

**UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL REI
ENGENHARIA CIVIL - *CAMPUS* ALTO PARAÓPEBA**

**ELISA RODRIGUES DE OLIVEIRA VIDAL
ISABELA MARIA SOUZA HENRIQUES**

**COMPARAÇÃO ENTRE TRÊS MÉTODOS DE PREENCHIMENTOS DE FALHAS PARA
SÉRIES DE DADOS DE VAZÕES**

Ouro Branco

2017

ELISA RODRIGUES DE OLIVEIRA VIDAL

ISABELA MARIA SOUZA HENRIQUES

**COMPARAÇÃO ENTRE TRÊS MÉTODOS DE PREENCHIMENTOS DE FALHAS PARA
SÉRIES DE DADOS DE VAZÕES**

Trabalho de Conclusão de Curso apresentado a Coordenação do Curso de Graduação em Engenharia Civil – Ênfase em Estruturas Metálicas, da Universidade Federal de São João Del-Rei, Campus Alto Paraopeba, como requisito parcial para obtenção do grau de bacharel em Engenharia Civil.

Orientador: Emmanuel Kennedy da Costa Teixeira.

Ouro Branco

2017



Universidade Federal de São João Del Rei
Campus Alto Paraopeba
Trabalho de Conclusão de Curso

Trabalho de Conclusão de Curso de Engenharia Civil intitulado Comparação entre três métodos de preenchimentos de falhas para séries de dados de vazões, de autoria de Elisa Rodrigues de Oliveira Vidal e Isabela Maria Souza Henriques, aprovada pela banca examinadora constituída pelos seguintes professores:

Emmanuel Kennedy da Costa Teixeira- UFSJ (Orientador)

Dr. David Rafael Quintão Rosa- IFSudeste MG

Eliane Prado Cunha Costa dos Santos- UFSJ

Data de aprovação: Ouro Branco, 9 de Agosto 2017

Aos nossos queridos e amados pais,
Domingos Sávio e Maria Tereza, Salvador
e Celizabel, aos nossos irmãos, Bernardo
e Larissa, Pedro, e a Deus, dedicamos.

AGRADECIMENTOS

Agradecemos as nossas famílias pelo apoio, carinho e paciência neste momento de nossas vidas.

Ao professor Emmanuel pelos ensinamentos, orientação, apoio e compreensão ao longo do desenvolvimento do trabalho.

Aos colegas Rafael, Pietro, Fred e Saymon pela parceria durante todo o processo de pesquisa.

Aos amigos que estiveram ao nosso lado durante toda essa caminhada.

A Universidade Federal de São João Del Rei pela oportunidade de realizar a graduação na área de Engenharia Civil.

RESUMO

Objetivou-se verticalizar dados de vazões de diversas estações fluviométricas da bacia do Rio Doce e analisar aquelas que se apresentam estacionárias através de médias mensais. Para as estações que apresentaram estacionariedade, coeficiente de correlação com a estação de apoio maior que 0,84 e uma sequência de pelo menos 10 anos coincidentes de dados, foram comparados três métodos de preenchimento de falhas, sendo utilizada uma única estação de apoio para esses preenchimentos. Foram criados 10% de falhas aleatórias e, posteriormente, aplicados os métodos estatísticos de Regressão linear entre dados da estação com falhas e estação de apoio; Regressão linear entre os meses de uma mesma estação e Preenchimento através das áreas de drenagem das estações. Após a aplicação de cada método calculou-se o erro quadrático médio e o desempenho dos métodos foram avaliados para preenchimento. Isso foi possível, pois houve a comparação dos dados reais com os dados gerados. Nesta análise constatou-se que o método de Preenchimento através das áreas de drenagem das estações pode ser utilizado apenas para estações que se encontram no mesmo rio. O método da Regressão linear entre os meses de uma mesma estação apresenta bons resultados com erros quadráticos médios próximos, apesar de não preencher todas as falhas. O método de Regressão linear entre dados da estação com falhas e estação de apoio apresentou melhores resultados quando realizado com estações também localizadas no mesmo rio, apesar de também poder ser aplicado em outras estações.

Palavras-chave: Preenchimento de falhas, Vazões, Bacia do Rio Doce.

LISTA DE FIGURAS

FIGURA 1- ESTAÇÕES FLUVIOMÉTRICAS DA BACIA DO RIO DOCE.....	15
FIGURA 2- MAPEAMENTO DAS ESTAÇÕES ESTACIONÁRIAS.....	31
FIGURA 3- MAPA DAS ESTAÇÕES ANALISADAS - REPRESENTADAS COM O SÍMBOLO ESTRELA	32
FIGURA 4- GRÁFICO DA REGRESSÃO LINEAR ENTRE RAUL SOARES E MATIPÓ	33
FIGURA 5- GRÁFICO DA REGRESSÃO LINEAR ENTRE RAUL SOARES E CACHOEIRA DA NEBLINA	34

LISTA DE TABELAS

TABELA 1 - MÉTODOS UTILIZADOS PARA O PREENCHIMENTO DE FALHAS	27
TABELA 2 - ESTAÇÕES FLUVIOMÉTRICAS COM SÉRIES ESTACIONÁRIAS DE VAZÕES DA BACIA DO RIO DOCE	29
TABELA 3 - ESTAÇÕES COM COEFICIENTE DE CORRELAÇÃO QUE ATENDEM DOS PARÂMETROS ANALISADOS	32
TABELA 4 - COEFICIENTES A E B DAS ESTAÇÕES	33
TABELA 5 - COEFICIENTES DA ESTAÇÃO DE MATIPÓ.....	34
TABELA 6 - COEFICIENTES DA ESTAÇÃO CACHOEIRA DA NEBLINA	34
TABELA 7 - ÁREA DE DRENAGEM DAS ESTAÇÕES ANALISADAS	35
TABELA 8 - PREENCHIMENTO DE DADOS (EM M ³ /S) DA ESTAÇÃO FLUVIOMÉTRICA DE MATIPÓ.....	35
TABELA 9 - ERRO QUADRÁTICO MÉDIO (EM M ³ /S) DA ESTAÇÃO FLUVIOMÉTRICA DE MATIPÓ.....	36
TABELA 10 - PREENCHIMENTO DE DADOS (EM M ³ /S) DA ESTAÇÃO FLUVIOMÉTRICA CACHOEIRA DA NEBLINA	36
TABELA 11 - ERRO QUADRÁTICO MÉDIO (EM M ³ /S) DA ESTAÇÃO FLUVIOMÉTRICA CACHOEIRA DA NEBLINA	37

LISTA DE EQUAÇÕES

EQUAÇÃO (1)	18
EQUAÇÃO (2)	19
EQUAÇÃO (3)	19
EQUAÇÃO (4)	19
EQUAÇÃO (5)	19
EQUAÇÃO (6)	19
EQUAÇÃO (7)	21
EQUAÇÃO (8)	21

SUMÁRIO

1. INTRODUÇÃO	12
2. OBJETIVOS	14
2.1 OBJETIVO GERAL	14
2.2 OBJETIVOS ESPECÍFICOS	14
3. REVISÃO DE LITERATURA	15
3.1 DESCRIÇÃO DA BACIA DO RIO DOCE	15
3.2 MÉTODOS UTILIZADOS PARA O PREENCHIMENTO DE FALHAS NAS SÉRIES HISTÓRICAS	18
3.2.1 Método da regressão linear	18
3.2.2 Método comparativo entre vazões e áreas de drenagem	20
3.2.3 Erro quadrático médio	21
3.3 TRABALHOS SOBRE PREENCHIMENTO DE FALHAS DE SÉRIES TEMPORAIS	22
4. MATERIAIS E MÉTODOS	26
4.1 OBTENÇÃO DOS DADOS E ANÁLISE PRELIMINAR	26
4.2. PREENCHIMENTO DE FALHAS	27
5. RESULTADOS E DISCUSSÕES	29
5.1. OBTENÇÃO DOS DADOS E ANÁLISE PRELIMINAR	29
5.2. PREENCHIMENTO DE FALHAS	33
6. CONCLUSÃO	38
7. REFERÊNCIAS BIBLIOGRÁFICAS	39

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

ANA- Agência Nacional das Águas

UPGRHs- Unidades de Planejamento e Gestão dos Recursos Hídricos

CBHs- Comitês de Bacias Hidrográficas

PIB- Produto Interno Bruto

SIGRH - Sistema Integrado de Gestão de Recursos Humanos

UHEs- Usinas Hidrelétricas

PCHs- Pequenas Centrais Hidrelétricas

ANEEL- Agência Nacional de Energia Elétrica

PIRH- Plano Integrado de Recursos Hídricos

IGAM- Instituto Mineiro de Gestão das Águas

1. INTRODUÇÃO

De acordo com o Conselho Nacional da Água (2017), a água cobre cerca de 70% da superfície da Terra e totaliza um volume de 1386 milhões de quilômetros cúbicos. Não obstante, a quantidade de água doce disponível para utilização humana é limitada pelas condições naturais do planeta. A água doce é geralmente definida como a água com uma salinidade inferior a 1% da observada nos oceanos. De fato, apenas 2,5% de toda a água existente na Terra é doce, sendo o resto salgada (a maior parte encontra-se nos oceanos). Destes 2,5%, a maior parte (1,8%) está retida em forma de gelo na Antártida, no Ártico e nos glaciares, não estando disponível para uso humano. As necessidades em água da humanidade e dos ecossistemas terrestres têm de ser satisfeitas com base nos restantes 0,7% da água doce existente no planeta, que totalizam cerca de 10,7 milhões de quilômetros cúbicos.

Ainda de acordo com esse Conselho, a água doce está distribuída de forma muito desigual na superfície da Terra. Menos de dez países concentram 60% dos recursos de água doce disponíveis: Brasil, Rússia, China, Canadá, Indonésia, USA, Índia, Colômbia e República Democrática do Congo.

A gestão cuidada da interferência humana no ciclo hidrológico é essencial para assegurar o futuro da própria humanidade. A quantidade de água disponível, desde que bem gerida, é globalmente suficiente para qualquer cenário de crescimento demográfico, mas para que a água não seja um fator limitante do desenvolvimento e, mesmo, de sobrevivência em vastas zonas do mundo, é essencial que seja gerida de uma forma integrada (envolvendo abordagens técnicas, ecológicas e socioeconômicas), num contexto adverso de alterações climáticas.

A água é um recurso natural renovável essencial à vida e ao bem estar da população. É o indicador de desenvolvimento econômico e social que está presente em quase todas as atividades. Por essa razão torna-se imprescindível a adoção de critérios de gestão dos recursos hídricos e o conhecimento do ciclo hidrológico.

Os Planos de Recursos Hídricos podem ser elaborados por bacia hidrográfica (de domínio federal ou estadual), por estado e para o país. Bacia hidrográfica é uma área drenada constituída por rio, afluente e subafluente. Seu relevo é responsável pela drenagem da água e determina para onde essa irá escoar. De acordo com a Agência Nacional das Águas, o Brasil possui doze bacias hidrográficas que totaliza 14% de toda água doce presente no mundo.

A Bacia hidrográfica do Atlântico leste possui 364.677 quilômetros quadrados e abrange os territórios de Sergipe, Bahia, Minas Gerais e Espírito Santo. Esta é subdividida em dez sub-bacias, sendo uma delas a Bacia do Rio Doce, que contém uma extensão de 879 quilômetros e abrange os territórios de Minas Gerais e Espírito Santo. Essa sub-bacia contém 480 estações monitoradas, sendo que desses apenas 148 contêm dados de vazões diárias.

A vazão é uma das principais variáveis hidrológicas, de tal forma que é registrada diariamente em postos ou estações fluviométricas nas quais são monitoradas pela Agência Nacional das Águas. Esse registro é chamado de série histórica e permite a análise das frequências e das ocorrências das vazões ao longo dos anos. Algumas vezes não há o registro de dados, seja por falta de operador, inacessibilidade do local da medição ou perda de dados.

O uso de registro com falhas influencia o estudo estatístico da série, gerando resultados não confiáveis, sendo assim, é necessário fazer um preenchimento dos dados. Com as séries completas, pode-se iniciar o estudo de planejamento, projeto e operação de qualquer obra relacionada com o aproveitamento e controle dos recursos hídricos em uma bacia hidrográfica e auxiliar nas definições das outorgas do direito de uso da água.

Neste contexto, avaliou-se com o presente trabalho metodologias de preenchimento de falhas que foram testadas nas séries históricas de vazões das estações fluviométricas do Rio Doce.

2. OBJETIVOS

2.1 Objetivo Geral

Comparar três métodos de preenchimentos de falhas de dados de vazões médias mensais de estações fluviométricas da bacia do rio Doce.

2.2 Objetivos Específicos

Os objetivos específicos são:

- Avaliar a disponibilidade e qualidade dos dados fluviométricos disponíveis no banco de dados da *Hidroweb*;
- Avaliar quais métodos de preenchimento de falhas podem ser utilizados na bacia do Rio Doce e escolher métodos para ser utilizados neste trabalho;
- Comparar os métodos de preenchimento de falhas: Regressão linear entre os dados da estação com falhas e a estação de apoio; regressão linear entre os meses de uma mesma estação; e preenchimento através das áreas de drenagem das estações.

3. REVISÃO DE LITERATURA

3.1 Descrição da Bacia do Rio Doce

De acordo com o Comitê da Bacia Hidrográfica do Rio Doce (2016), essa bacia possui área de drenagem de 86.715 Km², dos quais 86% estão no Leste mineiro e 14% no Nordeste do Espírito Santo. Em Minas Gerais, é subdividida em seis Unidades de Planejamento e Gestão dos Recursos Hídricos (UPGRHs), às quais correspondem as seguintes sub-bacias e seus respectivos Comitês de Bacia Hidrográfica (CBHs): Rio Piranga (DO1), Rio Piracicaba (DO2), Rio Santo Antônio (DO3), Rio Suaçuí (DO4), Rio Caratinga (DO5), Rio Manhuaçu (DO6). No Espírito Santo, não há subdivisões administrativas, existindo CBHs dos Rios Santa Maria do Doce, Guandu, Pontões e Lagoas do Rio Doce e Barra Seca e Foz do Rio Doce. No total, a bacia possui 480 estações fluviométricas distribuídas entre esses dois estados.

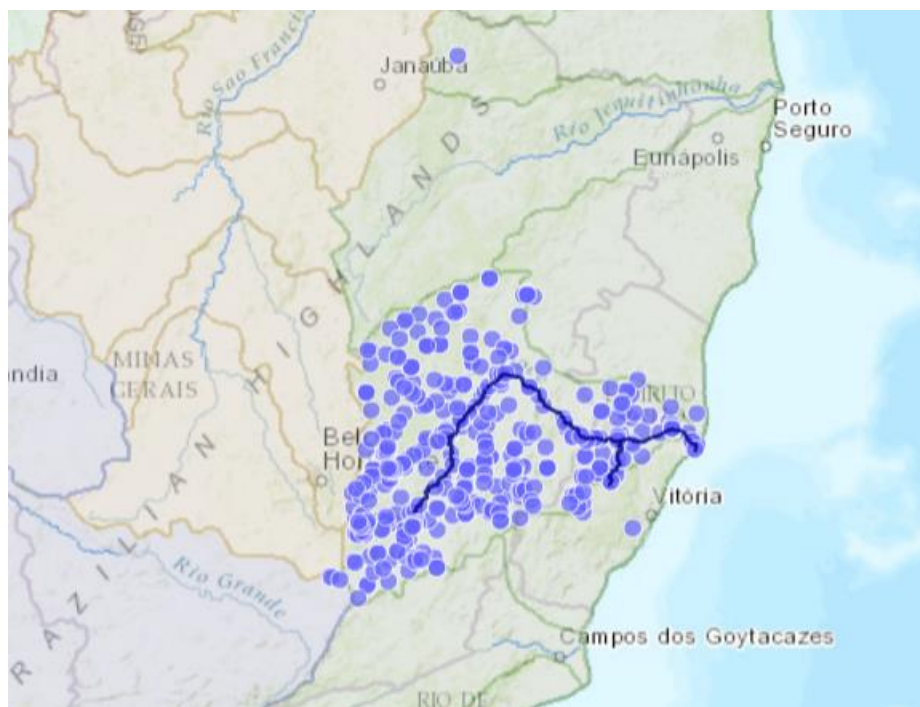


Figura 1- Estações Fluviométricas da bacia do Rio Doce

Fonte: snirh.gov.br/hidroweb/

O Rio Doce tem extensão de 879 Km e suas nascentes estão em Minas Gerais, nas Serras da Mantiqueira e do Espinhaço. O relevo da bacia é ondulado, montanhoso e acidentado. No passado, uma das principais atividades econômicas foi a extração de ouro, que determinou a ocupação da região e, ainda hoje, o sistema de drenagem é importante em sua economia, fornecendo água para uso doméstico, agropecuário, industrial e geração

de energia elétrica. Os rios da região funcionam, ainda, como canais receptores e transportadores de rejeitos e efluentes.

A população da Bacia do Rio Doce, estimada em torno de 3,5 milhões de habitantes, está distribuída em 228 municípios, sendo 202 mineiros e 26 capixabas. Mais de 85% desses municípios têm até 20 mil habitantes e cerca de 73% da população total da bacia concentra-se na área urbana, segundo dados de 2007. Nos municípios com até 10 mil habitantes, 47,75% da população vive na área rural. As bacias do Piranga e do Piracicaba, com o maior Produto Interno Bruto (PIB) industrial, concentram aproximadamente 48% da população total.

A atividade econômica na área é diversificada. Na agropecuária, lavouras tradicionais, cultura de café, cana de açúcar, criação de gado de corte e leiteiro, suinocultura, dentre outras. Na agroindústria, sobretudo a produção de açúcar e álcool. A região possui o maior complexo siderúrgico da América Latina, ao qual estão associadas empresas de mineração e reflorestadoras. Destacam-se, ainda, indústrias de celulose e laticínios, comércio e serviços voltados aos complexos industriais, bem como geração de energia elétrica, com grande potencial de exploração.

De acordo com Sistema Integrado de Gestão de Recursos Humanos (2017) a bacia possui relevância na geração hidrelétrica, sendo que atualmente encontram-se em operação 10 usinas hidrelétricas (UHEs), sendo 4 delas localizadas no rio Doce e 6 em seus afluentes. Há ainda 29 pequenas centrais hidrelétricas (PCHs) em operação e outros 148 outros aproveitamentos hidrelétricos na bacia, em funcionamento ou em outras fases de implantação, segundo dados da Agência Nacional de Energia Elétrica (ANEEL). Para o gerenciamento dessas UHEs e PCHs é fundamental ter o conhecimento de séries de dados hidrossedimentológicos confiáveis, o que muitas vezes não ocorrem por alguns motivos, como, por exemplo, dados faltantes. Assim, justifica-se a aplicação de metodologias de preenchimentos de falhas.

Uma das questões cruciais da bacia do rio Doce, destacadas no PIRH-Doce, diz respeito à qualidade da água. Em praticamente todos os pontos de monitoramento de qualidade da água do IGAM, ocorreram problemas com coliformes e parâmetros relativos à cor e turbidez, além de outros compostos tóxicos associados ao uso agrícola e à erosão do solo. O controle de geração de sedimentos passa, portanto, por um processo integrado de recuperação da bacia, envolvendo recuperação de matas ciliares, implantação de dispositivos de controle de enxurradas em estradas rurais e recuperação de áreas degradadas (inclusive pastagens).

Tendo em vista todos esses planos de gestão, o conhecimento das variáveis que compõem o regime hidrológico torna-se imprescindível para iniciar e melhorar as obras de condução, de desvios de cursos de água, de controle de cheias, aproveitamento hidrelétrico, projeto de captação para abastecimento, irrigação e beneficiamento industrial. Dessa maneira a Agência Nacional das Águas coleta os dados pelas estações de monitoramento e os utiliza para produzir estudos, definir políticas públicas e avaliar a disponibilidade hídrica. Por meio dessas informações, a ANA monitora eventos considerados críticos, como cheias e estiagens, disponibiliza informações para a execução de projetos, identifica o potencial energético, de navegação ou de lazer em um determinado ponto ou ao longo da calha do manancial, levanta as condições dos corpos d'água para atender a projetos de irrigação ou de abastecimento público, entre outros.

No monitoramento das séries hidrológicas, é comum a ocorrência de problemas mecânicos ou com operador, de modo que normalmente existem períodos sem registros ou com falhas nas observações. Como falhas, são designados dados cujos valores são incoerentes, denotam erros grosseiros e apresentam inconsistências.

De acordo com Horn (2016) a falta de informações hidrológicas é um problema para as diversas áreas de afinidades da hidrologia. A carência de informações hidrológicas deve-se, principalmente, à falta de estações de medição fluviométrica no local desejado. No entanto, quando estas existem, nem sempre apresentam dados confiáveis ou não possuem uma quantidade de informações considerada representativa para a região. Buscar informações sobre métodos que possam ser utilizados para geração de dados nas áreas carentes de monitoramento é de extrema importância. Pesquisadores do mundo todo tentam solucionar esse paradigma, substituindo dados provenientes de monitoramento por dados gerados a partir de métodos que utilizam informações reais.

Segundo Santos (2012), há uma necessidade de informações confiáveis, pois uma falha na série temporal pode comprometer a análise e a interpretação dos dados. Mais que isso, Silveira (1998) apresenta em seu trabalho que a falta de dados hidrológicos em pequenas bacias geram incertezas que comprometem o gerenciamento dos recursos hídricos. Como exemplo, pode-se citar a carência de energia para propriedades e aglomerados rurais devido à ausência de dados fluviométricos de pequenos mananciais para o desenvolvimento de avaliações de potencialidades hídricas e energéticas. Além disso, o abastecimento de água humano, a irrigação e o controle ambiental das bacias necessitam conhecer a distribuição temporal e espacial da vazão dos rios de pequenos mananciais.

Sabendo de todos esses problemas gerados pela falta de dados, para as bacias desprovidas desses registros sistemáticos, torna-se necessário desenvolver metodologias que possibilitem a transferência das informações hidrológicas existentes em outras bacias.

3.2 Métodos utilizados para o preenchimento de falhas nas séries históricas

Segundo Naghettini e Pinto (2007), o conhecimento das ferramentas de estatística é fundamental para a evolução e para a prática da hidrologia, onde se encontra diversificada gama de aplicações nas atividades rotineiras ligadas aos estudos e projetos de engenharia hidrológica, que necessitam das teorias probabilísticas para as suas soluções. Variáveis hidrológicas estarão sempre associadas a uma probabilidade de ocorrência. Em obras hidráulicas devem sempre ser dimensionadas para um determinado “risco” de falha. O objetivo da estatística é o de extrair informações significativas de uma dada massa de dados. As técnicas utilizadas em estatística aplicadas à hidrologia permitem avaliar a probabilidade de ocorrência de um fenômeno hidrológico com determinada magnitude.

3.2.1 Método da regressão linear

Existe um conjunto de métodos estatísticos que visam estudar a associação entre duas ou mais variáveis aleatórias. Dentre tais métodos, segundo Naghettini e Pinto (2007), a teoria da regressão e correlação ocupa lugar de destaque por ser o de uso mais difundido.

De acordo com o coeficiente de Correlação Linear de Pearson (NAGUETTINI e PINTO, 2007), duas variáveis apresentam uma correlação linear quando os pontos de diagrama de dispersão se aproximam de uma reta. O diagrama de dispersão é o primeiro fator de análise entre as observações simultâneas, permitindo visualizar o grau de associação entre os dados e a tendência de variação conjunta. A correlação pode variar de -1 a 1, sendo que quanto mais próximo desses valores maior será a tendência que as variáveis apresentam quanto à sua variação conjunta.

O coeficiente de correlação linear de Pearson é dado pela Equação 1.

$$r = \frac{S_{x,y}}{s_x s_y} \quad \text{Equação (1)}$$

onde, r é o coeficiente de correlação linear; $s_{x,y}$ é a covariância entre as variáveis; s_x e s_y são os desvios-padrão das amostras respectivamente.

A covariância é dada pela Equação 2.

$$s_{x,y} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}} \quad \text{Equação (2)}$$

Onde, $s_{x,y}$ é a covariância amostral entre as variáveis x e y ; \bar{x} e \bar{y} são as médias aritméticas de cada uma das variáveis; n é o tamanho da amostra; x_i e y_i são as observações simultâneas das variáveis.

Os desvios-padrão são calculados de acordo com as Equações 3 e 4.

$$s_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad \text{Equação (3)}$$

$$s_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}} \quad \text{Equação (4)}$$

Após o cálculo do coeficiente de correlação deve-se fazer um estudo da regressão cujo intuito é investigar e modelar a relação entre variáveis. A relação entre os dados de x e y pode se apresentar como: linear, quadrática, cúbico, exponencial, logarítmico, etc.

Entretanto, para esse estudo verificou-se que essa relação é linear, caracterizando-se como a equação de uma reta cuja equação é dada pela Equação 5.

$$y = \alpha + \beta x \quad \text{Equação (5)}$$

Onde, y é a variável dependente; x é a variável independente; α e β são os coeficientes do modelo.

Como maneira de verificar se o modelo proposto é adequado, deve-se calcular o coeficiente de determinação, que estabelece a parcela da variabilidade amostral de fato explicada pela reta de regressão. Esse coeficiente varia no intervalo de 0 a 1, onde os valores próximos de 1 indicam que o modelo proposto é adequado para análise. Esse coeficiente é determinado de acordo com a Equação 6.

$$r^2 = b^2 \frac{s_x^2}{s_y^2} \quad \text{Equação (6)}$$

Onde, r^2 é o coeficiente de determinação; s^2_x é a variância amostral de X; s^2_y é a variância amostral de Y; e b é o coeficiente angular da reta da regressão.

Segundo os estudos de Vieira *et. al.*(2016), as análises cujo coeficiente de determinação (r^2) foi superior a 0,8 indicaram uma relação linear satisfatória entre os valores observados e modelados.

De acordo com Santos *et. al.* (2015), a qualidade do ajuste na regressão está relacionada com coeficiente de determinação (r^2), quanto menor for este espalhamento, melhor a regressão representa o conjunto de valores observados. Seus estudos geraram resultados da matriz de correlação através de regressão linear múltipla considerando apenas as três melhores situações de defasagem encontradas dentre as estações analisadas. Por fim, foi indicada a existência de uma ótima correlação, $r = 0,87$ e $r = 0,84$. Também segundo Naghettini e Pinto (2007), para se evitar a multi-colinearidade, elimina-se uma entre cada conjunto de duas variáveis independentes da regressão múltipla, que apresentarem coeficiente de correlação inferior a 0,85. Por fim adotou-se um valor de 0,84 para o coeficiente de correlação nesse trabalho.

3.2.2 Método comparativo entre vazões e áreas de drenagem

Segundo Silva (2014), para utilizar o método de proporcionalidade de vazões, deve-se primeiramente analisar a região das estações e proceder ao enquadramento do local de interesse em relação às estações fluviométricas utilizadas no estudo, conforme um dos quatro casos preconizados a seguir: Local de interesse situado a montante de um posto com vazão conhecida; Local de interesse situado entre dois postos com vazão conhecida; Local de interesse situado à jusante de um posto com vazão conhecida; Local de interesse situado em um canal afluente, cuja foz está entre dois postos fluviométricos de um canal de ordem superior. No estudo cujo local de interesse está situado a montante de um posto com vazão conhecida pôde-se observar erros muitos grandes, tal que o maior é 76,6%. Além disso, nesse caso pôde-se observar também uma diferença de área de drenagem 4,4 vezes superior da seção de análise. Isso se justifica de acordo com Eletrobrás (1985b), que ressalta que esse método não deve ser aplicado quando a área de drenagem do posto fluviométrico em análise e a seção de interesse for superior a três.

Com base nessas informações e visto que nesse trabalho foram encontradas apenas duas estações comparativas, como apresentadas no item resultados, o método adotado

para análise foi montante e jusante, visto que o princípio adotado para ambos é o mesmo, e para os outros métodos são necessárias três estações.

Para estações situadas a montante ou jusante, o método comparativo entre vazões e áreas de drenagem, é utilizado uma simples regra de três. A vazão desconhecida da estação com dados faltantes é estimada pela razão da área de drenagem do seu próprio rio pela área de drenagem de uma estação com vazão conhecida (estação de apoio), utilizando a Equação 7.

$$Q_x = \frac{A_x}{A_y} Q_y \quad \text{Equação (7)}$$

Onde, Q_x é a vazão no local de interesse (m^3/s), Q_y é a vazão conhecida na estação de apoio (m^3/s), A_x é a área de drenagem no local onde se deseja conhecer a vazão (km^2) e A_y a área de drenagem da estação fluviométrica da estação de apoio (km^2).

3.2.3 Erro quadrático médio

Em estatística, o erro quadrático médio é uma forma de avaliar a diferença entre um estimador e o verdadeiro valor da quantidade estimada. Para uma amostra, define-se a variância amostral como o desvio quadrático médio, dada pela Equação 8.

$$S^2 = \frac{1}{N} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{Equação (8)}$$

Onde S é o desvio quadrático médio, \bar{x} é a média amostral, N é o número de amostras e x_i são as observações simultâneas das variáveis.

Segundo Lima (2009), o erro quadrático médio é muito sensível a grandes desvios entre os valores das séries comparadas, o que se torna mais relevante quando se trata de avaliação de grandes erros. Esse pode assumir qualquer valor não negativo e tem a mesma unidade de medida das séries. Quanto mais próximo de zero for seu valor, maior será a semelhança entre as séries simulada/prevista e a observada.

De acordo com Cavalcanti (2014), a raiz do Erro Quadrático Médio mede a amplitude do erro. Um erro quadrático médio de valor zero indica uma previsão perfeita, e conforme aumenta a diferença entre os valores estimados e observados o erro quadrático médio aumenta.

Para fins de estudo desse trabalho foi feita uma comparação entre os erros obtidos por cada método de preenchimento. Dessa forma não foi estabelecido um valor de erro no qual fosse considerado aceitável e sim analisado o menor dos valores encontrados para obtenção do método mais preciso.

3.3 Trabalhos sobre preenchimento de falhas de séries temporais

Como foi visto anteriormente, o uso de registro com falhas influencia o estudo estatístico da série, gerando resultados não confiáveis, sendo assim, é necessário fazer um preenchimento dos dados. Nesse contexto, vários foram os trabalhos que fizeram o preenchimento através de diferentes estudos, como os citados abaixo.

Santos (2012) analisou metodologias para avaliar a homogeneidade, a tendência e o preenchimento de falhas em séries de dados de temperatura máxima e mínima do ar e da precipitação em Minas Gerais. Na análise de tendência foram utilizados os testes de Mann-Kendall e análise de regressão. Para o preenchimento de falhas utilizou-se os modelos de SARIMA proposto por Box-Jenkins, que utilizam dados da própria estação com dados faltantes, e o modelo geoestatístico espaço-tempo que considera as estações vizinhas, ponderando tanto a dependência espacial como temporal dos dados. Comparando as duas metodologias, o modelo espaço-temporal está mais próximo da identidade apresentando menor dispersão em torno da reta estimada e maior coeficiente de determinação.

Pinheiro (2009) propôs e avaliou um método para calibração automática de um modelo chuva-vazão em bacias sem monitoramento hidrométrico, utilizando como paradigma do processo hidrológico a curva de permanência de longo período sintética, obtida a partir de um modelo estatístico regional. Como as curvas de permanência de longo período constituem funções características próprias do regime hidrológico da bacia hidrográfica, foram usadas como instrumento para calibração dos parâmetros de um modelo conceitual chuva-vazão. Uma vez obtidos os parâmetros que comandam a síntese hidrológica do modelo em questão, a simulação contínua de descargas, ao longo de um dado período de tempo, viabilizou avaliações hidrológicas diversas. Dessa maneira, foi proposto um método para regionalização de curvas de permanência de longo termo, permitindo a transferência dessa função hidrológica a locais não-monitorados, desde que esses se localizem na mesma região homogênea das bacias com dados. Depois foi realizada a calibração do modelo chuva-vazão RIO GRANDE, um modelo conceitual baseado na estrutura de balanço de umidade chamada Xinanjiang, amplamente conhecida, tendo como objetivo reproduzir as curvas de permanência sintetizadas a partir do modelo regional. Em ambas as fases, um experimento de validação cruzada de jack-knife foi

concebido a fim de calcular alguns índices de desempenho, o objetivo dos quais foi avaliar a confiabilidade proporcionada pelos procedimentos envolvidos.

Na tese de Barbosa (2004), o estudo se inicia com as vazões médias diárias que ocorrem nas diferentes seções dos cursos d'água naturais da bacia do rio do Carmo, e com as informações pluviométricas na forma de dados de chuvas diárias nesta bacia. Assim, foram construídas as séries anuais de vazões máximas diárias, de vazões mínimas de sete dias de duração e de vazões médias anuais bem como as séries anuais de chuvas máximas diárias, de totais anuais precipitados e de totais do semestre mais chuvoso. Complementarmente, foram construídas as curvas de permanência das vazões e as curvas de intensidade-duração-frequência das chuvas. Com base nestas séries, e tomando-se os resultados dos levantamentos das características fisiográficas (área de drenagem, comprimento de talvegue e densidade de drenagem) e das características climáticas (chuva, especialmente), são construídos os modelos matemáticos de regionalização das vazões médias, mínimas e máximas, por meio de modelos de regressão múltipla que consideram os efeitos da área de drenagem, do comprimento do curso d'água, da densidade de drenagem, da precipitação média, entre outros, sobre as vazões mencionadas. Com isso, disponibiliza-se uma ferramenta capaz de produzir as informações em áreas climaticamente homogêneas com ausência de dados de vazões e chuvas.

Oliveira *et. al.* (2010) compararam metodologias de preenchimento de falhas de séries históricas de precipitação pluvial anual. Utilizaram os métodos de ponderação regional (PR), regressões linear (RM) e potencial (RP) múltiplas, ponderação regional com base em regressões lineares (RS) e vetor regional (VR). Fizeram uma análise comparativa entre os métodos de preenchimento de falha e avaliaram os desvios relativos entre os valores observados e os estimados pelos métodos empregados. Pelos resultados obtidos pode-se concluir que, em ordem crescente, que os menores desvios relativos foram obtidos pelos métodos: RM, VR/RP, PR, VR/RM, VR/RS, RP, RS, VR/PR e VR. Com relação à metodologia do vetor regional, obteve-se uma melhoria nas estimativas do preenchimento das falhas quando se empregou a combinação com os demais métodos.

De acordo com o trabalho de Caldeira, Piyathisse e Nandalal (2016), o objetivo do estudo foi analisar diferentes métodos para preenchimento de falhas de dados de uma estação hidrológica. O estudo foi realizado em um rio localizado em uma área montanhosa da Sri Lanka. Foram recolhidos dados diários de 10 estações para ser comparado com os resultados analisados. A análise foi feita por sete diferentes métodos, sendo eles: Método da média aritmética, Método da relação normal, Método da distância inversa ponderada, Método da regressão linear, Método da regressão linear principal, Método da regressão

linear múltipla e Método probabilístico. Os resultados do estudo apresentado mostraram que para as estações que possuem apenas uma estação próxima, com alto coeficiente de correlação, o método probabilístico e o método de regressão linear apresentaram os melhores resultados. Para estações que apresentam baixos coeficientes de correlação com estações vizinhas identificadas, os métodos de distância inversa ponderada e método da relação normal apresentaram melhores resultados. Para obter dados precisos pelos métodos de regressão linear múltiplo e da regressão linear principal, deve-se obter estações que tenham um alto coeficiente de correlação com estações vizinhas disponíveis.

No artigo de Alves *et. al.*(2005) foi feito o estudo do município de Alfredo Wagner, que vem sofrendo constantemente com inundações devido à topografia acidentada da bacia hidrográfica na qual está inserido e ao manejo inadequado do solo, resultando no desmatamento de encostas e matas ciliares e no assoreamento do rio. O objetivo deste trabalho foi analisar os dados pluviométricos e fluviométricos na região a fim de contribuir em futuros estudos e planejamentos. Foram adquiridos dados de sete estações pluviométricas e uma fluviométrica. Algumas dessas estações possuíam falhas nas séries históricas. As falhas maiores que um mês foram descartadas, já as falhas de alguns dias foram preenchidas pelo método das razões dos valores normais (método das médias ponderadas) com os dados das outras seis estações. Além disso, foram gerados gráficos com curvas duplo-acumulativas para cada estação com o intuito de detectar eventuais erros de consistência nos dados históricos.

Fernandez (2004) analisou séries temporais com falhas das variáveis de temperatura máxima, temperatura mínima, temperatura média, umidade relativa e precipitação e foram usadas as médias mensais de treze estações do Rio Grande do Sul. Nas séries que estavam completas criou falhas artificiais e aplicou os métodos estatísticos de: análise de regressão múltipla, média simples, Steurer, média de três estações, proporção normal e análise harmônica a fim de prever os dados faltantes. Após a aplicação de cada método, calculou o erro absoluto, o erro quadrático e o erro percentual e avaliou o desempenho do método para o preenchimento em cada variável. Nestas análises constatou que, entre os métodos estudados, os que obtiveram melhores resultados para previsão de dados foram os de análise de regressão múltipla, Steurer e média de três estações.

De acordo com Presti *et. al.* (2008), as séries são constantemente afetadas com a presença de dados faltantes e, lidando com fatores estatísticos, deve-se preencher os valores vazios para esses serem considerados. No presente, há um grande número de técnicas estatísticas disponíveis para realizar esses estudos. Eles variam de métodos simples, como média das amostras, até métodos mais sofisticados, como imputação

múltipla. Uma nova metodologia para a estimativa de dados faltantes é proposta, na qual essa tenta misturar as vantagens óbvias das técnicas simples com a força das novas técnicas. O método proposto consiste em aplicar duas etapas consecutivas: Uma vez verificado que o monitoramento da estação específica é afetado pelos dados faltantes, a estação de monitoramento mais similar nas proximidades vizinhas é identificada, através de coeficientes de semelhança. Em segundo plano, um método regressivo é aplicado em ordem de estimular os dados faltantes. Nesse artigo, o preenchimento de vazios foi feito utilizando a precipitação pluvial das séries avaliadas na bacia do rio Candelaro, localizado no sul da Itália.

Segundo Kim *et. al.* (2008), a falta de registros de dados de precipitação diária é muito comum na prática de engenharia hidrológica. Contudo, esses devem ser substituídos por estimativas adequadas para serem utilizados de forma confiável em modelos hidrológicos. Nesse artigo é apresentada uma aplicação que desenvolva um novo modelo diário espacial de precipitação que se destina especificamente a preencher lacunas em um conjunto de dados diários de chuva. O modelo proposto é diferente do esquema convencional de geração de chuva diária, visto que tem como vantagem o uso de medidas simultâneas nos locais vizinhos para aumentar a precisão da estimativa. O modelo é baseado em uma abordagem em duas etapas para lidar com a ocorrência e a quantidade de chuvas diárias separadamente. Os testes aplicados revelaram que a abordagem neural probabilística é a melhor maneira de determinar a ocorrência de chuvas diárias e a regressão com transformação logarítmica é recomendada para estimar as quantidades diárias de precipitação.

4. MATERIAIS E MÉTODOS

4.1 Obtenção dos dados e análise preliminar

Inicialmente, foram retirados da plataforma da Agência Nacional das Águas (hidroweb.ana.gov.br acesso em setembro de 2016) dados de vazões diárias das estações presentes na bacia do Rio Doce.

Foram verticalizados os dados de vazões diárias das 480 estações, a fim de facilitar o desenvolvimento e análise da pesquisa. Ou seja, foram criadas apenas duas colunas de dados, uma contendo a data da vazão e a outra contendo o dado de vazão diária. Para isso, criou-se uma Macro no *software* Excel[®]. Pôde-se observar que as séries históricas apresentam dados faltosos ao longo dos anos. Observou-se também que os registros mais remotos de dados são do ano de 1927, sendo que algumas séries se estendem até 2016.

Segundo Naghettini e Pinto (2007), as séries históricas apresentam tendências ou periodicidade ao longo do tempo, devido a variações naturais do clima ou alterações induzidas pelo homem, quando ocorre esse fenômeno a série é dita não estacionária. A série hidrológica é dita estacionária quando suas propriedades estatísticas não se alteram ao longo do tempo. Séries hidrológicas devem ter como pré-requisito o atributo de estacionariedade, garantindo assim, uma confiabilidade no estudo.

Segundo Santos (2012), em análises de séries temporais o conceito de estacionariedade é fundamental. Uma série temporal é dita ser estacionária quando o comportamento da série não se altera com o passar do tempo. Isto é, quando suas propriedades estatísticas (média, variância e covariância são invariantes no tempo e a auto correlação depende apenas da defasagem). Dessa forma, a série é considerada fracamente estacionária, condição necessária para análises práticas de séries temporais.

Ainda, segundo Santos (2015), na previsão hidrológica em bacias hidrográficas sob diferentes condições, muitas vezes são desenvolvidos e aplicados modelos mensais de conversão. A vantagem dessa escala temporal está na simplificação da formulação de modelos hidrológicos, onde os números de parâmetros utilizados são menores do que naqueles desenvolvidos em base diária.

Dessa forma, para análise de estacionariedade das estações, foi desenvolvida uma função no *software* estatístico R[®] (APÊNDICE B), o qual é livre e apresenta código livre, com intuito de gerar séries de médias mensais para anos hidrológicos, sendo definido o início em outubro e o fim em setembro. Foram excluídos os anos que contem meses com mais de 10% de falhas (três dias de falhas), sendo utilizados para esse estudo apenas os

anos que contem um índice de dados mensais superior a 90%. Esse critério foi estabelecido para que o preenchimento de dados fosse realizado com valores mais condizentes com os efetivos. Se os meses com mais de 10% de falhas fossem considerados, essas médias seriam fictícias, gerando valores incoerentes.

Segundo Naghettini (2007), dentre os três testes não paramétricos existentes para análise da estacionariedade - Pettitt, Coeficiente de Correlação de Spearman e Mann-Kendall - o teste de Coeficiente de Correlação de Spearman é um dos mais usuais em estudos hidrológicos.

Dessa forma foi realizado o teste de estacionariedade de Spearman, que também foi feito no *software* estatístico R[®]. Assim, foram selecionadas a fim de estudo apenas as que apresentaram essas características.

Em seguida, as estações estacionárias foram comparadas entre si, a fim de encontrar duas com intervalo de dados coincidentes dentro de um mesmo período de pelo menos uma década. Foram realizadas diversas análises comparativas entre as estações, e o critério de seleção foi feito de acordo com o coeficiente de correlação encontrado pelo *Software* R[®]. Aqueles com coeficientes (r) maiores que 0,84 foram ditos válidos para preenchimento, assim como dito no item 3.2.1

4.2. Preenchimento de falhas

Para as estações fluviométricas que atendem os critérios estabelecidos no item anterior, foram aplicados os métodos de preenchimento de falhas listados no item 3.2.

Tabela 1 - Métodos utilizados para o preenchimento de falhas

Método 1	Regressão linear entre os dados da estação com falhas e a estação de apoio
Método 2	Regressão linear entre os meses de uma mesma estação
Método 3	Preenchimento através das áreas de drenagem das estações

O primeiro método aplicado foi o de Regressão Linear para duas estações correlacionáveis. Como será apresentado no item resultados, as estações Raul Soares e Cachoeira da Neblina, Raul Soares e Matipó apresentaram boas correlações entre as suas séries de médias mensais, sendo que Raul Soares serviu como estação de apoio em ambos os casos, e as estações Cachoeira da Neblina e Matipó foram as estações nas quais houve o preenchimento de falhas. Assim, foi realizado uma regressão linear no *software* R[®] com estes dados e os dados da estação de apoio (Raul Soares). Dentre os resultados fornecidos

pelo *software* R[®], puderam-se obter os coeficientes da reta (α e β) e assim foi feito o preenchimento das falhas pelo *software* Excel[®].

O segundo método foi o da Regressão linear entre os meses de uma mesma estação. Esse método foi feito para as estações Cachoeira da Neblina e Matipó. As falhas foram analisadas individualmente, sendo que o mês utilizado para o preenchimento foi aquele que apresentou o maior coeficiente de correlação. Assim, como no primeiro método, foram fornecidos os valores de regressão pelo *software* R[®], sendo possível o preenchimento das falhas pelo *software* Excel[®].

O terceiro método foi o de preenchimento através das áreas de drenagem das estações, feito após a verificação do coeficiente de correlação. Para este método foi utilizado apenas às áreas de drenagem das estações e a vazão da estação de apoio correspondente ao mês falho da estação a ser preenchida. Todos os dados conhecidos foram obtidos pelo site da Agência Nacional de Águas (ANA).

Para se verificar a precisão dos métodos de preenchimento foi desenvolvida uma função no *software* R[®] (APÊNDICE C), a qual gerava 10% de falhas aleatórias nas séries de médias mensais das estações Cachoeira da Neblina e Matipó, ou seja, essa função excluía valores reais conhecidos. Essas falhas geradas foram preenchidas pelos três métodos citadas anteriormente. O valor preenchido foi comparado com o valor conhecido que foi excluído. Para se determinar o quão próximos estavam os valores, utilizou-se o método estatístico do erro médio quadrático, o qual foi descrito no item 3.2.3. Foram utilizadas as mesmas falhas em todos os métodos para comparar os resultados de preenchimento obtidos, podendo assim escolher o método mais adequado.

5. RESULTADOS E DISCUSSÕES

5.1. Obtenção dos dados e análise preliminar

O teste de estacionariedade foi aplicado nas 148 estações fluviométricas, utilizando a função desenvolvida no *software* R®, apresentada no APÊNDICE A. Pôde-se verificar que das 148 estações que continham dados diários de vazão, apenas 72 se classificaram como estacionárias, sendo elas apresentadas na Tabela 1 abaixo.

Tabela 2 - Estações fluviométricas com séries estacionárias de vazões da bacia do rio Doce

ESTAÇÃO	CÓDIGO	PERÍODO	
		ANO INICIAL	ANO FINAL
Acesita	56695000	1975	1984
Afonso Cláudio	56991000	1939	2016
Afonso Cláudio Montante	56990990	1939	2016
Antonio Dias	56681000	1926	2016
Baixo Guandu	56992000	1938	2016
Barra De São Gabriel	56997000	1939	2016
Barra Do Capim	56990500	1939	2016
Barra Do Cuieté	56941000	1939	2016
Belo Oriente	56719998	1989	2014
Bicas Jusante	56336001	1939	2016
Bom Jesus Do Galho	56565000	1939	2016
Cachoeira Da Neblina	56974000	1938	1965
Cachoeira Do Guarara	56012000	1946	1963
Cachoeira Dos Óculos	56540001	1973	1973
Cachoeira Dos Óculos Montante	56539000	1975	2014
Cachoeira Escura	56720000	1941	1985
Capim Cheiroso	56635000	1955	1959
Caputira	56470000	1944	1957
Carrapato (Brumal)	56640000	1955	2013
Chalé	56985000	1939	1956
Chapada	56182000	1940	1964
Colatina Jusante	56994502	1976	1983
Conselheiro Pena	56943000	1945	1965
Coronel Fabriciano	56700000	1972	1974
Dom Cavati	56935000	1966	2014
Dores De Manhumirim	56983000	1939	2004
Fazenda Alegria	56590000	1972	1976
Fazenda Bragança	56976000	1972	2014
Fazenda Bretz	56847000	1972	1977
Fazenda Bretz - Montante	56846900	1997	2013

**Tabela 2- Estações fluviométricas com séries estacionárias de vazões da bacia do rio Doce
(continuação)**

ESTAÇÃO	CÓDIGO	PERÍODO	
		ANO INICIAL	ANO FINAL
Fazenda Do Retiro	56005000	1937	1954
Fazenda Florença	56780005	1951	1953
Fazenda Olaria	56670000	1941	1965
Fazenda Urupuca	56880000	1968	1991
Fazenda Vargem Alegre	56960005	1983	2016
Fojo	56185000	1939	1948
Governador Valadares	56850000	1969	2016
Ipanema	56988500	1938	2015
Itaguaçu – Jusante	56993002	1976	2010
Jequeri	56400000	1938	1965
Laranja Da Terra	56991500	1967	2005
Limoeiro	56155000	1940	2004
Linhares	56998000	1967	1994
Manhuaçu	56960000	1936	1983
Marimbondo	56150000	1940	1954
Mario De Carvalho	56696000	1986	2005
Matipó	56460000	1965	2005
Mutum	56989001	1973	2005
Mutum	56989000	1938	1966
Naque Velho	56825000	1974	2007
Nova Era - Montante	56659999	1974	1980
Nova Era Iv	56659998	1989	2005
Nova Era Telemetrica	56661000	2014	2016
Pai Tomás	56152000	1940	1954
Parada Da Independencia	56963000	1944	1952
Paraiso De Ipanema	56982000	1938	1953
Pingo D'água	56570000	1974	2005
Ponte Do Brumado	56633000	1938	1946
Ponte Do Caboclo	56195000	1938	1943
Ponte Do Carvalho	56105000	1952	1961
Ponte Do Pancas	56995500	1965	2005
Ponte Do Pancas	56995500	1965	2005
Ponte Do Peres	56430000	1938	1946
Ponte Itajuru	56631500	1937	1946
Ponte Itatiaia	56160000	1940	1946
Ponte Santa Rita	56158000	1940	1947

Tabela 2- Estações fluviométricas com séries estacionárias de vazões da bacia do rio Doce (continuação)

ESTAÇÃO	CÓDIGO	PERÍODO	
		ANO INICIAL	ANO FINAL
Raul Soares	56485000	1936	1977
Resplendor Jusante	56948005	1984	2004
São Miguel Do Anta	56385000	1965	2005
São Sebastião Da Encruzilhada	56990000	1938	1952
Senador Firmino	56065000	1941	2005
Usina Ana Florência	56130000	1950	1961

Como se pode perceber 48,65% das estações apresentaram-se estacionárias. As outras séries, portanto, apresentam uma tendência, ou um 'salto', ou uma periodicidade ao longo do tempo, como resultado de variações naturais do clima ou alterações induzidas pela ação do homem.

Essas estações estacionárias foram mapeadas de acordo com as coordenadas geográficas, e se apresentam na Figura 2 a seguir.

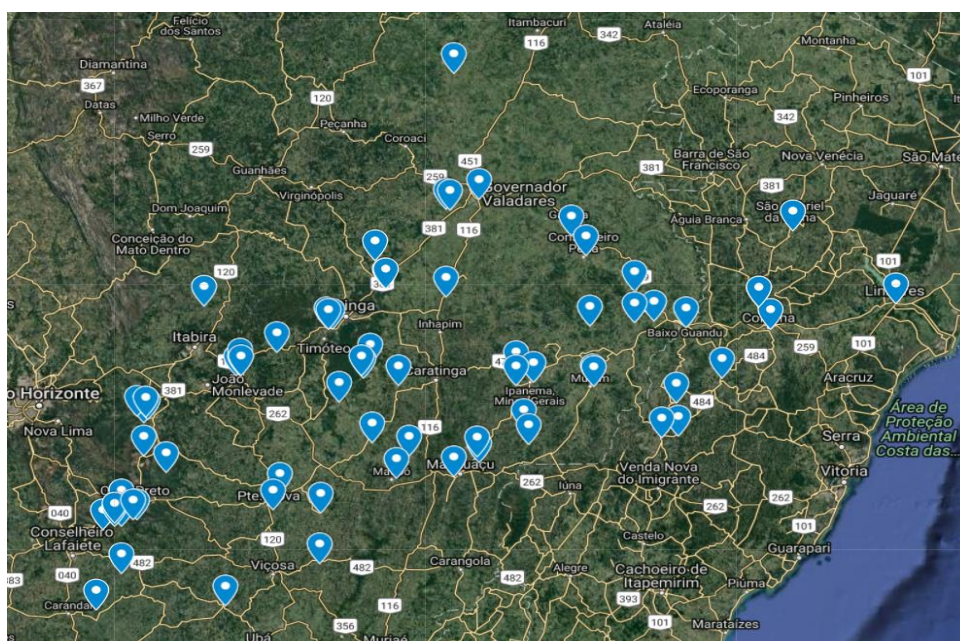


Figura 2- Mapeamento das estações estacionárias

FONTE: Google Maps

Como foi dito anteriormente, para que a não estacionariedade interferisse nas análises, trabalhou-se apenas com as estações estacionárias, as quais foram comparadas entre si para se obter a correlação entre elas. As estações que coincidiram dados de no mínimo 10 anos e obtiveram um coeficiente de correlação maior que 0,84 estão listadas na Tabela 2. Para realização da correlação entre as estações, desenvolveu-se uma função no R, qual está apresentada no APÊNDICE D.

Tabela 3 - Estações com coeficiente de correlação que atendem dos parâmetros analisados

Estação 1	Estação 2	Intervalo	Coeficiente de Correlação
Matipó	Raul Soares	1965-1977	0,9278
Cachoeira da Neblina	Raul Soares	1938-1965	0,9258

Como se pode observar na Tabela 3, foram poucas as estações que atenderam o conjunto de especificações a serem utilizados para o preenchimento de dados. Um dos motivos se deve ao fato de que o período de monitoramento das estações não coincidir, isto é, umas são recentemente monitoradas, enquanto outras foram há alguns anos desativadas. Além disso, a grande maioria dos coeficientes de correlação entre as estações não apresentaram bons resultados.

O mapeamento dessas estações selecionadas pode ser visualizado na Figura 3:

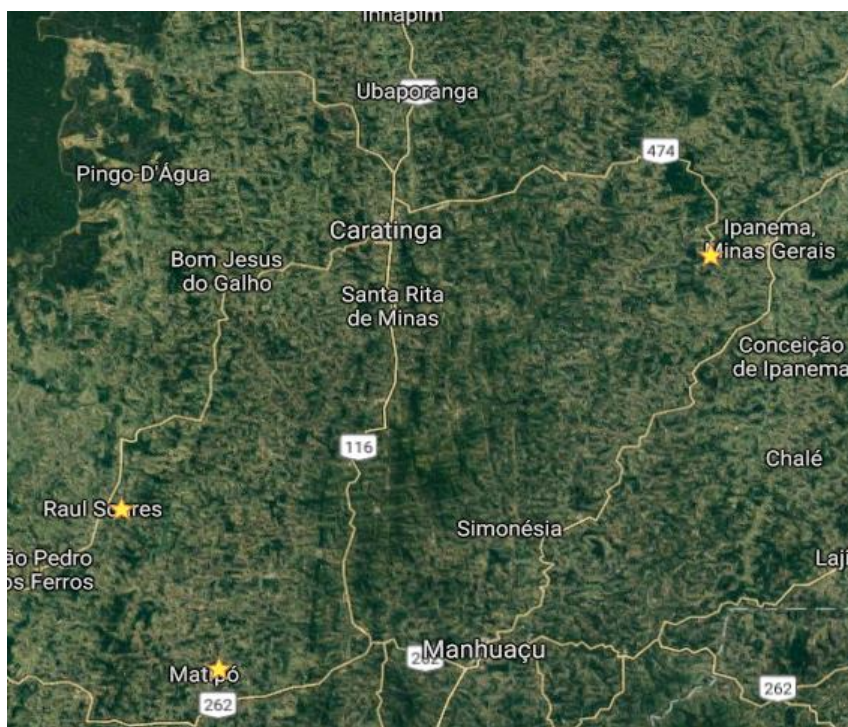


Figura 3- Mapa das estações analisadas - Representadas com o símbolo estrela

FONTE: Google Maps

A estação Raul Soares está localizada no município de Raul Soares, no rio Matipó. A estação Cachoeira da Neblina está localizada no município de Ipanema, no rio Manhuaçu, e a estação Matipó está localizada no município de Matipó, no rio Matipó. A entidade responsável pelas três estações é a ANA (Agência Nacional das Águas).

5.2. Preenchimento de falhas

Com as estações selecionadas, puderam-se realizar os três métodos de preenchimento de falhas. No primeiro método, fez-se a regressão linear entre a estação de apoio e a estação com falhas geradas. A função desenvolvida no R[®] está apresentada no APÊNDICE D. Os valores dos coeficientes gerados por essa função são apresentados na Tabela 3 abaixo.

Tabela 4 - Coeficientes α e β das estações

Estações	α	β
Matipó e Raul Soares	-0,1988	0,5336
Cachoeira da Neblina e Raul Soares	3,904	1,796

Nos gráficos apresentados nas Figuras 4 e 5 abaixo, pode-se observar os respectivos gráficos de dispersão e a reta de regressão. Lembra-se que a estação de apoio nos dois casos foi Raul Soares.

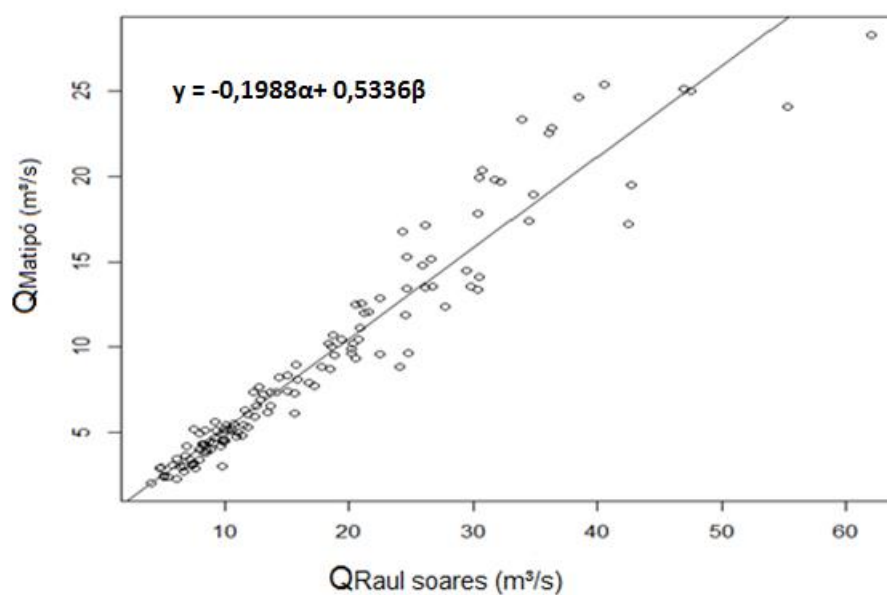


Figura 4- Gráfico da regressão linear entre Raul Soares e Matipó

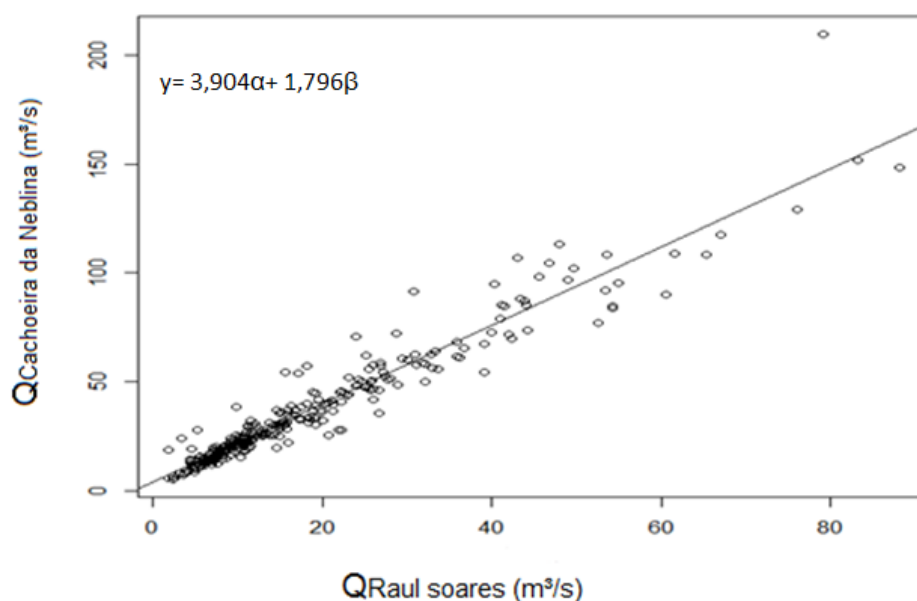


Figura 5- Gráfico da regressão linear entre Raul Soares e Cachoeira da Neblina

Com os coeficientes da reta, pôde-se fazer o preenchimento das falhas pelo *software* Excel®. Esse preenchimento pode ser visto na Tabela 8 e 10.

Para o segundo método, calculou-se os coeficientes de correlação entre os meses de uma mesma estação. A função desenvolvida para utilização desse método está no APÊNDICE E. Com os meses que apresentaram valores de coeficientes mais altos, fez-se uma regressão linear. Os coeficientes de correlação e os coeficientes α e β da reta estão apresentados na Tabela 5 e 6 abaixo:

Tabela 5 - Coeficientes da estação de Matipó

Mês com falha	Mês para regressão	r	α	β
3	8	0,830	6,146	-9,877
4	6	0,839	2,702	-3,828
5	6	0,939	1,496	-1,001
7	8	0,820	1,479	-0,728
8	6	0,875	0,601	0,677

Tabela 6 - Coeficientes da estação Cachoeira da Neblina

Mês com falha	Mês para regressão	r	α	β
5	6	0,977	1,304	-1,575
6	5	0,977	0,749	1,709
7	8	0,985	1,176	0,646
8	7	0,985	0,837	-0,292
9	8	0,870	0,858	1,366

Como se pode observar, não estão apresentados todos os meses nas tabelas, isso se justifica pelo fato de que os coeficientes de correlação, entre os meses com falhas e meses para regressão, serem menores que 0,84, o que implica em resultados não confiáveis. Com os valores das Tabelas 5 e 6, pôde-se fazer o preenchimento pelo *software* Excel®, e os resultados podem ser vistos nas Tabelas 8 e 10.

Para o terceiro método de preenchimento, foi necessária a utilização das áreas de drenagem disponíveis no *site* da ANA. A Tabela 7 abaixo apresenta a área de drenagem de cada uma das estações em estudo.

Tabela 7 - Área de drenagem das estações analisadas

Estação	Área de drenagem (km²)
Cachoeira da Neblina	2060
Matipó	616
Raul Soares	1390

Os resultados desse método podem ser analisados nas Tabelas 8 e 10.

Por fim, pode-se observar e comparar os valores reais de médias mensais de vazão com os valores encontrados pelos três métodos de preenchimento de dados da estação fluviométrica de Matipó, tais valores estão apresentados na Tabela 8 abaixo.

Tabela 8 - Preenchimento de dados (em m³/s) da estação fluviométrica de Matipó

Data	Valor Real	Método 1	Método 2	Método 3
ago-68	3.73	4.84	3.69	4.18
abr-69	6.27	6.08	12.38	5.22
mai-69	5.28	4.79	6.03	4.14
ago-69	3.58	3.56	3.50	3.12
ago-75	3.28	2.63	3.20	2.35
jul-77	3.96	4.69	11.11	4.06

Analisando a Tabela 8, é possível observar que os valores de preenchimento gerados por todos os métodos são bem próximos dos valores reais. O método 2, apesar de apresentar em alguns casos os melhores valores de preenchimento, pode-se observar alguns valores discrepantes, como em abril de 1969, onde o valor real é 6,27 m³/s e o método gerou um valor de 12,38 m³/s. Outro exemplo é julho de 1977, cuja vazão real é 3,96 m³/s e o valor imputado foi de 11,11 m³/s. Esses valores provavelmente são devidos a ocorrência de uma grande diferença entre a vazão média do mês com falha e o mês que foi utilizado para o preenchimento. Nesse caso a regressão gera resultados não satisfatórios.

Os Métodos 1 e 3 apresentaram bons resultados, sendo eles com valores próximos entre si e com um total preenchimento de falhas.

O erro quadrático encontrado pelos três métodos é mostrado na Tabela 9 a seguir.

Tabela 9 - Erro quadrático médio (em m³/s) da estação fluviométrica de Matipó

Método 1	Método 2	Método 3
0,56	3,38	1,55

O erro médio quadrático do Método 2 apresenta o pior resultado devido aos valores discrepantes gerados. No Método 3 o erro gerado foi bem pequeno, como já era de se esperar, pois como citado em Silva (2014) esse método é eficiente para estudos cuja a área de drenagem da estação de apoio (Raul Soares – 1390 km²) é até três vezes maior que a da estação em análise (Matipó – 616 km²). O Método 1 foi o que apresentou menor erro, portanto o melhor resultado. Esse método se apresentou o mais eficiente em vários trabalhos, como exemplo no estudo de Fernandez (2007), onde foi analisado o comportamento de cada método em cada variável e através do cálculo do erro médio, erro quadrático e do erro percentual o melhor método encontrado foi o de regressão linear.

Para a estação da Cachoeira da Neblina, o preenchimento de dados pelos três métodos é apresentado na Tabela 10 a seguir.

Tabela 10 - Preenchimento de dados (em m³/s) da estação fluviométrica Cachoeira da Neblina

Data	Valor Real	Método 1	Método 2	Método 3
mai/39	18.68	20.03	18.20	13.30
set/39	12.00	15.59	11.62	9.64
mar/41	55.85	61.72	65.88	47.71
mai/42	32.08	30.77	32.83	22.17
jun/44	25.94	27.34	24.62	19.34
ago/44	19.36	19.85	18.29	13.16
mai/45	44.16	44.24	43.65	33.28
jul/45	28.89	28.67	29.13	20.44
mar/47	37.18	38.85	36.12	28.84
jun/49	39.78	41.16	36.94	30.75
mai/50	25.77	27.21	26.40	19.23
ago/51	17.98	17.41	17.28	11.15
ago/52	24.63	26.86	23.64	18.95
set/52	24.17	27.81	21.76	19.73
jun/58	22.14	24.60	16.64	17.07
mar/61	48.29	55.63	42.51	42.68
jun/61	21.82	25.59	20.47	17.89

Tabela 10 - Preenchimento de dados (em m³/s) da estação fluviométrica Cachoeira da Neblina (continuação)

Data	Valor Real	Método 1	Método 2	Método 3
jun/63	21.87	17.39	19.92	11.12
jul/63	17.23	13.88	18.64	8.23
ago/63	15.73	13.01	15.28	7.51
jun/64	18.56	17.93	18.42	11.57

De uma maneira geral, os resultados dos Métodos 1 e 2 são os que mais se aproximam dos valores reais. No Método 3, os dados preenchidos se apresentaram mais distantes dos dados reais quando comparados com os métodos 1 e 2.

E o erro quadrático encontrado pelos três métodos é dado na Tabela 11.

Tabela 11 - Erro quadrático médio (em m³/s) da estação fluviométrica Cachoeira da Neblina

Método 1	Método 2	Método 3
2,44	3,03	5,99

O Método 3 obteve um grande erro. Esse valor é esperado já que as duas estações não estão localizadas no mesmo rio. O Método 2 obteve um erro relativamente bom se comparado com os outros valores. O método 1 mais uma vez apresentou o menor erro e melhores resultados de preenchimento.

Como se pode observar, o erro quadrático médio de Cachoeira da Neblina x Raul Soares, se apresenta no geral, maior que de Matipó x Raul Soares. Possivelmente, esses valores são maiores devido ao fato das estações não se apresentarem no mesmo rio e por haver uma grande variação de vazão em alguns meses que apresentam períodos mais chuvosos, o que ressalta mais ainda a teoria que as estações devem se apresentar no mesmo rio para obtenção de melhores resultados. Quando a estação é comparada com dados dela mesma (Método 2), o valor do erro gerado é semelhante.

6. CONCLUSÃO

Ao longo do desenvolvimento deste trabalho pode-se observar que a maioria das séries históricas de vazão apresentam períodos de falhas tornando-as inapropriadas para seu estudo. Nesse aspecto, torna-se fundamental o desenvolvimento de metodologias de preenchimento de falhas. Os preenchimentos foram feitos como apresentado anteriormente no projeto e pode-se fazer as seguintes conclusões.

O método de Regressão linear entre dados da estação com falhas e estação de apoio (método 1) apresentou bons resultados para ambas as análises. Para a estação de Matipó esses resultados foram melhores, o que é justificado pelo fato de estar presente no mesmo rio da estação de apoio Raul Soares. De maneira geral se apresentou como o melhor método para o preenchimento das falhas.

O método de Regressão linear entre os meses de uma mesma estação (método 2) se mostrou eficaz e apresenta erros relativamente baixos, porém não houve um total preenchimento de dados devido a correlação dos coeficientes não estar dentro dos padrões adequados para obter confiabilidade.

O método de Preenchimento através das áreas de drenagem das estações (método 3) apresentou melhores resultados para a estação de Matipó, o que já era de se esperar pois para uma boa eficiência do método as estações devem estar no mesmo rio.

Por fim, deve ser avaliado inicialmente o período das séries hidrológicas e a localização em relação à estação de apoio para escolher o melhor método a se utilizar. Sendo que em ambos os casos, com uma boa correlação, o método de Regressão linear entre dados da estação com falhas e estação de apoio pode ser utilizado.

7. REFERÊNCIAS BIBLIOGRÁFICAS

Agência Nacional das Águas (s.d.). Acesso em Junho de 2017, disponível em <http://www.snirh.gov.br/hidroweb/>

Agência Nacional das Águas (s.d.). Acesso em Julho de 2017, disponível em www.ana.gov.br/

ALVES. A, KOBIYAMA. M., SILVA. R.V, CHECCHIA. T – **Análise de dados hidrológicos na região do município de Alfredo Wagner/SC**, Simpósio de Recursos Hídricos do Sul, I Simpósio de Águas da AUGM, Florianópolis, 2005.

BARBOSA, S. E. S. – **Análise de dados hidrológicas e regionalização de vazões da bacia do Rio do Carmo**, Ouro Preto, UFOP, 2004.

CALDEIRA. H. P. M. G., PIYATHISSE. V. R. P. C., NANDALAL. K. D. W.- **A Comparison of Methods os Estimating Missing Daily Rainfall Data**, Sri Lanka, The Institution of Engineers, 2016.

CAVALCANTI. D.L.O – **Estudo comparativo de métodos de estimação de vazões máximas de projeto para grandes bacias hidrográficas**, São Carlos, 2014.

Comitê da Bacia Hidrográfica do Rio Doce (s.d.). Acesso em Julho de 2017, disponível em <http://www.cbhdoce.org.br/>

Conselho Nacional da Água (s.d.). Acesso em Julho de 2017, disponível em <http://conselhonacionaldaagua.weebly.com/aacutegua-no-planeta-terra.html>

ELETROBRÁS. Centrais Elétricas Brasileiras S.A. Manual de minicentrals hidrelétricas. Rio de Janeiro, 1985b.

FERNANDEZ. M. N. – **Preenchimento de falhas em séries temporais**, Rio Grande, UFRG, 2007.

HORN. J.F.C – **Estimativa de vazão com dados escassos: Novas hipóteses para o método Silveira**, Santa Maria, UFSM, 2016.

KIM. T.W., AHN. H. – **Spatial rainfall model using a pattern classifier for estimating missing daily rainfall data**, Springer Journal, 2008

LIMA. J.P.R., ALVES.J.M.B.- **Um estudo de downscaling dinâmico de precipitação intrasazonal acoplado a modelo chuva-vazão na bacia hidrográfica alto-médio São Francisco**, Revista Brasileira de Meteorologia, São Paulo, Setembro, 2009.

NAGHETTINI, M.; PINTO, E. J. A. **Hidrologia Estatística**. 1. ed. Belo Horizonte: CPRM – Companhia de Pesquisa de Recursos Minerais – Superintendência Regional de Belo Horizonte, 2007.

OLIVEIRA. L. F. C et al. – **Comparação de metodologias de séries históricas de precipitação pluvial anual**, Lavras, UFLA, Goiânia, UGF, 2010.

PAZ. A.R. – **Hidrologia aplicada**, Caxias do Sul, UERS, 2004.

PINHEIRO. V. B. – **Calibração de um modelo chuva-vazão em bacias sem monitoramento fluviométrico a partir de curvas de permanência sintéticas**, Belo Horizonte, UFMG, 2009.

PRESTI. R.L., BARCA. E., PASSARELLA. G. – **A methodology for treating missing data applied to daily rainfall data in the Candelaro River Basin**, Italia, 2008.

REVISTA BRASILEIRA DE RECURSOS HÍDRICOS, MOREIRA. M. C., SILVA. D. D.- **Análise de métodos para estimativa das vazões da bacia do Rio Paraopeba**, Volume 19, nº2- Abril/junho, 2014.

REVISTA BRASILEIRA DE RECURSOS HÍDRICOS, SILVEIRA. G.L., TUCCI. C.E.M., SILVEIRA. A.L.L.- **Quantificação de vazão em pequenas bacias sem dados**, Volume 3, nº3- Jul/Set, 1998.

SANTOS. F.A.A et al. - **Avaliação do uso de dados de precipitação gerados pelo CPC/NCEP na modelagem hidrológica em uma região hidrográfica do rio Araguaia**, Simpósio Brasileiro de Sensoriamento Remoto - SBSR, João Pessoa-PB, Brasil, Abril, 2015.

SANTOS, R.S. - **Homogeneidade e reconstrução de séries climatológicas para localidades no estado de Minas Gerais**, UFV, Viçosa, 2012.

Sistema Integrado de Gestão de Recursos Humanos (s.d.). Acesso em Junho de 2017, disponível em http://www.sigrh.sp.gov.br/public/uploads/ckfinder/EncarteRioDoce_2016.pdf.

VEIRA. J.P.B., ALMEIDA. R.C. - **Regionalização de Vazões com o Emprego de Redes Neurais Artificiais RBF**, UFP, Curitiba. 2016.

APÊNDICE A

```

# Essa função aplica o teste de estacionariedade.

N<-length(vazoes)
mt<-order(vazoes)
vazoesordenadas<-0
soma<-0
for (i in 1:N)
{
  vazoesordenadas[i]<-vazoes[(mt[i])]
} for (i in 1:N)
{
  for (j in 1:N)
  {
    if (vazoes[i]==vazoesordenadas[j])
    {
      soma<-soma+((j-i)**2)
    }
  }
}
rs<-(-1-((6*soma)/((N**3)-N)))
### para N > 10
E.rs<-0
Var.rs<-((1)/(N-1))
T<-((rs)/(sqrt(Var.rs)))
if (z < T) {
  print("Rejeitar H0, ou seja, Série Não Estacionária com 95% de Confiança")
} else {
  print("Nao rejeitar H0, ou seja, Série Estacionária com 95% de Confiança")
}

```

APÊNDICE B

```
# Essa função gera as médias mensais dos dados de vazões.

# 1) Copiar todo o código e colar no arquivo .Rdata que contem todos os dados de cada estação e
dar enter.

# 2) Verificar se os dados começam em 01/01 de qualquer ano.

# 3) Utilizar o código abaixo para exportar em .csv:

# write.csv2(testes(media,ano0,anof),"Nome da estação.csv")

testes<-function(dados,ano0,anof)

{

  {

    ### CALCULA MÉDIAS MENSAIS E RETORNA MATRIZ COM SEIS COLUNAS

    anos<-anof-ano0+1

    meses<-anos*12

    medmensal<-matrix(nrow=meses,ncol=6)

    # COLUNA 1 - ANOS DA SÉRIE

    # COLUNA 4 - MESES DO ANO

    {

      a<-0

      am<-1

      for (i in 1:anos)

      {

        for (j in 1:12)

        {

          medmensal[am,1]<-ano0+a

          medmensal[am,4]<-j

          am<-am+1

        }

        a<-a+1

      }

    }

  }

}
```

```

}

# COLUNA 2 - MÉDIAS ANUAIS

# COLUNA 3 - % DE FALHAS NAQUELE ANO

# COLUNA 5 - MÉDIAS MENSAIS

# COLUNA 6 - % DE FALHAS NAQUELE MÊS

{

    j<-0 # contador de dias

    l<-0 # contador de anos

    m<-1 # contador de meses

    for (i in ano0:anof)

    {

        if (i%%4==0 && (i%%400==0 || i%%100!=0))

        {

            l<-l+1

            x<-0

            y<-0

            for (k in 1:366)

            {

                x[k]<-dados[j+k]

                if (is.na(dados[j+k])==TRUE)

                {

                    y<-y+1

                }

            }

        }

        media<-mean(x,na.rm=TRUE)

        falhas<-((y/366)*100)

        for (z in m:(m+11))

```

```

{
    medmensal[z,2]<-media
    medmensal[z,3]<-falhas
}

# Calcula as médias mensais
x<-0
y<-0
for (k in 1:31) # Janeiro
{
    x[k]<-dados[j+k]
    if (is.na(dados[j+k])==TRUE)
    {
        y<-y+1
    }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/31)*100)
j<-j+31
m<-m+1
x<-0
y<-0
for (k in 1:29) # Fevereiro
{
    x[k]<-dados[j+k]
    if (is.na(dados[j+k])==TRUE)
    {
        y<-y+1
    }
}

```

```

medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/29)*100)

j<-j+29
m<-m+1

x<-0
y<-0
for (k in 1:31) # Março
{
    x[k]<-dados[j+k]
    if (is.na(dados[j+k])==TRUE)
    {
        y<-y+1
    }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/31)*100)

j<-j+31
m<-m+1

x<-0
y<-0
for (k in 1:30) # Abril
{
    x[k]<-dados[j+k]
    if (is.na(dados[j+k])==TRUE)
    {
        y<-y+1
    }
}

```

```
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/30)*100)
j<-j+30
m<-m+1
x<-0
y<-0
for (k in 1:31) # Maio
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/31)*100)
j<-j+31
m<-m+1
x<-0
y<-0
for (k in 1:30) # Junho
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
```

```
medmensal[m,6]<-((y/30)*100)
j<-j+30
m<-m+1
x<-0
y<-0
for (k in 1:31) # Julho
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/31)*100)
j<-j+31
m<-m+1

x<-0
y<-0
for (k in 1:31) # Agosto
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
```

```
medmensal[m,6]<-((y/31)*100)
j<-j+31
m<-m+1

x<-0
y<-0
for (k in 1:30) # Setembro
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/30)*100)
j<-j+30
m<-m+1

x<-0
y<-0
for (k in 1:31) # Outubro
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
```



```
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/31)*100)
j<-j+31
m<-m+1

x<-0
y<-0
for (k in 1:30) # Novembro
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/30)*100)
j<-j+30
m<-m+1

x<-0
y<-0
for (k in 1:31) # Dezembro
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
```

```

    }
    medmensal[m,5]<-mean(x,na.rm=TRUE)
    medmensal[m,6]<-((y/31)*100)
    j<-j+31
    m<-m+1
  }

else
{
  l<-l+1
  x<-0
  y<-0
  for (k in 1:365)
  {
    x[k]<-dados[j+k]
    if (is.na(dados[j+k])==TRUE)
    {
      y<-y+1
    }
  }
  media<-mean(x,na.rm=TRUE)
  falhas<-((y/365)*100)
  for (z in m:(m+11))
  {
    medmensal[z,2]<-media
    medmensal[z,3]<-falhas
  }

# Calcula as médias mensais

```

```
x<-0
y<-0
for (k in 1:31) # Janeiro
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/31)*100)
j<-j+31
m<-m+1

x<-0
y<-0
for (k in 1:28) # Fevereiro
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/28)*100)
j<-j+28
m<-m+1
```

```

x<-0
y<-0
for (k in 1:31) # Março
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/31)*100)
j<-j+31
m<-m+1
x<-0
y<-0
for (k in 1:30) # Abril
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/30)*100)
j<-j+30
m<-m+1

```

```
x<-0
y<-0
for (k in 1:31) # Maio
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/31)*100)
j<-j+31
m<-m+1
x<-0
y<-0
for (k in 1:30) # Junho
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/30)*100)
j<-j+30
m<-m+1
```

```
x<-0
y<-0
for (k in 1:31) # Julho
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/31)*100)
j<-j+31
m<-m+1
x<-0
y<-0
for (k in 1:31) # Agosto
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/31)*100)
j<-j+31
m<-m+1
```

```
x<-0
y<-0
for (k in 1:30) # Setembro
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/30)*100)
j<-j+30
m<-m+1

x<-0
y<-0
for (k in 1:31) # Outubro
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/31)*100)
j<-j+31
m<-m+1
```

```
x<-0
y<-0
for (k in 1:30) # Novembro
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/30)*100)
j<-j+30
m<-m+1

x<-0
y<-0
for (k in 1:31) # Dezembro
{
  x[k]<-dados[j+k]
  if (is.na(dados[j+k])==TRUE)
  {
    y<-y+1
  }
}
medmensal[m,5]<-mean(x,na.rm=TRUE)
medmensal[m,6]<-((y/31)*100)
j<-j+31
```



```
                                m<-m+1
                                }
                                }
                                }

medmensal<-medmensal[,-2]
medmensal<-medmensal[,-2]

# Coluna 1 - Ano
# Coluna 2 – Mês
# Coluna 3 - média mensal
# Coluna 4 - % de falha mensal

# Exclui os 9 primeiros meses
for (i in 1:9)
{
    medmensal<-medmensal[-1,]
}

# Exclui os 3 últimos meses
for (i in 1:3)
{
    N<-nrow(medmensal)
    medmensal<-medmensal[-N,]
}
}

return(medmensal)
}
```

APÊNDICE C

| # Essa função insere falhas aleatórias nos dados.

```
gera.NA<-function(dados,porcentagem)
{
  N<-length(dados)
  falhas<-floor(porcentagem*N)
  linhas<-sort(sample(1:N,falhas,r=F))

  for (i in 1:falhas)
  {
    for (j in 1:N)
    {
      if (linhas[i]==j)
      {
        dados[j]<-NA
      }
    }
  }
  return(dados)
}

# write.csv2(gera.NA(vazoes1,0.1),"Dados com falhas.csv")
```

APÊNDICE D

Essa função gera os coeficientes α e β da reta através da regressão linear.

```
reglin<-lm(vazoesx~vazoesy)
```

```
reglin
```

Função que gera o coeficiente de correlação linear de Pearson.

```
summary (reglin)
```

APÊNDICE E

```

# Essa função faz a regressão entre dos meses da mesma estação.

# 1) Copiar todo o código e colar no arquivo .Rdata que contem todos os dados de cada estação e
dar enter.

# 2) Verificar se os dados começam em 01/01 de qualquer ano.

# 3) Chamar a função da seguinte maneira:

#     testes(cotas,ano0,anof)
#     testes(vazoes,ano0,anof)

# Onde ano0 é o ano inicial da série e anof é o ano final da série.

testes<-function(medmensal,mesF,mesA,mesP)
{

    # Segundo teste: regressão entre dois meses da mesma estação (falha em outubro,
    regressão de outubro e setembro)

    N<-nrow(medmensal)

    # mês anterior

    xA<-0

    kA<-1

    # mês posterior

    xP<-0

    kP<-1

    # mês com falha

    y<-0

    k<-1

    for(i in 1:N)
    {

        if((medmensal[i,2])==mesF)
        {

            y[k]<-medmensal[i,3]

```

```
        k<-(k+1)
    }
    if((medmensal[i,2])==mesA)
    {
        xA[kA]<-medmensal[i,3]
        kA<-(kA+1)
    }
    if((medmensal[i,2])==mesP)
    {
        xP[kP]<-medmensal[i,3]
        kP<-(kP+1)
    }
}
print(y)
print(xA)
print(xP)

reglin<-lm(y~xA)
reglin2<-lm(y~xP)

print(summary(reglin))
print(summary(reglin2))
}
```