

**PREVIVAZ – MODELO DE PREVISÃO DE
VAZÕES SEMANAIS AFLUENTES AOS
APROVEITAMENTOS HIDROELÉTRICOS DO
SISTEMA BRASILEIRO**

Versão 6.1

MANUAL DE METODOLOGIA

Dezembro 2015

Índice

| | |
|---|-----------|
| 1. INTRODUÇÃO | 1 |
| 2. MÉTODOS DE PREVISÃO DE VAZÕES FLUVIAIS | 4 |
| 3. MODELAGEM ESTOCÁSTICA DE SÉRIES TEMPORAIS | 6 |
| 3.1 CARACTERÍSTICAS ESTATÍSTICAS DA SÉRIE HISTÓRICA | 6 |
| 3.2 AGRUPAMENTO DA ESTRUTURA DE CORRELAÇÃO | 8 |
| 3.3 MODELO AUTO-REGRESSIVO - AR(p) | 8 |
| 3.3.1 DESCRIÇÃO DO MODELO | 8 |
| 3.3.2 ESTIMAÇÃO DOS PARÂMETROS | 11 |
| 3.4 MODELO AUTO-REGRESSIVO MÉDIA-MÓVEL - ARMA(p,1) | 11 |
| 3.4.1 DESCRIÇÃO DO MODELO | 11 |
| 3.4.2 ESTIMAÇÃO DOS PARÂMETROS | 15 |
| 3.5 MODELO AUTO-REGRESSIVO - PAR(p) | 15 |
| 3.5.1 DESCRIÇÃO DO MODELO | 15 |
| 3.5.2 ESTIMAÇÃO DOS PARÂMETROS PELO MÉTODO DOS MOMENTOS | 18 |
| 3.5.3 ESTIMAÇÃO DOS PARÂMETROS POR REGRESSÃO LINEAR COM RESPEITO À ORIGEM DE PREVISÃO | 18 |
| 3.6 MODELO AUTO-REGRESSIVO MÉDIA-MÓVEL PERIÓDICO - PARMA(p,1) | 19 |
| 3.6.1 DESCRIÇÃO DO MODELO | 19 |
| 3.6.2 ESTIMAÇÃO DOS PARÂMETROS PELO MÉTODO DOS MOMENTOS | 22 |
| 3.6.3 ESTIMAÇÃO DOS PARÂMETROS POR REGRESSÃO LINEAR | 23 |
| 3.6.4 ESTIMAÇÃO DOS PARÂMETROS POR REGRESSÃO LINEAR COM RESPEITO À ORIGEM DE PREVISÃO | 24 |
| 4. TRANSFORMAÇÕES APLICADAS ÀS SÉRIES DE AFLUÊNCIAS | 25 |

| | |
|---|-----------|
| 5. PADRONIZAÇÃO DAS SÉRIES DE AFLUÊNCIAS | 27 |
| 6. INTERVALO DE CONFIANÇA DAS PREVISÕES | 28 |
| 7. ESCOLHA DO MELHOR ALGORITMO DE PREVISÃO | 30 |
| 7.1 APLICAÇÃO DE LIMITES NAS PREVISÕES | 33 |
| 7.1.1 APLICAÇÃO DE LIMITES NA ESCOLHA DA MELHOR ALTERNATIVA NO PREVIVAZ | 34 |
| 7.1.2 APLICAÇÃO DE LIMITES NA PREVISÃO DO MODELO PREVIVAZ | 37 |
| 8. CONCLUSÕES | 40 |
| 9. REFERÊNCIAS | 41 |

1. INTRODUÇÃO

A geração de energia elétrica no Brasil depende basicamente das vazões que naturalmente afluem aos aproveitamentos hidroelétricos brasileiros distribuídos por doze bacias hidrográficas do país, que hoje consistem em mais de uma centena. Esses aproveitamentos totalizam mais de 90% da capacidade instalada de geração do país. O processo natural de vazões fluviais tem como característica principal a sua inconstância, dependente que é da ocorrência de precipitações, fenômeno sabidamente intermitente e de comportamento sazonal ditado pelo clima. Por exemplo, no sudeste brasileiro, chuvas copiosas tendem a ocorrer com maior frequência no Verão, enquanto que no Inverno, as precipitações, além de pouco frequentes, tendem a serem pouco copiosas, [1].

Devido à irregularidade das vazões fluviais e a necessidade de manter a continuidade do fornecimento de energia elétrica, o sistema de geração de energia elétrica brasileiro, além de contar com usinas termoeletricas de complementação (o que implica em custos de combustíveis), conta também com um conjunto de reservatórios de acumulação, cuja função é essencialmente a regularização dos regimes fluviais, através do armazenamento de água nos períodos de maiores afluências naturais de vazões fluviais, para ser utilizada durante os períodos mais secos, quando são escassas as vazões fluviais. A capacidade de armazenamento hoje disponível permite não só a regularização intra-anual do sistema, como também fornece proteção contra ocorrência de seqüências de anos “secos”, configurando-se a chamada regularização plurianual do sistema.

A coordenação da operação do sistema de reservatórios do setor elétrico em conjunto com a operação do sistema de usinas termoeletricas de complementação, permite o melhor aproveitamento das vazões naturais, evitando o desperdício de água e gastos excessivos com combustíveis. Esta coordenação é feita no âmbito do chamado Planejamento da Operação do Sistema Interligado, executado atualmente no ONS. Vale também ressaltar que, além da função de regularização para efeito de geração de energia elétrica, parte da capacidade dos reservatórios do setor elétrico é utilizada para controle de cheias. Neste caso, reserva-se uma parcela superior do reservatório para ser ocupada somente durante a ocorrência de grandes cheias.

O plano de operação de um sistema hidrotérmico de energia elétrica deve definir para cada instante a geração de cada unidade de tal forma que a demanda seja atendida a um custo mínimo. A existência de reservatórios com capacidade de regularização significativa, torna o problema do planejamento da operação um problema não-separável no tempo, uma vez que qualquer decisão de deplecionamento destes reservatórios influencia a capacidade de geração do sistema no futuro. A capacidade de geração futura do sistema também é influenciada pelas afluições hidrológicas futuras, cuja natureza extrinsecamente aleatória deve ser considerada no planejamento da operação do sistema.

Em geral, o uso de previsões hidrológicas pode melhorar o desempenho da operação do sistema aumentando benefícios e confiabilidade e, reduzindo custos. Potencialmente, o sistema brasileiro de geração de energia elétrica, que contém uma grande capacidade de regularização, pode beneficiar-se largamente do desenvolvimento de melhores modelos de previsão hidrológica.

O planejamento da operação do sistema hidrotérmico brasileiro, devido a sua complexidade, é atualmente feito através de uma hierarquia. No topo desta hierarquia, um planejamento de médio prazo constrói anualmente uma política para ser usada ao início de cada mês para decidir quanto da demanda total de energia do próximo mês será atendida pelo parque térmico e quanto será atendida por cada sub-conjunto de hidroelétricas agrupadas por regiões. A cada mês os blocos de geração hidráulico e térmico definidos pela política de médio prazo são desagregados em metas de geração para cada usina do sistema. O planejamento hidráulico de curto prazo é feito considerando no primeiro mês uma discretização temporal semanal e para cada semana as afluições previstas são obtidas pelo modelo PREVIVAZ [2], para os demais meses do horizonte do estudo considera-se cenários de afluições mensais gerados pelo modelo GEVAZP [3]. Hoje é considerado um horizonte de apenas 2 meses.

O Modelo PREVIVAZ é um modelo de previsão de vazões semanais. Ele tem como objetivo a análise e teste de alternativas metodológicas para a modelagem estocástica das afluições semanais e cálculo em frequência mensal de previsões de até seis semanas a frente. Ele utiliza modelos lineares de séries temporais univariadas, sendo as previsões de afluições a um aproveitamento obtidas usando apenas observações passadas das afluições ao aproveitamento em questão, não sendo, portanto, utilizadas as informações sobre o

comportamento das afluições aos aproveitamentos próximos ou de postos pluviométricos na área de drenagem do aproveitamento em questão.

O item 2 deste relatório faz uma breve discussão do uso de modelos lineares de séries temporais univariadas para a obtenção de previsões de vazões mensais. O item 3 apresenta as equações dos diferentes modelos, assim como os métodos de estimação de parâmetros utilizados pelo PREVIVAZ. O item 4 descreve as transformações dos dados utilizados pelo modelo, enquanto que o item 5 descreve a padronização das afluições adotada pelo PREVIVAZ. O item 6 descreve como é calculado o intervalo de confiança para as previsões. O item 7 descreve o procedimento adotado no PREVIVAZ para escolher qual das combinações: modelos, método de estimação de parâmetros e transformação dos dados, será a adotada para fornecer a previsão de cada semana, assim como seu intervalo de confiança. Neste item foi acrescentada a descrição do novo critério, implementado na versão 5.3, para a redefinição do modelo adotado com base em limites de variação de vazões. Finalmente o item 8 apresenta as conclusões.

2. MÉTODOS DE PREVISÃO DE VAZÕES FLUVIAIS

A previsão de vazões fluviais apresenta alguns desafios naturais. Os modelos de previsão quantitativa de precipitação estão ainda em fase de nascimento e enfrentam dificuldades inerentes à previsão dos fenômenos turbulentos intervenientes na Circulação Geral da Atmosfera. Some-se ainda o desconhecimento dos processos físicos que ocorrem no interior das células de precipitação. Os mais recentes avanços nesta área são os chamados Modelos Regionais de Circulação da Atmosfera onde, as previsões de vento, temperatura, pressão e umidade do ar, de Modelos Gerais de Circulação da Atmosfera, são utilizadas como condições de contorno ao lado de Modelos Digitais de Terreno detalhados da região. Os chamados modelos chuva-vazão conceituais [5], de longa tradição na Hidrologia Aplicada, podem ser usados para a transformação de previsão de precipitações em previsões de vazões. Com efeito, esses modelos têm sido usados pelo setor de Operação de concessionárias do setor elétrico para obtenção de previsões de vazões diárias, partindo-se de previsões quantitativas e precipitação. Uma das grandes dificuldades de sistemas de previsão de vazões baseados em modelos chuva-vazão conceituais é necessidade constante de revisão da calibração e/ou das condições iniciais de volumes armazenados nos reservatórios conceituais do modelo.

Outra alternativa bastante utilizada para obtenção de previsões de vazões fluviais é o uso da parte determinística de modelos de séries temporais ajustados às séries de vazões fluviais como previsão. Neste caso, a previsão toma partido não só do comportamento sazonal das vazões fluviais referido anteriormente, como também considera o importante fenômeno denominado tendência hidrológica. A tendência hidrológica é o termo utilizado para se referir à tendência observada nas séries históricas de vazões fluviais naturais de valores superiores/inferiores ao valor normal do período serem seguidas/precedidas por valores também superiores/inferiores ao valor normal. Ou seja, se numa certa semana, a vazão do rio foi superior ao valor médio para aquela semana, é provável que na semana seguinte a vazão do rio deverá continuar acima da média. Da mesma forma, se na semana a vazão foi inferior à média da semana, na semana seguinte a vazão do rio deverá continuar abaixo da média. A ocorrência de tendência hidrológica tem sido atribuída ao fenômeno da infiltração de parcela do volume precipitado sobre a bacia hidrográfica. A parcela do volume precipitado que

infiltra depende de diversos fatores, dentre eles, o estado de umidade da superfície do terreno. Assim, se na semana passada foi uma semana com bastante precipitação, e digamos, com vazões fluviais altas, esta semana tende a ser de vazões altas, pois a superfície do terreno deve estar úmida e a infiltração será pequena, aumentando o escoamento direto. Além disto, a parcela que infiltrou na semana passada, pode estar chegando ao rio esta semana.

Na terminologia de séries temporais, a tendência hidrológica é conhecida como estrutura de dependência temporal, sendo quantificada pela função de autocorrelação estimada do registro de vazões. Essencialmente, qualquer estrutura de dependência temporal sazonal pode ser reproduzida por modelos de séries temporais lineares do tipo PARMA(p,q), o que faz deste tipo de modelo uma abordagem bastante flexível, e bastante popular para a modelagem estocástica de vazões fluviais [6]. Os modelo PREVIVAZ utiliza modelos lineares do tipo PARMA(p,q), acoplado a diferentes pré-transformações, tipo Box-Cox ou logarítmicas e a diferentes formas de estimação de parâmetros dos modelos, para obtenção de previsões de vazões semanais.

3. MODELAGEM ESTOCÁSTICA DE SÉRIES TEMPORAIS

Desde a década de 60 os modelos auto-regressivos são utilizados em hidrologia e recursos hídricos na modelagem de séries temporais estacionárias e periódicas. Sua aplicação teve início em 1962/63 com trabalhos de Thomas e Fiering [7] e Yevjevich [8]. Na década de 70 a publicação do livro de Box e Jenkins [9] motivou, também, a utilização dos modelos auto-regressivos com médias móveis.

3.1 CARACTERÍSTICAS ESTATÍSTICAS DA SÉRIE HISTÓRICA

Em geral, processos naturais hidrológicos tais como vazões naturais de um rio, quando analisados na escala semanal, apresentam um comportamento periódico, refletindo o ciclo das estações do ano. Cada semana apresenta um conjunto de características próprias descritas pela média, desvio-padrão e função de autocorrelação.

Seja uma série de vazões semanais observada em uma determinada seção de um rio $[z_1, z_2, \dots, z_{SN}]$, onde S é o número de semanas e N é o número de anos. Podemos estimar os seguintes índices estatísticos:

A média amostral de cada semana s é dada por:

$$\hat{\mu}_s = \frac{1}{N} \sum_{i=1}^N z_{(i-1)52+s} \quad s = 1,52 \quad (3.1)$$

De forma análoga, o desvio-padrão amostral de cada semana s é dado por:

$$\hat{\sigma}_s = \sqrt{\frac{1}{N} \sum_{i=1}^N z_{(i-1)52+s}^2 - \hat{\mu}_s^2} \quad s = 1,52 \quad (3.2)$$

A autocovariância *lag* 1 mede o grau de dependência linear entre observações contíguas de um processo estocástico. Neste caso, a autocovariância *lag* 1 indica a dependência linear entre a vazão de uma semana com a vazão da semana imediatamente anterior. Supondo-se que a estrutura de dependência temporal é estacionária, isto é, não varia de semana para semana ao longo do ano, a autocovariância *lag* 1 amostral é obtida por:

$$\hat{\gamma}_s = \frac{1}{SN} \sum_{i=2}^{SN} (z_i - \hat{\mu}_s)(z_{i-1} - \hat{\mu}_{s-1}) \quad (3.3)$$

A função de auto-covariância amostral é dada por $\hat{\gamma}(k)$, $k = 1, 2, \dots$, e cada valor é calculado por:

$$\hat{\gamma}(k) = \frac{1}{SN} \sum_{i=k}^{SN} (z_i - \hat{\mu}_s)(z_{i-k} - \hat{\mu}_{s-k}) \quad (3.4)$$

A autocovariância possui dimensão da variável Z^2 . Para se obter um índice de análise mais direto divide-se a função de autocovariância pelos desvios padrão das variáveis envolvidas, resultando na função de autocorrelação amostral:

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\sigma}_s \hat{\sigma}_{s-k}} \quad (3.5)$$

Se este valor for igual a 1 significa que a vazão de uma semana pode ser perfeitamente descrita pela vazão da semana anterior. Caso este valor seja igual a -1 significa que as variáveis são perfeitamente correlacionadas, mas o crescimento de uma implica no decréscimo da outra. Se este valor é zero significa que as variáveis não têm dependência linear.

Por outro lado, em processos semanais é comum definir um conjunto de valores que descreva a estrutura de autocorrelação linear de uma dada semana com as semanas anteriores, para cada semana. Esses valores podem ser definidos pela autocorrelação *lag* 1, que descreve a dependência linear da vazão de uma semana qualquer s com a vazão da semana imediatamente anterior ($s-1$); autocorrelação *lag* 2, que descreve a dependência linear da vazão da semana s com a vazão da semana ($s-2$); ...; autocorrelação *lag* k , que descreve a dependência linear da vazão da semana s com a vazão do semana ($s-k$), onde k é um valor qualquer. A esse conjunto de valores dá-se o nome de função de autocorrelação da semana s . Os valores amostrais desses índices estatísticos podem ser obtidos da seguinte forma:

$$\hat{\gamma}^s(k) = \frac{1}{N} \sum_{i=k}^N (z_{(i-1)52+s} - \hat{\mu}_s)(z_{(i-1)52+s-k} - \hat{\mu}_{s-k}) \quad (3.6)$$

$$\hat{\rho}^s(k) = \frac{\hat{\gamma}^s(k)}{\hat{\sigma}_s \hat{\sigma}_{s-k}} \quad (3.7)$$

3.2 AGRUPAMENTO DA ESTRUTURA DE CORRELAÇÃO

Nos modelos periódicos, cada período apresenta uma função de autocorrelação. Muitas vezes, algumas das estimativas dessas funções são similares entre si podendo-se, então, agrupá-las por períodos de tempo maiores que uma semana, de tal forma que a sua estimação torna-se mais robusta devido à maior disponibilidade de observações.

Por exemplo, uma função de auto-correlação válida para as semanas 1 a 4, representando o primeiro mês, pode ser estimada por:

$$\hat{\rho}^{1,4}(k) = \frac{1}{4N} \sum_{i=1}^N \sum_{j=1}^4 \frac{(z_{(i-1)52+J} - \hat{\mu}_j)}{\hat{\sigma}_j} \frac{(z_{(i-1)52+j-k} - \hat{\mu}_{j-k})}{\hat{\sigma}_{j-k}} \quad (3.8)$$

A esta representação damos o nome de agrupamento mensal da função de auto-correlação. O agrupamento trimestral da função de autocorrelação pode ser estimado da seguinte forma:

$$\hat{\rho}^{1,13}(k) = \frac{1}{13N} \sum_{i=1}^N \sum_{j=1}^{13} \frac{(z_{(i-1)52+J} - \hat{\mu}_j)}{\hat{\sigma}_j} \frac{(z_{(i-1)52+j-k} - \hat{\mu}_{j-k})}{\hat{\sigma}_{j-k}} \quad (3.9)$$

Analogamente, o agrupamento *semestral* da função de autocorrelação é dado por:

$$\hat{\rho}^{1,26}(k) = \frac{1}{26N} \sum_{i=1}^N \sum_{j=1}^{26} \frac{(z_{(i-1)52+J} - \hat{\mu}_j)}{\hat{\sigma}_j} \frac{(z_{(i-1)52+j-k} - \hat{\mu}_{j-k})}{\hat{\sigma}_{j-k}} \quad (3.10)$$

3.3 MODELO AUTO-REGRESSIVO - AR(p)

3.3.1 Descrição do Modelo

O modelo AR(p) pode ser descrito matematicamente por:

$$\left(\frac{Z_t - \mu_s}{\sigma_s} \right) = \phi_1 \left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}} \right) + \dots + \phi_p \left(\frac{Z_{t-p} - \mu_{s-p}}{\sigma_{s-p}} \right) + a_t \quad (3.11a)$$

Ou

$$\Phi(B) \left(\frac{Z_t - \mu_s}{\sigma_s} \right) = a_t \quad (3.11b)$$

onde:

Z_t é uma série sazonal de período s

S é o número de períodos ($S = 52$ para séries semanais)

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 s ($s = 1, 2, \dots, s$)

μ_s é a média sazonal de período s

σ_s é desvio-padrão sazonal de período s

$\Phi(B)$ é o operador auto-regressivo de ordem os

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

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

p é a ordem do operador auto-regressivo

a_t série de ruídos independentes com média zero e variância σ_a^2

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

$$\rho(k) = E \left[\left(\frac{Z_{t-1} - \mu_s}{\sigma_s} \right) \left(\frac{Z_{t-k} - \mu_{s-k}}{\sigma_{s-k}} \right) \right] \quad (3.12)$$

Para a derivação de $\rho(k)$ multiplica-se ambos os lados da equação (3.11a) por $\frac{Z_{t-k} - \mu_{s-k}}{\sigma_{s-k}}$ e

toma-se o valor esperado:

$$E\left[\left(\frac{Z_t - \mu_s}{\sigma_s}\right)\left(\frac{Z_{t-k} - \mu_{s-k}}{\sigma_{s-k}}\right)\right] = \phi_1 E\left[\left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}}\right)\left(\frac{Z_{t-k} - \mu_{s-k}}{\sigma_{s-k}}\right)\right] \\ + \dots + \phi_p E\left[\left(\frac{Z_{t-p} - \mu_{s-p}}{\sigma_{s-p}}\right)\left(\frac{Z_{t-k} - \mu_{s-k}}{\sigma_{s-k}}\right)\right] + E\left[a_t \left(\frac{Z_{t-k} - \mu_{s-k}}{\sigma_{s-k}}\right)\right] \quad (3.13)$$

obtendo-se:

$$\rho(k) = \phi_1 \rho(k-1) + \dots + \phi_p \rho(k-p) \quad k > 0 \quad (3.14)$$

Conhecidos os parâmetros de um modelo AR(p) a função $\rho(k)$ é dada pela solução de (3.14) e pode ser expressa por uma combinação de decaimentos exponenciais e/ou ondas senoidais, o que faz com que $\rho(k)$ tenda a zero à medida que k cresce.

Fixando-se s e variando k de 1 a p em (3.14), obtemos para cada período um conjunto de equações comumente denominado de equações de Yule-Walker. Para um período s qualquer:

$$\begin{bmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(p-1) \\ \rho(1) & 1 & \rho(2) & \dots & \rho(p-2) \\ \dots & \dots & \dots & \dots & \dots \\ \rho(p-1) & \rho(p-2) & \dots & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \dots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \dots \\ \rho_p \end{bmatrix} \quad (3.15)$$

Para $k = 0$, a expressão (3.13) fica:

$$1 = \phi_1 \rho(1) + \phi_2 \rho(2) + \dots + \phi_p \rho(p) + E\left[a_t \left(\frac{Z_t - \mu_s}{\sigma_s}\right)\right] \quad (3.16)$$

Multiplicando (3.11a) por a_t e tomando o valor esperado, obtemos:

$$E\left[a_t \left(\frac{Z_t - \mu_s}{\sigma_s}\right)\right] = \sigma_a^2 \quad (3.17)$$

Substituindo (3.17) em (3.16) temos:

$$\sigma_a^2 = 1 - \phi_1 \rho(1) - \phi_2 \rho(2) - \dots - \phi_p \rho(p) \quad (3.18)$$

Esta expressão é válida para qualquer semana s .

3.3.2 Estimação dos Parâmetros

Depois de fixada a ordem do modelo, p , é necessário obter estimativas para os diversos parâmetros do modelo. Para modelos auto-regressivos os estimadores de momentos são, em geral, bastante eficientes [10].

As médias e os desvios padrões sazonais são estimados pelas equações (3.1) e (3.2) respectivamente.

Os parâmetros ϕ_i , $i = 1, \dots, p$, são estimados substituindo-se em (3.15) $\rho(k)$, $k = 1, \dots, p$, por suas estimativas dadas pela expressão (3.5). O sistema resultante pode ser resolvido por Decomposição de Cholesky [11].

Finalmente, as estimativas de σ_a^2 podem ser obtidas usando-se a expressão (3.18).

É sempre possível estimar o único parâmetro de um modelo AR(1), pois $\phi_1 = \rho(1)$. No entanto, no caso dos modelos AR(p) para $p > 1$, pode não ser possível a resolução do sistema (3.15) ou, o modelo resultante pode ser não-estacionário. Quando isso ocorre, opta-se pela adoção de um modelo de ordem imediatamente inferior, AR(p-1).

3.4 MODELO AUTO-REGRESSIVO MÉDIA-MÓVEL - ARMA(p,1)

3.4.1 Descrição do Modelo

O modelo ARMA(p,1) pode ser descrito matematicamente por:

$$\left(\frac{Z_t - \mu_s}{\sigma_s} \right) = \phi_1 \left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}} \right) + \dots + \phi_p \left(\frac{Z_{t-p} - \mu_{s-p}}{\sigma_{s-p}} \right) - \theta_1 a_{t-1} + a_t \quad (3.19)$$

onde:

Z_t é uma série sazonal de período s

S o número de períodos ($S = 52$ para séries semanais)

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 s ($s = 1, 2, \dots, S$)

μ_s é a média sazonal de período s

σ_s é desvio-padrão sazonal de período s

$\Phi(B)$ é o operador auto-regressivo de ordem p

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

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

p é a ordem do operador auto-regressivo

θ_1 é o coeficiente média-móvel de ordem 1

a_t série de ruídos independentes com média zero e variância σ_a^2

Multiplicando-se ambos os lados da equação (3.19) por a_t e achando-se o valor esperado:

$$E\left[\left(\frac{Z_t - \mu_s}{\sigma_s}\right)a_t\right] = \phi_1 E\left[\left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}}\right)\right] + \dots + \phi_p E\left[\left(\frac{Z_{t-p} - \mu_{s-p}}{\sigma_{s-p}}\right)\right] - \theta_1 E[a_{t-1}a_t] + E[a_t a_t]$$

$$E\left[\left(\frac{Z_t - \mu_s}{\sigma_s}\right)a_t\right] = \sigma_a^2 \quad (3.20)$$

Multiplicando (3.19) por a_{t-1} e tomando o valor esperado, obtemos:

$$E\left[\left(\frac{Z_t - \mu_s}{\sigma_s}\right)a_{t-1}\right] = \phi_1 E\left[\left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}}\right)a_{t-1}\right] + \dots + \phi_p E\left[\left(\frac{Z_{t-p} - \mu_{s-p}}{\sigma_{s-p}}\right)a_{t-1}\right] - \theta_1 E[a_{t-1}a_{t-1}] + E[a_t a_{t-1}]$$

$$E\left[\left(\frac{Z_t - \mu_s}{\sigma_s}\right)a_{t-1}\right] = \phi_1 \sigma_a^2 - \theta_1 \sigma_a^2$$

ou,

$$E\left[\left(\frac{Z_t - \mu_s}{\sigma_s}\right)a_{t-1}\right] = (\phi_1 - \theta_1)\sigma_a^2 \quad (3.21)$$

Multiplicando (3.19) por Z_t e tomando o valor esperado, obtemos:

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

$$1 = \phi_1 \rho(1) + \dots + \phi_p \rho(p) + \sigma_a^2 - \theta_1 (\phi_1 - \theta_1) \sigma_a^2$$

ou,

$$\sigma_a^2 = \frac{1 - \phi_1 \rho(1) - \dots - \phi_p \rho(p)}{1 - \theta_1 \phi_1 + \theta_1^2} \quad (3.22)$$

Multiplicando (3.19) por Z_{t-1} e tomando o valor esperado, obtemos:

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

$$\rho(1) = \phi_1 + \phi_2 \rho(1) + \dots + \phi_p \rho(p-1) + \theta_1 \sigma_a^2$$

ou,

$$\theta_1 = \frac{\phi_1 - \rho(1) + \phi_2 \rho(1) + \dots + \phi_p \rho(p-1)}{\sigma_a^2} \quad (3.23)$$

Multiplicando-se ambos os lados da equação (3.19) por Z_{t-2} e achando-se os valores esperados:

$$\begin{aligned}
 E\left[\left(\frac{Z_t - \mu_s}{\sigma_s}\right)\left(\frac{Z_t - \mu_{s-2}}{\sigma_{s-2}}\right)\right] &= \phi_1 E\left[\left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}}\right)\left(\frac{Z_t - \mu_{s-2}}{\sigma_{s-2}}\right)\right] + \dots + \\
 \phi_p E\left[\left(\frac{Z_{t-p} - \mu_{s-p}}{\sigma_{s-p}}\right)\left(\frac{Z_t - \mu_{s-2}}{\sigma_{s-2}}\right)\right] &- \theta_1 E\left[a_{t-1}\left(\frac{Z_t - \mu_{s-2}}{\sigma_{s-2}}\right)\right] \\
 + E\left[a_t\left(\frac{Z_t - \mu_{s-2}}{\sigma_{s-2}}\right)\right] & \\
 \rho(2) = \phi_2 + \phi_1 \rho(1) + \dots + \phi_p \rho(p-2) & \quad (3.24)
 \end{aligned}$$

Multiplicando-se ambos os lados da equação (3.19) por Z_{t-3} e achando-se os valores esperados:

$$\begin{aligned}
 E\left[\left(\frac{Z_t - \mu_s}{\sigma_s}\right)\left(\frac{Z_t - \mu_{s-3}}{\sigma_{s-3}}\right)\right] &= \phi_1 E\left[\left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}}\right)\left(\frac{Z_t - \mu_{s-3}}{\sigma_{s-3}}\right)\right] + \dots + \\
 \phi_p E\left[\left(\frac{Z_{t-p} - \mu_{s-p}}{\sigma_{s-p}}\right)\left(\frac{Z_t - \mu_{s-3}}{\sigma_{s-3}}\right)\right] &- \theta_1 E\left[a_{t-1}\left(\frac{Z_t - \mu_{s-3}}{\sigma_{s-3}}\right)\right] \\
 + E\left[a_t\left(\frac{Z_t - \mu_{s-3}}{\sigma_{s-3}}\right)\right] & \\
 \rho(3) = \phi_1 \rho(2) + \phi_2 \rho(1) + \dots + \phi_p \rho(p-3) & \quad (3.25)
 \end{aligned}$$

Repetindo-se este procedimento até Z_{t-p+1} :

$$\begin{aligned}
 E\left[\left(\frac{Z_t - \mu_s}{\sigma_s}\right)\left(\frac{Z_t - \mu_{s-p+1}}{\sigma_{s-p+1}}\right)\right] &= \phi_1 E\left[\left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}}\right)\left(\frac{Z_t - \mu_{s-p+1}}{\sigma_{s-p+1}}\right)\right] + \dots + \\
 \phi_p E\left[\left(\frac{Z_{t-p} - \mu_{s-p}}{\sigma_{s-p}}\right)\left(\frac{Z_t - \mu_{s-p+1}}{\sigma_{s-p+1}}\right)\right] &- \theta_1 E\left[a_{t-1}\left(\frac{Z_t - \mu_{s-p+1}}{\sigma_{s-p+1}}\right)\right] \\
 + E\left[a_t\left(\frac{Z_t - \mu_{s-p+1}}{\sigma_{s-p+1}}\right)\right] & \\
 \rho(p+1) = \phi_1 \rho(p) + \phi_2 \rho(p-1) + \dots + \phi_p \rho(1) & \quad (3.26)
 \end{aligned}$$

Arrumando as equações (3.24), (3.25) e (3.26) na forma matricial:

$$\begin{bmatrix} \rho(1) & 1 & \dots & \rho(p-2) \\ \rho(2) & \rho(1) & \dots & \rho(p-3) \\ \dots & \dots & \dots & \dots \\ \rho(p) & \rho(p-1) & \dots & \rho(1) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \dots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \rho(2) \\ \rho(3) \\ \dots \\ \rho(p+1) \end{bmatrix} \quad (3.27)$$

3.4.2 Estimação dos Parâmetros

Os parâmetros auto-regressivos são obtidos pela solução do conjunto de equações (3.27). O parâmetro média-móvel é obtido a partir de um procedimento iterativo com as equações (3.22) e (3.23). Inicialmente, adota-se como uma primeira estimativa para a variância dos ruídos, σ_a^2 , o valor correspondente ao do modelo auto-regressivo de ordem 1, dado pela expressão (3.18). A seguir, obtém-se θ_1 pela expressão (3.23). Substituindo-se este valor em (3.22), obtém-se σ_a^2 associado à estimativa θ_1 . Caso o novo valor de σ_a^2 não seja significativamente diferente do anterior (a diferença entre eles não deve ultrapassar 0,1% do valor anterior), o processo de estimação está concluído. Caso contrário, deve-se obter novo valor de θ_1 pela expressão (3.23) até se obter a convergência.

No caso do conjunto de parâmetros auto-regressivos não resultar em um modelo estacionário, ou o parâmetro média-móvel ser superior a $|1|$, indicando que o processo não é invertível em um modelo auto-regressivo, adota-se um modelo de ordem imediatamente inferior, ARMA(p-1,1). Se o modelo ARMA(1,1) não puder ser estimado escolhe-se, então, o modelo AR(1).

3.5 MODELO AUTO-REGRESSIVO - PAR(p)

3.5.1 Descrição do Modelo

O modelo PAR(p_1, p_2, \dots, p_{12}) pode ser descrito matematicamente por:

$$\left(\frac{Z_t - \mu_s}{\sigma_s} \right) = \phi_1^s \left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}} \right) + \dots + \phi_p^s \left(\frac{Z_{t-p} - \mu_{s-p}}{\sigma_{s-p}} \right) + a_t \quad (3.28a)$$

ou,

$$\Phi^s(B) \left(\frac{Z_t - \mu_s}{\sigma_s} \right) = a_t \quad (3.28b)$$

onde:

Z_t é uma série sazonal de período s

S é o número de períodos ($S = 52$ para séries semanais)

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 s ($s = 1, 2, \dots, S$)

μ_s é a média sazonal de período m

σ_s é desvio-padrão sazonal de período m

$\Phi^s(B)$ é o operador auto-regressivo de ordem os

$$\Phi^s(B) = (1 - \phi_1^s B - \phi_2^s B^2 - \dots - \phi_p^s B^{ps})$$

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 série de ruídos independentes com média zero e variância $\sigma_a^{2(s)}$

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

$$\rho^s(k) = E \left[\left(\frac{Z_{t-1} - \mu_s}{\sigma_s} \right) \left(\frac{Z_{t-k} - \mu_{s-k}}{\sigma_{s-k}} \right) \right] \quad (3.29)$$

O conjunto de funções de autocorrelação $\rho^s(k)$ dos períodos $s = 1, \dots, S$, descrevem a estrutura de dependência temporal da série. Estas funções são obtidas por [10]:

Multiplicando-se ambos os lados da equação por $\frac{Z_{t-k} - \mu_{s-k}}{\sigma_{s-k}}$ e tomando o valor esperado

obtemos para cada período:

$$E\left[\left(\frac{Z_t - \mu_s}{\sigma_s}\right)\left(\frac{Z_{t-k} - \mu_{s-k}}{\sigma_{s-k}}\right)\right] = \phi_1^s E\left[\left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}}\right)\left(\frac{Z_{t-k} - \mu_{s-k}}{\sigma_{s-k}}\right)\right] + \dots + \phi_p^s E\left[\left(\frac{Z_{t-p} - \mu_{s-p}}{\sigma_{s-p}}\right)\left(\frac{Z_{t-k} - \mu_{s-k}}{\sigma_{s-k}}\right)\right] + E\left[a_t \left(\frac{Z_{t-k} - \mu_{s-k}}{\sigma_{s-k}}\right)\right] \quad (3.30)$$

Por exemplo, para $k = 1$ a expressão (3.30) resulta em:

$$\rho^s(1) = \phi_1^s + \phi_2^s \rho^{s-1}(1) + \dots + \phi_p^s \rho^{s-1}(p_s - 1) \quad (3.31)$$

Conhecidos os parâmetros de um modelo PAR(p) as funções $\rho^s(k)$ são dadas pela solução de (3.30) e podem ser expressas por uma combinação de decaimentos exponenciais e/ou ondas senoidais, o que faz com que cada $\rho^s(k)$ tenda a zero à medida que k cresce.

Fixando-se s e variando k de 1 a p em (3.30) obtemos para cada período um conjunto de equações comumente denominado de equações de Yule-Walker. Para um período s qualquer:

$$\begin{bmatrix} 1 & \rho^{s-1}(1) & \rho^{s-1}(2) & \dots & \rho^{s-1}(p-1) \\ \rho^{s-1}(1) & 1 & \rho^{s-2}(1) & \dots & \rho^{s-2}(p-2) \\ \rho^{s-1}(2) & \rho^{s-2}(1) & 1 & \dots & \rho^{s-3}(p-3) \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{s-1}(p-1) & \rho^{s-2}(p-2) & \rho^{s-3}(p-3) & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_1^s \\ \phi_2^s \\ \phi_3^s \\ \dots \\ \phi_p^s \end{bmatrix} = \begin{bmatrix} \rho^s(1) \\ \rho^s(2) \\ \rho^s(3) \\ \dots \\ \rho_p^s \end{bmatrix} \quad (3.32)$$

Para $k = 0$, a expressão (3.30) fica:

$$1 = \phi_1^s \rho^s(1) + \dots + \phi_p^s \rho^s(p) + E\left[a_t \left(\frac{Z_t - \mu_s}{\sigma_s}\right)\right] \quad (3.33)$$

Multiplicando (3.28a) por a_t e tomando o valor esperado, obtemos:

$$E\left[a_t \left(\frac{Z_t - \mu_s}{\sigma_s}\right)\right] = \sigma_a^{2(s)} \quad (3.34)$$

Substituindo (3.34) em (3.33) temos:

$$\sigma_a^{2(s)} = 1 - \phi_1^s \rho^s(1) - \phi_2^s \rho^s(2) - \dots - \phi_p^s \rho^s(p) \quad (3.35)$$

Esta expressão é válida para qualquer período s .

3.5.2 Estimação dos parâmetros pelo método dos momentos

Depois de fixado a ordem do modelo, p , é necessário obter estimativas para os diversos parâmetros. Como já mencionado anteriormente, os estimadores de momento para modelos auto-regressivos são, em geral, bastante eficientes.

As médias e os desvios padrões sazonais são estimados pelas equações (3.1) e (3.2) respectivamente.

Os parâmetros ϕ_i^s , $i = 1, \dots, p$, são estimados substituindo-se em (3.32) $\rho^s(k)$, $k = 1, \dots, p$, por suas estimativas dadas pela expressão (3.7). Considerando-se os agrupamentos mensal, trimestral e semestral substitui-se em (3.32) $\rho^s(k)$, $k = 1, \dots, p$ por suas estimativas dadas pelas expressões (3.8) e (3.10).

O sistema resultante pode ser resolvido por Decomposição de Cholesky.

Finalmente, as estimativas de $\sigma_a^{2(s)}$ podem ser obtidas usando-se a expressão (3.35).

O coeficiente do modelo AR(1) de um semana s qualquer é dado por $\phi_1 = \rho^s(1)$, o qual é sempre possível de ser obtido. No entanto, para $p > 1$ pode não ser possível a resolução do sistema (3.32). Quando isso ocorre, opta-se pela adoção de um modelo de ordem imediatamente inferior, AR($p-1$), para o semana s .

3.5.3 Estimação dos parâmetros por regressão linear com respeito à origem de previsão

Neste caso o modelo PAR(p) é descrito por uma regressão nas p observações conhecidas no instante da previsão.

Seja a origem de previsão no início de cada mês e chamando de s a primeira semana do mês, podemos reescrever a equação (3.28a) para Z do ano j e semana ($s+i$) como:

$$\left(\frac{Z_j^{s+i} - \mu_{s+i}}{\sigma_{s+i}} \right) = \phi_1^{s+i} \left(\frac{Z_j^{s-1} - \mu_{s-1}}{\sigma_{s-1}} \right) + \dots + \phi_p^{s+i} \left(\frac{Z_j^{s-p} - \mu_{s-p}}{\sigma_{s-p}} \right) + a_j^{s+i} \quad (3.36)$$

Desta forma, o sistema de equações referente à i-ésima semana do mês, supondo uma regressão nas observações conhecidas no instante da previsão (origem de previsão), é dado por:

$$\begin{bmatrix} z_1^{s+i} \\ z_2^{s+i} \\ \dots \\ z_N^{s+i} \end{bmatrix} = \begin{bmatrix} z_1^{s-1} & z_1^{s-2} & \dots & z_1^{s-p} \\ z_2^{s-1} & z_2^{s-2} & \dots & z_2^{s-p} \\ \dots & \dots & \dots & \dots \\ z_N^{s-1} & z_N^{s-2} & \dots & z_N^{s-p} \end{bmatrix} \begin{bmatrix} \phi_1^{s+i} \\ \phi_2^{s+i} \\ \dots \\ \phi_p^{s+i} \end{bmatrix} + \begin{bmatrix} a_1^{s+i} \\ a_2^{s+i} \\ \dots \\ a_N^{s+i} \end{bmatrix} \quad (3.37)$$

ou,

$$\mathbf{Z} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3.38)$$

Onde $\boldsymbol{\beta}$, por mínimos quadrados [10] é dada por:

$$\boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Z}) \quad (3.39)$$

3.6 MODELO AUTO-REGRESSIVO MÉDIA-MÓVEL PERIÓDICO - PARMA(p,1)

3.6.1 Descrição do Modelo

O modelo PARMA(p,1) pode ser descrito matematicamente por:

$$\left(\frac{Z_t - \mu_s}{\sigma_s} \right) = \phi_1^s \left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}} \right) + \dots + \phi_p^s \left(\frac{Z_{t-p} - \mu_{s-p}}{\sigma_{s-p}} \right) + \theta_1^s a_{t-1} + a_t \quad (3.40)$$

onde:

Z_t é uma série sazonal de período m

S é o número de períodos ($S= 52$ para séries semanais)

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 s ($s = 1, 2, \dots, S$)

μ_s é a média sazonal de período M

σ_s é desvio-padrão sazonal de período M

$\Phi^s(B)$ é o operador auto-regressivo de ordem os

$$\Phi^s(B) = (1 - \phi_1^s B - \phi_2^s B^2 - \dots - \phi_p^s B^{ps})$$

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

p é a ordem do operador auto-regressivo

θ_1^s é o coeficiente média-móvel de ordem 1

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

Multiplicando-se ambos os lados da equação (3.40) por a_t e achando-se os valores esperados:

$$E\left[\left(\frac{Z_t - \mu_s}{\sigma_s}\right)a_t\right] = \phi_1^s E\left[\left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}}\right)a_t\right] + \dots + \phi_p^s E\left[\left(\frac{Z_{t-p} - \mu_{s-p}}{\sigma_{s-p}}\right)a_t\right] - \theta_1^s E[a_{t-1}a_t] + E[a_t a_t] \quad (3.41)$$

Multiplicando (3.40) por a_{t-1} e tomando o valor esperado, obtemos:

$$E\left[\left(\frac{Z_t - \mu_s}{\sigma_s}\right)a_{t-1}\right] = \phi_1^s E\left[\left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}}\right)a_{t-1}\right] + \dots + \phi_p^s E\left[\left(\frac{Z_{t-p} - \mu_{s-p}}{\sigma_{s-p}}\right)a_{t-1}\right] - \theta_1^s E[a_{t-1}a_{t-1}] + E[a_t a_{t-1}]$$

$$E\left[\left(\frac{Z_t - \mu_s}{\sigma_s}\right)a_{t-1}\right] = \phi_1^s \sigma_{a_{t-1}}^2 - \theta_1^s \sigma_{a_{t-1}}^2$$

ou,

$$E\left[\left(\frac{Z_t - \mu_s}{\sigma_s}\right)a_{t-1}\right] = (\phi_1^s - \theta_1^s) \sigma_{a_{t-1}}^2 \quad (3.42)$$

Multiplicando (3.40) por Z_t e tomando o valor esperado, obtemos:

$$E\left[\left(\frac{Z_t - \mu_s}{\sigma_s}\right)\left(\frac{Z_t - \mu_s}{\sigma_s}\right)\right] = \phi_1^s E\left[\left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}}\right)\left(\frac{Z_t - \mu_s}{\sigma_s}\right)\right] + \dots + \phi_p^s E\left[\left(\frac{Z_{t-p} - \mu_{s-p}}{\sigma_{s-p}}\right)\left(\frac{Z_t - \mu_s}{\sigma_s}\right)\right] - \theta_1^s E\left[a_{t-1}\left(\frac{Z_t - \mu_s}{\sigma_s}\right)\right] + E\left[a_t\left(\frac{Z_t - \mu_s}{\sigma_s}\right)\right]$$

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

ou,

$$\sigma_{a_t}^2 = 1 - \phi_1^s \rho^s(1) - \dots - \phi_p^s \rho^s(p) + \theta_1^s (\phi_1^s - \theta_1^s) \sigma_{a_{t-1}}^2 \quad (3.43)$$

Para obtermos a função de autocorrelação *lag* um basta (3.40) por Z_{t-1} e tomarmos o valor esperado, obtendo:

$$E\left[\left(\frac{Z_t - \mu_s}{\sigma_s}\right)\left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}}\right)\right] = \phi_1^s E\left[\left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}}\right)\left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}}\right)\right] + \dots + \phi_p^s E\left[\left(\frac{Z_{t-p} - \mu_{s-p}}{\sigma_{s-p}}\right)\left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}}\right)\right] - \theta_1^s E\left[a_{t-1}\left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}}\right)\right] + E\left[a_t\left(\frac{Z_{t-1} - \mu_{s-1}}{\sigma_{s-1}}\right)\right]$$

$$\rho^s(1) = \phi_1^s + \phi_2^s \rho^{s-1}(1) + \dots + \phi_p^s \rho^{s-1}(p-1) - \theta_1^s \sigma_{a_{t-1}}^2 \quad (3.44)$$

ou,

$$\theta_1^s = \frac{\phi_1^s - \rho^s(1) + \phi_2^s \rho^{s-1}(1) + \dots + \phi_p^s \rho^{s-1}(p-1)}{\sigma_{a_{t-1}}^2} \quad (3.45)$$

Multiplicando-se ambos os lados da equação (3.40) por Z_{t-2} e achando-se os valores esperados, obtemos uma expressão para o coeficiente auto regressivo *lag* dois:

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

$$\rho^s(2) = \phi_1^s \rho^{s-1}(1) + \phi_2^s + \dots + \phi_p^s \rho^{s-2}(p-2) \quad (3.46)$$

Repetindo-se este procedimento até Z_{t-p-1} , obtemos uma expressão para o coeficiente auto-regressivo lag p+1:

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

$$\rho^s(p+1) = \phi_1^s \rho^{s-1}(p) + \phi_2^s \rho^{s-2}(p-1) + \dots + \phi_p^s \rho^{s-p}(1) \quad (3.47)$$

Arrumando as equações e na forma matricial:

$$\begin{bmatrix} \rho^{s-1}(1) & 1 & \dots & \rho^{s-2}(p-2) \\ \rho^{s-1}(2) & \rho^{s-2}(1) & \dots & \rho^{s-3}(p-3) \\ \dots & \dots & \dots & \dots \\ \rho^{s-1}(p) & \rho^{s-2}(p-1) & \dots & \rho^{s-4}(1) \end{bmatrix} \begin{bmatrix} \phi_1^s \\ \phi_2^s \\ \dots \\ \phi_p^s \end{bmatrix} = \begin{bmatrix} \rho^s(2) \\ \rho^s(3) \\ \dots \\ \rho^s(p+1) \end{bmatrix} \quad (3.48)$$

3.6.2 Estimação dos Parâmetros pelo Método dos Momentos

Os parâmetros auto-regressivos de cada semana são obtidos pela solução do conjunto de equações (3.48). As estimativas de $\rho^s(k)$, $k = 1, \dots, p$ são obtidas pelas equações (3.7) ou (3.8) ou (3.9) ou (3.10), considerando-se agrupamento da função de auto-correlação, por semana ou, por trimestre ou, por semestre, respectivamente.

Caso não seja possível resolver o conjunto de equações de uma semana qualquer, adota-se o conjunto de parâmetros auto-regressivos do modelo ARMA(p-1,1) daquela mesma semana.

O parâmetro média-móvel de cada semana é obtido a partir de um procedimento iterativo com as equações (3.43) e (3.45). O processo iterativo tem início com a primeira semana do ano. Adota-se como uma primeira estimativa para a variância dos ruídos do período imediatamente anterior (por exemplo, a última semana do ano), $\sigma_{a_{t-1}}^{2(s)}$, o valor correspondente ao do modelo auto-regressivo de ordem 1, dado pela expressão (3.35). A seguir, obtém-se θ_1^s pela expressão (3.45). Substituindo-se este valor em (3.43), obtém-se $\sigma_{a_n}^{2(s)}$ associado à estimativa θ_1^s . Este procedimento é realizado para todas as semanas do processo. Se a nova estimativa de $\sigma_a^{2(s)}$ da última semana for semelhante à primeira, o processo finaliza. Caso contrário deve-se repetir o procedimento até obter a convergência do algoritmo.

Durante o processo iterativo é possível obter-se em alguma semana um valor de $\sigma_{a_t}^{2(s)}$ negativo. Neste caso, o programa adota o modelo PARMA (p-1,1) daquela semana.

3.6.3 Estimação dos parâmetros por regressão linear

Chamando de z_i^s , o valor observado na semana s do i -ésimo ano, podemos escrever o seguinte conjunto de equações de regressão na forma matricial:

$$\begin{bmatrix} z_1^s \\ z_2^s \\ \dots \\ z_N^s \end{bmatrix} = \begin{bmatrix} z_1^{s-1} & \dots & z_1^{s-p} & a_1^{s-1} \\ z_2^{s-1} & \dots & z_2^{s-p} & a_2^{s-1} \\ \dots & \dots & \dots & \dots \\ z_N^{s-1} & \dots & z_N^{s-1} & a_N^{s-1} \end{bmatrix} \begin{bmatrix} \phi_1^s \\ \dots \\ \phi_p^s \\ \theta_1^s \end{bmatrix} + \begin{bmatrix} a_1^s \\ a_2^s \\ \dots \\ a_N^s \end{bmatrix} \quad (3.49)$$

ou,

$$\mathbf{Z} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3.50)$$

Note que cada equação deste sistema representa a equação (3.40) escrita para cada uma das observações disponíveis da semana s .

A solução desse sistema de equações por mínimos quadrados [10] é dada por:

$$\boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Z}) \quad (3.51)$$

3.6.4 Estimação dos parâmetros por regressão linear com respeito à origem de previsão

Neste caso o modelo PARMA(p,1) é descrito por uma regressão nas p observações e no ruído conhecidos no instante da previsão.

Seja a origem de previsão no início de cada mês e chamando de s a primeira semana do mês, podemos reescrever a equação (3.40) para Z do ano j e semana (s+i) como:

$$\left(\frac{Z_j^{s+i} - \mu_{s+i}}{\sigma_{s+i}} \right) = \phi_1^s \left(\frac{Z_j^{s-1} - \mu_{s-1}}{\sigma_{s-1}} \right) + \dots + \phi_p^s \left(\frac{Z_j^{s-p} - \mu_{s-p}}{\sigma_{s-p}} \right) - \theta_1^s a_j^{s-1} + a_j^{s+i} \quad (3.52)$$

Desta forma, o sistema de equações referente à i-ésima semana do mês, supondo uma regressão nas observações conhecidas no instante da previsão (origem de previsão), é dado por:

$$\begin{bmatrix} z_1^{s+i} \\ z_2^{s+i} \\ \dots \\ z_N^{s+i} \end{bmatrix} = \begin{bmatrix} z_1^{s-1} & z_1^{s-2} & \dots & z_1^{s-p} \\ z_2^{s-1} & z_2^{s-2} & \dots & z_2^{s-p} \\ \dots & \dots & \dots & \dots \\ z_N^{s-1} & z_N^{s-2} & \dots & z_N^{s-p} \end{bmatrix} \begin{bmatrix} \phi_1^{s+i} \\ \phi_2^{s+i} \\ \dots \\ \phi_p^{s+i} \end{bmatrix} + \begin{bmatrix} a_1^{s+i} \\ a_2^{s+i} \\ \dots \\ a_N^{s+i} \end{bmatrix} \quad (3.53)$$

ou,

$$\mathbf{Z} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3.54)$$

Onde $\boldsymbol{\beta}$, por mínimos quadrados [10] é dada por:

$$\boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Z}) \quad (3.55)$$

4. TRANSFORMAÇÕES APLICADAS ÀS SÉRIES DE AFLUÊNCIAS

Box e Jenkins [3] sugeriram uma metodologia bastante elaborada para ajuste de modelos estocásticos da família ARIMA (p,d,q) a séries temporais, que pode ser estendida para modelos da família PAR (p). Nesta metodologia, a estratégia de seleção do modelo é dividida em três etapas. A primeira etapa, denominada por Box e Jenkins de identificação do modelo, consiste em escolher, tentativamente, a ordem do modelo baseando-se em estimativas das funções $\rho^s(k)$ e Φ_k^s obtidas a partir da série temporal. Na modelagem auto-regressiva periódica isso consiste em escolher o vetor p. A segunda etapa refere-se à estimação do modelo, ou seja, estimação dos seus parâmetros. A terceira etapa diz respeito à verificação do modelo, isto é, verificar, através de testes estatísticos, se as hipóteses assumidas durante as etapas anteriores são atendidas. Se as hipóteses não são verificadas deve-se retornar à primeira etapa até que os resultados sejam satisfatórios.

Uma hipótese assumida neste tipo de modelagem é que a distribuição de probabilidades dos ruídos é Normal. Uma tentativa de garantir a obtenção de ruídos Normais é usar a transformação Box-Cox, Box et al. [12].

$$X_t^{\lambda_s} = \frac{(X_t)^{\lambda_s} - 1}{\lambda_s} \quad \lambda_s \neq 0 \quad (4.1)$$

$$X_t^{\lambda_s} = \ln X_t \quad \lambda_s = 0 \quad (4.2)$$

onde λ_s é o expoente da transformação Box-Cox do período s.

O expoente da transformação de cada período s, $s = 1, \dots, S$, é obtido de modo que quando aplicado à série $\{X_t\}$ a assimetria resultante seja em cada período nula.

Assim o modelo passa a ser descrito por:

$$\Phi_S(B) \cdot \frac{X_t^{\lambda_s} - \mu_s}{\sigma_s} + a_t \quad (4.3)$$

As variáveis μ_s e σ_s passam a representar os momentos de primeira e segunda ordem da série no período m , após a transformação Box-Cox.

Quando $\lambda_s = 0$, a transformação Box-Cox se reduz a equação (4.2) que representa a transformação Logarítmica.

No modelo PREVIVAZ $\{X_t\}$ é a série de vazões semanais, s a semana, e S o número de semanas.

5. PADRONIZAÇÃO DAS SÉRIES DE AFLUÊNCIAS

O modelo PREVIVAZ utiliza séries de vazões padronizadas. Esta padronização, entretanto, não está associado à hipótese de normalidade dos ruídos como no caso das transformações Box-Cox/Logarítmica. O tratamento aplicado às séries de vazões consiste na padronização das séries em função da área de drenagem do posto fluviométrico correspondente de acordo com a equação (5.1), de forma que as vazões de todos os postos fluviométricos, assim como suas estatísticas, tenham mesma ordem de grandeza.

$$X_i = \frac{X_i \cdot Cte}{Ad} \quad (5.1)$$

Onde Ad é a área de drenagem do posto fluviométrico em km² e Cte é a constante de conversão de unidades.

6. INTERVALO DE CONFIANÇA DAS PREVISÕES

Seja a função densidade de probabilidade $f(x)$, cujo valor esperado é $E[x]=\mu$ e com variância $V(x)=\sigma^2$.

$$f(x) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \quad \text{para } \sigma > 0; \quad -\infty < \mu < \infty \quad (6.1)$$

cuja probabilidade de conter o verdadeiro valor da população é determinada através da integração numérica da função $f(x)$, como a seguir:

$$P_{\text{int } c} = \int_{-\infty}^{z_{\alpha/2}} \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} dx \quad (6.2)$$

Possuindo a série de vazões normalizada determina-se a probabilidade $1 - \alpha$ de o intervalo de confiança (int c) conter o verdadeiro valor do parâmetro populacional da seguinte forma:

$$p_{\text{int } c} = [1 - ((1 - (\text{int } c / 100)) / 2)] \quad (6.3)$$

Um intervalo de confiança (ou estimativa intervalar) é a amplitude de valores que tem probabilidade de conter o verdadeiro valor da população, sendo o grau de confiança a probabilidade $1 - \alpha$ de o intervalo de confiança conter o verdadeiro valor do parâmetro populacional.

Uma vez determinada a probabilidade que é a integração da função distribuição de probabilidade acumulada, obtém-se o valor crítico $Z_{\alpha/2}$ necessário para o cálculo da margem de erro $\delta = Z_{\alpha/2} \sigma$, com a qual se determina os limites inferior e superior do intervalo de confiança.

$$\begin{aligned} \text{Lim.inf} &= \bar{x} - \delta \\ \text{Lim.sup} &= \bar{x} + \delta \end{aligned} \quad (6.4)$$

Onde \bar{x} é a média amostral.

No modelo PREVIVAZ o intervalo de confiança da previsão (limites inferior e superior) é calculado segundo (6.4) substituindo-se x pela previsão da semana e σ pelo desvio padrão dos ruídos do modelo selecionado. O desvio padrão dos ruídos é função do modelo utilizado, sendo calculado como descrito no item 3.

7. ESCOLHA DO MELHOR ALGORITMO DE PREVISÃO

Na abordagem clássica de uso de modelos estocásticos para previsão (ver por exemplo [9]), as previsões são obtidas por equações matemáticas cujas estruturas e número de termos, foram estatisticamente identificados e cujos coeficientes eficientemente estimados utilizando toda a informação disponível. Usando o ponto de vista prático, procura-se adotar tanto na identificação, quanto na estimação, métodos “robustos”, ou seja, métodos resistentes às deficiências de modelagem.

Com efeito, expressões matemáticas utilizadas para descrever fenômenos da natureza são sempre aproximações da realidade, e devem sempre ser julgadas tendo em vista a sua utilização. Dooge (1972), [13], lembra que um modelo é algo para ser usado e não algo em que se deve acreditar. Colocado diante de dois possíveis modelos de previsão, ou ainda de duas diferentes estimativas de um dos parâmetros do modelo, o previsor deve procurar desenvolver uma medida de precisão com a qual tomará a decisão sobre qual modelo utilizar para fazer a previsão.

No modelo PREVIVAZ os algoritmos de previsão, definido por um modelo, método de estimação específico e transformação da série de vazões, são testados por um esquema onde cada série é dividida em duas partes. Inicialmente apenas a primeira parte da série é utilizada para a estimação dos parâmetros (estimação) e a segunda parte apenas para o cálculo de erros de previsão (verificação). Em seguida a estimação dos parâmetros passa a ser feita com a segunda parte da série, ficando a primeira parte apenas para cálculo de erros de previsão. Para cada parte da série é computado o erro padrão de previsão (raiz quadrada da média dos quadrados dos erros de previsão - EQM) um passo à frente obtendo-se, a seguir, a média dos dois valores. A Figura 7.1 ilustra o procedimento.

O modelo PREVIVAZ, utiliza a cada semana o método (algoritmo) de menor erro médio quadrático de previsão um passo a frente, dentre os enumerados na Tabela 7.1.

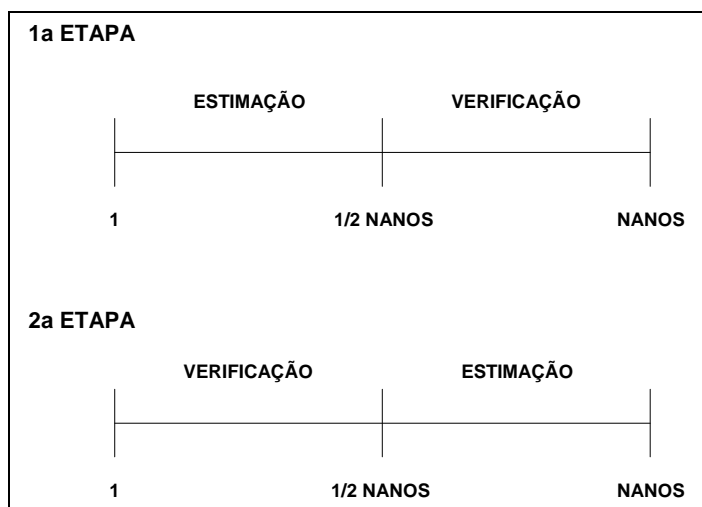


Figura 7.1 – Procedimento de escolha do algoritmo de previsão

Tabela 7.1 - Algoritmos de Previsão

| 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 semanal | momentos |
| PAR(p)-G2 ($1 \leq p \leq 4$) * | estrutura de correlação sazonal mensal | momentos |
| PAR(p)-G3 ($1 \leq p \leq 4$) * | estrutura de correlação sazonal trimestral | momentos |
| PAR(p)-G4 ($1 \leq p \leq 4$) * | estrutura de correlação sazonal semestral | momentos |
| PAR(p)-RO ($1 \leq p \leq 4$) * | estrutura de correlação sazonal | regressão simples a partir |

| Modelo de Previsão | Característica | Método de Estimação |
|---------------------------------------|--|--|
| | | da origem de previsão |
| PARMA(p,1)-G1 ($1 \leq p \leq 3$) * | estrutura de correlação sazonal semanal | momentos |
| PARMA(p,1)-G2 ($1 \leq p \leq 3$) * | estrutura de correlação sazonal mensal | momentos |
| PARMA(p,1)-G3 ($1 \leq p \leq 3$) * | estrutura de correlação sazonal trimestral | momentos |
| PARMA(p,1)-G4 ($1 \leq p \leq 3$) * | estrutura de correlação sazonal semestral | momentos |
| PARMA(p,1)-RO ($1 \leq p \leq 3$) * | estrutura de correlação sazonal | regressão simples a partir da origem de previsão |
| 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.

Na implementação do programa PREVIVAZ são consideradas algumas exceções no critério de escolha do melhor algoritmo de previsão:

1. Séries com menos de 20 anos de observação: Neste caso, os algoritmos que consideram estrutura de correlação sazonal não são utilizados, pois o número de observações disponível ao se dividir a série em duas partes para estimar a estrutura de correlação sazonal seria muito reduzida (no máximo 10 anos), o que poderia conduzir a estimativas pouco robustas.
2. Modelos CONSTANTE e SAZONAL: No caso de algum destes dois modelos ser o de menor erro médio quadrático, se a diferença entre seu erro médio quadrático e o do segundo modelo for menos que 5%, então se abandona o modelo CONSTANTE ou

SAZONAL, conforme o caso, e seleciona-se o segundo melhor modelo. Este critério baseia-se no fato que ambos são abordagens muito simples para se fazer previsões.

3. Modelos (PAR e PARMA) com estimação dos parâmetros através de Regressão em Relação à Origem: Se este algoritmo apresentar o de menor erro médio quadrático e a diferença entre seu erro médio quadrático e o do segundo modelo for menos que 5%, então este modelo é abandonado e seleciona-se o segundo melhor modelo.

7.1 APLICAÇÃO DE LIMITES NAS PREVISÕES

A principal característica do modelo PREVIVAZ na obtenção de previsões de vazões fluviais é o uso da parte determinística de modelos de séries temporais ajustados às séries de vazões fluviais como previsão. Neste caso, a previsão toma partido não só do comportamento sazonal das vazões fluviais referido anteriormente, como também considera o importante fenômeno denominado tendência hidrológica. Na terminologia de séries temporais, a tendência hidrológica é conhecida como estrutura de dependência temporal, sendo quantificada pela função de autocorrelação estimada do registro de vazões.

No entanto, em bacias que apresentam fraca estrutura de dependência temporal e grande variabilidade das vazões, modelos de séries temporais como o PREVIVAZ tendem a fornecer previsões de valores futuros próximos da média, independente das condições passadas. Pode-se assim obter valores previstos muito superiores ou inferiores aos últimos valores observados, que apresentem razões entre as vazões das semanas $t+1$ e t pouco frequentes no histórico de aflúências.

Outro problema verificado ocorre quando se adota transformação das séries histórica de aflúências. Tem-se verificado que na previsão de séries transformadas pode ocorrer uma atenuação ou um aumento acentuado nos valores de previsão quando ocorre uma combinação entre o valor do parâmetro da transformação adotada e um período de alta variabilidade de vazões. Esta combinação pode levar a que a previsão da série transformada se localize na região assintota da função de transformação, fazendo com que o valor obtido na inversão da função de transformação seja muito grande e hidrológicamente improvável para a bacia na semana a ser prevista.

Diante do exposto, a metodologia propõe que durante o processo de escolha do melhor algoritmo e no cálculo das previsões, as previsões calculadas sejam revistas considerando limites conforme descrito nos itens 7.1.1 e 7.1.2.

7.1.1 Aplicação de limites na escolha da melhor alternativa no PREVIVAZ

De acordo com o algoritmo apresentado no item 7 percebe-se que os algoritmos são testados por um esquema onde cada série é dividida em duas partes. Inicialmente apenas a primeira parte da série é utilizada para a estimação dos parâmetros (estimação) e a segunda parte apenas para o cálculo de erros de previsão (verificação). Em seguida a estimação dos parâmetros passa a ser feita com a segunda parte da série, ficando a primeira parte apenas para cálculo de erros de previsão. Para cada parte da série é computado o erro padrão de previsão (raiz quadrada da média dos quadrados dos erros de previsão - REQM) um passo à frente obtendo-se, a seguir, a média dos dois valores. A partir da versão 5.3 do PREVIVAZ permite-se que o algoritmo escolhido segundo o critério acima descrito seja alterado.

Têm-se inicialmente todas as previsões para todos os algoritmos utilizando a primeira parte da série para estimação dos parâmetros, como feito atualmente, entretanto, as previsões obtidas serão avaliadas para verificar se seu valor extrapola limites previamente definidos para a bacia em estudo. Estes limites são calculados considerando a parte da série usada para estimar os parâmetros dos algoritmos (primeira metade). O mesmo procedimento é aplicado quando se utiliza a segunda metade da série para estimar os parâmetros.

Para a definição dos limites calculam-se as razões entre as vazões da semana $t+1$ (semana a ser prevista) com a da semana t (semana anterior à semana a ser prevista) da parte da série utilizada para a estimação dos parâmetros. Constrói-se, então, a distribuição de frequência acumulada de não-excedência das razões entre as vazões das semanas $t+1$ e t . A partir dessa distribuição de frequência se estabelecem os limites superior e inferior para as previsões associados às probabilidades de não-excedência definidas pelo usuário. Assim, por exemplo, se a probabilidade utilizada para definir o limite inferior foi de 20% e o limite superior foi de 70%, significa que o limite inferior para a previsão da semana $t+1$ é o valor da vazão observada na semana t multiplicada pela razão associada à probabilidade de não-excedência

de 20% e o limite superior é o valor da vazão observada na semana t multiplicada pela razão associada à probabilidade de não-excedência de 70%.

De posse destes limites os valores de previsão calculados que localizarem-se fora deste intervalo serão substituídos pelos limites mais próximos. Desta forma, os algoritmos que tiveram seus erros médios quadráticos influenciados por valores extremos e que, portanto deixaram de ser competitivas, após a consideração dos limites para previsão podem inclusive apresentar erros médios quadráticos inferiores a outras, que no critério original estavam em posição de vantagem.

As figuras 7.1.1 e 7.1.2 mostram um exemplo das distribuições de frequência acumuladas de não-excedência das razões entre as afluições das semanas $t+1$ e t para as duas metades de uma série histórica de afluições, as setas indicam os valores das razões que multiplicadas pela vazão observada da semana t fornecerá os limites para os valores previstos para a semana $t+1$. A figura 7.1.3 ilustra as previsões para uma semana qualquer $t+1$ e os valores fora e dentro dos limites das previsões.

O conjunto dos valores que serão utilizados para a construção da distribuição de frequência acumulada de não-excedência poderá ser constituído apenas dos valores observados para a semana de interesse ou pode-se agregar a estes valores os demais valores do mês, do trimestre ou do semestre que contém a semana a ser prevista. A distribuição de frequência poderá ainda estar associada a diferentes magnitudes de vazão. Neste caso a magnitude será dada pelos centis segundo o número de subdivisões escolhidas. Assim, se o interesse for adotar distribuições de frequências associadas a 2 faixas de magnitudes da vazão, então o centil considerado será de 50% (mediana), se for associada a 3 faixas os centis serão 33% e 66%, e se for a 4 faixas os centis serão 25%, 50% e 75%.

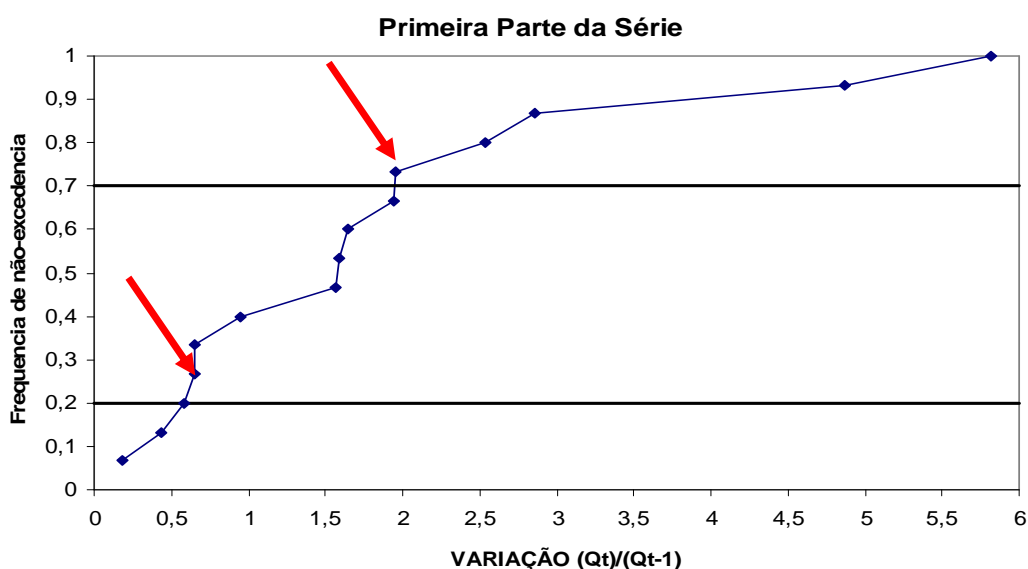


Figura 7.1.1 – Distribuição de frequência acumulada não-excedência da razão entre as semanas t+1 e t da primeira parte de uma série.

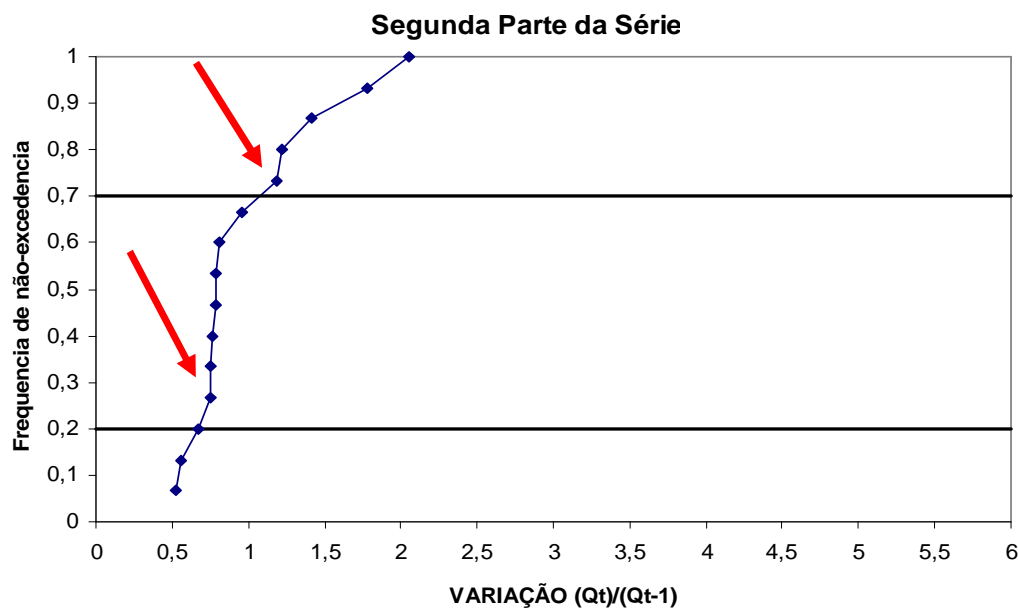


Figura 7.1.2 – Distribuição de frequência acumulada não-excedência da razão entre as semanas t+1 e t da segunda parte de uma série.

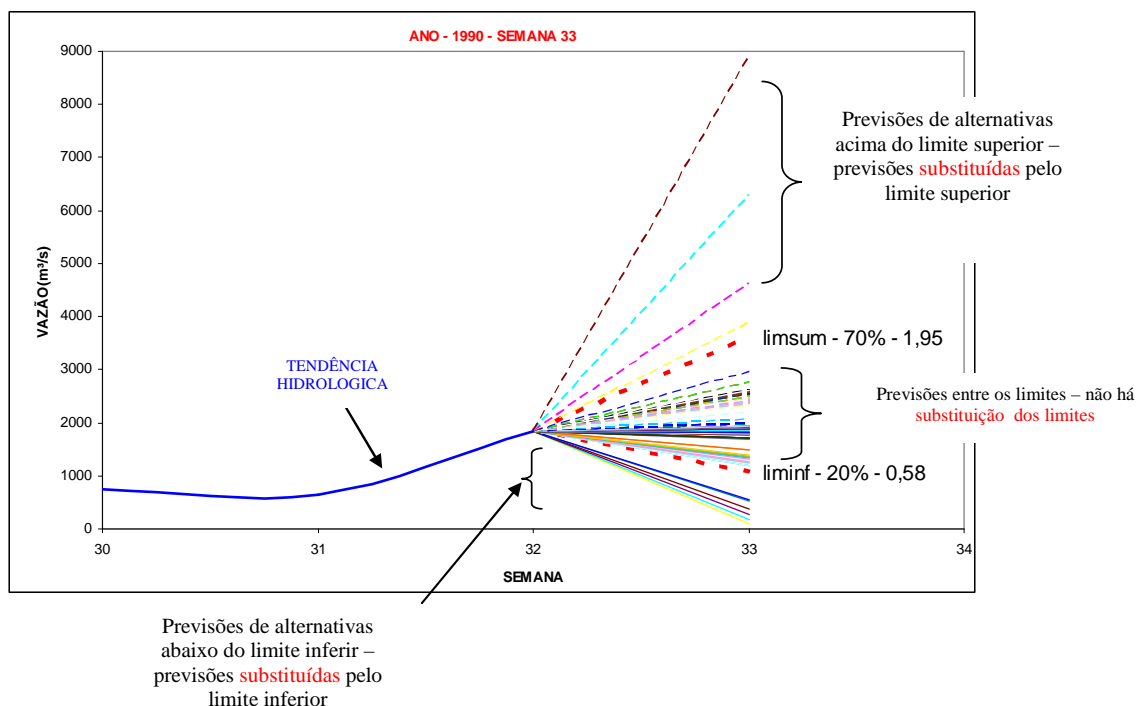


Figura 7.1.3 – Conjunto de previsões para a semana 33 obtidas com todos os algoritmos e limites inferior (associada a probabilidade de não-excedência de 20%) e superior (70%)

7.1.2 Aplicação de limites na previsão do modelo PREVIVAZ

Na etapa de previsão é utilizado o mesmo critério para definição dos limites adotados na alteração da escolha do melhor algoritmo descrito no item 7.1.1, entretanto, nesta etapa utiliza-se toda a série histórica de aflúências. É também respeitada a hierarquia dos algoritmos definida pelo menor erro médio quadrático. Deste modo, não há substituição do valor previsto, mas a alteração do melhor algoritmo caso o valor da previsão obtida por este não se encontre dentro dos limites para os valores previstos. As etapas para alteração do algoritmo de previsão são:

1. Para a semana da previsão (t+1) calcular a previsão utilizando o melhor algoritmo (aquele que resultou no menor erro médio quadrático, considerando o procedimento descrito no item 7.1.1), ou seja, o algoritmo de ordem $k=1$;

2. Estabelecer a distribuição de frequência acumulada de não-excedência para a semana t+1 das razões $(Q_i(t+1)/Q_i(t), i=1,...,N)$, levando-se em conta a série histórica

completa, onde N corresponde ao número de anos da série histórica se o conjunto de valores considerados para construção da distribuição de frequência forem apenas os da semana $t+1$;

3. Definir os limites a partir da distribuição de frequência acumulada de não-excedência e das probabilidades de não-excedência pré-estabelecidas;

4. Verificar se a razão entre a previsão da semana $t+1$ com o valor observado (ou previsto) da semana t é menor que o limite inferior ou maior que o limite superior;

5. Se a situação da etapa 4 for falsa a previsão da semana $t+1$ continua a mesma; se for verdadeira a previsão é substituída pela previsão do algoritmo de ordem $(k+1)$ e a verificação descrita na etapa 4 é refeita. A previsão é substituída até que o valor previsto por um algoritmo se encontre entre os limites inferior e superior;

6. Repetir as etapas 1 a 5 para as 6 semanas do horizonte de previsão.

Pode ocorrer na verificação dos valores das previsões obtidas pelos k algoritmos que todas estejam fora dos limites pré-estabelecidos. Neste caso existem três possibilidades:

- (i) caso todas as previsões estejam abaixo do limite inferior, a previsão escolhida será a do algoritmo que apresentou o maior valor (mais próximo do limite inferior);
- (ii) caso todas as previsões estejam acima do limite superior, será utilizada a previsão obtida pelo algoritmo que teve a menor previsão (mais próximo do limite superior); e
- (iii) finalmente, na situação de todos os valores das previsões encontrarem-se fora dos limites, porém alguns acima do limite superior e os outros abaixo do limite inferior, determina-se, então, os números de previsões acima do limite superior e abaixo do limite inferior e escolhe-se aquele que estiver mais próxima do limite além do qual se encontra o maior número de previsões.

A figura 7.1.4 mostra o conjunto de previsões para uma semana qualquer $t+1$ com um indicativo dos algoritmos que apresentaram os 4 menores erros médios quadráticos na etapa de escolha do melhor algoritmo. Observa-se nesta figura que utilizando a opção de alteração do algoritmo de previsão, o valor da previsão será o do algoritmo de 4ª ordem (4º menor erro

médio quadrático na etapa de escolha do melhor algoritmo) visto que verificando as previsões dos algoritmos de 1^a, 2^a e 3^a ordem todas estão fora dos limites pré-estabelecidos.

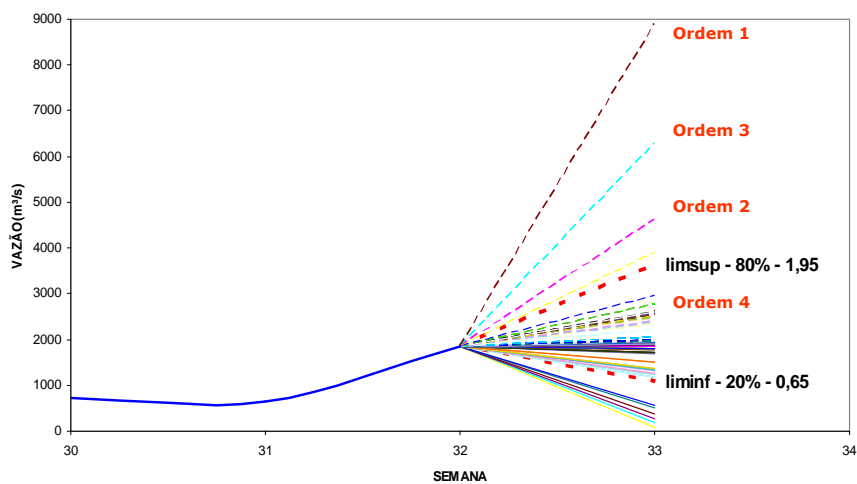


Figura 7.1.4 – Conjunto de previsões para uma semana t+1 com algoritmos que extrapolam ou não os limites pré-definidos.

8. CONCLUSÕES

Este relatório descreve a metodologia implementada no modelo PREVIVAZ. O objetivo do modelo é fornecer as previsões semanais de aflúncias aos aproveitamentos do Sistema Hidroelétrico Brasileiro para um horizonte de até 6 semanas.

A abordagem utilizada baseia-se em modelos lineares de séries temporais univariadas. Neste caso, as previsões de um aproveitamento são obtidas utilizando-se apenas observações passadas das aflúncias ao aproveitamento em questão, não sendo, portanto, utilizadas as informações sobre o comportamento das aflúncias aos aproveitamentos próximos, ou de postos pluviométricos na área de drenagem do aproveitamento em questão.

Cabe ressaltar que devido a sua modelagem, apesar do modelo permitir fazer previsões para um horizonte de até 6 semanas, as previsões mais afastadas da origem da previsão, para as quais a tendência hidrológica passa a ter menos influência, tendem à vazão média de longo termo (MLT) da respectiva semana, preservando o comportamento sazonal.

A consideração de limites nos valores das previsões que permitem a alteração na etapa de escolha do melhor algoritmo de previsão melhora a capacidade preditiva do modelo.

9. REFERÊNCIAS

- [1] NIMER, E., Clima, In: Geografia do Brasil, Região Sudeste, Fundação Instituto Brasileiro de Geografia e Estatística – IBGE, Rio de Janeiro, 1977.
- [2] CEPEL, Modelo de Previsão de Vazões Semanais Aplicado ao Sistema Hidroelétrico Brasileiro – Modelo PREVIVAZ, Relatório Técnico 125/97 CEPEL, Rio de Janeiro, Brasil, 1997.
- [3] MACEIRA, M.E.P., JARDIM, D.L.D.D., 2000, Modelo de Geração de Séries Sintéticas de Energias e Vazões - GEVAZP - Manual de Referência, Relatório Técnico DPP/PEN 83/2000, CEPEL, Rio de Janeiro, 2000.
- [4] CEPEL, Modelo de Previsão de Vazões Mensais Aplicado ao Sistema Hidroelétrico brasileiro – Modelo PREVIVAZM – Manual do Usuário, CEPEL, 2002, Rio de Janeiro, Brasil.
- [5] TUCCI, C.E.M, Modelos Determinísticos, In: Modelos Para Gerenciamento de Recursos Hídricos, NOBEL/ABRH, Rio de Janeiro, 1987
- [6] HIPEL, K.W., Macleod, A.I., Time Series Modelling of Water Resources and Environmental Systems, Elsevier, 1994.
- [7] THOMAS, H. A., FIERING, M. B., *Mathematical Synthesis of Streamflow Sequences for the Analysis of River Basins by Simulation*, In Design of Water Resources Systems, A. Mass et al., Harvard University Press, Cambridge, Massachusetts ,1962.
- [8] YEVJEVICH, V., *Fluctuations of Wet and Dry Years Part I*, Research Data Assembly and Mathematical Models, Hydology Paper 1, Colorado State University, Fort Collins, Colorado, 1963.
- [9] BOX, G. E. P., JENKINS, G. M., *Time Series Analysis-Forecasting and Control*, Holden-Day, 1970.
- [10] HARVEY, A. C., *Time Series Models*, P. Allan, 1981
- [11] GERE, J. M., WEAVER Jr. W., *Matrix Algebra for Engineers*, PWS Engineering, Boston, Massachusetts, 1983.

[12] BOX G.E.P., COX, D.R., “An Analysis of Transformations”, Journal of the Royal Statistical Society, A127, 211-252, 1964.

[13] DOOGE, J. C. I., Mathematical Models of Hydrologic Series, In: Biswas, A.K. (Ed.) Modelling of Water Resource Systems, Ed. A. Biswas Montreal, Harvest House, Vol 1, 1972.