goldszmidt, M., "Bayesian network classifiers". *Machine learning*, 29:131-163, 1997.
- Guilhon, L.G.F. "Modelo Heurístico de Previsão de Vazões Naturais Médias Semanais Aplicado à Usina de Foz do Areia", - Dissertação de Mestrado, UFRJ, 2003
- Witten, I.H., Frank, E., "Data Mining: Practical machine learning tools and techniques with Java implementations", *Morgan Kaufmann Publishers*, 2000.

Use of Data Mining Techniques to Complement Univariate Stochastic Forecasts of Natural Flow Studied Case by Case in the Iguaçu River Basin

ABSTRACT

This paper presents the results obtained by using software in the public domain that, through Data Mining and Neural Networks with Bayesian training, can lay the foundation to select the most appropriate natural inflow forecast used in the PREVIVAZ stochastic modeling system. This technique utilizes forecast and observed information on precipitation, as well as natural inflow data recorded

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)