

UNIVERSIDADE FEDERAL DO PARANÁ

ELIAS ALBERTO DA SILVA

DETECÇÃO E TRATAMENTO DE *OUTLIERS* EM SÉRIES TEMPORAIS  
APLICADO AO CONTROLE ESTATÍSTICO DE INSTRUMENTAÇÃO DE  
BARRAGENS

CURITIBA

2018

ELIAS ALBERTO DA SILVA

DETECÇÃO E TRATAMENTO DE *OUTLIERS* EM SÉRIES TEMPORAIS  
APLICADO AO CONTROLE ESTATÍSTICO DE INSTRUMENTAÇÃO DE  
BARRAGENS

Monografia apresentada como requisito parcial à obtenção do título de Especialista em Métodos Numéricos ao Programa de Pós-Graduação de Métodos Numéricos em Engenharia, na Área de Concentração de Programação Matemática e na Linha de Pesquisa em Métodos Estatísticos Aplicados à Engenharia, da Universidade Federal do Paraná – UFPR.

Orientador: Prof. Dr. Emerson Lazzarotto  
Coorientador: Prof. Dr. Eduardo Cesar Dechechi

CURITIBA

2018

## FOLHA/TERMO DE APROVAÇÃO

ELIAS ALBERTO DA SILVA

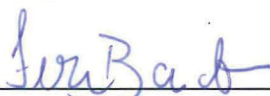
### DETECÇÃO E TRATAMENTO DE OUTLIERS EM SÉRIES TEMPORAIS APLICADO AO CONTROLE ESTATÍSTICO DE INSTRUMENTAÇÃO DE BARRAGENS

Monografia aprovada como requisito parcial à obtenção do título de Especialista, Curso de Especialização em Métodos Numéricos em Engenharia, Setor de Tecnologia/Setor de Ciências Exatas, Universidade Federal do Paraná, pela seguinte banca examinadora:



Prof. Dr. Emerson Lazzarotto

Prof. Dr. Emerson Lazzarotto  
Orientador – Centro de Engenharias e Ciências Exatas - UNIOESTE



Prof. Dr. Fernando Mucio Bando  
Centro de Engenharias e Ciências Exatas - UNIOESTE



Prof. Dr. Luciano Panek  
Centro de Engenharias e Ciências Exatas - UNIOESTE

Dedico aos meus familiares e amigos que sempre estiveram ao meu lado.

## **AGRADECIMENTOS**

Na realização deste trabalho tive muitos colaboradores e recebi diversos apoios e é com satisfação que agradeço:

Ao Prof. Dr. Emerson Lazzarotto, pelo grande trabalho de orientação, e pela confiança desde o tempo da minha graduação.

Ao meu coorientador, Prof. Dr. Eduardo Cesar Dechechi, pela colaboração, disponibilidade e sugestões.

Aos professores que desempenharam com muita competência as aulas ministradas, em especial a Profa. Dra. Liliana e ao Prof. Dr. Anselmo pelos esforços dedicados para a realização da especialização.

Ao CEASB-FPTI/BR e à Itaipu Binacional por viabilizarem a abertura de uma turma de Especialização fora de sede do PPGMNE e pela cessão de dados e autorização para divulgação. Em especial à Alexandra, pela confiança, apoio e incentivo, e ao engenheiro Dimilson, por sua disponibilidade e pelas informações sobre Itaipu.

A todos os meus colegas de turma, pela amizade e pela ajuda nos momentos de dificuldade.

À minha supervisora Cristiane Fraga e ao meu gerente Celso Torino, pela compreensão e permitirem minha liberação do trabalho para assistir as aulas quando foi preciso.

A todos os colegas de trabalho que me apoiaram e incentivaram durante todo esse tempo.

Aos meus grandes amigos Marcos, Eder e Márcio, pelos momentos de descontração e pela amizade de longa data, sem os quais tudo seria mais difícil.

A vida nunca está completa sem seus desafios.

Stan Lee

## RESUMO

Na análise de dados de instrumentação de barragem, além da sazonalidade, é comum encontrar a presença de observações influenciadas por eventos externos e de forma não sistemática ou erros de medição. Neste trabalho avalia-se o efeito de *outliers* na modelagem de algumas séries temporais de instrumentos que monitoram o trecho E da barragem de ITAIPU Binacional, situada no Rio Paraná, entre o Brasil e Paraguai. No estudo, realiza-se o ajustamento de um modelo ARIMA Sazonal a série de observações do instrumento. Para validar os modelos ajustados considera-se a capacidade das suas previsões em termo do erro cometido. A avaliação do efeito de *outliers* na série se dá pela comparação do teste de normalidade e a carta de controle para a média dos resíduos das previsões, com e sem a presença de valores atípicos. Os resultados indicam que *outliers* influenciam significativamente na normalidade e na identificação de valores fora de controle estatístico dos resíduos das previsões da série temporal do instrumento utilizado. Portanto, é fundamental a identificação e o tratamento de valores atípicos.

Palavras-chave: *Outliers*; monitoramento numérico de barragens; gráficos de controle; séries temporais; modelos SARIMA.

## ABSTRACT

In the analysis of dam instrumentation data, in addition to the seasonality, it is common to find the presence of observations influenced by external events and in a non-systematic way or measurement errors. This work evaluates the effect of outliers on the modeling of the monitoring instruments time series in the section E of the ITaipu Binacional dam, located on the Paraná River between Brazil and Paraguay. In this study, a SARIMA model is adjusted in some time series of instrument observations. In order to validate the adjusted models, it is considered the capacity of their forecasts in order of the committed error. The evaluation of the effect of outliers in the time series is done by comparing the normality test and the control chart for the mean of the predicted residuals, with and without the presence of atypical values. The results indicate that outliers significantly influence the normality and the identification of values outside statistical control of the time series residuals predictions of the used instrument. Therefore, the identification and treatment of outliers is very important.

Key-words: Outliers; numerical monitoring instruments; control charts; time series; SARIMA models.



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>10</b>
<b>1.1</b>	<b>O PROBLEMA.....</b>	<b>10</b>
<b>1.2</b>	<b>OBJETIVOS.....</b>	<b>11</b>
1.2.1	Objetivo Geral.....	11
1.2.2	Objetivos Específicos.....	12
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>13</b>
<b>2.1</b>	<b>Segurança e Instrumentação de Barragens.....</b>	<b>13</b>
<b>2.2</b>	<b>Cartas de Controle.....</b>	<b>17</b>
<b>2.3</b>	<b>Séries Temporais e Modelagem SARIMA.....</b>	<b>20</b>
<b>2.4</b>	<b><i>Outliers</i> em Séries Temporais.....</b>	<b>26</b>
<b>3</b>	<b>MATERIAL E MÉTODOS.....</b>	<b>34</b>
<b>3.1</b>	<b>MATERIAL.....</b>	<b>34</b>
<b>3.2</b>	<b>MÉTODOS.....</b>	<b>38</b>
<b>4</b>	<b>RESULTADOS.....</b>	<b>40</b>
<b>4.1</b>	<b>Extensômetro 1.....</b>	<b>40</b>
<b>4.2</b>	<b>Extensômetro 2.....</b>	<b>44</b>
<b>5</b>	<b>CONCLUSÕES.....</b>	<b>48</b>
	<b>REFERÊNCIAS.....</b>	<b>49</b>

## 1 INTRODUÇÃO

### 1.1 O PROBLEMA

A Usina Hidrelétrica Binacional de Itaipu, localizada no Rio Paraná, na divisa entre o Brasil e o Paraguai, possui mais de dois mil instrumentos para a análise da auscultação da barragem com vistas à garantia da contínua manutenção de sua operação. Esses instrumentos geram uma grande quantidade de dados, que precisam ser tratados e transformados em informações úteis para auxiliar na tomada de decisão dos profissionais responsáveis pela segurança da barragem.

Na análise de dados de instrumentação de barragem, além da sazonalidade, é comum encontrar a presença de observações influenciadas por eventos externos e de forma não sistemática ou mesmo erros de medição. Essas observações são conhecidas na literatura como valores atípicos ou *outliers*. O conceito de *outliers* em séries temporais foi introduzido por Fox<sup>1</sup> em 1972, considerando dois tipos: *outliers* aditivos (AO – *additive outlier*), *outlier* de inovação (IO – *innovation outlier*). Como extensão do trabalho de Fox, num artigo de referência na área, Chen e Liu (1993), consideraram as seguintes alterações na estrutura da série: mudança de nível permanente (LS – *level shift*) e mudança temporária ou transitória (TC – *temporary change*).

Estes quatro tipos de *outliers*, AO, IO, LS e TC, interferem na série de modo diferente, ora causando impacto somente no instante da sua ocorrência, ora afetando toda série subjacente ao instante que ocorre a perturbação. Chen e Liu (1993) mostraram que a presença de dados atípicos afeta os procedimentos convencionais de análise estatística, além de prejudicar as estimativas dos parâmetros e as possíveis previsões obtidas a partir dos modelos ARMA ajustados.

De acordo com a natureza do *outlier*, seus efeitos podem ser expressivos sobre os resultados obtidos, influenciando de forma significativa inferências realizadas sobre os modelos ajustados e, conseqüentemente, podendo levar a erros de análises de previsões, bem como na avaliação do controle estatístico do conjunto

---

<sup>1</sup> FOX, A. J. Outliers in Time Series. Journal of the Royal Statistical Society. Series B. Vol. 34, No. 3 (1972), pp. 350-363.

de dados. A presença de valores atípicos tem efeito na estimação dos parâmetros do modelo ARIMA e, conseqüentemente, na especificação correta do modelo e nas previsões de valores futuros (MORETTIN e TOLOI, 2006).

Estes eventos são tipicamente não sistemáticos e não podem ser capturados por modelos básicos de séries temporais. Uma análise preliminar dos dados pode revelar a existência de eventos excepcionais que não seguem aos padrões capturados pelos modelos de séries temporais, isto é, efeitos que são exógenos ao modelo (LACALLE, 2016).

Esta análise pode ser trabalhosa ou inviável se o conjunto de dados for grande. Além disso, às vezes, esses efeitos não se destacam nas representações gráficas e permanecem mascaradas com a dinâmica geral dos dados. Se existe algum conhecimento *a priori* sobre esses efeitos, então uma variável explicativa pode ser explicitamente definida e incluída no modelo por meio de variáveis de intervenção. Quando este efeito exógeno não é atribuído a qualquer evento conhecido, diz-se que há um *outlier* na série temporal (LACALLE, 2016).

Este trabalho aplica a metodologia proposta por Chen e Liu (1993) para localizar e eliminar os efeitos de valores discrepantes nos dados das séries temporais de dois instrumentos de monitoramento da barragem da usina hidrelétrica Itaipu Binacional com o objetivo de elaborar cartas de controle e seus limites estatísticos para as médias amostrais, além de reduzir os erros nas previsões.

## 1.2 OBJETIVOS

### 1.2.1 Objetivo Geral

Identificar e tratar *outliers* nos dados com métodos conhecidos da literatura, diminuindo o erro cometido pelo modelo de série temporal ajustada à série histórica da leitura de instrumentos que monitoram a barragem da usina hidrelétrica Itaipu Binacional.

### 1.2.2 Objetivos Específicos

- a) Elaborar cartas de controle e seus limites estatísticos para a média amostral da série temporal de alguns instrumentos de monitoramento da barragem da usina hidrelétrica Itaipu Binacional;
- b) Determinar o período da série que será considerado a fim de se obter um modelo com menor erro de previsão.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 Segurança e Instrumentação de Barragens

As barragens, por serem grandes obras de engenharia, estão sujeitas a grandes acidentes. A ruptura de uma barragem pode ocasionar extensos alagamentos a jusante e causar enorme impacto social, econômico e ambiental. Não basta uma construção segura, é necessário assegurar que sua operação também o seja. São necessários o acompanhamento e a avaliação permanente do comportamento da barragem. Isto se dá através de um processo, denominado auscultação da barragem, que consiste de inspeções visuais e por instrumentação (FUSARO, 2007).

A Lei 12.334 (CASA CIVIL DA PRESIDÊNCIA DA REPÚBLICA, 2010) estabelece a Política Nacional de Segurança de Barragens do Brasil (PNSB), define em seu artigo 2º parágrafo III que: “segurança de barragem é uma condição que visa manter a sua integridade estrutural e operacional e a preservação da vida, da saúde, da propriedade e do meio ambiente”, já no artigo 4º, destaca-se dentre outros fundamentos da PNSB, o inciso V: “a segurança de uma barragem influi diretamente na sua sustentabilidade e no alcance de seus potenciais efeitos sociais e ambientais”.

Deve ser dada atenção contínua a operação e a manutenção da segurança de uma barragem de usina. A ferramenta para isto é a instrumentação da barragem. O conceito básico da instrumentação é o monitoramento da infiltração, pressão e deslocamentos. Por conta das particularidades de cada usina hidrelétrica, não há procedimentos universais aplicáveis a todas as usinas no que se refere à avaliação da instrumentação. A avaliação e o julgamento das informações disponíveis por uma experiente equipe de engenharia são a melhor forma de contribuir para uma tomada de decisão e a escolha da melhor ação a ser executada (USACE, 1995, *apud* LAZZAROTTO, 2016).

A instrumentação é um conjunto de dispositivos instalados nas barragens ou perto delas para efetuar medições que podem ser usadas para avaliar o comportamento estrutural e os parâmetros de desempenho da estrutura (ICOLD - CIGB, 2008).

Dentre outros requisitos técnicos e administrativos, as barragens devem possuir instrumentação adequada para monitoramento do seu desempenho, o objetivo é permitir o controle da operação da barragem com segurança sob qualquer condição. As condições climáticas, afluência, nível de reservatório, vazão e níveis do rio a jusante devem ser monitoradas. A instrumentação instalada deve permitir a medição do comportamento estrutural e a condição física da barragem (USACE, 1995, *apud* LAZZAROTTO, 2016).

Um programa de instrumentação contribui muito para a segurança de uma barragem. Para a correta compreensão do comportamento de uma barragem, especialmente com o seu envelhecimento, é importante aplicar o conhecimento em quantidade e qualidade que for exigido. Deve haver pessoal qualificado, equipamentos de qualidade e informação adequada e rápida para avaliação (USACE, 1995, *apud* LAZZAROTTO, 2016). Em especial, esta é a proposta do CEASB – Centro Avançado de Estudo de Segurança de Barragem da usina hidrelétrica de Itaipu Binacional, localizada no Rio Iguazu, entre o Brasil e o Paraguai.

Como a terra e as rochas são formações naturais, ao contrário de outros materiais da engenharia, tais como aço e concreto, elas estão mais sujeitas a incertezas, portanto, observações visuais e a instrumentação adequada fornecem informações importantes sobre a manutenção das premissas de um projeto e a avaliação do desempenho e a segurança de uma barragem de usina. Quando uma falha, uma condição severa, mudança visualmente observada ou uma infiltração tenha ocorrido em uma barragem, a instrumentação pode ser extremamente útil na determinação das causas do evento, bem como, quando ações corretivas são implantadas e se deseja avaliar o efeito do tratamento (USACE, 1995, *apud* LAZZAROTTO, 2016).

Outro benefício da análise dos dados de instrumentação é detectar um período de desempenho satisfatório. Isto pode servir para que em uma futura variação dos dados seja indicado algum potencial problema. Os dados também podem ser úteis na previsão do desempenho sob condições máximas de operação como, por exemplo, no caso de enchentes. A análise do desempenho de uma barragem e dos dados de instrumentação gerados durante sua fase de operação

pode ser usada no planejamento e na construção de novas barragens mais seguras e econômicas (LAZZAROTTO, 2016).

Outra importante etapa de um sistema de instrumentação é o gerenciamento dos dados. O gerenciamento consiste na coleta, redução e processamento e apresentação. Em especial na apresentação, gráficos são necessários para: fornecer comparações visuais entre o comportamento real e o previsto, facilitar a detecção de erros na aquisição de dados, determinar tendências ou efeitos cíclicos, comparar o comportamento com outros instrumentos, prever o comportamento futuro, determinar eventual necessidade de manutenção da instrumentação. Os gráficos permitem que os dados sejam comparados facilmente com eventos que causam alterações nos dados, tais como mudanças ambientais, bem como, avaliar o comportamento imprevisto e determinar a eficácia de uma ação corretiva (LAZZAROTTO, 2016).

A análise dos dados consiste em sua avaliação considerando sua influência sob várias condições. Deve haver consciência da existência de dados inválidos e cálculos indevidos a fim de que não se incorra em interpretações incorretas. Uma análise adequada deve abordar dois aspectos básicos de monitoramento de segurança de barragens: o desempenho do sistema de instrumentação e o desempenho da estrutura ou característica que está sendo monitorada. Deve haver precaução sobre a precisão de gráficos de controle gerados automaticamente por computadores. Estes não devem ser utilizados e interpretados sem uma cuidadosa revisão de pessoal especializado (USACE, 1995, *apud* LAZZAROTTO, 2016).

A interpretação dos dados deve ser realizada em tempo hábil, preferencialmente logo após a coleta dos dados. Programas de computador que geram gráficos de controle da leitura da instrumentação normalmente possuem limites de alerta baseados em critérios determinísticos, isto é, estabelecidos com base em critérios de projeto e que há poucos sistemas informatizados de alerta baseados em critérios estatísticos (FUSARO, 2007).

Os valores de controle para a leitura da instrumentação na barragem de Itaipu foram estabelecidos com base no método de elementos finitos (método numérico usado na determinação da solução de sistemas estruturais complexos, como

barragens, modelos reduzidos estruturais e geomecânicos) e critérios de projeto (ITAIPU BINACIONAL, 1984).

Os valores obtidos de análises teóricas ou modelos físicos são baseados em certos parâmetros e hipóteses simplificadoras, salientando ainda que deva haver a permanente confrontação entre os valores estabelecidos no projeto com aqueles obtidos nas leituras, não impedindo que eventuais leituras ultrapassem os limites de projeto (ITAIPU BINACIONAL, 1999\_a).

Uma definição para valores de controle é que:

Valores de controle são aqueles mais próximos da realidade física, considerando as influências térmicas ambientais atuantes sobre as estruturas e modelos reológicos mais realistas do concreto e da fundação. Os valores de controle são utilizados para o acompanhamento do desempenho das estruturas, por meio das leituras efetuadas “in situ”, e constituem sinais de alerta para situações anormais do comportamento estrutural. Estes valores, para o caso de Itaipu, já foram determinados anteriormente na fase de projeto e do enchimento do reservatório, não sendo, em muitos casos, mais aplicáveis na fase de operação (ITAIPU BINACIONAL, 1999\_b, p. 1).

De acordo com Fusaro (2007), as grandezas físicas avaliadas por instrumentos de auscultação de barragens geram séries temporais que podem ser tratadas por métodos estatísticos em substituição a modelos determinísticos ou analíticos que, neste caso, devido à geometria e ao contorno e à variabilidade nos parâmetros dos materiais de construção, tornam a implementação extremamente complexa não sendo possível eliminar todas as incertezas sobre os parâmetros geotécnicos.

Os valores da leitura de um instrumento de monitoramento de barragem possuem uma variabilidade inerente, associada ao próprio instrumento, a fatores humanos e a fatores de instalação e utilização (FUSARO, 2007).

Dentre as principais anomalias que a instrumentação pode detectar estão: a infiltração (*piping*) através dos maciços de terra e suas fundações, eventuais trincas com infiltrações no concreto, movimentos bruscos entre os blocos da barragem, subpressões excessivas na cabeça dos blocos a montante, ou a jusante da Casa de Força, movimentos excessivos na região da calha do vertedouro (ITAIPU BINACIONAL, 2002).



## 2.2 Cartas de Controle

Os gráficos ou cartas de controle servem para avaliar a qualidade em termos da variabilidade de uma característica de interesse, neste caso, as leituras de instrumentos. Consiste de uma linha central que representa o alvo, caso não houvesse variação e duas outras linhas, o limite superior e inferior de controle, determinados com base estatística (MONTGOMERY, 2013). Quando somente causas comuns (inerentes) de variação estiverem presentes no processo de controle ele é dito estável ou sob controle estatístico e quando há causas especiais de variação o processo é instável ou fora de controle estatístico (MONTGOMERY, 2013).

Um modelo geral para um gráfico de controle de uma estatística amostral  $w$  que mede alguma característica de qualidade de interesse com média  $\mu_w$  e desvio padrão  $\sigma_w$  terá como limite superior de controle (LSC), valor da linha central (LC) e limite inferior de controle (LIC) os valores

$$\begin{aligned} LSC &= \mu_w + L\sigma_w, \\ LC &= \mu_w, \\ LIC &= \mu_w - L\sigma_w, \end{aligned}$$

onde  $L$  é a distância dos limites de controle à linha central expressa em unidades de desvio padrão. Como estes princípios foram estabelecidos inicialmente por Walter S. Shewhart, tais gráficos de controle são chamados de gráficos de controle de Shewhart (LAZZAROTTO, 2016).

Quando o tamanho das amostras disponíveis é de somente uma unidade, caso da aplicação aqui realizada, dois dos mais importantes gráficos de controle executados são da média e da variabilidade do processo (MONTGOMERY, 2013). Estas variáveis são representadas, no caso de amostras individuais, respectivamente, pela variável  $x_t = x(t)$  e pela variável  $R$  que estima a variabilidade do processo por meio da amplitude móvel de duas observações consecutivas, isto é,  $R_i = |x_i - x_{i-1}|$ . Os limites de controle para uma probabilidade de 99,73%, no caso de amostras individuais, presumindo normalidade, do gráfico  $x$  são (MONTGOMERY, 2013):

$$\begin{aligned}
 LSC &= \bar{x} + 2,6595 \cdot \bar{R} \\
 \text{Linha Central} &= \bar{x} \\
 LIC &= \bar{x} - 2,6595 \cdot \bar{R}.
 \end{aligned}
 \tag{1}$$

O gráfico de controle avalia a hipótese nula de que o processo está sob controle. Um falso alarme significa rejeitar que o processo está sob controle num instante quando, na verdade, está sob controle. Este tipo de erro, chamado erro tipo I e denotada por  $\alpha$ , pode ser calculado por meio de uma probabilidade. Da mesma forma, se está sujeito ao erro tipo II, que significa aceitar que o processo está sob controle, quando, na verdade, não está. Conforme a necessidade, pode-se fixar o valor do múltiplo do desvio padrão e calcular o valor de  $\alpha$  ou vice-versa (MONTGOMERY, 2013).

O uso padrão do gráfico de controle envolve as aplicações das fases I e II, com dois diferentes objetivos. Na fase I, um conjunto de dados do processo é coletado e analisado de uma vez, em uma análise retrospectiva, construindo-se limites de controle de teste para se determinar se o processo estava sob controle durante o período em que os dados foram coletados, e para se verificar se limites de controle confiáveis podem ser estabelecidos para o monitoramento de produção futura. Essa é, tipicamente, a primeira coisa a ser feita quando se aplicam gráficos de controle a qualquer processo (MONTGOMERY, 2013).

Na fase I, os gráficos de controle ajudam a levar o processo para um estado de controle estatístico. A fase II se inicia após termos um conjunto “limpo” de dados de processo coletados sob condições estáveis e representativas do desempenho do processo sob controle. Na fase II, usamos o gráfico de controle para monitorar o processo, comparando a estatística amostral para cada amostra sucessiva, à medida que ela é extraída do processo, com os limites de controle (MONTGOMERY, 2013).

É bastante característico assumir-se, na fase I, que o processo esteja inicialmente fora de controle, de modo que o objetivo da análise é levar o processo a um estado de controle estatístico. Pontos que se situam fora dos limites de controle são investigados, à procura de causas atribuíveis. Pontos fora dos limites de controle são, então, eliminados e um novo conjunto de limites de controle revisado é

calculado. Então, novos dados são coletados e comparados com esses limites revisados (MONTGOMERY, 2013).

Algumas vezes, esse tipo de análise exigirá vários ciclos de análise nos quais o gráfico de controle é usado, causas atribuíveis são detectadas e corrigidas, limites de controle revisados são calculados, e o plano de ação para situação fora de controle é atualizado e expandido. Eventualmente, o processo se estabiliza, e obtém-se um conjunto limpo de dados que representam o desempenho do processo sob controle para uso na fase II. Geralmente, gráficos de controle de Shewhart são muito eficazes na fase I por serem de fácil construção e interpretação, e porque são eficazes na detecção tanto de mudanças grandes, contínuas nos parâmetros do processo, quanto de valores atípicos (desvios únicos que podem ser resultados de causas atribuíveis de curta duração), erros de medida, erros de registro e/ou transmissão de dados, e outros semelhantes (MONTGOMERY, 2013).

A suposição mais importante relativa aos gráficos de controle é a de independência das observações, porque os gráficos de controle convencionais não funcionam bem se a característica da qualidade apresentar níveis de correlação, ainda que baixos, ao longo do tempo. Especificamente, esses gráficos de controle darão resultados enganosos sob a forma de demasiados alarmes falsos se os dados são correlacionados (MONTGOMERY, 2013).

O uso de gráficos de controle para dados influenciados por variáveis ambientais fornece muitos falsos alarmes devido ao fato dos dados serem correlacionados. Este fenômeno é típico em instrumentos de monitoramento de barragens, cujas leituras são influenciadas por variações térmicas e do nível do reservatório (ROSSO *et al.*, 1995). Uma abordagem que se tem revelado útil com dados autocorrelacionados consiste em se modelar diretamente a estrutura de autodependência com um modelo de série temporal apropriado, usar esse modelo para remover a autocorrelação dos dados, e aplicar gráficos de controle aos resíduos (MONTGOMERY, 2013).

## 2.3 Séries Temporais e Modelagem SARIMA

Uma série temporal é uma sequência de dados observados em intervalos regulares de tempo. O estudo das séries temporais pode ser motivado pelo interesse em investigar o mecanismo gerador de um conjunto de dados observados ao longo do tempo, para descrever sua dinâmica, com o objetivo de gerar previsões acerca do seu comportamento futuro (MORETTIN e TOLOI, 2006).

Uma série temporal é considerada uma realização de um processo estocástico que pode ser descrito como um fenômeno estatístico que se desenvolve no tempo de acordo com leis probabilísticas e que é caracterizado por uma equação estocástica (modelo), a partir da qual pode se obter alguns dos seus momentos (média e variância). Em processos estocásticos o futuro é incerto. Os modelos de processos estocásticos tentam capturar a estrutura de dependência existente no passado da série para realizar previsões. O grau de acerto das previsões depende assim de dois fatores chaves: i) que o modelo seja uma boa aproximação da estrutura de dependência da série (testes de especificação) e ii) que a estrutura de dependência identificada no passado permaneça razoavelmente estável no futuro (testes de previsão fora da amostra) (FERNANDES, 2006, *apud* LAZZAROTTO, 2016).

A maioria dos métodos estatísticos trata de amostras aleatórias de observações independentes. Uma característica especial da análise de séries temporais é que observações sucessivas geralmente não são independentes e guardam uma autocorrelação, nesse caso os valores futuros podem ser previstos das observações passadas. Se a série pode ser prevista exatamente ela é dita ser determinística, mas na maioria das vezes apenas parcialmente se pode prever valores futuros dos valores passados e nesse caso ela é estocástica e se substitui a ideia da previsão exata dos valores futuros pela ideia que estes têm uma distribuição de probabilidades condicionada pelo conhecimento dos valores passados (CHATFIELD, 1996, *apud* LAZZAROTTO, 2016).

Diz-se que uma série temporal é estacionária se ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, reflexo de alguma forma de equilíbrio estável. As séries podem apresentar várias formas de não

estacionariedade, como tendência linear, em que flutua ao redor de uma reta com inclinação não nula ou tendência exponencial (LAZZAROTTO, 2016). Mais especificamente, costuma-se adotar estacionariedade de 2ª ordem (fraca) de uma série  $y_t$  se a média e a variância (incondicionais) do processo permanecem invariantes ao longo do tempo.

A metodologia mais utilizada para a modelagem de séries temporais é a apresentada por Box e Jenkins em 1970, a qual é baseada na utilização de modelos que consideram as partes autorregressivas e de médias móveis. Esses modelos são conhecidos como modelos ARMA (Autorregressivos e de Médias Móveis) e são amplamente utilizados em diversas áreas do conhecimento.

Os modelos autorregressivos integrados de médias móveis (ARIMA) são modelos lineares capazes de descrever satisfatoriamente séries estacionárias e não estacionárias que não apresentem comportamento explosivo, caso em que, por exemplo, a série flutua ao redor de um nível, depois muda de nível e flutua ao redor de outro nível, e assim por diante (MORETTIN e TOLOI, 2006).

Seja  $y_t (t = 1, \dots, T)$  uma série temporal estacionária (ou não estacionária que possa ser estacionária mediante algum tipo de transformação como, por exemplo, diferenciações). Denotando o operador de translação ao passado por  $B$ , pode-se escrever que  $By_t = y_{t-1}$ ,  $B^m y_t = y_{t-m}$  e o operador diferença  $\nabla$  usada para obter estacionariedade, de modo que  $\nabla y_t = y_t - y_{t-1} = (1 - B)y_t$ , os modelos lineares estacionários  $ARMA(p, q)$  tem a forma

$$y_t = \mu + \sum_{i=1}^p \phi_i y_{t-i} + a_t - \sum_{i=1}^q \theta_i a_{t-i}, \quad (2)$$

onde  $\mu$  é a média,  $\phi_i$  os coeficientes da parte autorregressiva,  $\theta_i$  os coeficientes da parte de médias móveis e  $a_t \sim N(0, \sigma_a^2)$  é um ruído branco (BOX, JENKINS e REINSEL, 2008).

Quando a série não se desenvolve ao longo do tempo ao redor de uma média constante, situação comum na prática, modelos ARMA não são apropriados para descrevê-las. Séries não estacionárias  $Z_t$  em que ao tomar-se um número finito de diferenças  $d$ , tornam-se estacionárias, são chamadas de não estacionárias

homogêneas. Seja  $W_t = Z_t - Z_{t-1} = (1 - B)Z_t = \nabla Z_t$  e, mais geralmente, se  $W_t = \nabla^d Z_t$  é estacionária, então  $W_t$  pode ser representada por um modelo  $ARMA(p, q)$ , isto é,  $\phi(B)W_t = \theta(B)a_t$  (BOX, JENKINS e REINSEL, 2008).

Se  $W_t$  é uma diferença de  $Z_t$ , diz-se que  $Z_t$  é uma integral de  $W_t$ , por isso se diz que  $Z_t$  segue um modelo autorregressivo integrado de médias móveis (ARIMA) de ordem  $(p, d, q)$ ,  $\phi(B)\nabla^d Z_t = \theta(B)a_t$ , onde  $p$  e  $q$  são as ordens de autorregressão e de médias móveis, respectivamente. É usual representar na forma

$$\phi(B)Z_t = \underbrace{\phi(B)\nabla^d}_{\phi} Z_t = \phi(B)(1 - B)^d Z_t = \theta(B)a_t.$$

O modelo também pode ser escrito na forma  $AR(\infty)$  como  $\pi(B)Z_t = a_t$ , onde

$$\pi(B) = \frac{\nabla^d \phi(B)}{\theta(B)} = 1 - \pi_1 B - \pi_2 B - \dots, \quad (3)$$

ou na forma  $MA(\infty)$  como  $Z_t = \psi(B)a_t$ , onde

$$\psi(B) = \frac{\theta(B)}{\nabla^d \phi(B)}. \quad (4)$$

As séries temporais frequentemente mostram comportamento periódico (MORETTIN e TOLOI, 2006). Uma série temporal tem um padrão periódico quando se repete em todo período  $s, s > 1$ . O mais comum tipo de periodicidade é a sazonal.

Quando as séries temporais são do tipo com correlação serial entre e dentro dos períodos sazonais tem-se a seguinte formulação geral:

$$\Phi(B^s)\phi(B)\nabla_s^D \nabla^d Z_t = \Theta(B^s)\theta(B)a_t. \quad (5)$$

Onde

$$\Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps},$$

$$\Theta(B^s) = 1 - \theta_1 B^s - \dots - \theta_Q B^{Qs},$$

com  $\Phi_i, \theta_i = 0$  para  $i = 1, 2, \dots, s - 1, s + 1, \dots$ ,  $\Phi_i, \theta_i \neq 0$  para  $i = s, 2s, \dots, Ps/Qs$ ,  $\nabla_s^D = (1 - B^s)^D$  e  $s$  é o comprimento do período sazonal (MORETTIN e TOLOI, 2006).

A notação usual para o modelo multiplicativo sazonal SARIMA é:

$$ARIMA(p, d, q) \times (P, D, Q)_s. \quad (6)$$

Segundo Fernandes (2006) *apud* LAZZAROTTO (2016), a função de autocorrelação (FAC) representa uma medida de associação linear entre o presente do processo e o seu passado, ou seja, uma medida da “memória” do processo. A FAC é o procedimento padrão para investigar a dependência linear relacionada por um modelo de série. A ausência de autocorrelação de um processo estocástico não implica que este seja independente. Entretanto, se o processo estocástico for independente, a sua FAC, e qualquer outra medida de autoassociação será nula. A FAC de um processo estocástico é como se fosse uma “assinatura” deste processo. Processos de uma mesma “família” geram FAC’s do mesmo tipo. Portanto, a FAC pode ser utilizada como um procedimento para identificar o processo estocástico que gera uma série temporal.

Quando houver uma relação linear perfeita entre  $y_t$  e  $y_{t-k}$ , ou seja, entre uma observação no instante  $t$  e  $k$  instantes de tempo antes, então  $|\rho_k| < 1$  e se  $\rho_k = 0$ , pode haver algum outro tipo de relação entre  $y_t$  e  $y_{t-k}$  (relação não-linear) onde

$$\rho_k = \frac{\sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}. \quad (7)$$

Em séries estacionárias o correlograma decai rapidamente para zero com o aumento de  $k$  ou é estatisticamente nulo para todo  $k$ . Em séries não estacionárias o correlograma decai lentamente para zero com o aumento de  $k$ .

Alguns processos estocásticos não podem ser distinguidos através de suas FAC’s, pois estas apresentam formas similares. Este comportamento faz com que dada uma série temporal real, a partir da sua FAC estimada (correlograma) não se pode identificar se o modelo gerador da série é um AR(1) ou AR(2), por exemplo. É importante um tipo de função de autocorrelação que seja única para processos AR(p). Esta função existe e é denominada de FACP (Função de Autocorrelação Parcial).

Para estimar a dependência entre observações distantes de  $k$  unidades de tempo, deve se eliminar da dependência total, a contribuição dada pelas

observações intermediárias:  $y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-k+1}$ . O coeficiente de correlação parcial calcula a correlação “líquida” entre  $y_t$  e  $y_{t-k}$  eliminando as influências de  $y_{t-1}, y_{t-2}, y_{t-k+1}$ . É possível mostrar que o coeficiente de correlação parcial de ordem  $k$  coincide com o último parâmetro de um modelo  $AR(k)$ , onde  $\rho_k = \text{corr}(y_t, y_{t-k})$  e, mais geralmente, se conclui que a forma geral da FACP para um processo  $AR(p)$ , será  $\varphi_k = \varphi_{kk} = 0, k > p$  e  $\varphi_k = \varphi_{kk} \neq 0, k \leq p$ . Esta é a FACP teórica, obtida a partir do modelo. A FACP de ordem  $k$  estimada a partir de uma série temporal é obtida estimando-se o último coeficiente de um processo  $AR(k)$ . Se o valor deste coeficiente estiver dentro do intervalo de confiança de 95%, então, ao nível de 5%, não podemos rejeitar a hipótese de que  $\varphi_{kk} = 0$  (FERNANDES, 2006, *apud* LAZZAROTTO, 2016).

A maioria dos procedimentos em análise estatística de séries temporais presume a hipótese da estacionariedade. Quando esta não existir, uma alternativa é a transformação dos dados originais para obtenção de uma série estacionária. A transformação mais comum é tomar diferenças sucessivas da série original e denota-se a primeira diferença de  $Z(t)$  por

$$\nabla Z(t) = Z(t) - Z(t - 1) \quad (8)$$

e a segunda diferença dada por

$$\nabla^2 Z(t) = \nabla(\nabla Z(t)) = \nabla(Z(t) - Z(t - 1)) = Z(t) - 2Z(t - 1) + Z(t - 2) \quad (9)$$

e, assim sucessivamente, normalmente é suficiente tomar uma ou duas diferenças para que a série se torne estacionária.

A abordagem de Box e Jenkins (BOX, JENKINS e REINSEL, 2008) é uma metodologia bastante utilizada na análise de modelos paramétricos e consiste do ajuste de modelos autorregressivos integrados de médias móveis ( $ARIMA(p, d, q)$ ) aos dados disponíveis. Sua construção se baseia num ciclo iterativo cujos estágios estão representados na Figura 1. No caso da análise dos resíduos indicar a estrutura do modelo como inadequada, deve-se retomar a etapa da identificação.

Muitas vezes são identificados vários modelos que passam pelos estágios de estimação e verificação e escolhe-se aquele que apresenta o menor erro quadrático médio de previsão.





**Figura 1: Estágios do ciclo iterativo da construção de um modelo de séries temporais**  
**FONTE: Adaptado de Morettin e Toloí (2006)**

A identificação de modelo(s) ARIMA para ajustar um conjunto de dados tem o objetivo de determinar os valores de  $p$ ,  $d$  e  $q$  e estimar preliminarmente os parâmetros a serem usados na etapa de estimativas. A identificação é feita, sobretudo, com base nas autocorrelações e autocorrelações parciais estimadas, as quais se espera representarem adequadamente seus valores teóricos que são desconhecidos (MORETTIN e TOLOI, 2006).

As etapas da identificação são:

- Analisar a necessidade de aplicar transformação nos dados da série original com vistas a estabilizar sua variância, tornar os dados mais simétricos e próximos de uma distribuição normal;
- Aplicar diferenças na série para obtenção de estacionariedade e redução a busca de um modelo  $ARMA(p, q)$ ;
- Fazer a identificação dos valores de  $p$  e  $q$  com base nas autocorrelações e autocorrelações parciais.

Após o ajuste de um modelo a uma série temporal é recomendável verificar se o modelo realmente fornece uma descrição adequada dos dados. Isto é feito através da análise do resíduo que é a diferença entre o valor da observação e o valor ajustado pelo modelo. Para um 'bom' modelo se esperam resíduos 'aleatórios' e próximos de zero em que cada coeficiente de autocorrelação é aproximadamente normalmente distribuído (CHATFIELD, 1996, apud LAZZAROTTO, 2016).

## 2.4 *Outliers* em Séries Temporais

Os valores de uma série temporal podem, muitas vezes, serem afetados por eventos inesperados tais como mudanças de política ou crises econômicas, ondas inesperadas de frio ou calor, erros de medida, erros de digitação, etc. A consequência da ocorrência desses tipos de eventos é a criação de observações espúrias que são inconsistentes com o restante da série; tais observações são denominadas valores atípicos ou “*outliers*”. A presença de valores atípicos tem efeito na estimação dos parâmetros do modelo ARIMA e, conseqüentemente, na especificação correta do modelo e nas previsões de valores futuros (MORETTIN e TOLOI, 2006). Serão considerados três tipos de *outliers*, conforme o QUADRO 1.

QUADRO 1: Tipos de *outliers*.

Tipo	Tradução	Abreviatura
<i>Additive outlier</i>	<i>Outlier</i> aditivo	AO
<i>Level Shift</i>	Mudança de Nível	LS
<i>Temporary Change</i>	Mudança Temporária	TC

FONTE: o autor.

No caso de um *outlier* aditivo, ocorre um efeito imediato e pontual na série observada. Um *outlier* de mudança de nível produz uma mudança abrupta e permanente no nível da série (CHEN e LIU, 1993). A Figura 2 mostra uma simplificação dos três tipos de *outliers* em uma série temporal estacionária.

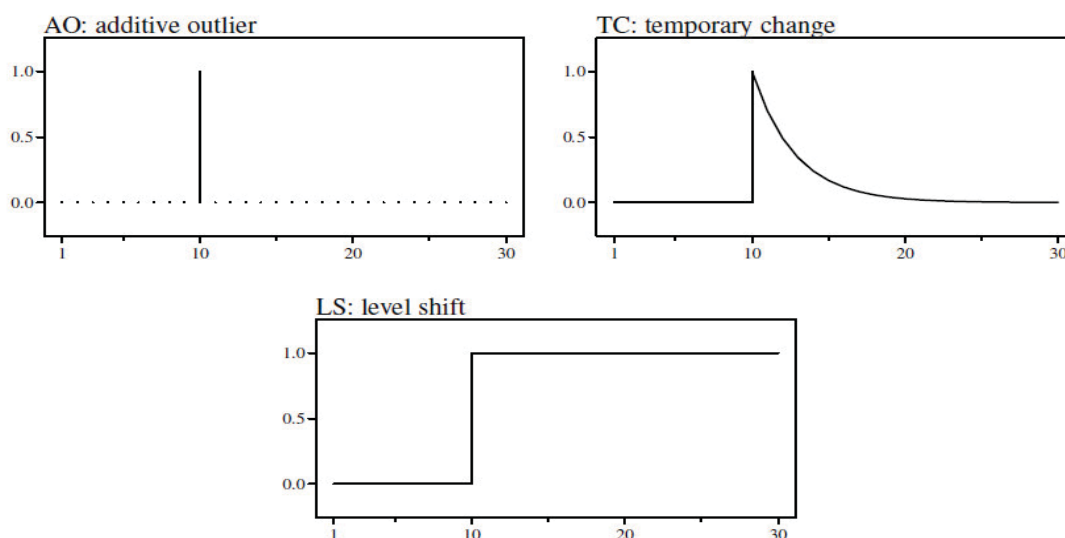


FIGURA 2: Tipos de *outliers* em séries temporais.  
FONTE: Adaptado de LACALLE (2016).

A representação geral de um *outlier* em uma série temporal pode ser dada pela expressão

$$L(B)I(T),$$

onde  $L(B)$  é um polinômio de operadores de retardo e  $I(T)$  é uma variável indicadora, ou função impulso, que assume o valor 1 se o ponto em  $t = T$  for um *outlier* e valor 0 em caso contrário. Assim, para descrever uma série temporal sujeita a um evento não-repetitivo, o seguinte modelo é considerado:

$$Z_t = Y_t + \omega L(B)I_t(T), \quad (10)$$

onde  $Y_t$  segue um processo  $ARMA(p, q)$  geral, descrito em (2), e  $\omega$  é a magnitude do efeito do *outlier*. O polinômio  $L(B)$  para cada tipo de *outlier* é definido como

$$LS: \quad L(B) = \frac{1}{(1 - B)} \quad (11)$$

$$AO: \quad L(B) = 1 \quad (12)$$

$$TC: \quad L(B) = \frac{1}{(1 - \delta B)}. \quad (13)$$

O valor de  $\delta$  pode ser especificado pelo analista para identificar um *outlier* de mudança temporária. CHEN e LIU (1993) recomendam utilizar  $\delta = 0.7$ . É importante destacar que os *outliers* do tipo aditivo e de mudança de nível são dois casos extremos de um *outlier* de mudança temporária, com  $\delta = 0$  e  $\delta = 1$ , respectivamente.

Um *outlier* aditivo pode ser interpretado em termos gerais como um erro de medida no instante  $T, 1 \leq T \leq n$ , ou como um efeito impulso devido a causas exógenas, sem afetar os valores futuros da série (PEÑA, TIAO e TSAY, 2001). Um *outlier* aditivo pode ter efeitos muito sérios nas propriedades da série temporal observada. Ele afetará os resíduos estimados e também as estimativas dos valores dos parâmetros (média e variância). *Outliers* aditivos também podem ter um forte efeito nas estimativas dos parâmetros do modelo (PEÑA, TIAO e TSAY, 2001).

Um *outlier* aditivo corresponde a um erro externo ou mudança exógena do valor observado na série temporal em um instante de tempo em particular. Isto é, ao

invés de observarmos a série  $Y_t$ , observa-se uma nova série  $Z_t$ , que é relacionada com a original por

$$Z_t = \begin{cases} Y_t, & t \neq T \\ Y_t + \omega_A, & t = T. \end{cases} \quad (14)$$

Uma representação alternativa pode ser feita através do modelo

$$Z_t = \omega_A I_t^{(T)} + \psi(B)a_t. \quad (15)$$

Um *outlier* de mudança de nível corresponde a mudança da média local ou nível do processo, começando de um ponto específico e continua até o final do período de tempo observado. Em um processo estacionário, uma mudança de nível implica uma mudança na média depois de algum ponto, e o processo se torna não-estacionário (PEÑA, TIAO e TSAY, 2001).

Quando um *outlier* de mudança de nível ocorre, a série observada é relacionada a série original por

$$Z_t = \begin{cases} Y_t, & t < T \\ Y_t + \omega_L, & t \geq T. \end{cases} \quad (16)$$

Um *outlier* de mudança de nível pode ser visto como uma sequência de *outliers* aditivos de mesmo tamanho começando a partir de algum ponto e permanecendo até o final do período de tempo observado (PEÑA, TIAO e TSAY, 2001). O modelo para este tipo de *outlier* é

$$Z_t = \frac{\omega_L}{1-B} I_t^{(T)} + \psi(B)a_t, \quad (17)$$

que pode ser reescrita como  $Z_t = \omega_L S_t^{(T)} + \psi(B)a_t$ , onde  $S_t^{(T)}$  é uma função degrau que assume o valor 0 antes de  $T$  e 1 quando  $t \geq T$ . Segundo Peña, Tiao e Tsay (2001), esta função está relacionada com a função impulso da seguinte forma

$$S_t^{(T)} = \frac{1}{1-B} I_t^{(T)}.$$

Considere que o efeito de um *outlier* de mudança de nível decresce com o tempo até desaparecer. Um meio de modelar isto é por um *outlier* de Mudança Temporária ou Transitória (TC – *Transitory Change*) como um *outlier* de mudança de

nível que decresce de forma exponencial (PEÑA, TIAO e TSAY, 2001) e seu modelo é definido como:

$$Z_t = \frac{\omega_{TC}}{(1-\delta B)} I_t^{(T)} + \psi(B) a_t. \quad (18)$$

Em um *outlier* de mudança temporária, é produzido um efeito inicial  $\omega$  no instante  $T$  e este efeito diminui gradualmente com o tempo. O parâmetro  $\delta$  é designado para modelar a cadência da dinâmica do efeito de amortecimento.

De uma forma geral, uma série temporal pode conter  $k$  valores atípicos de diferentes tipos e pode ser representada pelo modelo

$$Z_t = \sum_{i=1}^k \omega_i L_i(B) I_t^{(T_i)} + Y_t. \quad (19)$$

Quando o instante  $T$  de ocorrência de um valor atípico é conhecido, pode-se utilizar a análise de intervenção, para modelar o efeito de tal ocorrência. Entretanto, no início de uma análise, a presença ou não de valores atípicos dificilmente é conhecida, o que faz necessário o desenvolvimento de procedimentos para detectá-los e, posteriormente, incluí-los no modelo ou, então, removê-los (MORETTIN e TOLOI, 2006).

Para eliminar o efeito de um *outlier* em uma dada série temporal, é necessário:

- (1) determinar o ponto no qual ocorre;
- (2) identificar o tipo e
- (3) remover o seu efeito estimando um modelo no qual o *outlier* é incorporado.

O problema é complicado, pois inicialmente não se conhece nem a localização nem o tipo de *outlier* (PEÑA, TIAO e TSAY, 2001).

Suponha primeiramente que os parâmetros do modelo ARIMA são conhecidos, de modo que será analisado como estimar a magnitude do *outlier* em um dado ponto  $T$ , quando o tipo de *outlier* (AO, LS, TC) é conhecido. O modelo da série observada  $Z_t$  é dado por (19), onde  $L(B)$  é 1,  $1/(1-B)$  ou  $1/(1-\delta B)$ , dependendo do tipo de *outlier*. Sejam

$$\hat{e}_t = \pi(B) Z_t$$

os resíduos estimados da série observada, com  $\pi(B)$  definido em (3). Pode-se reescrever o modelo como

$$\hat{e}_t = \omega_i x_t + a_t,$$

onde

$$\omega_i = \omega_A \text{ e } x_t = \pi(B)I_t^{(T)} \text{ para } \textit{outliers} \text{ aditivos,}$$

$$\omega_i = \omega_L \text{ e } x_t = \pi(B)/(1 - B) \text{ para } \textit{outliers} \text{ de mudança de nível e}$$

$$\omega_i = \omega_{TC} \text{ e } x_t = \pi(B)/(1 - \delta B) \text{ para } \textit{outliers} \text{ de mudança temporária.}$$

Assumindo que os parâmetros do modelo  $Z_t$  são conhecidos, os efeitos  $\omega_i$  podem ser estimados por mínimos quadrados (PEÑA, TIAO e TSAY, 2001), levando a

$$\hat{\omega}_i = \frac{\sum_{t=T}^n \hat{e}_t x_t}{\sum_{t=T}^n x_t^2}. \quad (20)$$

No caso do *outlier* aditivo, a estimativa é dada por

$$\hat{\omega}_A = \rho_A^2 \pi(F) e_T,$$

sendo  $F$  um operador de avanço definido por  $F(Z_t) = Z_{t+1}$ , e

$$\rho_A^2 = \frac{1}{1 + \pi_1^2 + \dots + \pi_{n-T}^2}.$$

Perceba que este resultado é consistente com a propriedade de que todos os resíduos depois de  $T$  são afetados pelo *outlier* e, portanto, todos possuem informação sobre ele (PEÑA, TIAO e TSAY, 2001).

No caso de *outliers* de mudança de nível, todos os resíduos após a mudança contém informação, e a estimativa é

$$\hat{\omega}_L = \rho_L^2 l(F) \hat{e}_t,$$

em que  $l(F) = \frac{\pi(F)}{1-F}$  e

$$\rho_L^2 = \frac{1}{1 + l_1^2 + \dots + l_{n-T}^2},$$

Onde os  $l_i$  são os coeficientes de  $l(F)$ .

Finalmente, no caso de *outliers* de mudança temporária:

$$\hat{\omega}_{TC} = \rho_{TC}^2 \beta(F) e_T,$$

onde

$$\rho_{TC}^2 = \frac{1}{1 + \beta_1^2 + \dots + \beta_{n-T}^2},$$

no qual os  $\beta_i$  são os coeficiente de  $\beta(F) = \frac{\pi(F)}{1-\delta F}$ .

Para testar se um *outlier* de tipo conhecido ocorreu no ponto T, o teste padrão é

$$H_0: \omega_j = 0$$

$$H_1: \omega_j \neq 0$$

onde  $j = AO, LS, TC$ .

As estatísticas do teste da razão de verossimilhança para as observações atípicas são dadas por

$$H_0 \times H_1: \lambda_{j,t} = \frac{\hat{\omega}_{j,t}}{\rho_{j,t} \sigma_a} \quad (21)$$

em que  $\rho_{j,t} = \sqrt{\rho_j^2}$  no ponto  $t$ , e a distribuição de  $\lambda_{j,t}$  é t de Student (PEÑA, TIAO e TSAY, 2001). Para calcular o teste estatístico para *outliers* é necessário estimar  $\sigma_a$ . A determinação de *outliers* pode ser sensível a esta estimativa. Na presença de valores atípicos, os resíduos são contaminados; portanto  $\sigma_a$  pode ser superestimado se o desvio padrão usual for utilizado (CHEN e LIU, 1993). O desvio padrão residual é estimado inicialmente pelo desvio mediano absoluto,  $\hat{\sigma}_a$ , definido por

$$\hat{\sigma}_a = 1.483 \times \text{mediana}(|\hat{e}_t - \tilde{e}|), \quad (22)$$

onde  $\tilde{e} = \text{mediana}(\hat{e}_t)$ .

Na prática, o instante  $T$  de ocorrência de uma observação atípica, bem como os parâmetros do modelo são desconhecidos (MORETTIN e TOLOI, 2006). Chen e

Liu (1993) propõe um procedimento iterativo para identificar observações atípicas de forma conjunta e ajustar um modelo apropriado.

O procedimento consiste em três fases: na primeira, todos os potenciais *outliers*,  $t_j$  e  $L_j(B)$  são identificados, baseados nos parâmetros do modelo ARIMA preliminar. Na segunda fase, são estimados conjuntamente os parâmetros do modelo e os efeitos dos *outliers*, usando as informações acumuladas na primeira fase. Na terceira fase, *outliers*,  $t_j$  e  $L_j(B)$  são identificados e estimados novamente, baseado nas estimativas menos contaminadas dos parâmetros do modelo obtidos na segunda fase.

Os passos desse procedimento são:

- 1) Ajustar um modelo para a série original  $Z_t$ , supondo que não existem observações atípicas e calcular os resíduos do modelo. Na primeira iteração, a série original é usada para iniciar o procedimento; depois da primeira iteração a série ajustada é utilizada.
- 2) Para cada  $t = 1, \dots, n$ , calcular  $\lambda_{A,t}$ ,  $\lambda_{L,t}$  e  $\lambda_{TC,t}$  usando os resíduos obtidos no passo 1 e considere

$$\eta_t = \max\{|\lambda_{A,t}|, |\lambda_{L,t}|, |\lambda_{TC,t}|\}.$$

Se  $\max \eta_t = |\lambda_{j,t_1}| > C$ , onde  $C$  é um valor crítico predeterminado, então há um possível *outlier* do tipo  $j$  em  $t_1$ , com efeito estimado  $\hat{\omega}_{j,t_1}$ . Essa observação pode ser eliminada utilizando

$$\tilde{Z}_t = Z_t - L_j(B)\hat{\omega}_{j,t_1}I_t^{(t_1)}.$$

Baseado em resultados de simulações, Chen e Liu (1993) recomendam para séries com 100 a 200 observações, um valor crítico  $C = 3$ . Para séries menores, valores entre 2,5 e 2,9 são recomendados. Pode-se considerar um valor crítico maior que 3 para séries com mais de 200 observações. Na prática, é recomendado que mais de um valor crítico seja usado na análise, para permitir um exame na sensibilidade dos resultados para a escolha de  $C$ . Quando os valores críticos são muito pequenos, o método tende a superajustar a série e produz resultados insatisfatórios. Por outro lado, se o valor crítico for muito grande, poucos ou nenhum valor atípico será encontrado.



- 3) Quando um valor atípico é identificado, os resíduos são corrigidos levando em conta o tipo de *outlier*. Assim, se um *outlier* aditivo for identificado, um novo resíduo é definido nesse ponto por

$$\check{e}_t = \hat{e}_t - \hat{\omega}_A \pi(B) I_t^{(T)},$$

para  $t \geq T$ ; e se um *outlier* de mudança de nível (ou temporária) é detectado, tem-se

$$\check{e}_t = \hat{e}_t - \hat{\omega}_{(L \text{ ou } TC)} g(B) I_t^{(T)},$$

onde  $g(B) = l(B)$  para *outlier* de mudança de nível e  $g(B) = \beta(B)$  para *outlier* de mudança temporária;

- 4) Usando os novos resíduos  $\check{e}_t$ , obtêm-se  $\check{\sigma}_e$  e os  $\lambda_{j,t}$  são calculados novamente utilizando esses valores. A identificação de *outliers* e o cálculo de novos resíduos são repetidos até que nenhum outro ponto apareça como *outlier*;
- 5) Suponha que tenham sido identificados preliminarmente  $k$  observações atípicas nos instantes  $t_1, t_2, \dots, t_k$ . Assumindo esses instantes de tempo conhecidos, pode-se estimar simultaneamente os parâmetros das observações atípicas  $\omega_1, \omega_2, \dots, \omega_k$  e os parâmetros da série temporal, utilizando o modelo

$$Z_t = \sum_{i=1}^k \omega_i L_i(B) I_t^{(t_i)} + \frac{\theta(B)}{\nabla^d \phi(B)} a_t. \quad (23)$$

A estimação do modelo leva a novos resíduos

$$\hat{e}_t^{(1)} = \hat{\pi}^{(1)}(B) \left[ Z_t - \sum_{i=1}^k \hat{\omega}_i L_i(B) I_t^{(t_i)} \right] \quad (24)$$

e a uma nova estimativa para  $\sigma_a$ .

- 6) Repetem-se os passos de (2) a (5) até que todas as observações atípicas sejam identificadas e seus impactos estimados simultaneamente.

Assim, o modelo ajustado com observações atípicas é dado por

$$Z_t = \sum_{i=1}^k \hat{\omega}_i L_i(B) I_t^{(t_i)} + \frac{\hat{\theta}(B)}{\nabla^d \hat{\phi}(B)} a_t, \quad (25)$$

com  $\hat{\omega}_i$ ,  $\hat{\theta}(B)$  e  $\hat{\phi}(B)$  obtidos na iteração final.

### 3 MATERIAL E MÉTODOS

Neste capítulo são apresentadas as justificativas das escolhas das técnicas e os materiais utilizados neste trabalho, que incluem os dados e os *softwares* utilizados para aplicação dos métodos descritos no Capítulo 2.

Este trabalho limitou-se ao monitoramento do gráfico de controle da média ou do vetor de médias de um processo, muito embora existam gráficos para o controle da variância, da amplitude, entre outros.

Além disso, também se limitou aos gráficos de Shewhart, em detrimento de outros gráficos de controle univariados como os de somas cumulativas (CUSUM) e de médias móveis exponencialmente ponderadas (EWMA) mais indicados quando se deseja detectar mudanças no processo da ordem de  $1,5\sigma$  ou menos, tendo em vista que tais procedimentos são mais indicados para detectar erro tipo II.

Para o tratamento paramétrico escolheu-se a modelagem SARIMA por ser consagrada na modelagem do comportamento linear de séries temporais e por permitir tratar com séries temporais eventualmente não estacionárias, que apresentem estrutura de autocorrelação, com sazonalidades e variância constantes.

#### 3.1 MATERIAL

Foram usados dados provenientes da leitura de instrumentos de monitoramento da barragem da usina hidrelétrica de Itaipu Binacional, localizada no Rio Paraná, entre o Brasil e o Paraguai. A Itaipu Binacional é líder mundial em produção de energia limpa e renovável, tendo produzido mais de 2,5 bilhões de MWh desde o início de sua operação, em 1984. Com 20 unidades geradoras e 14.000 MW de potência instalada, fornece cerca de 15% da energia consumida no Brasil e 86% no Paraguai<sup>2</sup>.

De acordo com Rosso *et al.*(1995, p. 2-3) a barragem principal de Itaipu é constituída por blocos tipo gravidade aliviada com 34 m de largura e 612 m de extensão total e na região do leito do Rio Paraná atinge a altura máxima de 196 m. Ao lado desta estrutura, a barragem principal de Itaipu é integrada por 83 blocos tipo

---

<sup>2</sup> [www.itaipu.gov.br/energia/geracao](http://www.itaipu.gov.br/energia/geracao)

contraforte, com 17 m de largura, altura média de 60 m e extensão total de 1411 m. Ambas as estruturas possuem instrumentação detalhada e acompanhamento pormenorizado do seu comportamento durante o período operacional da barragem com vistas a verificar as hipóteses de projeto e supervisionar sua segurança.

Devido à grande extensão da barragem de Itaipu, por sugestão da equipe de segurança, como o trecho E é constituído de 6 blocos e o bloco E-6 é considerado um bloco chave (mais instrumentado), este trabalho limitou-se a este bloco. O bloco E-6 está localizado na ligação direita da barragem principal, na região indicada na Figura 3. O bloco E-6 é um dos blocos que compõe a ligação entre os blocos de contraforte e a barragem principal e possui alta densidade de instrumentação. Sua altura a montante vai da elevação 140 m até 225 m e à jusante da elevação 125 m até 225 m, portanto, possui 85 m de altura a montante e 100 m de altura a jusante e a largura da ‘cabeça’ do bloco é de 17 m.



**Figura 3: Barragem da Usina de Itaipu e o Trecho E em destaque**  
**FONTE: Itaipu Binacional**

Para validar as ideias de detecção e ajuste de *outliers* a série temporal de dados da leitura de instrumento de monitoramento de barragem, foi aplicada uma metodologia aos dados de dois extensômetros que monitoram o bloco 6 do trecho E da barragem da usina hidrelétrica de Itaipu. Na Figura 4 pode-se ver uma seção do bloco e toda sua instrumentação.

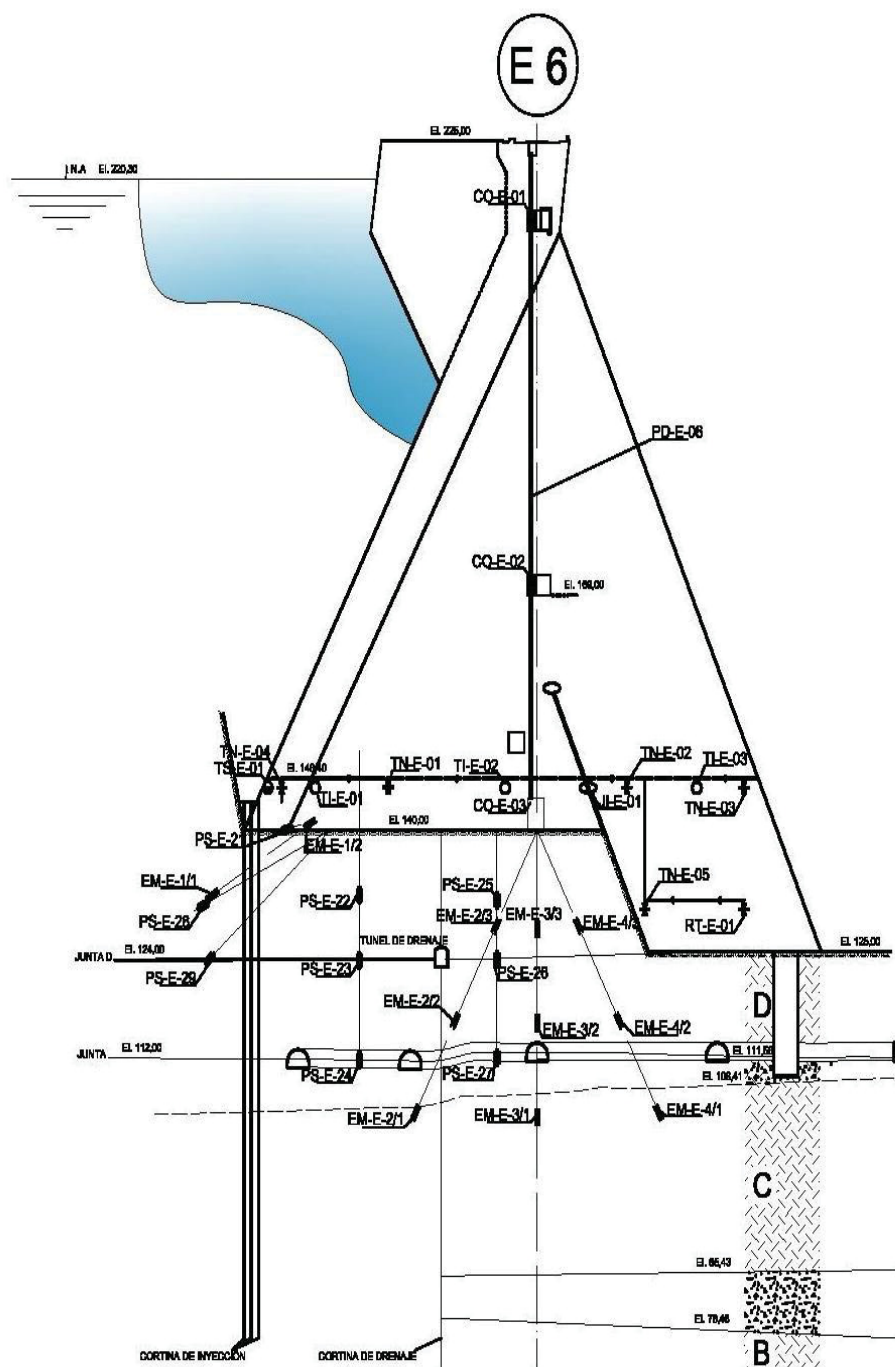


Figura 4: Instrumentação do Bloco E6 em Itaipu.  
 FONTE: Equipe de segurança de barragem de Itaipu

A função do extensômetro é medir os deslocamentos (deformações) verticais (recalques) e inclinados da barragem. Na Figura 5 é apresentado o esquema de funcionamento de um extensômetro múltiplo constituído de três hastes em Itaipu.

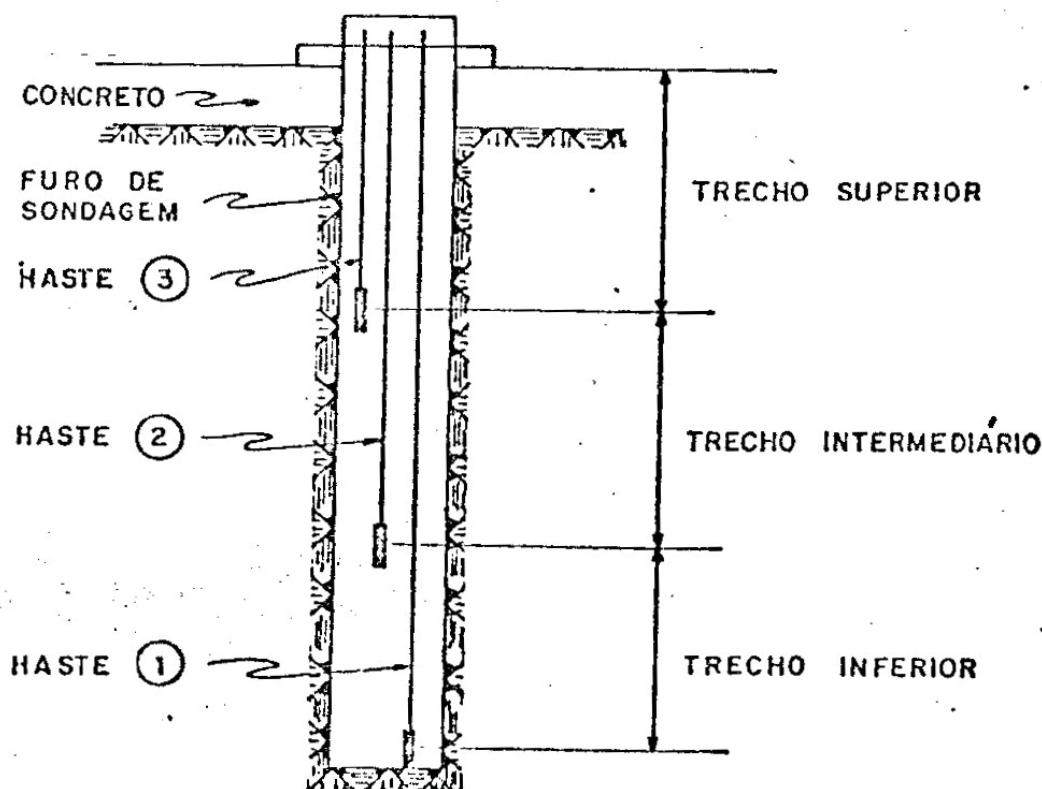


Figura 5: Esquema de um extensômetro múltiplo em Itaipu.  
 FONTE: Equipe de segurança de barragem de Itaipu

O primeiro extensômetro analisado é de haste única, enquanto o segundo corresponde a um extensômetro de três hastas. Estavam disponíveis (inicialmente) no primeiro extensômetro, doravante denominado extensômetro 1, havia 889 dados desde novembro de 1981 até janeiro de 2018. No segundo, doravante denominado extensômetro 2, havia 850 dados de março de 1982 a janeiro de 2018. A periodicidade da realização das leituras variou durante este período, bem como, o comportamento inicial do instrumento durante a etapa de enchimento do reservatório. Desta forma, adotou-se como referência a média mensal das leituras dos instrumentos, o que compreende 436 observações no primeiro extensômetro e 432 no segundo. A unidade de medida é o milímetro.

Neste trabalho foram utilizados alguns *softwares* com diversas finalidades. A seguir segue uma breve descrição do *software* e da sua finalidade.

- O *STATGRAPHICS Centurion XVI*®, de acordo com Statgraphics Centurion XVI (2014), é um *software* de para análise estatística, previsão de séries temporais, controle de qualidade e possui mais de 180 procedimentos

estatísticos, dentre eles, métodos multivariados e técnicas não paramétricas. O programa foi utilizado, sobretudo, para construção de gráficos de controle.

- O *software* livre *R* é uma linguagem de programação e ambiente para computação estatística e gráfica. O *R* oferece uma grande variedade de técnicas estatísticas (modelagem linear e não-linear, testes estatísticos clássicos, análise de séries temporais, classificação, *clustering*, etc.) e gráficas e é altamente extensível. Este programa foi utilizado para a detecção de *outliers* nas séries temporais, por meio do pacote de extensão *tsoutliers*.
- As planilhas eletrônicas *Excel*<sup>®</sup> foram utilizadas para armazenamento de dados das séries temporais para que se pudessem exportar dados para outros *softwares*, bem como, receber resultados obtidos de outros *softwares*.

### 3.2 MÉTODOS

O método utilizado para a detecção e remoção de *outliers* foi proposto por Chen e Liu (1993) e é aquele que foi apresentado anteriormente. Sua escolha se deve ao fato de ter sido amplamente utilizado (PEÑA, TIAO e TSAY, 2001) e já estar implementado no software estatístico R, usado neste trabalho.

A primeira etapa do trabalho consistiu em ajustar um modelo ARIMA para descrever cada série observada, não levando em consideração a presença de *outliers*. Os dados foram separados em uma amostra de treinamento, usada para construção e estimativa dos parâmetros dos modelos e uma amostra de validação, para avaliar o desempenho em termos da capacidade preditiva. O objetivo desse modelo é comparar as previsões e cartas de controle da série observada com a série livre dos efeitos dos *outliers*. Os modelos encontrados para as duas séries inicialmente foram os ARIMA Sazonais  $(3,1,3) \times (0,1,1)_{12}$  e  $(1,0,2) \times (0,1,1)_{12}$ .

A seguir, localizam-se os *outliers* das séries utilizando o software R. O valor crítico foi determinado automaticamente pelo software em  $C = 3.965$ , utilizando a fórmula proposta por Lacalle (2016):

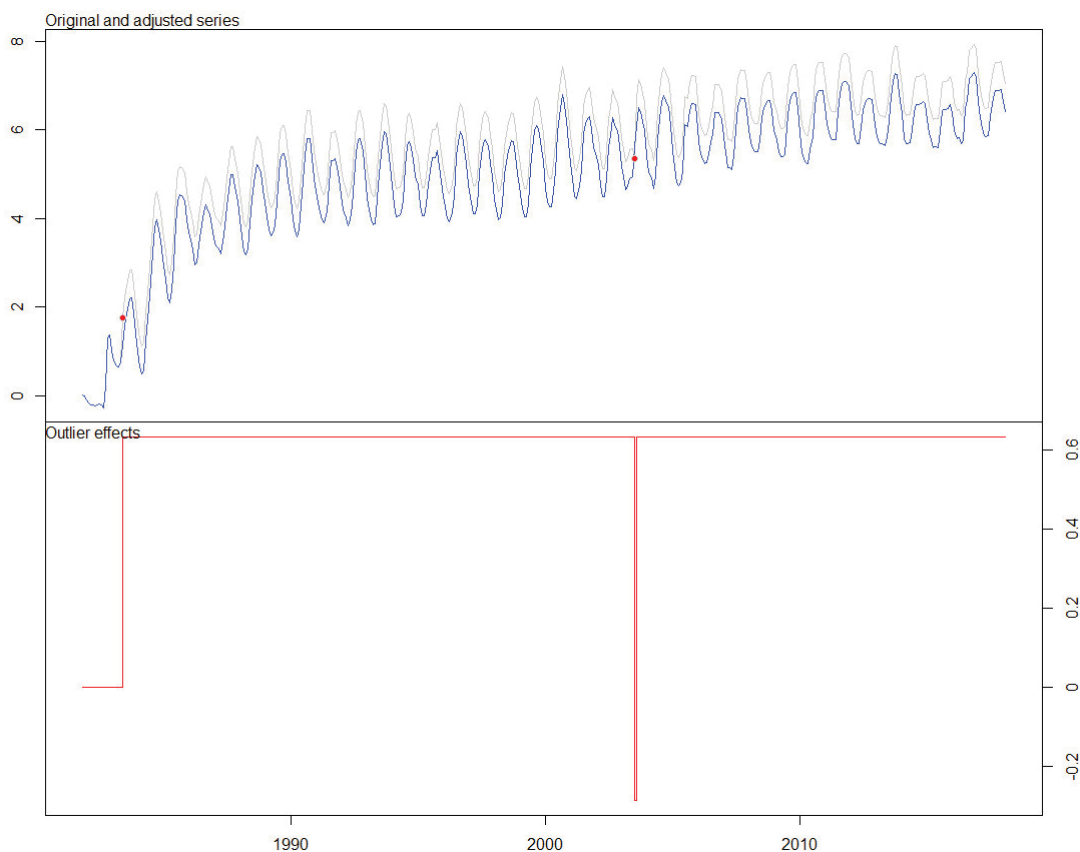
$$C = 3 + 0,0025(n - 50), 50 \leq n \leq 450.$$

O desempenho do método fez-se pela comparação dos erros de previsão dos modelos e da análise dos resíduos. A qualidade do ajuste é medida em termos do erro cometido pelo modelo, onde o erro, em cada instante de tempo  $t$ , é a diferença entre o valor real e o previsto. O período considerado para as previsões e validação foram os últimos doze meses observados. As formas de avaliar o erro foram: raiz do erro quadrático médio (RMSE), erro absoluto médio (MAE), o erro absoluto percentual médio (MAPE), e o erro médio (ME).

## 4 RESULTADOS

### 4.1 Extensômetro 1

No primeiro extensômetro, a presença de valores atípicos é identificada no gráfico nos pontos em vermelho, conforme Figura 6.



**Figura 6: Identificação gráfica da localização e tipo dos *outliers* no Extensômetro 1.**  
**FONTE: o autor.**

A linha cinza representa a série original, enquanto a linha azul mostra como seria a série sem os efeitos de *outliers* (a mudança de nível na série). O segundo gráfico em vermelho representa os tipos de *outliers* encontrados. A TABELA 1 mostra os valores atípicos encontrados.

**TABELA 1: *Outliers* identificados na série do extensômetro 1.**  
**FONTE: o autor.**

Índice	Data	Tipo de <i>outlier</i>	Efeito ( $\omega$ )	Estatística ( $\lambda$ )
20	06/1983	LS	0.6337	3.979
261	07/2003	AO	-0.9218	-9.624



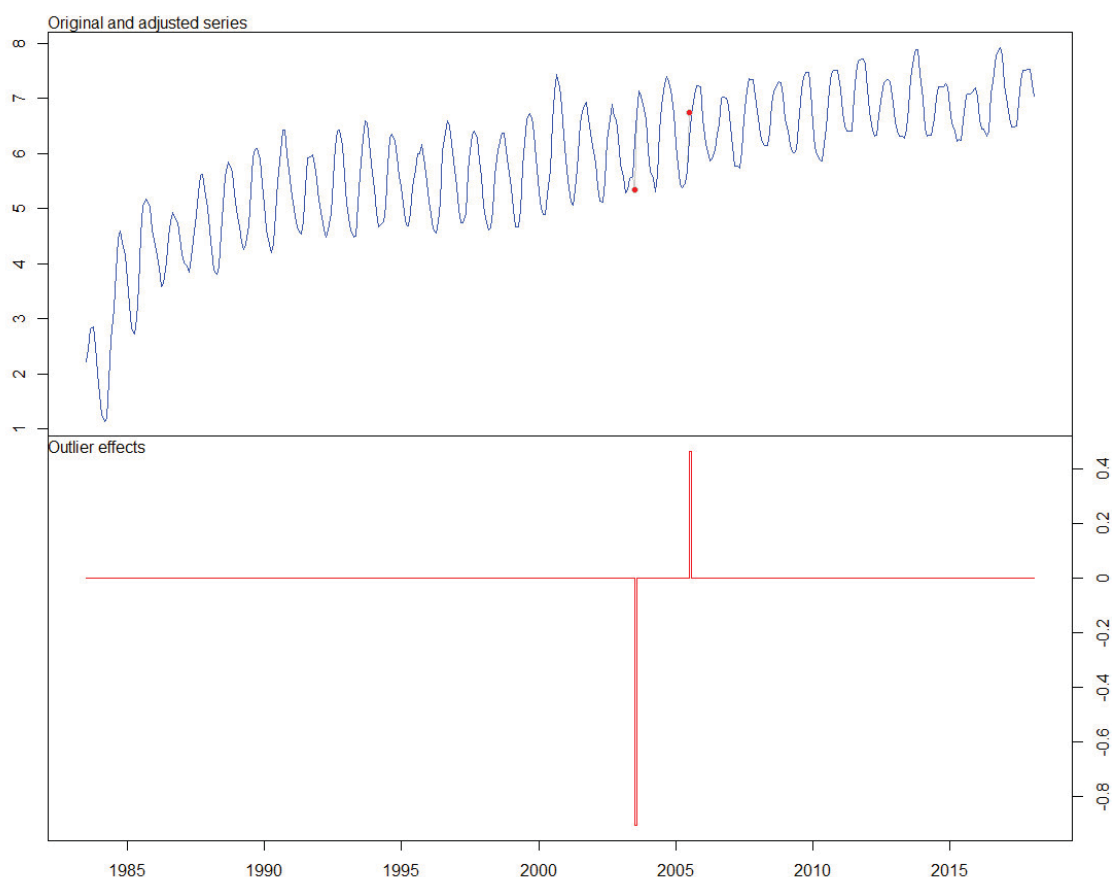
O *outlier* de mudança de nível identificado em junho de 1983 pode ser explicado pelo enchimento do reservatório ocorrido no ano de 1982. Já o *outlier* aditivo encontrado foi detectado por estar significativamente abaixo do previsto pelo modelo ARIMA ajustado.

Como o comportamento da série pode ser consideravelmente diferente antes e depois do enchimento do reservatório, optou-se utilizar para análise uma nova série, com início em julho de 1983. O modelo encontrado para a nova série foi um ARIMA Sazonal  $(4,1,6) \times (2,1,1)_{12}$ . A nova detecção de *outliers* na série foi feita com  $C = 4$ , obtendo-se os resultados da TABELA 2.

**TABELA 2: *Outliers* da nova série (Extensômetro 1).**  
**FONTE: o autor.**

Índice	Data	Tipo de <i>outlier</i>	Efeito ( $\omega$ )	Estatística ( $\lambda$ )
241	07/2003	AO	-0.9218	-9.624
265	07/2005	AO	0.4660	5.353

O gráfico com as novas localizações e efeitos dos *outliers* pode ser visto na Figura 7.



**Figura 7: Localização e efeitos dos *outliers* (Extensômetro 1).**  
**FONTE: o autor.**

Verifica-se que além do mesmo *outlier* aditivo encontrado em julho de 2003, foi encontrado outro valor atípico, também do tipo aditivo, em julho de 2005, significativamente acima do valor previsto pelo modelo ARIMA ajustado. Uma vez removidos os efeitos dos *outliers*, um novo modelo ARIMA Sazonal  $(4,1,4) \times (2,1,1)_{12}$  foi ajustado a série.

Na TABELA 3 são exibidos os erros obtidos com e sem a presença de *outliers* no período de validação dos modelos.

**TABELA 3: Comparativo dos erros dos modelos ARIMA (EXTENSÔMETRO 1).**  
**FONTE: o autor.**

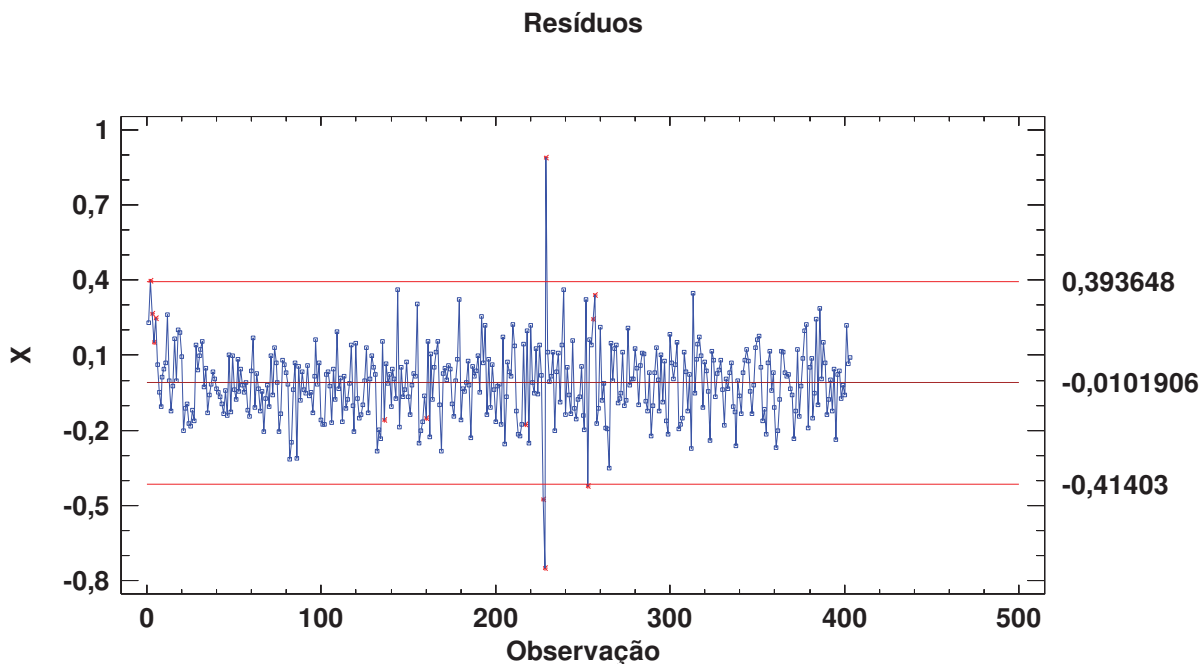
Início	11/1983	07/1983	07/1983
Outliers	Sim	Sim	Não
Modelo	ARIMA(3, 1, 3)x(0, 1, 1) <sub>12</sub>	ARIMA(4, 1, 6)x(2, 1, 1) <sub>12</sub>	ARIMA(4, 1, 4)x(2, 1, 1) <sub>12</sub>
RMSE	0,1139	0,1099	0,0905
MAE	0,0894	0,0828	0,0694
MAPE	-	1,1845	0,9887
ME	-0,0260	-0,0029	-0,0099

Em relação a série temporal dos resíduos, a TABELA 4 apresenta o valor-p do teste de Shapiro-Wilk para testar a normalidade ou o ajuste a distribuição normal. Conforme se observa pelo valor-p, não se pode rejeitar a hipótese de que os resíduos da série sem *outliers* vêm de uma distribuição normal com 95% de confiança.

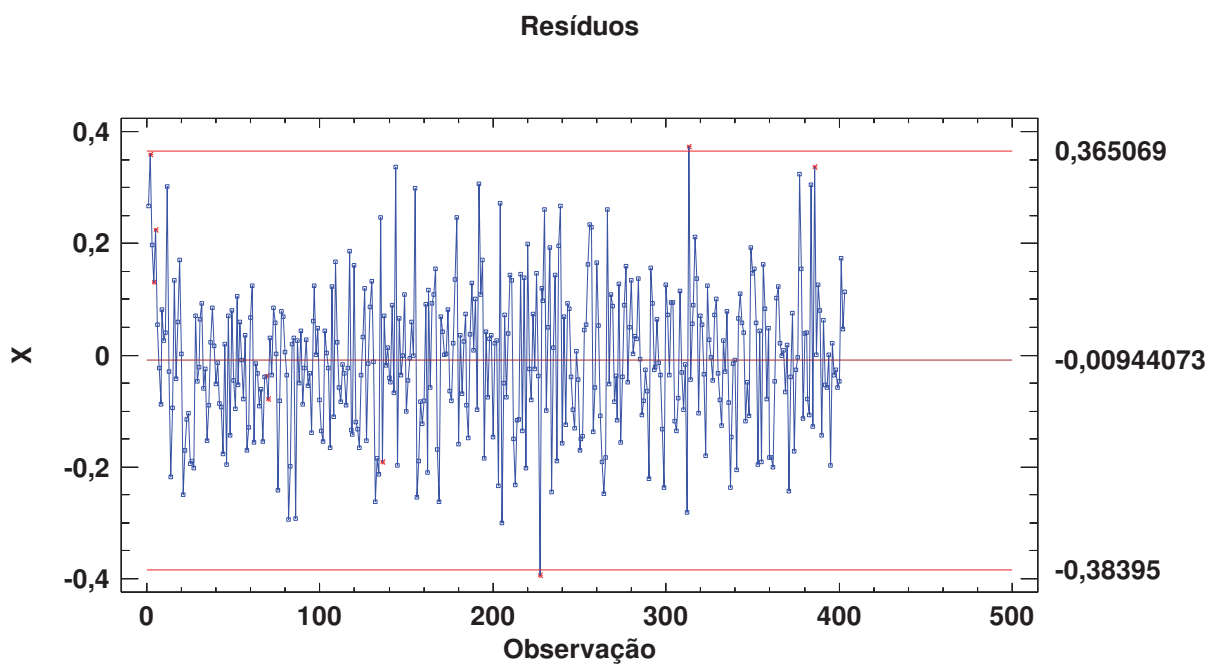
**TABELA 4: Teste de Shapiro-Wilk para a normalidade dos resíduos (extensômetro 1).**  
**FONTE: o autor.**

Série	Valor-p
Com <i>outliers</i> (11/1981)	$5.821 \times 10^{-12}$
Com <i>outliers</i> (07/1983)	$2.092 \times 10^{-8}$
Sem <i>outliers</i> (07/1983)	0.1827

Uma das motivações do estudo foi investigar o efeito de valores atípicos nas cartas de controle para a média nos resíduos das previsões. As Figuras 8 e 9 a seguir mostram as cartas de controle X da média dos resíduos com e sem *outliers*.



**Figura 8: Carta de Controle dos resíduos com *outliers* no extensômetro 1.  
FONTE: o autor.**



**Figura 9: Carta de Controle dos resíduos sem *outliers* no extensômetro 1.  
FONTE: o autor.**

Na TABELA 5 é exibida a quantidade de observações Fora dos Limites de Controle (FLC)  $6\sigma$  para o gráfico dos resíduos do modelo ARIMA ajustado com a presença de outliers, ou seja, antes da intervenção e sem a presença de outliers (após o tratamento realizado para o extensômetro 1).

**TABELA 5: Observações Fora dos Limites de Controle.**  
**FONTE: o autor.**

	Com outliers	Sem outliers
FLC	5	2

Observa-se que a remoção dos efeitos dos dois valores atípicos encontrados reduziu em módulo os limites de controle  $6\sigma$ , e o número de observações fora de controle passou de cinco para dois.

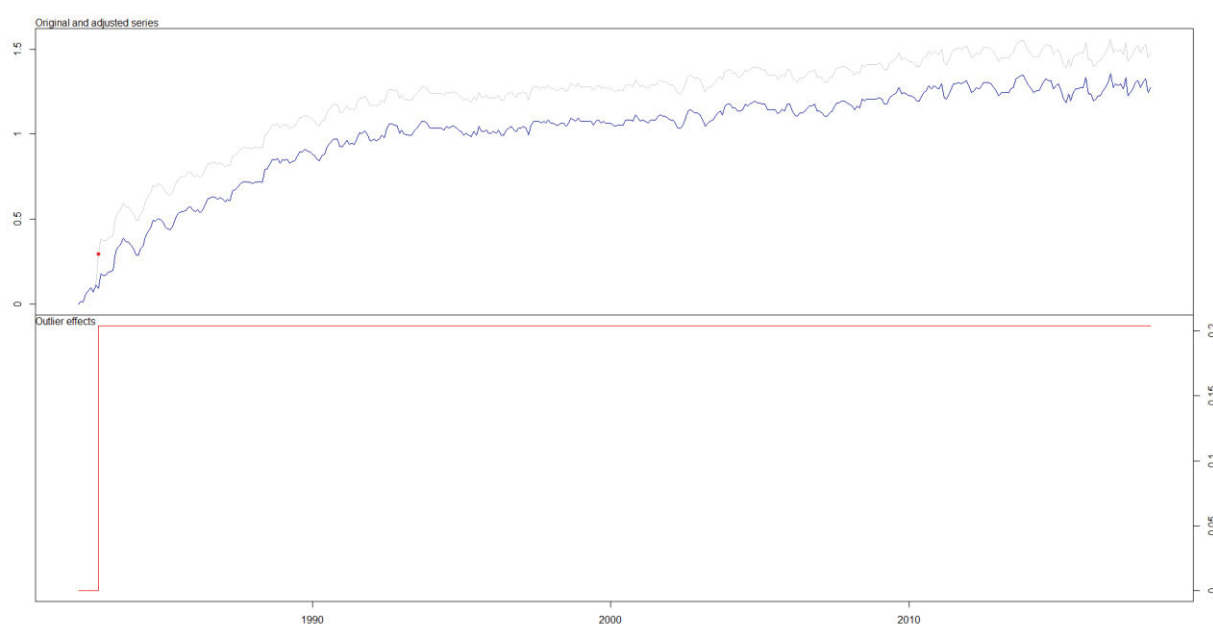
## 4.2 Extensômetro 2

No segundo extensômetro, a TABELA 6 mostra os valores atípicos encontrados.

**TABELA 6: *Outlier* identificado na série do extensômetro 2.**  
**FONTE: o autor.**

Índice	Data	Tipo	Efeito ( $\omega$ )	Estat. ( $\lambda$ )
9	11/1982	LS	0.2041	6.854

E a presença de valores atípicos é identificada no gráfico no ponto em vermelho, conforme Figura 10.



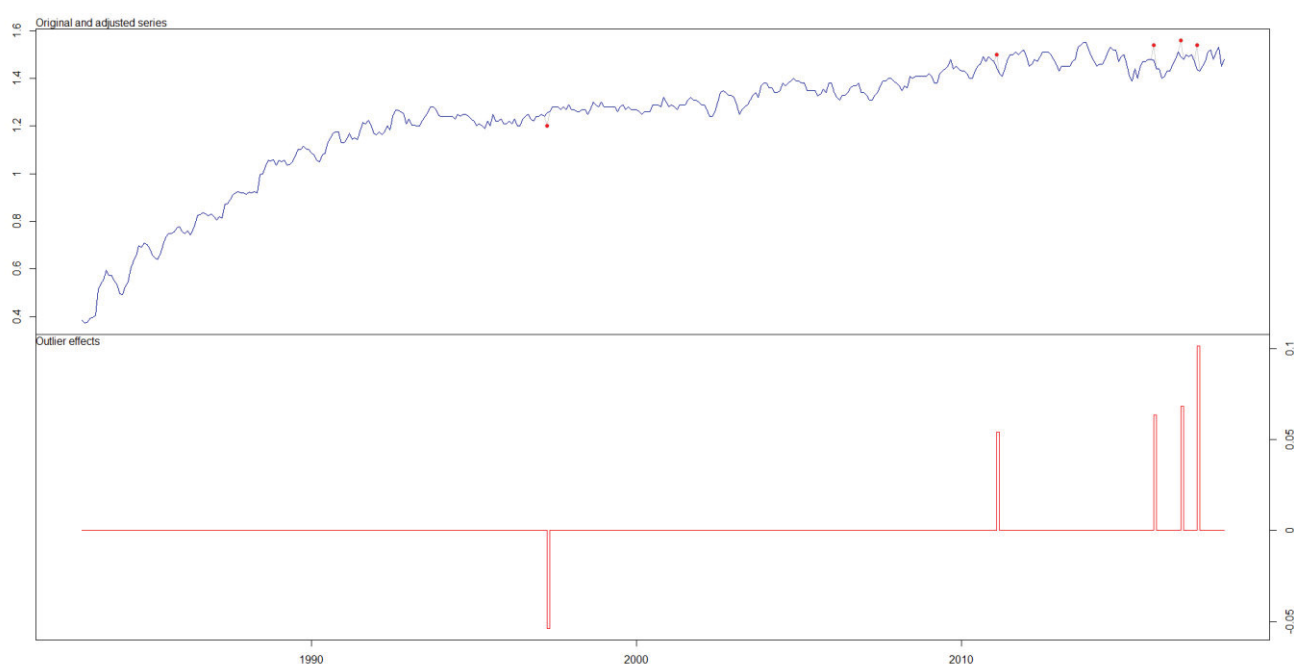
**Figura 10: Identificação gráfica da localização e tipo dos outliers no Extensômetro 2.**  
**FONTE: o autor.**

O *outlier* de mudança de nível identificado em novembro de 1982 também pode ser explicado pelo enchimento do reservatório. Novamente, optou-se utilizar para análise uma nova série, com início em dezembro de 1982. O modelo encontrado para a nova série foi um ARIMA Sazonal  $(0,1,1) \times (0,1,1)_{12}$ . A nova detecção de *outliers* na série foi feita com  $C = 4$ , obtendo-se os resultados da TABELA 7:

**Tabela 7: OUTLIERS DA NOVA SÉRIE (EXTENSÔMETRO 2).**  
**FONTE: o autor.**

Índice	Data	Tipo	Efeito ( $\omega$ )	Estat. ( $\lambda$ )
173	04/1997	AO	-0.05391	-4.121
339	02/2011	AO	0.05430	4.132
397	12/2015	AO	0.06358	4.835
407	10/2016	AO	0.06850	5.009
413	04/2017	AO	0.10160	7.024

O gráfico com as novas localizações e efeitos dos *outliers* pode ser visto na Figura 11.



**Figura 11: Localização e efeitos dos outliers (Extensômetro 2).**  
**FONTE: o autor.**

Uma vez removidos os efeitos dos *outliers*, um novo modelo ARIMA Sazonal  $(1,1,0) \times (1,1,1)_{12}$  foi ajustado a série. Na TABELA 8 são exibidos os erros obtidos com e sem a presença de *outliers*.

TABELA 8: Comparativo dos erros dos modelos ARIMA. (EXTENSÔMETRO 2).  
 FONTE: o autor.

Início	03/1982	12/1982	12/1982
Outliers	Sim	Sim	Não
Modelo	ARIMA(1, 0, 2)x(0, 1, 1) <sub>12</sub>	ARIMA(0, 1, 1)x(0, 1, 1) <sub>12</sub>	ARIMA(1, 1, 0)x(1, 1, 1) <sub>12</sub>
RMSE	0,0491	0,0443	0,0285
MAE	0,0355	0,0341	0,0228
MAPE	-	2,3031	1,5493
ME	-0,0026	-0,0039	-0,0051

Em relação a série temporal dos resíduos, a TABELA 9 apresenta o valor-p do teste de Shapiro-Wilk para testar a normalidade ou o ajuste a distribuição normal. Conforme se observa pelo valor-p, não se pode rejeitar a hipótese de que os resíduos da série sem *outliers* vêm de uma distribuição normal com 95% de confiança.

Tabela 9: Teste de Shapiro-Wilk para a normalidade dos resíduos (extensômetro 2).  
 FONTE: o autor.

Série	Valor-p
Com <i>outliers</i> (03/1982)	$6.214 \times 10^{-9}$
Com <i>outliers</i> (12/1982)	$6.115 \times 10^{-6}$
Sem <i>outliers</i> (12/1982)	0.1372

As Figuras 12 e 13 a seguir mostram as cartas de controle X da média dos resíduos com e sem *outliers*, respectivamente.

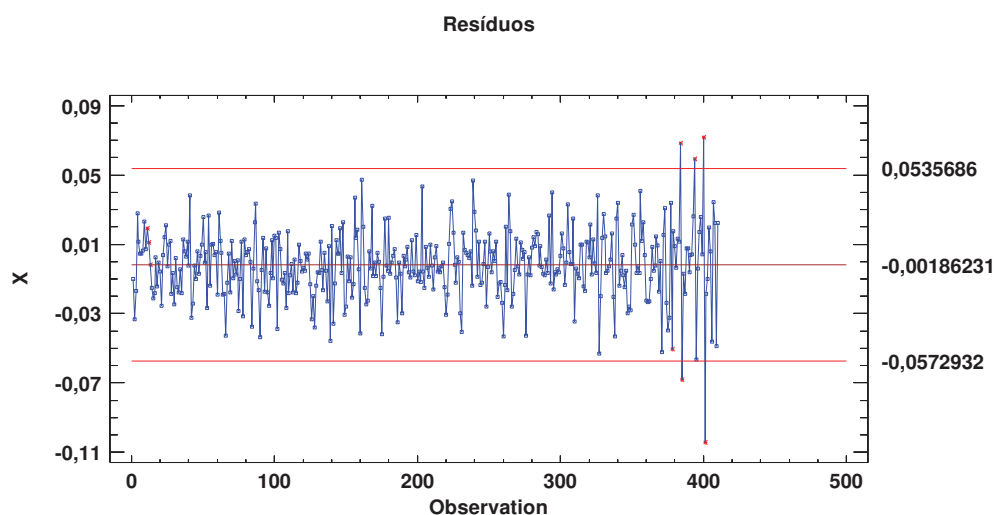
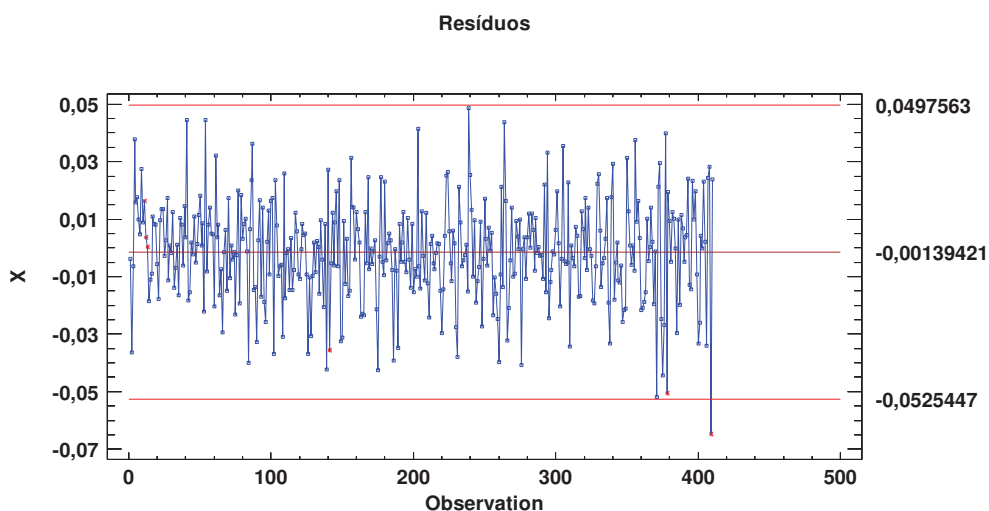


Figura 12: Carta de Controle dos resíduos com *outliers* no extensômetro 2.  
 FONTE: o autor.



**Figura 13: Carta de Controle dos resíduos sem outliers no extensômetro 2.  
FONTE: o autor.**

Na TABELA 10 é exibida a quantidade de observações Fora dos Limites de Controle (FLC)  $6\sigma$  para o gráfico dos resíduos do modelo ARIMA ajustado com a presença de outliers, ou seja, antes da intervenção e sem a presença de outliers (após o tratamento realizado para o extensômetro 2).

**Tabela 10: Observações Fora dos Limites de Controle.  
FONTE: o autor.**

	<b>Com outliers</b>	<b>Sem outliers</b>
FLC	5	1

Observa-se nessa série que a remoção dos efeitos dos valores atípicos encontrados também reduziu em módulo os limites de controle  $6\sigma$ , e o número de observações fora de controle passou de cinco para um.

Cabe ressaltar que sob o aspecto da análise da equipe de engenharia de segurança da barragem, o aspecto da engenharia que este instrumento avalia está sob controle.

## 5 CONCLUSÕES

Neste trabalho aplicou-se uma metodologia para estimar e remover os efeitos de valores atípicos em dados referentes a série temporal das leituras de dois instrumentos de monitoramento da barragem da usina hidrelétrica de Itaipu.

Foi assegurado que os modelos SARIMA estabelecidos se ajustam bem aos dados das observações e que, estatisticamente, os instrumentos indicam que os deslocamentos no local da barragem que eles mensuram estão apresentando apenas uma variabilidade inerente proveniente possivelmente, segundo a equipe de engenharia de segurança da barragem, da variação térmica do ambiente e que não há causas especiais de variação.

As principais contribuições do trabalho são: determinar limites de controle baseados em dados históricos que independem de modelos físicos e podem ser mais realísticos; ajustar um modelo a série de dados do instrumento que realiza previsões acuradas; alertar para a análise previa dos dados (escolha da periodicidade, análise de dados discrepantes, conhecer razões de variabilidade excessiva temporária); e avaliação de pressupostos teóricos fundamentais como a normalidade dos dados.

Com a remoção dos efeitos dos *outliers* nas séries, observou-se redução nos erros de previsão, normalidade dos resíduos estatisticamente significativa e redução de pontos fora dos limites da carta de controle. Esses resultados mostram a importância de se considerar o efeito de *outliers* na modelagem de séries temporais.

Observou-se que utilizar os dados coletados depois do período de enchimento do reservatório também contribui para um melhor ajuste dos dados à distribuição normal e na redução dos erros das previsões, se comparado aos modelos que utilizam todos os dados.

No presente momento o método tem sido aplicado a outras séries de instrumentos, obtendo-se resultados semelhantes. Futuramente, será estudado o efeito de *outliers* nas cartas R da variabilidade dos resíduos das previsões e investigado com a equipe de segurança de barragem de Itaipu as possíveis causas das ocorrências dos valores atípicos encontrados.



## REFERÊNCIAS

BOX, G. E. P.; JENKINS, G. W.; REINSEL, G. C. **Time Series Analysis Forecasting and Control**. 4a. ed. Hoboken-NJ: John Wiley & Sons, Inc., 2008.

CASA CIVIL DA PRESIDÊNCIA DA REPÚBLICA. Lei nº 12.334. **Site da Casa Civil**, Brasília, 20 setembro 2010. Disponível em: <[www.planalto.gov.br/ccivil\\_03/\\_Ato2007-2010/2010/Lei/L12334.htm](http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2010/Lei/L12334.htm)>. Acesso em: 07 abril 2014.

CHATFIELD, C. **The analysis of time series**. 5a. ed. London: Chapman & Hall/CRC, 1996.

CHEN, C., LIU, L., "Joint Estimation of Model Parameters and Outlier Effects in Time Series", **Journal of the American Statistical Association**, Vol. 88, No. 421, pp.284-297, March 1993.

CIGB-ICOLD. **80 Years - Dams for human sustainable development**. Paris: International Commission on Large Dams , 2008.

FERNANDES, C. **Análise de séries temporais**. Rio de Janeiro: Laboratório de Estatística Computacional - PUC-RJ, 2006.

FUSARO, T. C. **Estabelecimento estatístico de valores de controle para instrumentação de barragens de terra - Estudo de caso das barragens de Emborcação e Piau**. Ouro Preto: Dissertação de Mestrado Profissional em Engenharia Geotécnicas da UFPOF, 2007.

ICOLD - CIGB. **Dams & The World's Water - An educational book that explains how dams help to manage the world's water**. Paris: International Commission on Large Dams/Commission Internationale des Grand Barrages, 2008.

ITAIPU BINACIONAL. **Síntese das Principais Informações de Projeto para o Controle da Instrumentação da Barragem Principal - Trechos E, F e I (4106.50.06001)**. Foz do Iguaçu. 1984.

ITAIPU BINACIONAL. **Aspectos teóricos dos Estudos realizados - Estudos para a reavaliação dos valores de controle para a instrumentação civil das estruturas e fundações (2660.50.15402.P)**. Foz do Iguaçu. 1999\_a.

ITAIPU BINACIONAL. **Valores de Controle para os instrumentos integrantes do "ADAS" e critérios para a geração de alarmes**. Foz do Iguaçu. 2002.

LACALLE, J. L. **“tsoutliers - R Package for Detection of Outliers in Time Series”**, 2016. Disponível em: <<https://www.jalobe.com/doc/tsoutliers.pdf>>. Acesso em: 01 fevereiro 2018.

LAZZAROTTO, E. **Obtenção de limites estatísticos de controle em gráficos de controle univariados e multivariados aplicados a dados de instrumentação de barragem**. Tese de Doutorado. UFPR, 2016.

MONTGOMERY, D. C. **Introdução ao Controle Estatístico de Qualidade**. 4a. ed. Rio de Janeiro: LTC, 2013.

MORETTIN, P. A.; TOLOI, C. M. **Análise de séries temporais**, 2ª edição rev. e ampl. São Paulo: Edgard Blucher, 2006.

PEÑA, D., TIAO, G. C., TSAY, R. S. **A Course on time Series Analysis**, Toronto: John Wiley & Sons, 2001.

ROSSO, J. A. *et al.* **Barragem de Itaipu - Lições aprendidas com o comportamento térmico das estruturas tipo gravidade aliviada e contraforte**. Foz do Iguaçu. 1995.

USACE. **Engineering and Design - Instrumentation of Embankment Dams and Levees**. U.S.Army Corps of Engineers. Washington, DC. 1995.