Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Ajuste de modelos e comparação de séries temporais para dados
de vazão específica em microbacias pareadas

Marcus Vinicius Silva Gurgel do Amaral

Dissertação apresentada para obtenção do título
de Mestre em Ciências. Área de concentração:
Estatística e Experimentação Agronômica

Piracicaba
2014

Marcus Vinicius Silva Gurgel do Amaral
Engenheiro Florestal

**Ajuste de modelos e comparação de séries temporais para dados de vazão
específica em microbacias pareadas**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientadora:
Prof^a Dr^a **TACIANA VILLELA SAVIAN**

Dissertação apresentada para obtenção do título
de Mestre em Ciências. Área de concentração:
Estatística e Experimentação Agronômica

**Piracicaba
2014**

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Amaral, Marcus Vinicius Silva Gurgel do

Ajuste de modelos e comparação de séries temporais para dados de vazão específica em microbacias pareadas / Marcus Vinicius Silva Gurgel do Amaral -- versão revisada de acordo com a resolução CoPGr 6018 de 2011. -- Piracicaba, 2014. 91 p: il.

Dissertação (Mestrado) -- Escola Superior de Agricultura "Luiz de Queiroz", 2014.

1. Microbacias pareadas 2. Séries temporais 3. Comparação de séries temporais
4. Imputação de dados I. Título

CDD 551.483
A485a

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte -O autor"

DEDICATÓRIA

*Aos meus pais,
Alcides Gurgel do Amaral e
Renata Aparecida Alves
da Silva Gurgel do Amaral*

*Ao meu irmão,
Matheus Silva Gurgel do Amaral*

*À minha companheira,
Simone Silmara Werner*

Com amor, DEDICO.

AGRADECIMENTOS

Agradeço inicialmente a Deus, por minha crescente fé e principalmente por minha vida em toda sua perfeita imperfeição.

A minha orientadora, Prof^ª. Dr^ª. Taciana Villela Savian, pela orientação, dedicação, incentivo e paciência ao longo do desenvolvimento desta pesquisa.

Ao Prof. Dr. Sílvio Ferraz, pelo apoio, concessão do banco de dados utilizado neste trabalho e valiosas sugestões.

Aos meus pais, Alcides Gurgel do Amaral e Renata Aparecida Alves da Silva Gurgel do Amaral pela contribuição na formação de um indivíduo que sequer se tornará metade do ser humano que eles são.

Aos professores do programa de Pós-graduação em Estatística e Experimentação Agronômica, Dr. Carlos Tadeu dos Santos Dias, Dr. César Gonçalves de Lima, Dr^ª. Clarice Garcia Borges Demétrio, Dr. Cristian Villegas Lobos, Dr. Décio Barbin, Dr. Gabriel Adrián Sarriés, Dr. Idemauro Antonio Rodrigues de Lara, Dr. Paulo Justiniano Ribeiro Jr., Dr^ª. Renata Alcarde, Dr^ª. Roseli Aparecida Leandro, Dr. Sílvio Sandoval Zocchi, Dr^ª. Sônia Maria de Stefano Piedade, por todos os ensinamentos e amizade.

Aos funcionários do Departamento, Luciane Brajão, Mayara Segatto, Solange Sabadin, Eduardo Bonilha, Jorge Wiendll e Rosni Honofre.

Aos amigos Djair Durand, Maria Cristina, Pedro Amoedo, Ricardo Klein e Thiago Oliveira pela ajuda em momentos de dificuldade e amizade.

Em especial agradeço ao amigo Rafael Moral, pelos conselhos, ajuda e amizade incondicionais. Por sua mãe, Estela Moral, que ao me acolher em sua casa, não poupou esforços desde o primeiro momento para me fazer sentir como um membro da família.

Trago as melhores lembranças do nosso projeto, a banda "Produto de Kronecker", essa família de dois irmãos (Rafael e Ricardo) e um tio (José Guilherme).

Guardo os maiores agradecimentos e meus melhores sentimentos para Simone Silmara Werner, minha esposa, eterna companheira, amiga e confidente, que com sua sinceridade nunca deixou de confiar em mim, mesmo quando minhas forças pareciam não serem suficientes para encontrar uma solução.

A todos aqueles que indiretamente colaboraram para a conclusão desta dissertação.

A CAPES pela concessão da bolsa de mestrado.

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE FIGURAS	13
LISTA DE TABELAS	17
1 INTRODUÇÃO	19
2 REVISÃO BIBLIOGRÁFICA	21
2.1 Contexto Hidrológico	21
2.2 Bacias Hidrográficas	21
2.3 Imputação de dados perdidos	22
2.3.1 Tipos de dados perdidos	23
2.3.2 Bootstrap	23
2.3.3 Amelia II (EMB)	24
2.4 Séries Temporais	26
2.4.1 Processos estocásticos	27
2.4.1.1 Processos estocásticos estacionários	27
2.4.2 Função de autocovariância e autocorrelação	29
2.4.3 Decomposição clássica de uma série temporal	30
2.4.4 Estacionariedade	31
2.4.4.1 Teste de Cox-Stuart	32
2.4.4.2 Teste de Fisher	33
2.4.5 Modelos	35
2.4.5.1 Modelos autorregressivos (AR)	36
2.4.5.2 Modelos de médias móveis (MA)	36
2.4.5.3 Modelos integrados autorregressivos de médias móveis (ARIMA)	37
2.4.5.4 Função de autocorrelação (fac)	38
2.4.5.5 Função de autocorrelação parcial (facp)	39
2.5 Testes para comparação de séries temporais	41
2.5.1 Teste de Igualdade das Funções de Autocorrelação	41
2.5.2 Método de comparação de Séries Temporais proposto por Silva, Ferreira e Sáfyadi (2000)	42
2.5.2.1 Teste de Box-Pierce	43

3 METODOLOGIA	45
3.1 Material	45
3.2 Método	47
4 RESULTADOS E DISCUSSÃO	51
4.1 Imputação direta de dados para as séries temporais das microbacias Colônia e Mortandade	52
4.2 Análise das séries temporais para valores de vazão específica ($L/s.km^2$) em microbacias pareadas	55
4.3 Teste de igualdades das funções de autocorrelação para a comparação das séries de vazão específica para as microbacias Colônia e Mortandade	62
4.4 Teste de comparação de séries temporais proposto por Silva, Ferreira e Sáfadi (2000)	65
4.5 Imputação de dados em intervalos para as séries temporais das microbacias Colônia e Mortandade	68
4.6 Teste de igualdades das funções de autocorrelação para a comparação das séries de vazão específica para as microbacias Colônia e Mortandade	71
4.7 Teste de comparação de séries temporais proposto por Silva, Ferreira e Sáfadi (2000)	72
5 CONCLUSÕES	75
REFERÊNCIAS	77
APÊNDICE	81

RESUMO

Ajuste de modelos e comparação de séries temporais para dados de vazão específica em microbacias pareadas

A crescente preocupação com o meio ambiente pressiona a sociedade como um todo para a uma mudança rumo a hábitos mais sustentáveis. No setor produtivo, o impulso se dá pelo desenvolvimento de técnicas mais eficientes de produção, embasados em pesquisas e experimentos de campo. No setor florestal, além da preocupação com a técnicas de manejo e com o solo, o principal recurso a ser preservado é a água. Por meio do monitoramento de rios em bacias hidrográficas, séries históricas são coletadas, possibilitando o uso da teoria de séries temporais para ajuste de modelos pela metodologia Box e Jenkins. Em casos de monitoramentos de microbacias pareadas, existe a possibilidade de se comparar séries temporais, como descrito no presente trabalho. Em duas microbacias pareadas localizadas na região centro-leste do estado do Paraná, em uma fazenda no município de Telêmaco Borba, dados correspondendo a duas séries temporais distintas de vazão específica foram coletados. Devido a presença de falhas nos conjuntos de dados, uma metodologia para imputação foi utilizada de duas maneiras diferentes, possibilitando a posterior comparação das duas séries temporais pela metodologia de séries temporais. De acordo com os resultados, verifica-se que ambas as séries são diferentes tanto para o teste de comparação das funções de autocorrelação, quanto para o teste de comparação de séries temporais proposto por Silva, Ferreira e Sáfiadi (2000). Portanto, segundo a caracterização dos estudos em microbacias pareadas, pode-se constatar que o manejo florestal empregado nos dois locais influenciam de forma diferente no comportamento da variável avaliada.

Palavras-chave: Microbacias pareadas; Séries temporais; Comparação de séries temporais; Imputação de dados

ABSTRACT

Fitting of models and comparison of time series for specific flow data in paired catchments

The growing concern for the environment presses society as a whole for a change towards sustainable habits. Regarding the production systems, more efficient production techniques based on research and field experiments are needed. As for forestry, besides the concern with management techniques and with soil preparation, the main resource to be preserved is water. Time series are collected by monitoring rivers in drainage basins, making possible the use of time series theory for fitting models based on Box and Jenkins methodology. When studying paired drainage basins, it is possible to compare time series, as described in this work. Two time series consisting of specific flow data were collected in a farm situated in the municipality of Telêmaco Borba, Eastern Paraná state, in two paired drainage basins. Because there were missing data, imputation techniques were used, making it possible to compare the two time series. Results showed that the time series are different for the comparison of the autocorrelation test and the time series comparison test proposed by Silva, Ferreira e Sáfiadi (2000). Therefore, according to studies involving paired drainage basins, different forest management techniques influence differently the behavior of the response variable in the different drainage basins.

Keywords: Paired catchments; Time series; Comparison of time series; Data imputation

LISTA DE FIGURAS

Figura 1 - Processo estocástico interpretado como uma família de variáveis aleatórias (Fonte: MORETTIN; TOLOI, 2006)	28
Figura 2 - Processo estocástico interpretado como uma família de trajetórias (Fonte: MORETTIN; TOLOI, 2006)	28
Figura 3 - Mapa da fazenda Monte Alegre com especificação das microbacias Mor- tandade e Colônia (extraído de Voigtlander, 2007)	45
Figura 4 - Valores diários de vazão específica ($L/s.km^2$), para microbacia Colônia, no período de 13/07/2005 a 28/02/2013	51
Figura 5 - Valores diários de vazão específica ($L/s.km^2$), para microbacia Mortan- dade, no período de 13/07/2005 a 28/02/2013	52
Figura 6 - Valores diários de vazão específica imputados de forma direta e seus res- pectivos intervalos de 90% confiança para a microbacia Colônia	53
Figura 7 - Valores diários de vazão específica imputados de forma direta e seus res- pectivos intervalos de 90% confiança para a microbacia Mortandade	54
Figura 8 - Densidade dos dados imputados e dos dados observados para microbacia Colônia	54
Figura 9 - Densidade dos dados imputados e dos dados observados para microbacia Mortandade	55
Figura 10 - Valores diários de vazão específica ($L/s.km^2$) para as microbacias (a) Colô- nia e (b) Mortandade, no período de 13/07/2005 a 28/02/2013	55
Figura 11 - Representação dos valores amplitude versus média para a série de valores diários de vazão específica para as microbacias (a) Colônia e (b) Mortandade	56
Figura 12 - Representação dos valores amplitude versus média para o logaritmo da série de valores diários de vazão específica para as microbacias (a) Colônia e (b) Mortandade	57
Figura 13 - Logaritmo da série de valores diários de vazão específica para microbacia Colônia	57
Figura 14 - Logaritmo da série de valores diários de vazão específica para microbacia Mortandade	58
Figura 15 - Série de valores diários de vazão específica transformada e diferenciada para microbacia Colônia	58

Figura 16 -Série de valores diários de vazão específica transformada e diferenciada para microbacia Mortandade	59
Figura 17 -Função de autocorrelação para microbacia colônia calculada na série log .	59
Figura 18 -Função de autocorrelação para microbacia mortandade calculada na série log	60
Figura 19 -Função de autocorrelação para microbacia colônia calculada na série diferenciada	60
Figura 20 -Função de autocorrelação para microbacia mortandade calculada na série diferenciada	60
Figura 21 -Periodograma para a série transformada e diferenciada de valores diários de vazão específica para microbacia Colônia	61
Figura 22 -Periodograma para a série transformada e diferenciada de valores diários de vazão específica para microbacia Mortandade	61
Figura 23 -Função de autocorrelação para a série estacionária de valores diários de vazão específica para as microbacias (a) Colônia e (b) Mortandade	62
Figura 24 -Função de autocorrelação parcial para a série estacionária de valores diários de vazão específica para microbacia Colônia	63
Figura 25 -Função de autocorrelação parcial para a série estacionária de valores diários de vazão específica para microbacia Mortandade	63
Figura 26 -Função de autocorrelação parcial estimada comum	63
Figura 27 -Função de autocorrelação residual para a série da microbacia Colônia . . .	64
Figura 28 -Função de autocorrelação residual para a série da microbacia Mortandade	64
Figura 29 -Série da diferença entre as séries de valores para as microbacias Colônia e Mortandade	65
Figura 30 -Função de autocorrelação para a série da diferença entre as microbacias Colônia e Mortandade	66
Figura 31 -Função de autocorrelação parcial para a série da diferença entre as microbacias Colônia e Mortandade	66
Figura 32 -Periodograma para a série da diferença entre as séries de valores para as microbacias Colônia e Mortandade	67

Figura 33 - Valores diários de vazão específica imputados por intervalos e seus respectivos intervalos de 90% confiança para as microbacias Colônia (a) e Mortandade (b)	68
Figura 34 - Valores diários de vazão específica para as microbacias Colônia (a) e Mortandade (b)	69
Figura 35 - Valores diários de vazão específica para as microbacias Colônia (a) e Mortandade (b)	69
Figura 36 - Representação dos valores amplitude versus média para a série de valores diários de vazão específica para as microbacias (a) Colônia e (b) Mortandade	70
Figura 37 - Representação dos valores amplitude versus média para logaritmo das séries de valores diários de vazão específica para as microbacias (a) Colônia e (b) Mortandade	70
Figura 38 - Séries temporais transformadas para os valores diários de vazão específica das microbacias (a) Colônia e (b) Mortandade	70
Figura 39 - Séries temporais transformadas para os valores diários de vazão específica das microbacias (a) Colônia e (b) Mortandade	71
Figura 40 - Funções de autocorrelação das séries temporais diferenciadas dos valores diários de vazão específica para as microbacias (a) Colônia e (b) Mortandade	71
Figura 41 - Funções de autocorrelação parcial residual das séries temporais de valores diários de vazão específica para as microbacias (a) Colônia e (b) Mortandade	72
Figura 42 - Função de autocorrelação parcial estimada comum	72
Figura 43 - Série da diferença entre as séries de valores para as microbacias Colônia e Mortandade	73
Figura 44 - Valores diários de vazão específica para as microbacias Colônia (a) e Mortandade (b)	73

LISTA DE TABELAS

Tabela 1 - Equações para o cálculo da vazão ($m^3 s^{-1}$) para a cota H (m) da microbacia

Colônia 46

Tabela 2 - Equações para o cálculo da vazão ($m^3 s^{-1}$) para a cota H (m) da microbacia

Mortandade 47

1 INTRODUÇÃO

As florestas, sejam naturais ou plantadas, distribuídas pelos seis diferentes biomas do território brasileiro, desempenham diferentes e importantes funções. Além de se tratarem de fontes de produtos florestais, madeireiros ou não, os processos ecológicos desses ecossistemas auxiliam na conservação de recursos hídricos e da biodiversidade, influenciam a estabilidade climática e possuem valores culturais. De acordo com o resumo atualizado de 2013 do serviço florestal brasileiro (Serviço Florestal Brasileiro, 2013), a área total de florestas plantadas corresponde a 1,55% da área ocupada por florestas no Brasil, totalizando 7.185.943 ha.

O setor florestal brasileiro tem apresentado crescimento devido a características ambientais favoráveis e aplicação de diferentes técnicas de melhoramento genético de sementes e clonagem de espécies florestais. Isso leva o Brasil a se destacar em produtividade tanto para os plantios de *Eucalyptus* quanto para os plantios de *Pinus* (Serviço Florestal Brasileiro, 2013). Uma vez que estes são plantios homogêneos de rápido crescimento, é clara não só a modificação da paisagem em termos da estrutura visual de cobertura do solo, como também a mudança na dinâmica de ciclagem de nutrientes e consumo de água (POGGIANI, 1985; SWANK; MINER, 1968)

Faz-se necessário, dada a importância dos recursos hídricos, o monitoramento sistemático de área plantadas quanto à alteração da qualidade e quantidade de água produzida. Assim, na hidrologia florestal, é comum a instalação de experimentos em microbacias hidrográficas com a intenção de verificar a qualidade ambiental das práticas de manejo florestal, bem como suas consequências para os ecossistemas (ANDRÉASSIAN, 2004).

Como alternativa experimental para a avaliação de impactos no uso do solo, segundo Hewlett (1971), o princípio de microbacias pareadas é utilizado na tentativa de se obter uma forma de comparação entre as características de microbacias próximas e com características semelhantes. Nesse sentido, diferentes tipos de experimentos têm sido instalados nessas condições a fim de estudar os diferentes processos ecológicos e bioquímicos envolvidos nos ecossistemas naturais, florestais, agrícolas e áreas de intensa atividade antrópica (ROTHACHER, 1970; HIBBERT, 1965).

Um dos objetivos em hidrologia florestal é o monitoramento de microbacias ao longo do tempo, de tal forma, que é comum a obtenção de séries históricas de dados

(WATSON et al., 2001). Estas são caracterizadas como séries temporais das variáveis observadas nos experimentos. Quando são obtidas em experimentos realizados em microbacias pareadas, um dos objetivos é o estudo das diferenças entre as séries de dados. Por meio de testes de comparações de séries temporais, é possível avaliar se duas séries diferentes são resultado de um mesmo processo natural (ECHEVERRY; TOLOI, 2000). Esse conhecimento é de grande importância para a hidrologia florestal devido à mudança de uso do solo, que interfere diretamente na ecologia da paisagem.

Os principais objetivos deste trabalho de pesquisa são, além de comparar dados de monitoramento de vazão específica de microbacias pareadas em uma fazenda de plantio florestal pertencente à empresa Klabin Paraná, apresentar e aplicar diferentes métodos para a comparação de séries temporais e utilizar um método para lidar com conjuntos de dados incompletos que foi empregado de duas maneiras diferentes.

2 REVISÃO BIBLIOGRÁFICA

2.1 Contexto Hidrológico

Desde de antiguidade há relatos da importância da água e da manutenção das florestas para o equilíbrio dinâmico existente entre as duas partes. Andréassian (2004) cita os aspectos históricos do estudo de interações entre florestas e recursos hídricos. No início, os debates eram predominantemente embasados em verificações empíricas. Segundo o autor, a França presenciou um debate com argumentos “românticos e históricos” a respeito das mudanças hidroclimatológicas em regiões afetadas pelo corte raso, ou seja, extração de madeira em grandes áreas de uma única vez sem qualquer tipo de seleção ou manejo, que precedeu debates científicos sobre o assunto.

Os debates empíricos gerados nesse momento histórico foram sucedidos por debates científicos baseados em comparações entre áreas com diferentes proporções de cobertura vegetal e monitoramentos com diferentes durações (ANDRÉASSIAN, 2004).

Watson et al. (2001) sugeriu comparações entre microbacias pareadas com a utilização de modelos de séries temporais envolvendo regressões múltiplas. Nesse caso, um modelo de regressão é ajustado entre a microbacia que não sofre alteração e a microbacia em que houve tratamento para a verificação de correlação entre os comportamentos de ambas, incluindo componentes sazonais.

Brown et al. (2013) propõe a utilização da mesma metodologia utilizada por Watson et al. (2001), que no entanto não considera variações sazonais entre as microbacias próximas já que tal modelo não indicou incremento significativo para o modelo proposto.

2.2 Bacias Hidrográficas

Uma bacia hidrográfica pode ser caracterizada como uma área ou região drenada por um rio e seus afluentes. Nesses locais as terras mais altas do relevo são caracterizadas como divisores de águas que, ou escoarão superficialmente as águas das chuvas, ou favorecerão o processo de infiltração no solo formando nascentes e abastecendo o lençol freático. O escoamento superficial até as partes mais baixas do terreno, dependendo do volume, será caracterizado como rio, ou riachos, formados nas partes superiores e mais íngremes chamadas cabeceiras, descendo todo o relevo e aumentando de volume até a chamada foz, área

do rio que desemboca no oceano (BARELLA et al., 2000).

Um conceito mais formal considera uma bacia hidrográfica como um sistema geomorfológico aberto. Lima (1986), identifica um equilíbrio dinâmico e inconstante responsável pelo delicado equilíbrio em que se encontram os sistemas hidrológicos. Dessa forma, fatores como relevo, regime de chuvas e cobertura vegetal são diretamente responsáveis pelo deflúvio e quantidade de água produzida.

Microbacia é definida como a menor unidade territorial da porção geográfica que compreende os canais de confluência delimitada por divisores naturais (SABANÉS, 2002). A ocorrência de uma microbacia geralmente está delimitada geograficamente em uma extensão territorial compreendida entre áreas com maior altitude, de onde a água captada parte convergindo para um mesmo local, formando assim rios de menor volume.

O conceito de microbacias pareadas envolve dois ou mais cursos d'água com características semelhantes, sejam elas, solo, área, vegetação, clima e regime de chuvas, e o mais importante, devem estar próximas uma da outra. O pareamento das microbacias proporciona a possibilidade de comparação por meio da utilização de métodos estatísticos Brown et al. (2013).

2.3 Imputação de dados perdidos

Um problema decorrente da obtenção ou coleta de dados é a presença de falhas. Seja por problemas em aparelhos, falta de experiência de pessoal de campo ou erros de medição, os dados perdidos são um problema real que gera dificuldades na análise de séries temporais, já que a correlação existente entre os dados é o principal motivo da utilização de tal abordagem.

A imputação de dados possibilita, em seus mais diversos mecanismos, a suposição de observações que não foram coletadas. No entanto, as metodologias de imputação podem, cada uma com seu grau de precisão, fornecer valores de observações com uma boa confiabilidade (RUBIN, 1976).

Geralmente a análise de dados é realizada assumindo que as falhas podem simplesmente ser ignoradas. No entanto os problemas de coleta que incorrem em falhas não são sempre os mesmo, resultando em tipos diferentes de dados perdidos.

2.3.1 Tipos de dados perdidos

A forma de deleção dos dados, ou seja, motivo pelo qual as observações não foram obtidas, deve ser avaliada uma vez que dados perdidos podem comprometer consideravelmente as análises pela possibilidade de introdução de viés. Segundo Little e Rubin (2002) existem três definições para dados perdidos. Supondo X e Y variáveis aleatórias:

- OAR - *Observed at Random* - essa definição de dados faltantes pode ser avaliada caso falhas em Y estiverem relacionadas ao comportamento de X , ou seja, se X não determina o padrão de deleção em Y , elas serão OAR .
- MAR - *Missing at Random* - os dados são MAR se não há algum padrão em Y que possa ditar falhas no próprio Y .
- $MCAR$ - *Missing Completely at Random* - caso o padrão de falhas não seja função de Y ou de X , ou seja, caso os dados perdidos sejam OAR e MAR ao mesmo tempo, eles serão $MCAR$.

A caracterização de falhas permite avaliar qual a forma mais adequada de lidar com o conjunto de dados de interesse. Em primeira instância, classificar dados perdidos pela definição $MCAR$ ou no mínimo MAR implica dizer que o processo gerador de deleção pode ser ignorado. Caso contrário não deve ser ignorado.

Dentre as várias abordagens atuais para lidar com dados faltantes, o algoritmo EM (Expectation Maximization) pode ser citado como uma alternativa para dados MAR (RUBIN, 1976). Esse algoritmo consiste na estimação de parâmetros de modelos em duas etapas. Inicialmente o conjunto é separado em dados faltantes e não faltantes e os parâmetros estimados a partir do conjunto de dados observado são utilizados na predição dos valores faltantes. Em seguida utilizando o novo conjunto de dados, novos parâmetros são estimados. Os passos são repetidos até a convergência do processo.

2.3.2 Bootstrap

O *Bootstrap* como método de reamostragem, introduzido por Efron (1979), é baseado na construção de novas amostras a partir de uma amostra inicial retirada de uma população de estudo. Com utilidade tanto inferencial quanto para diagnósticos, o processo de reamostragem *bootstrap* tem funcionalidade na avaliação de erro padrão, viés e

distribuição de probabilidade de um estimador. Trabalhos propostos por Freedman (1981) e Wu (1986) tratam sobre características dos estimadores, precisão de coeficientes estimados e influência do viés gerado pelos processos de reamostragem. Aplicações do método de reamostragem bootstrap em séries temporais são apresentados em Efron e Tibshirani (1986).

Segundo Efron (1979), dada uma amostra aleatória $x = (x_1, \dots, x_n)$ observada de uma distribuição de probabilidade F não especificada, o objetivo do método é estimar um parâmetro de interesse $\theta = t(F)$ com base em x , para tanto, calcula-se uma estimativa $\hat{\theta} = s(\mathbf{x})$.

Para cada amostra bootstrap, denotada por $x^* = (x_1^*, \dots, x_n^*)$, obtém-se $\hat{\theta}^* = s(\mathbf{x}^*)$, cuja estimativa será definida como $se_{\hat{F}}(\hat{\theta}^*)$, chamada de *estimativa ideal bootstrap* do erro padrão de $s(\mathbf{x})$.

Supondo S o número de amostras bootstrap utilizadas, temos os seguintes passos realizados pelo algoritmo proposto por Efron (1979):

- Seleciona-se S amostras bootstrap independentes cada uma com n valores obtidos de uma amostragem com reposição de $\mathbf{x}$;
- Obter $\hat{\theta}^*(b) = s(\mathbf{x}_b^*)$, $b = 1, \dots, S$;
- Estimar o erro padrão $se_F(\hat{\theta})$ por meio dos erros padrões amostrais das B replicações como segue:

$$\widehat{se}_S = \left[\frac{1}{S-1} \sum_{b=1}^S \{\hat{\theta}^*(b) - \hat{\theta}^*(.)\}^2 \right] \quad (1)$$

em que:

$$\hat{\theta}^*(.) = S^{-1} \sum_{b=1}^S \hat{\theta}^*(b) \quad (2)$$

2.3.3 Amelia II (EMB)

Uma alternativa que possibilita a inserção de valores não observados é a utilização do pacote **amelia** (HONAKER; KING; BLACKWELL, 2011). Como um pacote com um algoritmo de imputação múltipla de propósito geral, o Amelia II possibilita imputação em dados *cross-sectional* e séries temporais.

De maneira geral, utilizando o algoritmo EMB (expectation-maximization

with bootstrapping), são criados K bancos de dados completos devido a imputação de K valores para cada observação perdida. Todos os bancos de dados criados possuem os mesmos dados observados, diferindo quanto as observações imputadas.

O modelo de imputação utilizando o pacote **amelia** assume que o conjunto de dados completo D (isso inclui observações e falhas) tem distribuição normal multivariada com vetor de médias μ e matriz de covariâncias Σ , ou seja,

$$D \sim N(\mu, \Sigma). \quad (3)$$

O real problema da imputação reside no fato de que não se observa todo D , mas apenas D^{ob} . Assume-se portanto que as falhas são aleatórias do tipo MAR e constituem uma matriz K de falha que depende apenas de D^{ob} . Tem-se que o conjunto D é constituído pelos dados efetivamente observados (D^{ob}) e pelas falhas (K), e a imputação múltipla irá considerar o parâmetros dos dados completos, $\theta = (\mu, \Sigma)$, assim, sob a pressuposição de que as falhas são do tipo MAR e que a matriz K de falhas é independente dos parâmetros μ e Σ dos dados completos, a verossimilhança é desmembrada em:

$$p(D^{ob}, K|\theta) = p(K|D^{ob})p(D^{ob}|\theta). \quad (4)$$

Como se quer a inferência em relação aos dados completos, dado que a matriz de falha K depende apenas de D^{ob} , a verossimilhança, que pode ser escrita como $L(\theta|D^{ob}) \propto p(D^{ob}|\theta)$, pela lei iterated expectations pode ser reescrita como:

$$p(D^{ob}|\theta) = \int p(D|\theta)dK. \quad (5)$$

Com a verossimilhança definida por 4 e uma distribuição a *priori* não informativa em θ , teremos a distribuição a *posteriori* como:

$$L(\theta|D^{ob}) \propto p(D^{ob}|\theta) = \int p(D|\theta)dK. \quad (6)$$

Uma vez que se tem sorteios da posteriori dos dados completos, a imputação é realizada sorteando valores de K da distribuição condicional de D^{ob} e os sorteios de θ .

Para combinar os resultados obtidos dos m bancos de dados, uma determinada quantia q , como a média, coeficiente de regressão ou uma primeira diferença, deve ser en-

contrada. Em seguida cada um dos m bancos de dados é analisado seguindo o modelo apropriado como se não houvessem observações perdidas (HONAKER; KING; BLACKWELL, 2011).

Na prática, pode-se, por exemplo, tanto sortear o valor q dentre todos os valores m simulados para cada observação perdida, como combinar as imputações múltiplas de forma a calcular a média dos m valores, obtendo assim, uma estimativa $\bar{q}$.

2.4 Séries Temporais

Séries temporais, também denominadas séries históricas ou cronológicas, são conjuntos de observações coletadas de forma ordenada ao longo do tempo. Embora as séries temporais sejam comumente avaliadas por meio de uma única variável resposta vinculada ao tempo, também é possível avaliar mais do que uma variável resposta (séries multivariadas) e em função de outras variáveis além do tempo (séries multidimensionais) (MORETTIN; TOLOI, 2006).

As séries temporais propriamente ditas além de possuírem cada observação vinculada a um único ponto num dado intervalo de tempo, são tratadas pela existência de correlação entre as observações próximas, exigindo uma forma determinada de análise. Podem ser discretas ou contínuas, distinguindo-se em relação ao período avaliado, em que, tanto maiores intervalos entre observações quanto menores intervalos podem ser estudado. Como exemplos, dados de séries coletados a cada segundo podem ser considerados contínuos, enquanto dados diários podem ser considerados discretos. A caracterização de uma série temporal quanto à sua continuidade está relacionada aos objetivos de pesquisa. No entanto, é comum tomar uma série discreta a partir de uma série contínua. Esse processo pode ser realizado tomando-se médias ou totais acumulados em períodos maiores, tomados intervalos iguais. Um exemplo exposto em Morettin e Toloi (2006) é o cálculo dos valores horários de registros de marés no porto de Santos, acumulando-se os valores coletados a cada segundo.

As séries temporais podem ser avaliadas de diferentes formas sendo tanto com interesse de análise segundo uma perspectiva de como a trajetória do fenômeno estudado se dá ao longo do tempo como sob a perspectiva da frequência de ocorrência de tais fenômenos. Dessa forma, séries temporais podem ser analisadas por modelos tanto paramétricos, quando desenvolvidas no domínio do tempo, quanto por modelos não-paramétricos, quando desenvolvidas no domínio da frequência. No primeiro caso tem-se a utilização de modelos

Autorregressivos (AR), Médias Móveis (MA), Autorregressivos Integrados e Médias Móveis (ARIMA), Autorregressivos Integrados e Médias Móveis Sazonais (SARIMA), etc. No segundo, a análise espectral, a qual possui muitas aplicações em áreas cujo interesse é avaliar a periodicidade dos fenômenos avaliados, como Meteorologia e Oceanografia (MORETTIN; TOLOI, 2006).

Todos os conceitos, teoremas e definições a cerca da teoria de séries temporais que serão apresentados a seguir, são necessários ao desenvolvimento do presente trabalho e são expostos em Morettin e Toloi (2006).

2.4.1 Processos estocásticos

Séries temporais são trajetórias de um processo físico determinado, regidas por leis probabilísticas. Processos estocásticos, como o conjunto de todas as possíveis trajetórias de um determinado evento, podem ser avaliados segundo alguma forma de modelagem.

Uma série temporal $Z(t)$, é uma trajetória individual de um processo estocástico Z . Para cada observação $Z(t)$ ao longo do tempo T , o seu valor particular (relacionado à unidade da variável observada), tem determinada probabilidade de ocorrência ω , que é regida pela distribuição de probabilidade do fenômeno observado Ω em um tempo específico $t \in T$.

De maneira formal tem-se a seguinte definição:

Definição 2.1 *Dado um conjunto arbitrário T , um processo estocástico é uma família $Z = \{Z(t), t \in T\}$, em que $Z(t)$ é uma variável aleatória para cada $t \in T$.*

Com mais rigor, cada $Z(t)$ é tomado como função de dois argumentos, ou seja, $Z(t, \omega)$, com $t \in T$ e $\omega \in \Omega$. Desse modo, cada $Z(t, \omega)$, é uma variável aleatória com distribuição de probabilidade, para cada $t \in T$ fixado, como apresentado na Figura 1.

Outra visualização é possível ao fixar $\omega \in \Omega$. Observa-se assim, uma trajetória específica $Z(t)$ dado ω , dentre todas as possíveis realizações de determinado fenômeno $Z(t, \omega)$, como mostrado na Figura 2.

2.4.1.1 Processos estocásticos estacionários

Em geral, na tentativa de modelagem e descrição de séries temporais, a suposição de estacionariedade é avaliada para possibilitar o ajuste de modelos paramétricos.

Figura 1 – Processo estocástico interpretado como uma família de variáveis aleatórias (Fonte: MORETTIN; TOLOI, 2006)

Figura 2 – Processo estocástico interpretado como uma família de trajetórias (Fonte: MORETTIN; TOLOI, 2006)

O conceito de estacionariedade, portanto, é formalizado para o processo estocástico, dado que $Z(t)$ é uma trajetória particular de $Z(t, \omega)$.

Intuitivamente um processo estocástico $Z(t, \omega)$ é estacionário quando sua trajetória se desenvolve de forma aproximadamente constante em torno de uma média (MORETTIN; TOLOI, 2006). Assim, pode-se dizer que, independente do ponto de partida em uma determinada série, suas características não mudam caso a mesma seja estacionária, ou seja, permanecem as mesmas sob translações no tempo.

O processo estocástico $Z = \{Z(t), t \in T\}$, estará especificado caso as distribuições finito dimensionais $F(z_1, \dots, z_n; t_1, \dots, t_n) = P\{Z(t_1) \leq z_1, \dots, Z(t_n) \leq z_n\}$, com $t_1, t_2, \dots, t_n$, para $n > 1$, sejam conhecidas.

Especificado o processo estocástico, pode-se definir processos estritamente

estacionários pela seguinte definição:

Definição 2.2 *Se todas as distribuições finito dimensionais de um processo estocástico $Z = \{Z(t), t \in T\}$, permanecem as mesmas sob translações τ no tempo, ou seja:*

$$F(z_1, \dots, z_n; t_1 + \tau, \dots, t_n + \tau) = F(z_1, \dots, z_n; t_1, \dots, t_n)$$

o processo diz-se estritamente estacionário para quaisquer $t_1, t_2, \dots, t_n, \tau$ de T .

A definição de estacionariedade de segunda ordem, ou fraca, indica que todas as distribuições unidimensionais são invariantes sob translações no tempo, consequentemente, $\mu(t) = \mu$ e $V(t) = \sigma^2$ são invariantes para todo $t \in T$. Tem-se ainda que todas as distribuições bidimensionais dependem de $t_2 - t_1$. Como $\gamma(t_1, t_2) = \gamma(t_1 + t, t_2 + t)$, qualquer que seja t , para $t = -t_2$ ou $t = -t_1$ tem-se $\gamma(t_1, t_2) = \gamma(\tau)$, com $\tau = |t_1 - t_2|$, ou seja, função de apenas um argumento.

Na realidade, a verificação de estacionariedade estrita não é realizada, uma vez que a intensão é avaliar um processo estocástico por meio de um número pequeno de distribuições finito dimensionais ou momentos, portanto, tem-se um processo fracamente estacionário definido como segue:

Definição 2.3 *Um processo estocástico $Z = \{Z(t), t \in T\}$ é fracamente estacionário se e somente se:*

- $E\{Z(t)\} = \mu(t) = \mu$, constante, para todo $t \in T$;
- $E\{Z^2(t)\} < \infty$, para todo $t \in T$;
- $\gamma(t_1, t_2) = \text{Cov}\{Z(t_1), Z(t_2)\}$ é uma função de $|t_1 - t_2|$

Assim, todo o processo definido simplesmente como *processo estacionário*, corresponde aos processos fracamente estacionários.

2.4.2 Função de autocovariância e autocorrelação

Uma vez que, dado um processo estocástico qualquer, pode-se avaliar a dependência entre as observações próximas por meio da verificação da variação apresentada entre elas ao longo da trajetória. Para a verificação da relação entre as observações, tem-se

o cálculo da função de autocovariância, e por meio desta, o cálculo da função de autocorrelação, utilizada para verificação e seleção de modelos paramétricos.

A função de autocovariância (facv) de um processo estacionário discreto $\{Z_t, t \in Z\}$ de média zero é definida como $\gamma_\tau = E\{Z_t, Z_{t+\tau}\}$ com $\tau = |t_1 - t_2|$. O cálculo da estimativa c_j da função de autocovariância γ_τ é dada pela seguinte função:

$$c_j = \frac{1}{N} \sum_{t=1}^{N-j} [(Z_t - \bar{Z})(Z_{t+j} - \bar{Z})], j = 0, 1, \dots, N-1$$

em que, $\bar{Z} = \frac{1}{N} \sum_{t=1}^N Z_t$ é a média amostral.

A estimativa ρ_j da função de autocorrelação (fac), calculada por meio da função de autocovariância, será então dada por:

$$\rho_j = \frac{c_j}{c_0}, j = 0, 1, \dots, N-1.$$

2.4.3 Decomposição clássica de uma série temporal

Um modelo de decomposição para séries temporais supõe que uma determinada série $\{Z_t, t = 1, 2, \dots, N\}$, é composta por três componentes distintos não observáveis como segue:

$$Z_t = T_t + S_t + a_t, \text{ com } t = 1, 2, \dots, N,$$

em que, T_t é a componente tendência, S_t corresponde à componente sazonalidade e o termo a_t é definido como ruído branco, isto é, erros aleatórios, independentes e identicamente distribuídos com a média seja nula ($\mu = 0$) e variância constante σ^2 . No entanto, pode-se eventualmente tomar a_t apenas como um processo estacionário, indicando, geralmente, a não estacionariedade de Z_t (MORETTIN; TOLOI, 2006).

Em casos cujas amplitudes sazonais variam conforme a tendência, o modelo mais adequado é o multiplicativo, representado da seguinte forma:

$$Z_t = T_t S_t a_t \text{ em que } t = 1, 2, \dots, N.$$

Nesses casos, efetua-se a aplicação do logaritmo para transformação da série.

O resultado é um modelo aditivo (COSTA; SÁFADI, 2010):

$$\begin{aligned} \log(Z_t) &= \log(T_t S_t a_t) \\ \log(Z_t) &= \log(T_t) + \log(S_t) + \log(a_t), t = 1, 2, \dots, N \end{aligned}$$

2.4.4 Estacionariedade

Como citado anteriormente, a estacionariedade de uma série temporal está relacionada ao seu desenvolvimento ao longo do tempo em torno de uma média (MORETTIN; TOLOI, 2006). A não estacionariedade em uma série temporal, cuja ocorrência é bastante comum, decorre da presença de fatores inerentes ao fenômeno estudado e são descritos nos modelos teóricos como componentes sazonais, tendência e casos que apresentam heterogeneidade de variâncias.

O procedimento mais simples para tratar a tendência consiste na diferenciação, onde a diferença entre uma determinada observação e sua anterior é tomada sequencialmente de forma a construir um novo conjunto de dados de tamanho $N - 1$. Tal manipulação pode ser realizada de forma sucessiva até que a série se torne estacionária, dado que em geral, uma ou no máximo duas diferenças são suficientes para essa finalidade. Do mesmo modo a sazonalidade pode ser retirada tomando-se a diferença na defasagem sazonal, ou seja, na amplitude em que a série apresentar periodicidade.

De forma geral, a n -ésima diferença de $Z(t)$ é definida como:

$$\Delta^n Z(t) = \Delta[\Delta^{n-1} Z(t)] \quad (7)$$

Quanto ao que se refere à heterogeneidade, o procedimento recomendado é aplicação de uma transformação não linear qualquer com a intenção de estabilizar essa variabilidade. A transformação logarítmica é comumente utilizada, principalmente em dados econômicos, e deve ser aplicada à série original.

Segundo Lütkepohl e Xu (2012), a transformação logarítmica é empregada para a obtenção tanto da homogeneidade de variâncias quanto de uma distribuição dos dados mais próxima da distribuição normal. Tal transformação proporciona previsões mais acuradas dos modelos ARIMA quando comparadas às previsões realizadas com dados não transformados.

A verificação da existência de heterocedasticidade é realizada de forma visual, com a representação gráfica da média versus amplitude ou desvio padrão versus amplitude do conjunto de dados. A série original é comumente dividida em subconjuntos para o cálculo de tais estatísticas para a construção do gráfico (MORETTIN; TOLOI, 2006).

Os testes para a verificação de tendência e sazonalidade devem ser realizados para avaliar a necessidade de diferenciação das séries temporais. Os testes apresentados a seguir são alguns dentre os vários teste possíveis para cada característica que deve ser avaliada.

2.4.4.1 Teste de Cox-Stuart

O teste de Cox-Stuart é um dos testes recomendados para a verificação da existência de tendência monótonas, isto é, crescentes ou decrescentes, em séries temporais Morettin e Toloi (2006). O teste se baseia nas seguintes hipóteses:

$$H_0 : P(Z_i < Z_{i+c}) = P(Z_i > Z_{i+c}), \forall_i : \text{não existe tendência};$$

$$H_1 : P(Z_i < Z_{i+c}) \neq P(Z_i > Z_{i+c}), \forall_i : \text{existe tendência};$$

A aplicação do teste consiste inicialmente em agrupar os dados em pares $(Z_1, Z_{1+c}), (Z_2, Z_{2+c}), \dots, (Z_{N-c}, Z_N)$, tal que, N é o número de observações da série e $c = \frac{N}{2}$ para N par, ou $c = \frac{N+1}{2}$ caso N seja ímpar. Calcula-se as diferenças $(Z_{1+c} - Z_1), (Z_{2+c} - Z_2), \dots, (Z_{N-c} - Z_N)$ associando os sinais $-$ para valores negativos e $+$ para valores positivos das diferenças. Para que haja tendência positiva, espera-se um número maior de valores positivos T_2 . Caso $T_2 > n - t$, rejeita-se H_0 , indicando a presença de tendência. Para $n \leq 20$, o valor t é obtido da distribuição binomial com parâmetro $p = \frac{1}{2}$. Para $n > 20$, utiliza-se a aproximação normal com média np e variância npq .

O teste de Cox-Stuart é uma alternativa para a verificação de tendência de fácil interpretação. É um teste bastante utilizado na literatura, em diferentes áreas aplicadas, como melhoramento de plantas (NASCIMENTO et al., 2008), ambiental (COSTA; SÁFADI, 2010) e produção animal (NOGUEIRA; SÁFADI, 2011).

2.4.4.2 Teste de Fisher

O teste de Fisher é uma das alternativas para a avaliação da existência de periodicidade, também denominada sazonalidade, em dados de séries temporais. Como o teste consiste na avaliação do período por meio da verificação de periodogramas, faz-se necessária a apresentação de alguns conceitos referentes à análise espectral. Como exposto em Morettin e Toloi (2006), o objetivo geral da análise espectral reside na necessidade de se estudar a periodicidade dos dados.

Considerando um processo estocástico $\{Z(t), t \in T\}$ estacionário com média zero e autocovariância decrescente à medida em que os valores se afastam entre si, ou seja, condição de independência assintótica expressa na forma $\sum_{\tau=-\infty}^{\infty} |\gamma(\tau)| < \infty$, tem-se a função de densidade espectral $f(\lambda)$, também definida como transformada de Fourier de $\gamma(t)$, ou seja, da função de autocovariância no ponto t , da seguinte forma:

$$f(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma(\tau) \exp -i\lambda\tau, \text{ em que: } -\infty \leq \lambda \leq \infty,$$

com $\exp i\lambda = \cos \lambda + i \sin \lambda$, $i = \sqrt{-1}$ e λ o valor de uma frequência que será utilizada para a verificação de um período a ser avaliado.

A função densidade espectral $f(\lambda)$ é periódica com período 2π , e par, sendo conveniente representá-la no intervalo $[0, \pi]$.

Como o interesse é verificar períodos em séries temporais discretas e finitas, define-se a transformada de Fourier finita de um processo estacionário com média zero $\{Z_t, t = 1, \dots, N\}$ como:

$$d^{(N)}(\lambda) = \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^N Z_t \exp -i\lambda t, \text{ em que: } -\infty < \lambda < \infty.$$

Na prática utiliza-se a transformada de Fourier discreta, tomando-se o cálculo para frequências com $\lambda_j = \frac{2\pi j}{N}$ e $-\lceil \frac{N-1}{2} \rceil \leq j \leq \lfloor \frac{N}{2} \rfloor$ com j o índice de cada observação de uma trajetória Z_t qualquer. A transformada de Fourier discreta é apresentada como segue:

$$\begin{aligned} d_j^{(N)} &= \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^N Z_t \exp \left(\frac{-i2\pi jt}{N} \right) = \\ &= \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^N Z_t \cos \left(\frac{2\pi jt}{N} \right) + i \frac{2\pi jt}{N} \sum_{t=1}^N Z_t \sin \left(\frac{2\pi jt}{N} \right), \end{aligned}$$

com $j = 0, 1, \dots, \left\lfloor \frac{N}{2} \right\rfloor$

Dada a transformada de Fourier discreta, a intensão é encontrar um estimador para $f(\lambda_j)$, para uma dada realização $\{Z_t, t = 1, \dots, N\}$ do processo estacionário que será denominado *periodograma*:

$$I^{(N)}(\lambda) = |d_j^{(N)}|^2 = \frac{1}{\sqrt{2\pi N}} \left| \sum_{t=1}^N Z_t \exp(-i\lambda_j t) \right|^2 \quad (8)$$

Como descrito por Morettin e Toloi (2006), temos o seguinte teorema:

Teorema 2.1 *As ordenadas do periodograma $I_j^{(N)}$ são variáveis aleatórias assintoticamente independentes com distribuição assintótica múltipla de uma variável aleatória qui-quadrado, isto é:*

$$I_j^{(N)} \xrightarrow{D} \begin{cases} \frac{1}{2}f(\lambda_i)\chi_2^2, j \neq 0, N/2, \\ f(\lambda_i)\chi_1^2, j = 0, N/2. \end{cases}$$

Dado o periodograma, como descrito anteriormente, utiliza-se o teste de Fisher para verificação de periodicidades na série temporal de interesse, como apresentado por Morettin e Toloi (2006). A hipótese a ser testada é:

H_0 : Não há periodicidade, para todo $I_j^{(N)}$

H_1 : Há periodicidade para algum $I_j^{(N)}$

A estatística do teste é escrita como

$$g = \frac{\max_j I_j^{(N)}}{\sum_{j=1}^{\frac{N}{2}} I_j^{(N)}}.$$

em que $I_j^{(N)}$ é o valor do periodograma na ordenada j e N é o número de observações da série temporal.

A aproximação apresentada por Morettin e Toloi (2006) é dada por:

$$P(g > \alpha) \equiv n(1 - \alpha)^{n-1}.$$

Caso o p-valor seja menor que o nível de significância proposto, rejeita-se H_0 para o nível de significância estabelecido, indicando que há sazonalidade na série temporal,

sendo o período correspondente à ordenada j em questão.

2.4.5 Modelos

Uma das formas de modelagem de dados de séries temporais é com a utilização de modelos paramétricos. Box e Pierce (1970), propuseram a metodologia conhecida atualmente como modelagem ARIMA que consiste na construção de modelos de ciclo iterativos em que a escolha do modelo se baseia no próprio conjunto de dados.

Para a caracterização de tais modelos modelos é comum a formalização do conceito de *operadores*. Um operador segundo Morettin e Toloi (2006), nada mais é que uma notação empregada para facilitar a manipulação dos modelos que são estudados. Para o modelos autorregressivo temos, portanto, a notação *operador translação para o passado*, denotado por B e definido como segue:

$$BZ_t = Z_{t-1}, \quad (9)$$

$$B^m Z_t = Z_{t-m}. \quad (10)$$

Uma determinada série temporal Z_t é o resultado da passagem de um ruído a_t , por um sistema linear $\psi(B)$ como se segue (MORETTIN; TOLOI, 2006):

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \mu + \psi(B)a_t, \quad (11)$$

em que $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$, determina a função de transferência e μ é o nível ou média da série temporal avaliada.

Dada uma determinada série $\tilde{Z}_t = Z_t - \mu$, tem-se de (11), $\tilde{Z}_t = \psi(B)a_t$ para a exemplificação dos modelos que serão expostos a seguir.

A ideia de um filtro linear corresponde às maneiras como tais processos serão descritos pelo modelo. Pela inclusão de termos autorregressivos e termos de médias móveis, os modelos de séries temporais são capazes de descrever o fenômeno em estudo e efetuar previsões. Para tanto se faz necessário compreender os termos de um modelo ARIMA (p,d,q).

2.4.5.1 Modelos autorregressivos (AR)

Dada a série temporal $\tilde{Z}_t$, podemos definir um processo por meio de um modelo autorregressivo de ordem p denotado por $AR(p)$, como consta em Morettin e Toloi (2006), da seguinte forma:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t. \quad (12)$$

Aqui, a função que determina o filtro linear $\psi(B)$, é descrita para os modelos autorregressivos com a inclusão de parâmetros ϕ .

Isolando a_t de (12), tem-se:

$$a_t = \tilde{Z}_t - \phi_1 \tilde{Z}_{t-1} - \phi_2 \tilde{Z}_{t-2} - \dots - \phi_p \tilde{Z}_{t-p}. \quad (13)$$

Definindo um operador para o modelos autorregressivo de ordem p a partir de (10), tem-se:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p. \quad (14)$$

Por meio do operador descrito em (14), pode-se reescrever a_t dado em (13) como:

$$a_t = \phi(B) \tilde{Z}_t. \quad (15)$$

Dessa forma, como descrito anteriormente, pode-se observar a função de transferência descrita por um operador definido para um modelo autorregressivo.

2.4.5.2 Modelos de médias móveis (MA)

Da mesma forma que para um modelo autorregressivo, um determinado processo pode ser descrito por um modelo de médias móveis de ordem q , de forma que a função de transferência seja descrita pela inclusão de parâmetros θ , denotado como:

$$Z_t = \mu + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}. \quad (16)$$

Pode-se, do mesmo modo como para um modelo autorregressivo segundo Morettin e Toloi (2006), definir um operador de médias móveis de ordem q , descrito por:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q. \quad (17)$$

Dado $\tilde{Z}_t = Z_t - \mu$, pode-se reescrever (16) com a utilização do operador definido em (17), resultando na função de transferência descrita por um operador definido para um modelo de médias móveis

$$\tilde{Z} = \theta(B)a_t. \quad (18)$$

2.4.5.3 Modelos integrados autorregressivos de médias móveis (ARIMA)

De forma geral, pode-se dizer que um modelos ARIMA é uma classe de modelos mais geral, dado que são integrados a partir de um modelo autorregressivo de médias móveis (ARMA). Um modelo ARMA (p,q), portanto, pode ser descrito como:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \dots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}, \quad (19)$$

que com a utilização dos operadores $\phi(B)$ e $\theta(B)$, pode ser reescrito como:

$$\phi(B)\tilde{Z}_t = \theta(B)a_t. \quad (20)$$

A utilização desses modelos pressupõe que a série temporal seja estacionária. Como, em geral, o procedimento mais simples é a tomada da diferença da série para torná-la estacionária, um termo que calcula a diferença da série foi inserido no modelo ARMA, resultando em uma nova classe, os modelos ARIMA ou *integrados*. Pode-se, portanto, definir que para uma série Z_t , não estacionária, uma nova série W_t pode ser calculada tomando se a primeira diferença, resultando em uma série estacionária como segue:

$$W_t = Z_t - Z_{t-1} = (1 - B)Z_t = \Delta Z_t. \quad (21)$$

Caso $W_t = \Delta^d Z_t$ seja estacionária, tomando-se o número de diferenças neces-

sárias d , a representação de um modelo ARMA (p,q) será:

$$\phi(B)W_t = \theta(B)a_t. \quad (22)$$

De uma forma simples, utilizando a mesma estratégia de modelagem ARIMA, caso uma determinada série temporal Z_t seja estacionária, como não é necessária a tomada de diferença, teríamos um ARIMA (p,0,q), que é o mesmo que um ARMA (p,q).

Dessa forma, um modelo ARIMA pode ser representado a partir de um modelo ARMA com a inclusão de um operador de diferença, dado que W_t é resultado da tomada da diferença em uma série não estacionária Z_t . Assim, um ARIMA (p,d,q), pode ser descrito como segue:

$$\phi(B)\Delta^d Z_t = \theta(B)a_t. \quad (23)$$

Segundo Morettin e Toloi (2006), o interesse na utilização dos modelos ARIMA em algumas áreas na prática, reside tanto na realização de previsões quanto na estimação do tempo de recorrência de determinados fenômenos com a utilização dos estimadores autorregressivos. Na prática, utiliza-se tais modelos quando é necessário o ajustes com um número menor de parâmetros.

2.4.5.4 Função de autocorrelação (fac)

Dado o processo descrito por um modelo autorregressivo de ordem p , a função de autocovariância pode ser obtida com a multiplicação da equação obtida para o modelo AR descrita em (12), por $\tilde{Z}_{t-j}$, com j sendo a defasagem temporal, e tomando a esperança como segue:

$$E(\tilde{Z}_t \tilde{Z}_{t-j}) = \phi_1 E(\tilde{Z}_{t-1} \tilde{Z}_{t-j}) + \dots + \phi_p E(\tilde{Z}_{t-p} \tilde{Z}_{t-j}) + E(a_t \tilde{Z}_{t-j}). \quad (24)$$

Como $\tilde{Z}_t$ pressupõe ruído branco até a_{t-j} , não correlacionados com a_t , e ainda, $E(a_t \tilde{Z}_{t-j}) > 0, j > 0$ com $E(\tilde{Z}_t \tilde{Z}_{t-j}) = \gamma_j$, a expressão da autocovariância descrita para o modelo AR (p) é dada segundo a expressão:

$$\gamma_j = \phi_1 \gamma_{j-1} + \dots + \phi_p \gamma_{j-p}. \quad (25)$$

Efetuada a divisão em ambos os lados da expressão (25) por $\gamma_0 = \text{Var}(Z_t)$, é obtida a função de autocorrelação:

$$\rho_j = \phi_1 \rho_{j-1} + \dots + \phi_p \rho_{j-p}. \quad (26)$$

Considerando $j = 1, 2, \dots, p$, dado que $\rho_0 = 1$ e $\rho_{-\tau} = \rho_\tau$, obtém-se as chamadas equações de *Yule – Walker*:

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

As equações de *Yule – Walker*, que também podem ser expressas na forma matricial, como:

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} \quad (27)$$

possibilitam a estimação dos coeficientes $\phi_1, \phi_2, \dots, \phi_p$ do modelo $AR(p)$ com a utilização das estimativas das fac ρ_p , calculadas como exposto na seção que descreve as funções de autocovariância e autocorrelação.

A função de autocorrelação auxilia no processo de identificação da ordem dos modelos que serão escolhidos. A ordem de um modelo pode ser verificada pelo número de defasagens fora do intervalo apresentado nas representações gráficas.

2.4.5.5 Função de autocorrelação parcial (facp)

Como uma proposta adicional para auxiliar a identificação de modelos de séries temporais, Box, Jenkins e Reinsel (1994), propuseram a utilização da *função de autocorrelação parcial* (facp), como descrita em Morettin e Toloi (2006). Considerando ϕ_{kj} e ϕ_{kk} o j -ésimo coeficiente e o último coeficiente do modelo $AR(k)$, respectivamente, pode-se

obter a partir de (26) a seguinte expressão:

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \dots + \phi_{kk}\gamma_{j-k}, j = 1, \dots, k \quad (28)$$

desenvolvendo-se as equações de *Yule – Walker* na forma matricial como segue:

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad (29)$$

A resolução das equações para $k = 1, 2, 3, \dots$ resulta em:

$$\begin{aligned} \phi_{11} &= \rho_1 \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} \\ \phi_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \end{aligned}$$

cuja forma geral é dada por:

$$\phi_{kk} = \frac{|P_k^*|}{|P_k|}, \quad (30)$$

em que, ϕ_{kk} , é a função de autocorrelação parcial, P_k^* é a matriz de autocorrelações P_k com a última coluna substituída pelo vetor de autocorrelações.

Para um modelo autorregressivo de ordem p [AR(p)], tem-se a função de autocorrelação parcial (facp) como $\phi_{kk} \neq 0$, para $k \leq p$ e $\phi_{kk} = 0$, para $k > p$. Tem-se ainda que, as facp estimadas de ordem $p+1, p+2, \dots$, são, aproximadamente, independentemente

distribuídas com $Var(\hat{\phi}_{jj}) \simeq \frac{1}{N}, j > p$.

A partir de $\rho(j)$, calcula-se a função de autocorrelação parcial estimada, $\hat{\Phi}_{kk}$, com a qual identifica-se a ordem p do modelo autorregressivo ajustado.

Como exposto em Morettin e Toloi (2006), sob a hipótese de que o processo é AR(p) com N suficientemente grande, $\hat{\phi}_{kk}$ tem distribuição aproximadamente normal.

2.5 Testes para comparação de séries temporais

Em áreas práticas do conhecimento a utilização da teoria de séries temporais nem sempre é suficiente para determinados estudos. A comparação dessas séries é muitas vezes o maior interesse de estudos de monitoramento de séries coletadas em locais próximos. A verificação da igualdade entre duas séries temporais pode acarretar numa redução de custos operacionais na coleta de dados, e para tanto, existem maneiras de comparar se duas séries diferentes, foram geradas pelo mesmo processo estocástico.

Para tanto Quenouille (1958) e Coates e Diggle (1986), propuseram testes univariados para a comparação da função de autocorrelação e função de densidade espectral, respectivamente.

Echeverry e Toloi (2000), aplicaram este teste em séries de temperatura e salinidade provenientes de estações de monitoramento instaladas na Colômbia.

Um segundo teste para comparação de séries temporais, proposto por Silva, Ferreira e Sáfadi (2000), utilizado por Costa e Sáfadi (2010) para a comparação de séries de material particulado, também será descrito.

2.5.1 Teste de Igualdade das Funções de Autocorrelação

O teste proposto por Quenouille (1958), tem a finalidade de verificar se duas séries temporais distintas tem a mesma estrutura de correlação. Para tanto, considerando o ajuste de um modelo autorregressivo para duas séries $Z_1(t)$ e $Z_2(t)$, e suas respectivas funções de autocorrelação, $\rho_1(j)$ e $\rho_2(j)$ as hipóteses testadas são:

$$H_0 : \rho_1(j) = \rho_2(j), \text{ para todo } j = 1, 2, \dots, J;$$

$$H_1 : \rho_1(j) \neq \rho_2(j), \text{ para algum } j = 1, 2, \dots, J.$$

O procedimento para a aplicação do teste de Quenouille, como descrito se-

gundo Echeverry e Toloi (2000), consiste em inicialmente calcular as funções de autocorrelação $\hat{\rho}_1(j)$ e $\hat{\rho}_2(j)$ das séries $Z_1(t)$ e $Z_2(t)$, com $j = 1, 2, \dots, J$.

Em seguida, obtém-se a função de autocorrelação comum de $\hat{\rho}_1(j)$ e $\hat{\rho}_2(j)$ por meio de uma média ponderada de forma $\rho(j) = \frac{n_1\hat{\rho}_1(j) + n_2\hat{\rho}_2(j)}{n_1 + n_2}$, onde n_1 e n_2 são os tamanhos das séries $Z_1(t)$ e $Z_2(t)$.

A partir de $\rho(j)$, calcula-se a função de autocorrelação parcial estimada comum entre as duas séries, $\hat{\Phi}_{kk}$, com a qual identifica-se a ordem p do modelo autorregressivo ajustado.

Resolvendo as equações de *Yule – Walker*, estima-se os p coeficientes do modelo $AR(p)$, ajustando para $Z_1(t)$ e $Z_2(t)$, o modelo $AR(p)$ correspondente aos coeficientes comuns encontrados, obtendo-se assim, as chamadas séries residuais $\hat{a}_1$ e $\hat{a}_2$.

Por fim, calcula-se as facp v_j e v'_j para as séries residuais $\hat{a}_1$ e $\hat{a}_2$, respectivamente, para a utilização da estatística teste dada por:

$$SQ = \sum_{j=1}^J \frac{(v_j - v'_j)^2}{\frac{1}{n_1-j} + \frac{1}{n_2-j}} \sim \chi_J^2 \quad (31)$$

A hipótese H_0 é rejeitada, a um determinado nível de significância α , caso $SQ > C_\alpha$, em que C_α é tal que $P(\chi_J^2 > C_\alpha) = \alpha$, com j igual ao número de graus de liberdade da distribuição qui-quadrado.

2.5.2 Método de comparação de Séries Temporais proposto por Silva, Ferreira e Sáfiadi (2000)

O procedimento para comparação de duas séries temporais $Z_1(t)$ e $Z_2(t)$, como proposto por Silva, Ferreira e Sáfiadi (2000), foi descrito por Costa e Sáfiadi (2010).

Inicialmente calcula-se a diferença entre as duas séries dada por $Z_d = Z_1(t) - Z_2(t)$, para $t = 1, 2, \dots$. Aplica-se em seguida a Z_d o teste de Cox-Stuart para verificar a existência de tendência, o teste de Fisher para verificar se existe sazonalidade e para verificar se os resíduos são independentes e identicamente distribuídos, com média zero e variância constante, aplica-se o teste de Box e Pierce.

Caso Z_d seja estacionária, ou seja, não apresente tendência nem sazonalidade, e os resíduos sejam ruído branco, conclui-se que as duas séries são resultado do mesmo

processo estocástico, ou seja, são iguais no período analisado.

2.5.2.1 Teste de Box-Pierce

Dadas as autocorrelações estimadas, Box e Pierce (1970) sugeriram um teste para a verificação das autocorrelações dos resíduos segundo a suposição de ruído branco. Segundo o teste proposto, caso o modelo escolhido seja adequado, não haverá autocorrelação nos resíduos. Dessa forma, tem-se um teste para as seguintes hipóteses:

H_0 : Os resíduos do modelo ajustado são independentes

H_1 : Não existe independência entre os resíduos do modelo ajustado.

A estatística do teste de Box-Pierce é dada por:

$$Q(k) = N(N+2) \sum_{j=1}^k \frac{\hat{p}_j^2}{(N-j)} \quad (32)$$

Caso $Q > \chi_{(k-p-q)}^2$, rejeita-se a hipótese nula. Os valores k, p e q são respectivamente, número de defasagem, ordem do termo autorregressivo, e ordem do termo de médias móveis

3 METODOLOGIA

3.1 Material

O presente trabalho foi desenvolvido com a utilização de dados provenientes de estudos hidrológicos realizados em uma fazenda de plantio florestal pertencente à Klabin Florestal Paraná.

A Figura 3 mostra a disposição das duas microbacias dentro da fazenda Monte Alegre.

Figura 3 – Mapa da fazenda Monte Alegre com especificação das microbacias Mortandade e Colônia (extraído de Voigtlander, 2007)

Com uma área total de 220 mil hectares, a Klabin Florestal Paraná possui 120 mil hectares de floresta plantada e 85 mil hectares de Área de Preservação Permanente (APP) e Reserva Legal (RL). A principal espécie plantada é o *Pinus taeda* ocupando 66 mil hectares (54,9%). Há ainda plantios de outras espécies como *Eucalyptus spp.* com 34 mil hectares (29,1%), *Pinus elliottii* com 10 mil hectares (8,7%) e 9 mil hectares restantes (7,3%) ocupados com outras espécies (BLOOD, 2005).

O estudo, como descrito por Voigtlaender (2007), foi realizado em duas microbacias, ambas com instrumentos instalados em março de 2004, presentes na Fazenda Monte Alegre, cada qual com características distintas de cobertura vegetal, uma com fragmento de floresta nativa e a outra com plantio florestal.

A Fazenda, pertencente à Klabin Florestal Paraná, está situada no Município de Telêmaco Borba, entre as latitudes 24°02'02" e 24°27'48" e as longitudes 50°17' e 50°55',

região centro-leste do Estado do Paraná. A 250 km de distância de Curitiba, em uma altitude média de 885 metros, a área pertence à região de abrangência de um dos afluentes do Rio Paranapanema.

O clima, subtropical transicional para o temperado, úmido mesotérmico, sem estação seca definida (Köppen é do tipo Cfa/Cfb), possui verões quentes com tendência a concentração de chuvas e há pequena possibilidade de ocorrência de geadas no inverno. Com médias anuais em torno de $19,3^{\circ}\text{C}$, apresenta temperaturas por volta de $22,2^{\circ}\text{C}$ em janeiro a fevereiro e temperatura de $15,6^{\circ}\text{C}$ em junho e julho. A precipitação média anual de 1500 mm, apresenta em torno de 767 mm excedente hídrico (BLOOD, 2005).

As microbacias avaliadas foram descritas por Voigtlaender (2007). A primeira, microbacia Colônia (Pinus), com uma área de 135,81 ha, foi reflorestada com 70% de *Pinus taeda*, 8% *Eucalyptus grandis* e 22% ocupados com Área de Preservação Permanente. A segunda, microbacia Mortandade (Mata Atlântica), possui uma área de 36,71 ha, com vegetação caracterizada pela transição entre Floresta Ombrófila Mista e Floresta Estacional Semidecidual, pertencente ao bioma Mata Atlântica.

Os conjuntos de dados correspondem a séries temporais de vazão específica definida como o escoamento de lâmina d'água através da secção de um vertedouro triangular corrigida para a área da bacia. Esses dados foram coletados nas microbacias Mortandade e Colônia no período de 13/07/2005 a 28/02/2013, em intervalos de 15 minutos entre cada observação. As médias diárias foram então calculadas, totalizando 2788 observações.

Para a microbacia Mortandade, o cálculo da vazão para um vertedouro de-sassoreado foi realizado por meio de uma conversão com a utilização de uma fórmula para vertedouro composto apresentada na Tabela 2. Para a microbacia Colônia, situada em plantio de Pinus, foi efetuado o cálculo para um *vertedouro de crista curta*, cuja equação é apresentada na Tabela 1.

Tabela 1 – Equações para o cálculo da vazão (m^3s^{-1}) para a cota H (m) da microbacia Colônia

Equação para cálculo da vazão	H
$Q = 0,001(3547,1H^3 + 605,16H^2 - 7,1822H + 0,0366)$	$0,030 \leq H \leq 0,317$
$Q = 0,001(4078,2H^3 - 2623,6H^2 + 2279H - 417,1)$	$> 0,317$

Extraído de Voigtlaender (2007)

As medições foram realizadas a partir da leitura de cota (altura da lâmina d'água) das estações linimétricas localizadas sobre poço tranquilizador acoplado ao tanque principal, cuja função principal é diminuir a agitação da água para que a medição do

Tabela 2 – Equações para o cálculo da vazão ($m^3 s^{-1}$) para a cota H (m) da microbacia Mortandade

Equação para cálculo da vazão	H
$Qd = 0,001(6173,7H^3 + 1574,8H^2 - 49,36H + 0,1361)$	$0,050 < H < 0,240$
$Qd = 0,001(-974,37H^3 + 348,7H^2 + 274,88H - 90,626)$	$0,241 \leq H \leq 0,461$
$Qd = 0,001(-2571,4H^3 + 10480H^2 - 4345,8H + 705,42)$	$0,461 \leq H \leq 1,15$

Extraído de Voigtlaender (2007)

aparelho seja possível. Foram também registradas informações de precipitação, com o auxílio de um pluviógrafo, de modo que os registros foram armazenados em intervalos de 10 minutos por um sistema digital (*datalogger*).

Para a finalidade desse trabalho, os dados utilizados foram transformados em médias diárias, resultando em séries diárias de vazão. As séries temporais de ambos os monitoramentos apresentam falhas em períodos distintos, sejam ele decorrentes de problemas com equipamento ou erros de leitura.

A microbacia Colônia (Pinus) (Figura 3), apresenta falhas nos seguintes períodos: de 13 de fevereiro a 02 de abril de 2007, 25 a 30 de setembro de 2010; de 10 de janeiro a 01 de maio de 2011; de 28 de maio a 08 de junho de 2011; de 23 de julho a 02 de agosto de 2011; de 23 de junho a 04 de julho de 2012; de 30 de outubro a 03 de dezembro de 2012; de 30 de janeiro a 20 de fevereiro de 2013;

A microbacia Mortandade (Mata Atlântica) (Figura 3), apresenta as seguintes falhas: de 13 a 24 de julho de 2005; de 13 de fevereiro a 02 de abril de 2007; de 30 de setembro a 08 de novembro de 2007 e de 23 de novembro a 12 de dezembro de 2007; de 14 de abril e 01 de maio de 2011; de 28 de maio e 08 de junho de 2011; de 23 de julho a 02 de agosto de 2011; de 30 de outubro a 08 de novembro de 2012; 01 a 31 de janeiro de 2013; de 26 a 28 de fevereiro de 2013.

3.2 Método

A metodologia estatística aplicada para comparar as séries provenientes das duas microbacias foi realizada seguindo a ordem de análise apresentada para avaliação de séries temporais.

Inicialmente foi feita a análise descritiva para a caracterização inicial das duas séries temporais obtidas. Foram calculadas medidas de dispersão e posição, verificando-se a existência de falhas, respectivas posições em relação ao conjunto de dados e número de observações perdidas consecutivamente.

Uma vez verificada a presença de falhas, as séries temporais foram submetidas ao mesmo processo de imputação, por meio do algoritmo Amelia II (EMB), com a utilização do pacote `amelia` utilizado no *software* R, de duas maneiras diferentes.

O processo de imputação direta foi realizado para as duas séries completas, sem fracionamento ou separação em subconjuntos. Foi efetuado o truncamento da distribuição dos dados observados para que no processo de reamostragem não houvessem valores negativos. Os resultados da análise descritiva inicial foram utilizados nessa etapa de forma a fornecer informações sobre o máximo valor que poderia ser obtido na reamostragem para cada uma das duas série temporais. A fim de obter uma imputação mais próxima dos dados reais, foi utilizado ainda o ajuste de polinômios, que foram posteriormente avaliados de forma gráfica por meio da representação dos dados imputados e dos dados observados.

O processo de imputação de forma fracionada foi realizado inicialmente com a avaliação da posição das falhas. Posteriormente as duas séries temporais foram separadas e subconjuntos de forma a compreender cada uma das lacunas separadamente incorporando 200 as observações anteriores à falha e 200 as observações posteriores à falha. Para cada um dos subconjuntos foi efetuado o processo de imputação com o algoritmo AMELIA II e as novas imputações foram então utilizadas para compor as séries completas. Da mesma forma que no processo anterior, a distribuição dos dados foi truncada em zero e no máximo apresentado pelas duas séries e o polinômio de grau três também foi utilizado.

Após o processo de imputação as séries geradas pelos dois processos de imputação foram avaliadas da mesma maneira segundo a teoria de séries temporais com a posterior comparação como descrito a seguir:

- i. a heterogeneidade de variâncias foi corrigida fazendo-se a transformação logarítmica dos dados;
- ii. o teste do sinal de Cox-Stuart foi utilizado para detectar tendências significativas e, caso necessário, a retirada da tendência foi feita ao diferenciar a série;
- iii. o teste G de Fisher foi utilizado para testar se havia um lag sazonal significativo na série e, caso houvesse, a série seria diferenciada no lag sazonal significativo;
- iv. o teste de Igualdade das Funções de Autocorrelação proposto por Quenouille (1958) e o método de comparação de Séries Temporais proposto por Silva, Ferreira e Sáfiadi

(2000) foram utilizados, para a verificação de igualdade entre as duas séries temporais testadas.

4 RESULTADOS E DISCUSSÃO

As séries temporais utilizadas nesse trabalho foram constituídas de valores diários de vazão específica ($L/s.km^2$), para as microbacias Mortandade, com vegetação nativa, e Colônia, com plantio de Pinus, no período de 13/07/2005 a 28/02/2013, com um total de 2788 observações cada série.

Para a caracterização das séries, foi realizada inicialmente a análise descritiva de cada uma delas. A série temporal referente à microbacia Colônia (Figura 4) apresentou valor mínimo de vazão específica igual a $0,2L/s.km^2$, um valor médio $4,6L/s.km^2$ e um valor máximo de $176,3L/s.km^2$. O valor máximo observado para vazão específica, corresponde a um valor discrepante, que pode ser visualizado na Figura 4, é consequência de uma chuva incomum que ocorreu na data de 30 de janeiro de 2010. Essa precipitação inesperada também afetou a série temporal referente à microbacia Mortandade (Figura 5), incorrendo no valor máximo da série igual a $451,9L/s.km^2$. O valor mínimo de vazão específica para a microbacia Mortandade foi de $1,2L/s.km^2$ e o valor médio $10,2L/s.km^2$. As duas séries temporais apresentaram diferentes números de falhas, sendo 260 observações perdidas para a microbacia Colônia e 207 para a microbacia Mortandade. As falhas observadas nas duas séries foram distintas em relação ao período em que ocorreram e também em relação ao tamanho da falha.

Figura 4 – Valores diários de vazão específica ($L/s.km^2$), para microbacia Colônia, no período de 13/07/2005 a 28/02/2013

Figura 5 – Valores diários de vazão específica ($L/s.km^2$), para microbacia Mortandade, no período de 13/07/2005 a 28/02/2013

Uma vez que para a utilização da metodologia de séries temporais é necessário que os conjuntos de dados não apresentem falhas e que as observações sejam igualmente espaçadas, foi utilizada a metodologia de imputação.

4.1 Imputação direta de dados para as séries temporais das microbacias Colônia e Mortandade

Conforme apresentado na Seção 3.1, as séries temporais apresentaram falhas em períodos distintos, sejam eles decorrentes de problemas com equipamento ou erros de leitura. Uma vez verificada a presença dessas falhas, as séries foram submetidas ao processo de imputação por meio do algoritmo AMELIA II (HONAKER; KING; BLACKWELL, 2011). As duas séries temporais foram organizadas em dois conjuntos de dados, disposto em três colunas identificadas como *Data*, *Bacia* e *Vazao*. A coluna *Bacia* é utilizada na programação como a forma de identificação da série ao longo do processo de imputação.

Para gerar uma série temporal completa para cada microbacia utilizando o pacote `amelia` (HONAKER; KING; BLACKWELL, 2011), é necessário fazer uma reamostragem da série, sendo que a nova amostra foi restringida a valores positivos, ou seja, truncando a distribuição do processo de reamostragem, obtendo apenas valores acima de zero e abaixo dos valores máximos obtidos para cada uma das microbacias. Isso se deve ao

fato de que a natureza dos dados estudados impossibilita a obtenção de valores negativos, já que se trata de valores de vazão específica em $L/s.km^2$, além de utilizar informações de natureza prática para informar qual o valor máximo que poderá ser obtido pelo processo de imputação.

As Figuras 6 e 7 correspondem à representação gráfica dos valores de vazão específica em $L/s.km^2$, para as microbacias Colônia e Mortandade, respectivamente, considerando a imputação dos dados.

Figura 6 – Valores diários de vazão específica imputados de forma direta e seus respectivos intervalos de 90% confiança para a microbacia Colônia

A escolha de um polinômio de ordem três se dá segundo a representação gráfica da densidade dos dados imputados e dos dados observados. Tal visualização possibilita avaliar a concentração das imputações em relação ao conjunto de dados. No caso apresentado, as Figuras 8 e 9, representam a melhor imputação com a utilização de uma polinômio de grau três para ambas as séries temporais.

Segundo Honaker, King e Blackwell (2011), ao realizar imputações o ideal é que se calcule um intervalo de confiança para cada observação obtida. Outro ponto a ser avaliado é a importância das informações de cunho prático que auxiliarão no processo de imputação indicando informações como amplitude da reamostragem que será efetuada.

Cabe lembrar, que o processo de imputação gera novas e diferentes amostras, pelo processo *bootstrap* e a cada execução do programa têm-se novos e diferentes conjuntos de dados completos. Os dados observados serão sempre os mesmos, no entanto, as lacunas

Figura 7 – Valores diários de vazão específica imputados de forma direta e seus respectivos intervalos de 90% confiança para a microbacia Mortandade

Figura 8 – Densidade dos dados imputados e dos dados observados para microbacia Colônia

serão preenchidas com dados diferentes. Portanto é importante, após o processo de execução da rotina computacional, exportar o novo conjunto de dados gerado, de forma a se utilizar sempre as mesmas informações para que não haja erros de análise.

Considerando o conjunto de dados completo constituído pelo valores observados de vazão específica ($L/s.km^2$) e os valores selecionados do processo de imputação para cada uma das microbacias, o próximo passo é analisá-los pela metodologia de séries temporais.

Figura 9 – Densidade dos dados imputados e dos dados observados para microbacia Mortandade

4.2 Análise das séries temporais para valores de vazão específica ($L/s.km^2$) em microbacias pareadas

Foi realizada uma nova análise descritiva das séries temporais e a representação gráfica das duas séries temporais pode ser visualizada na Figura 10.

Figura 10 – Valores diários de vazão específica ($L/s.km^2$) para as microbacias (a) Colônia e (b) Mortandade, no período de 13/07/2005 a 28/02/2013

Para a microbacia Colônia Figura 10(a), verificou-se um valor mínimo de $0,051L/s.km^2$, correspondendo a um dos valores imputados, uma média de $4,837L/s.km^2$ e o mesmo valor máximo igual a $176,300L/s.km^2$. Para a microbacia Mortandade Figura 10(b), o valor mínimo foi $0,015L/s.km^2$, valor médio igual a $10,470L/s.km^2$, sem alteração para o valor máximo de $451,900L/s.km^2$.

Em seguida, como sugerido por Morettin e Toloi (2006), foi realizada a avaliação dos gráficos de amplitude versus média para cada uma das séries temporais avaliadas.

Segundo os autores, para se ter uma ideia do tipo de tratamento adequando aos dados (por exemplo tipo de transformação necessária), pode-se utilizar um gráfico que traz no eixo das abscisas a média de subconjuntos de observações da série original, e no eixo das ordenadas, a amplitude de cada um desses subconjuntos. No presente trabalho, foram criados subconjuntos de cinco observações, e o gráfico de amplitude versus média para esses subconjuntos pode ser observado nas Figuras 36(a) e 36(b).

Figura 11 – Representação dos valores amplitude versus média para a série de valores diários de vazão específica para as microbacias (a) Colônia e (b) Mortandade

O resultado apresentado em ambas as figuras revela a necessidade de uma transformação logarítmica, pois a amplitude e a média são proporcionais, ou seja, as duas aumentam simultaneamente (MORETTIN; TOLOI, 2006). Considerando o logaritmo das séries pode-se visualizar indícios de que essa transformação foi adequada, uma vez que aparentemente a variabilidade de ambas foi estabilizada, como apresentado nas Figuras 12(a) e 12(b). Na apresentação da variabilidade das duas séries, pode-se visualizar em ambas, os pontos referentes às observações atípicas. De forma geral, a variabilidade da série pode ser verificada, devido à forma cônica com que se apresentam.

Nas Figuras 13 e 14 são apresentadas as séries transformadas com o logaritmo natural.

Para verificar a existência de tendência, foi aplicado o teste do sinal Morettin e Toloi (2006). O teste do sinal foi aplicado às duas séries transformadas para verificação da existência de tendência para a série transformada. Considerando a série com 2788

Figura 12 – Representação dos valores amplitude versus média para o logaritmo da série de valores diários de vazão específica para as microbacias (a) Colônia e (b) Mortandade

Figura 13 – Logaritmo da série de valores diários de vazão específica para microbacia Colônia

observações conclui-se, por meio do teste, que ambas as séries apresentam tendência positiva ($p - valor < 0,0001$), sendo necessário obter a primeira diferença para cada uma das séries. As séries transformadas e diferenciadas para cada uma das microbacias podem ser visualizadas nas Figuras 15 e 16.

Após diferenciadas, as séries foram novamente testadas quanto a presença de tendência por meio do teste do sinal e verificou-se que a microbacia Colônia não apresenta tendência ($p - valor = 0,7736$) assim como a microbacia Mortandade ($p - valor = 0,6757$).

Figura 14 – Logaritmo da série de valores diários de vazão específica para microbacia Mortandade

Figura 15 – Série de valores diários de vazão específica transformada e diferenciada para microbacia Colônia

No entanto, segundo a verificação das funções de autocorrelação e funções de autocorrelação parcial antes (Figuras 17 e 18) e depois (Figuras 19 e 20) da diferença, podemos observar indícios de excesso de diferenças, visível pela inversão do comportamento de tais funções (valores abaixo de zero).

Apesar de o valor do teste de Cox-Stuart indicar a ausência de tendência para as séries diferenciadas, optou-se por trabalhar com as séries transformadas no logaritmo.

Para a verificação da existência de sazonalidade foi utilizado o teste de Fisher

Figura 16 – Série de valores diários de vazão específica transformada e diferenciada para microbacia Mortandade

Figura 17 – Função de autocorrelação para microbacia colônia calculada na série log

para algum valor de defasagem (*lag*) (MORETTIN; TOLOI, 2006). O teste baseado na transformada de Fourier apresentada pelo periodograma mostrado na Figura 21, revela a inexistência de sazonalidade para a microbacia Colônia, dado o $p - valor = 0,0561$ para a estatística g de Fisher utilizada, considerando o nível de significância $\alpha = 0,05$.

O mesmo pode ser avaliado segundo o periodograma calculado para a microbacia Mortandade (Figura 22), $p - valor = 0,0856$, indicando a ausência de periodicidade na série temporal avaliada.

Como nenhuma diferença é necessária para o ajuste sazonal das séries transformadas para a microbacia Colônia e Mortandade, há indícios de que ambas as séries são

Figura 18 – Função de autocorrelação para microbacia mortandade calculada na série log

Figura 19 – Função de autocorrelação para microbacia colônia calculada na série diferenciada

Figura 20 – Função de autocorrelação para microbacia mortandade calculada na série diferenciada

estacionárias, indícios também visíveis pelo decaimento exponencial nos gráficos de função de autocorrelação para as duas séries avaliadas.

Dadas as duas séries estacionárias, o próximo procedimento é a aplicação dos

Figura 21 – Periodograma para a série transformada e diferenciada de valores diários de vazão específica para microbacia Colônia

Figura 22 – Periodograma para a série transformada e diferenciada de valores diários de vazão específica para microbacia Mortandade

testes para a comparação e posterior verificação da igualdade ou diferença entre as duas séries temporais em estudo.

4.3 Teste de igualdades das funções de autocorrelação para a comparação das séries de vazão específica para as microbacias Colônia e Mortandade

As séries temporais utilizadas para esse teste foram constituídas pelos valores transformados de vazão específica ($L/s.km^2$), com total de 2788 observações cada. O teste de igualdade das funções de autocorrelação proposto por Quenouille (1958), utiliza as funções de autocorrelação (fac) das séries estacionárias, que no presente estudo, são apresentadas nas Figuras 23(a) e 23(b).

Figura 23 – Função de autocorrelação para a série estacionária de valores diários de vazão específica para as microbacias (a) Colônia e (b) Mortandade

A partir das duas funções de autocorrelação parcial (Figuras 24 e 25), obtidas das funções de autocorrelação (23), o autor propõe o cálculo da função de autocorrelação parcial conjunta, utilizada para a identificação da ordem do termo $AR(p)$.

Por meio da função de autocorrelação parcial conjunta, que pode ser visualizada na Figura 26, pode-se sugerir o ajuste de um modelo autorregressivo de ordem quatro. Um modelo $AR(10)$ é sugerido a partir da avaliação da figura devido às autocorrelações significativas, ou seja, acima do intervalo, anteriores à primeira autocorrelação não significativa. Isso se deve ao fato de se escolher um modelo $AR(p)$ completo e mais parcimonioso, embora existam ordens superiores que poderiam ser inseridas no modelo.

A partir do modelo ajustado foram obtidas as séries residuais $\hat{a}_1$ e $\hat{a}_2$ para as microbacias Colônia e Mortandade, e suas respectivas funções de autocorrelação parcial apresentadas nas Figuras 27 e 28.

O valor da estatística do teste de comparação das funções de autocorrelação

Figura 24 – Função de autocorrelação parcial para a série estacionária de valores diários de vazão específica para microbacia Colônia

Figura 25 – Função de autocorrelação parcial para a série estacionária de valores diários de vazão específica para microbacia Mortandade

Figura 26 – Função de autocorrelação parcial estimada comum

encontrado é $SQ = 64,3020$. Para o $p - valor = 0,0003$, a hipótese de igualdade das funções de autocorrelação é rejeitada, indicando que, para o teste proposto por Quenouille

Figura 27 – Função de autocorrelação residual para a série da microbacia Colônia

Figura 28 – Função de autocorrelação residual para a série da microbacia Mortandade

(1958), as séries não são formadas pelo mesmo processo estocástico.

Como a maioria dos problemas de ajuste necessitam de estimação, o teste de Quenouille (1958), apresenta uma alternativa viável para comparação de séries baseada no ajuste de modelos por processos autorregressivos. No entanto, uma das suposições sob as quais o teste é proposto e apresentadas por Echeverry (1999), é que se o modelo ajustado é adequado, as autocorrelações dos resíduos são assintoticamente independentes, o que revela necessidade de atenção para a escolha da ordem do termo autorregressivo que será utilizado no teste. Ao escolher uma ordem muito baixa para o modelo AR, pode-se incorrer em erros ao classificar as séries como geradas pelo mesmo processo estocástico, dado que o ajuste do modelo de baixa ordem, pode não representar o comportamento real da série temporal.

4.4 Teste de comparação de séries temporais proposto por Silva, Ferreira e Sáfadi (2000)

Com o intuito de verificar se duas séries são formadas pelo mesmo processo estocástico, Silva, Ferreira e Sáfadi (2000) propuseram uma alternativa de comparação que foi utilizada para avaliar séries de preços ao consumir na cidade de Lavras. Este teste foi utilizado nesse trabalho para comparar séries de vazão específica de microbacias pareadas.

O procedimento de realização do teste proposto pelos autores consiste, inicialmente, em tomar a diferença entre as duas séries estacionárias para as microbacias Colônia e Mortandade (Figura 29) para o período de 13/07/2005 a 28/02/2013.

Figura 29 – Série da diferença entre as séries de valores para as microbacias Colônia e Mortandade

Para a escolha das ordens do modelo ajustado, foi calculada a função de autocorrelação e a partir dela, a função de autocorrelação parcial para a série resultante da diferença apresentadas nas Figuras 30 e 31.

Como proposto pelo autores, a verificação de estacionariedade e consequente a comprovação de que as duas séries são originárias de um mesmo processo estocástico, são necessárias as verificações de inexistência de tendência, sazonalidade e confirmação de que os resíduos do modelo ajustado são ruído branco. Para a verificação da presença de tendência na série resultante, aplicou-se o teste do sinal e concluiu-se pela rejeição da hipótese nula ($p\text{-valor} < 0,0001$), ou seja, a série da diferença apresenta tendência. Ainda para a série diferença, foi realizado o teste de Fisher, que se baseia no periodograma da

Figura 30 – Função de autocorrelação para a série da diferença entre as microbacias Colônia e Mortandade

Figura 31 – Função de autocorrelação parcial para a série da diferença entre as microbacias Colônia e Mortandade

série diferença (Figura 32), para a verificação da existência de sazonalidade. Para o teste proposto rejeita-se a hipótese nula de inexistência de sazonalidade ($p - \text{valor} < 0,0001$).

Para o presente estudo o teste de Box e Pierce (1970) resultou em um $p - \text{valor} < 0,0001$, levando à rejeição da hipótese nula de que a série diferença possui resíduos independentes e identicamente distribuídos com média zero e variância constante.

Embora pelo teste do sinal a série diferença não tenha apresentado tendência significativa, as pressuposições de ausência de sazonalidade e de que a série residual seja

Figura 32 – Periodograma para a série da diferença entre as séries de valores para as microbacias Colônia e Mortandade

ruído branco, conclui-se que as duas séries temporais referentes às microbacias Colônia e Mortandades não são resultantes de um mesmo processo estocástico, ou seja, as séries de valores de vazão específica da duas microbacias são diferentes segundo o teste proposto por Silva, Ferreira e Sáfadi (2000).

Silva, Ferreira e Sáfadi (2000), utilizaram o teste na comparação de séries de Índice de preços ao consumidor (IPC), e em sua abordagem, a mudança de plano monetário foi incluída no processo de modelagem, uma vez que tal característica interferiu drasticamente na política econômica do país. A inclusão dessa variável resultou em processo de comparação mais detalhado devido a possibilidade de separação da séries em dois períodos distintos, ou seja, período antecessor e período sucessor ao plano real.

Para dados ambientais, como o caso dos dados hidrológico utilizados nesse estudo, não há indícios de tendência ou mudanças de comportamento abruptas para as séries. Não foi testado, portanto, a influência de dados atípicos pontuais, que não constituem processos de memória longa como aqueles que modificam o comportamento do conjunto de dados.

4.5 Imputação de dados em intervalos para as séries temporais das microbacias Colônia e Mortandade

A primeira maneira de utilizar o método de imputação de dados sem a divisão em subconjuntos menores possibilita que ponto extremos influenciem a distribuição dos dados imputados, fazendo com que as imputações tenham valores mais elevados. Dessa forma a realização de imputações em intervalos de observações próximas possibilita um maior controle do processo de reamostragem, fazendo com que os intervalos de confiança das imputações sejam menores e mais coerentes com as observações obtidas em campo.

Assim, foram obtidas por meio do processo de imputação realizado com o algoritmo AMELIA II as seguintes séries (Figuras 33(a) e 33(b)):

Figura 33 – Valores diários de vazão específica imputados por intervalos e seus respectivos intervalos de 90% confiança para as microbacias Colônia (a) e Mortandade (b)

A observação da representação gráfica da distribuição das imputações e dos dados observados (Figuras 34(a) e 34(b)) indica uma melhor aproximação das imputações quando utilizado o processo de imputação em intervalos e reamostragem utilizando um polinômio de grau 3, em relação à imputação direta.

O resultado das imputações geraram as seguintes séries temporais sem falhas como pode ser visualizado nas Figuras 35(a) e 35(b).

Segundo a análise exploratória efetuada, agora para as séries com dados imputados em intervalos, obtém-se para a microbacia Colônia, Figura 35(a), um valor mínimo de $0,266L/s.km^2$, correspondendo a um dos valores imputados, uma média de $4,656L/s.km^2$ e o valor máximo inalterado de $176,300L/s.km^2$. Para a microbacia Mortandade Figura 35(b), o valor mínimo foi $0,041L/s.km^2$, valor médio igual a $10,070L/s.km^2$, também sem

Figura 34 – Valores diários de vazão específica para as microbacias Colônia (a) e Mortandade (b)

Figura 35 – Valores diários de vazão específica para as microbacias Colônia (a) e Mortandade (b)

alteração para o valor máximo de $451,900 L/s.km^2$.

Novamente as séries foram avaliadas para a verificação de variabilidade. A visualização da Figura 36 indica a presença de variabilidade para as séries temporais de ambas as microbacias.

Com a transformação logarítmica observa-se na Figura 37 que a variabilidade foi estabilizada.

As séries transformadas com o logaritmo, Figura 38, foram então avaliadas segundo a presença de tendência com a aplicação do teste de Cox-Stuart.

O teste de Cox-Stuart apresentou $p - valor < 0,0001$ para ambas as séries temporais, indicando que as duas apresentam tendência. No entanto a verificação das funções de autocorrelação das séries diferenciadas, Figura 40, indica um excesso de diferenças, de forma que as series transformadas não foram diferenciadas para dar sequência às análises.

A visualização da funções de autocorrelação das séries transformadas, Figura

Figura 36 – Representação dos valores amplitude versus média para a série de valores diários de vazão específica para as microbacias (a) Colônia e (b) Mortandade

Figura 37 – Representação dos valores amplitude versus média para logaritmo das séries de valores diários de vazão específica para as microbacias (a) Colônia e (b) Mortandade

Figura 38 – Séries temporais transformadas para os valores diários de vazão específica das microbacias (a) Colônia e (b) Mortandade

39, indica a estacionariedade das séries temporais, uma vez que há indícios de decaimento exponencial.

Figura 39 – Séries temporais transformadas para os valores diários de vazão específica das microbacias (a) Colônia e (b) Mortandade

Figura 40 – Funções de autocorrelação das séries temporais diferenciadas dos valores diários de vazão específica para as microbacias (a) Colônia e (b) Mortandade

Para a verificação de sazonalidade foi aplicado o teste de Fisher obtendo-se $p - valor = 0,0503$ para microbacia Colônia e $p - valor = 0,0894$ microbacia Mortandade, ou seja, nenhuma das duas séries temporais apresenta periodicidade.

Dado indício de séries temporais estacionárias foram realizados os testes de comparações de séries temporais propostos.

4.6 Teste de igualdades das funções de autocorrelação para a comparação das séries de vazão específica para as microbacias Colônia e Mortandade

Para o teste proposto por Quenouille (1958), foram utilizadas as séries transformadas de vazão específica ($L/s.km^2$), Figura 38, com total de 2788 observações cada.

Figura 41 – Funções de autocorrelação parcial residual das séries temporais de valores diários de vazão específica para as microbacias (a) Colônia e (b) Mortandade

A escolha da ordem do modelo autorregressivo de deu pela observação do gráfico da função de autocorrelação parcial comum, Figura 42, obtendo-se o valor de um modelo AR(6).

Figura 42 – Função de autocorrelação parcial estimada comum

O valor da estatística do teste de comparação das funções de autocorrelação encontrado é $SQ = 23.03234$. Para o p -valor = 0.8139, a hipótese de igualdade das funções de autocorrelação não foi rejeitada, indicando que, para o teste proposto por Quenouille (1958), as séries são formadas pelo mesmo processo estocástico.

4.7 Teste de comparação de séries temporais proposto por Silva, Ferreira e Sáfadi (2000)

Com o intuito de verificar se duas séries são formadas pelo mesmo processo estocástico com auxílio do teste proposto por Silva, Ferreira e Sáfadi (2000), obteve-se

a série temporal resultante da diferença entre as duas séries temporais das microbacias Colônia e Mortandade, resultado na Figura 43.

Figura 43 – Série da diferença entre as séries de valores para as microbacias Colônia e Mortandade

Foram calculadas ainda as funções de autocorrelação e autocorrelação parcial para a série da diferença, resultando na Figura 44.

Figura 44 – Valores diários de vazão específica para as microbacias Colônia (a) e Mortandade (b)

Segundo a metodologia proposta, foram efetuados os testes de Cox-Stuart para verificação de tendência, de Fisher para verificação de estacionariedade e o de Box e Pierce (1970) para verificação dos resíduos. Os resultados dos testes foram, respectivamente, presença de tendência ($p\text{-valor} < 0,0001$), presença de sazonalidade ($p\text{-valor} < 0,0001$) e indicação de que os resíduos não são ruído branco. De acordo com o teste proposto por Silva, Ferreira e Sáfyadi (2000), as séries não são formadas pelo mesmo processo estocástico.

5 CONCLUSÕES

A partir da utilização do algoritmo presente no pacote AMELIA II, foi possível avaliar duas formas de imputar dados faltantes para posterior análise utilizando a teoria de séries temporais. Observou-se uma significativa melhora da qualidade das imputações quando esta foi realizada de forma indireta, ou seja, imputações realizadas nos subconjuntos criados a partir das séries originais. Tomando a avaliação das duas formas distintas, tem-se portanto, uma melhor qualidade das imputações, sendo essa uma possível validação do método utilizado.

Os resultados de comparação revelaram diferenças entre os processos de imputação, de forma que o único teste que indicou igualdade entre as séries temporais foi o teste proposto por Quenouille (1958) aplicado nas séries imputadas por intervalos. A diferença apresentada pelo teste proposto por Silva, Ferreira e Sáfiadi (2000) pode ser um reflexo da natureza dos dados, que possuem variações e periodicidades não consideradas antes da aplicação no mesmo.

Considerando a imputação por intervalos dados seus melhores resultados, uma das questões práticas levantadas é sobre a existência de um fator do comportamento das duas séries temporais que as tornam diferentes entre si, uma vez que para um dos testes houve a verificação de que o mesmo processo é gerador das duas séries temporais. Como a teoria de microbacia pareadas trata condições físicas e ambientais como semelhantes, diferindo apenas em relação à cobertura vegetal, há indícios, segundo o teste proposto por Quenouille (1958), que o plantio de *Pinus* da microbacia Colônia e a vegetação nativa da microbacia Mortandade interferem de maneiras diferentes no regime de vazão de seus respectivos rios. Dada essa possibilidade, tem-se que a cobertura vegetal pode ser responsável pela diferença de nível entre as duas séries temporais avaliadas.

O presente estudo não contempla a caracterização dos processos naturais responsáveis pela ocorrência de observações atípicas, possibilitando futuros estudos que compreendam a avaliação de pontos extremos, assim como modelagem com a inclusão de variáveis auxiliares.

REFERÊNCIAS

- ANDRÉASSIAN, V. Waters and forests: from historical controversy to scientific debate. **Journal of hydrology**, Amsterdam, v. 291, n. 1, p. 1–27, 2004. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022169403005171>>. Acesso em: 21 out. 2013.
- BARELLA, W.; PETRERE, M.; SMITH, W. S.; MONTAG, L. A. As relações entre as matas ciliares, os rios e os peixes. In: RODRIGUES, R. R.; LEITÃO, H. F. (Ed.). **Matas Ciliares: Conservação e recuperação**. 2. ed. São Paulo: Edusp, 2000. p. 320.
- BLOOD, E. F. Y. **Plano de manejo florestal**: versão 2004. Telêmaco Borba, 2005. 253 p.
- BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis: Forecasting and control**. California: Holden-Day, 1976.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time serie Analysis: Forecasting and control**. New York: Prestice Hall, 1994. 598 p.
- BOX, G. E. P.; PIERCE, D. A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. **Journal of the American Statistical Association**, Alexandria, v. 65, n. 332, p. 1509–1526, 1970. Disponível em: <<http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1970.10481180>>. Acesso em: 30 jul. 2013.
- BROWN, A. E.; WESTERN, A. W.; MCMAHON, T. A.; ZHANG, L. Impact of forest cover changes on annual streamflow and flow duration curves. **Journal of Hydrology**, Amsterdam, v. 483, p. 39–50, 2013. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022169412011146>>. Acesso em: 02 jan. 2014.
- COATES, D. S.; DIGGLE, P. J. Tests for comparing two estimated spectral densities. **Journal of Time Series Analysis**, Hoboken, v. 7, n. 1, p. 7–20, 1986. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9892.1986.tb00482.x/abstract>>. Acesso em: 13 fev. 2014.
- COSTA, F. M.; SÁFADI, T. Comparação estatística de duas séries de material particulado (mp10) na cidade de são paulo. **Revista Brasileira de Biometria**, São Paulo, v. 28, n. 3, p. 23–38, 2010. Acesso em: 21 jul. 2013.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American statistical association**, Alexandria, v. 74, n. 366a, p. 427–431, 1979. Disponível em: <<http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1979.10482531>>. Acesso em: 13 fev. 2014.
- ECHEVERRY, G. H. S. **Métodos de comparação de séries temporais**. 1999. 131 p. Dissertação (Mestrado em Estatística) — Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 1999.

ECHEVERRY, G. H. S.; TOLOI, C. M. C. Testes para comparação de séries temporais: uma aplicação a séries de temperatura e salinidade da água, medidas em profundidades diferentes. **Revista Brasileira de Estatística**, São Paulo, v. 61, n. 215, p. 51–80, 2000. Acesso em: 19 maio 2013.

EFRON, B. Bootstrap methods: another look at the jackknife. **The annals of Statistics**, Philadelphia, v. 7, n. 1, p. 1–26, 1979. Acesso em: 12 fev. 2014.

EFRON, B.; TIBSHIRANI, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. **Statistical science**, Cleveland, p. 54–75, 1986. Acesso em: 12 fev. 2014.

FREEDMAN, D. A. Bootstrapping regression models. **The Annals of Statistics**, Cleveland, v. 9, n. 6, p. 1218–1228, 1981. Disponível em: <<http://projecteuclid.org/euclid.aos/1176345638>>. Acesso em: 13 fev. 2014.

FULLER, W. A. **Introduction to statistical time series**. 2 nd. ed. New York: John Wiley & Sons, 2009. 728 p.

HEWLETT, J. D. Comments on the catchment experiment to determine vegetal effects on water yelds. **Journal of the American Water Resources Association**, Hoboken, v. 7, n. 2, p. 376–381, 1971. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1752-1688.1971.tb05920.x/abstract>>. Acesso em: 21 jan. 2014.

HIBBERT, A. R. Forest treatment effects on water yield. In: SOPPER, W. E.; LULL, H. W. (Ed.). Oxford: Pergamon Press, 1965. p. 527–543.

HONAKER, J.; KING, G.; BLACKWELL, M. Amelia ii: A program for missing data. **Journal of Statistical Software**, Los Angeles, v. 45, n. 7, p. 1–47, 2011. Acesso em: 01 jan. 2014.

LIMA, W. P. **Princípios de hidrologia florestal para o manejo de bacias hidrográficas**. Piracicaba, 1986. 242 p.

LITTLE, R. J. A.; RUBIN, D. B. **Statistical Analysis With Missing Data**. 2 nd. ed. New York: JSTOR, 2002. 408 p. Disponível em: <<http://www.jstor.org/stable/1165119>>.

LÜTKEPOHL, H.; XU, F. The role of the log transformation in forecasting economic variables. **Empirical Economics**, Berlin, v. 42, n. 3, p. 619–638, 2012. Disponível em: <<http://link.springer.com/article/10.1007/s00181-010-0440-1>>. Acesso em: 14 mar. 2014.

MORETTIN, P. A.; TOLOI, C. **Análise de séries temporais**. 2. ed. São Paulo: E. Blucher, 2006. 564 p.

NASCIMENTO, M.; CRUZ, C. D.; PETERNELLI, L. A.; CAMPANA, A. C. M.; PINTO, D. S.; FERREIRA, R. P. Teste dos sinais para tendência: Uma aplicação em melhoramento de plantas. **Revista Brasileira de Biometria**, São Paulo, v. 26, n. 4, p. 19–30, 2008. Acesso em: 02 jan. 2014.

NOGUEIRA, D. A.; SÁFADI, T. Previsão de preços na bovinocultura de corte.

Organizações Rurais & Agroindustriais, Lavras, v. 2, n. 1, 2011. Disponível em: <<http://200.131.250.22/revistadae/index.php/ora/article/viewArticle/288>>. Acesso em: 07 fev. 2014.

POGGIANI, F. **Ciclagem de nutrientes em ecossistemas de plantações florestais de Eucalyptus e Pinus**: Implicações silviculturais. 1985. 211 p. Tese (Doutorado em Silvicultura) — Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1985.

QUENOUILLE, M. H. The comparison of correlations in time-series. **Journal of the Royal Statistical Society**, London, v. 20, n. 1, p. 158–164, 1958. Acesso em: 18 mar. 2014.

ROTHACHER, J. Increases in water yield following clear-cut logging in the pacific northwest. **Water Resources Research**, Washington, v. 6, n. 2, p. 653–658, 1970. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1029/WR006i002p00653/full>>. Acesso em: 30 jul. 2013.

RUBIN, D. B. Inference and missing data. **Biometrika**, Oxford, v. 63, n. 3, p. 581–592, 1976. Disponível em: <<http://biomet.oxfordjournals.org/content/63/3/581.short>>. Acesso em: 10 jan. 2014.

SABANÉS, L. **Manejo sócio-ambiental de recursos naturais e políticas públicas**: um estudo comparativo dos projetos “paraná rural” e “microbacias”. 2002. 175 p. Dissertação (Mestrado em Desenvolvimento Rural) — Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.

SAID, S. E.; DICKEY, D. A. Testing for unit roots in autoregressive-moving average models of unknown order. **Biometrika**, Oxford, v. 71, n. 3, p. 599–607, 1984. Disponível em: <<http://biomet.oxfordjournals.org/content/71/3/599.short>>. Acesso em: 14 mar. 2014.

Serviço Florestal Brasileiro. **Florestas do Brasil em resumo - 2013**: dados de 2007-2013. Brasília, 2013. Disponível em: <<http://www.florestal.gov.br/publicacoes/tecnico-cientifico/florestas-do-brasil-em-resumo-2013>>. Acesso em: 06 maio 2014.

SILVA, R. B. V.; FERREIRA, D. F.; SÁFADI, T. Modelos de séries temporais aplicados à série dos índices de preços ao consumidor na região de lavras, mg, no período de 1992 a 1999. **Organizações Rurais & Agroindustriais**, Lavras, v. 2, n. 2, p. 44–55, 2000. Disponível em: <<http://revista.dae.ufla.br/index.php/ora/article/viewArticle/283>>. Acesso em: 21 jan. 2014.

SWANK, W. T.; MINER, N. H. Conversion of hardwood-covered watersheds to white pine reduces water yield. **Water Resources Research**, Washington, v. 4, n. 5, p. 947–954, 1968. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1029/WR004i005p00947/full>>. Acesso em: 18 mar. 2014.

VOIGTLAENDER, M. **Caracterização hidrológica e biogeoquímica de microbacias**: uma comparação entre mata atlântica e pinus taeda l. 2007. 75 p. Dissertação (Mestrado em Recursos Florestais) — Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2007.

WATSON, F.; VERTESSY, R.; MCMAHON, T.; RHODES, B.; WATSON, I. Improved methods to assess water yield changes from paired-catchment studies: application to the maroondah catchments. **Forest Ecology and Management**, Amsterdam, v. 143, n. 1, p. 189–204, 2001. Disponível em: <http://www.sciencedirect.com/science/article/pii/S037811270000517X>. Acesso em: 03 mar. 2013.

WU, C. J. Jackknife, bootstrap and other resampling methods in regression analysis. **The Annals of Statistics**, Cleveland, v. 14, n. 4, p. 1261–1295, 1986. Disponível em: <http://projecteuclid.org/euclid.aos/1176350142>. Acesso em: 21 fev. 2014.

APÊNDICE

APÊNDICE A - Script do *software* R

```
#####

# Pacotes Utilizados

tseries; Amelia

#####

# Análise Preliminar

#####

dados  <- read.csv("dados.csv", h=T); dados
dados$Data <- as.Date(dados$Data, format="%d/%m/%Y")
vazpin <- ts(dados$Pinus, start=2005, freq=365); vazpin
vaznat <- ts(dados$Nativa, start=2005, freq=365); vaznat

#####

vazpin1 <- ts(na.omit(dados[,5]), start=2005, freq=365)
vaux1 <- rep(1, each = length(vazpin1))
summary(vazpin)
DPpin<- tapply(vazpin1,vaux1,sd); DPpin

vaznat1 <- ts(na.omit(dados[,6]), start=2005, freq=365)
vaux2 <- rep(1, each = length(vaznat1))
summary(vaznat)
DPnat<- tapply(vaznat1,vaux2,sd); DPnat

#####

ts.plot(vazpin, ylab="Vazão Específica (L/s.Km)", xlab="Anos", main = "Série de vazão específica
para microbacia Colônia")
ts.plot(vaznat, ylab="Vazão Específica (L/s.Km)", xlab="Anos", main = "Série de vazão específica
para microbacia Mortandade")

#####

# Processo de Imputação AMELIA II (EMB)

#####

require(Amelia)

##### COLONIA

#### Imputação Direta

ame1  <- read.csv("colonia.csv", h=T); ame1
ame1$Bacia <- as.character(ame1$Bacia)
ame1$time <- 1:2788

vaz_imp_colonia <- ame1[,c(2,3,5)]
bds1 <- matrix(c(2, 0, 50), nrow = 1, ncol = 3)
out_vaz1 <- amelia(vaz_imp_colonia, m = 1, ts = "time", cs = "Bacia", p2s = 2,
intercs = TRUE, polytime = 3, bounds=bds1, max.resample=1000)

tscsPlot(out_vaz1, cs = "Pinus", var = "Vazao", ylab="Valores imputados de vazão
específica", xlab="Observações", main = "", ylim=c(0,500))
compare.density(out_vaz1, var = 2, main="Valores imputados e observados de vazão
específica para microbacia Colônia") # para plot direto de apenas uma variável

##### Separando as Falhas em intervalo com 200 observações antes e depois de falhas

ame_i1c <- ame1[(581-200):(629+200),]
vaz_imp1c <- ame_i1c[,c(2,3,5)]
out_vaz1c <- amelia(vaz_imp1c, m = 1, ts = "time", cs = "Bacia", p2s = 2,
intercs = TRUE, polytime = 3, bounds=bds1, max.resample=1000)
```

```

tsctsPlot(out_vaz1c, cs = "Pinus", var = "Vazao", ylab="Valores imputados de
vazão específica", xlab="Observações", main = "", ylim=c(0,500))

IMP1c <- cbind(out_vaz1c$imputations$imp1[,2],out_vaz1c$missMatrix[,2])

ame_i2c <- ame1[(1094-200):(1094+200),]
vaz_imp2c <- ame_i2c[,c(2,3,5)]
out_vaz2c <- amelia(vaz_imp2c, m = 1, ts = "time", cs = "Bacia", p2s = 2,
  intercs = TRUE, polytime = 3, bounds=bds1, max.resample=1000)
tsctsPlot(out_vaz2c, cs = "Pinus", var = "Vazao", ylab="Valores imputados
de vazão específica", xlab="Observações", main = "", ylim=c(0,500))

IMP2c <- cbind(out_vaz2c$imputations$imp1[,2],out_vaz2c$missMatrix[,2])

ame_i3c <- ame1[(1901-200):(2007),]
vaz_imp3c <- ame_i3c[,c(2,3,5)]
out_vaz3c <- amelia(vaz_imp3c, m = 1, ts = "time", cs = "Bacia", p2s = 2,
  intercs = TRUE, polytime = 3, bounds=bds1, max.resample=1000)
tsctsPlot(out_vaz3c, cs = "Pinus", var = "Vazao", ylab="Valores imputados
de vazão específica", xlab="Observações", main = "", ylim=c(0,500))

IMP3c <- cbind(out_vaz3c$imputations$imp1[,2],out_vaz3c$missMatrix[,2])

ame_i4c <- ame1[(1907-200):(2119+200),]
vaz_imp4c <- ame_i4c[,c(2,3,5)]
out_vaz4c <- amelia(vaz_imp4c, m = 1, ts = "time", cs = "Bacia", p2s = 2,
  intercs = TRUE, polytime = 3, bounds=bds1, max.resample=1000)
tsctsPlot(out_vaz4c, cs = "Pinus", var = "Vazao", ylab="Valores imputados de vazão
específica", xlab="Observações", main = "", ylim=c(0,500))

IMP4c <- cbind(out_vaz4c$imputations$imp1[,2],out_vaz4c$missMatrix[,2])[250:430,]

ame_i5c <- ame1[(2146-200):(2157+200),]
vaz_imp5c <- ame_i5c[,c(2,3,5)]
out_vaz5c <- amelia(vaz_imp5c, m = 1, ts = "time", cs = "Bacia", p2s = 2,
  intercs = TRUE, polytime = 3, bounds=bds1, max.resample=1000)
tsctsPlot(out_vaz5c, cs = "Pinus", var = "Vazao", ylab="Valores imputados de vazão
específica", xlab="Observações", main = "", ylim=c(0,500))

IMP5c <- cbind(out_vaz5c$imputations$imp1[,2],out_vaz5c$missMatrix[,2])[200:220,]

ame_i6c <- ame1[(2158):(2212+200),]
vaz_imp6c <- ame_i6c[,c(2,3,5)]
out_vaz6c <- amelia(vaz_imp6c, m = 1, ts = "time", cs = "Bacia", p2s = 2,
  intercs = TRUE, polytime = 3, bounds=bds1, max.resample=1000)
tsctsPlot(out_vaz6c, cs = "Pinus", var = "Vazao", ylab="Valores imputados de vazão
específica", xlab="Observações", main = "", ylim=c(0,500))

IMP6c <- cbind(out_vaz6c$imputations$imp1[,2],out_vaz6c$missMatrix[,2])[1:200,]

ame_i7c <- ame1[(2538-200):(2549+200),]
vaz_imp7c <- ame_i7c[,c(2,3,5)]
out_vaz7c <- amelia(vaz_imp7c, m = 1, ts = "time", cs = "Bacia", p2s = 2,
  intercs = TRUE, polytime = 3, bounds=bds1, max.resample=1000)
tsctsPlot(out_vaz7c, cs = "Pinus", var = "Vazao", ylab="Valores imputados de vazão
específica", xlab="Observações", main = "", ylim=c(0,500))

IMP7c <- cbind(out_vaz7c$imputations$imp1[,2],out_vaz7c$missMatrix[,2])[1:300,]

ame_i8c <- ame1[(2550):2758,]
vaz_imp8c <- ame_i8c[,c(2,3,5)]
out_vaz8c <- amelia(vaz_imp8c, m = 1, ts = "time", cs = "Bacia", p2s = 2,
  intercs = TRUE, polytime = 3, bounds=bds1, max.resample=1000)
tsctsPlot(out_vaz8c, cs = "Pinus", var = "Vazao", ylab="Valores imputados de vazão
específica", xlab="Observações", main = "", ylim=c(0,500))

IMP8c <- cbind(out_vaz8c$imputations$imp1[,2],out_vaz8c$missMatrix[,2])

ame_i9c <- ame1[(2702):2788,]
vaz_imp9c <- ame_i9c[,c(2,3,5)]
out_vaz9c <- amelia(vaz_imp9c, m = 1, ts = "time", cs = "Bacia", p2s = 2,
  intercs = TRUE, polytime = 3, bounds=bds1, max.resample=1000)
tsctsPlot(out_vaz9c, cs = "Pinus", var = "Vazao", ylab="Valores imputados de vazão

```

```

    especifica", xlab="Observações", main = "", ylim=c(0,500))

IMP9c <- cbind(out_vaz9c$imputations$imp1[,2],out_vaz9c$missMatrix[,2])

IMPc <- c(IMP1c[IMP1c[,2]==1,][,1],IMP2c[IMP2c[,2]==1,][,1],IMP3c[IMP3c[,2]==1,][,1],
    IMP4c[IMP4c[,2]==1,][,1],IMP5c[IMP5c[,2]==1,][,1],
    IMP6c[IMP6c[,2]==1,][,1],IMP7c[IMP7c[,2]==1,][,1],IMP8c[IMP8c[,2]==1,][,1],
    IMP9c[IMP9c[,2]==1,][,1])

IMP <- cbind(out_vaz1$imputations$imp1[,2],out_vaz1$missMatrix[,2])
IMP[IMP[,2]==1,][,1] <- IMPc

out_vaz1$imputations$imp1[,2] <- IMP[,1]

tscsPlot(out_vaz1, cs = "Pinus", var = "Vazao", ylab="Valores imputados de vazão
    especifica", xlab="Observações", main = "", ylim=c(0,500))
compare.density(out_vaz1, var = 2, main="Valores imputados e observados de vazão
    especifica para microbacia Colônia") # para plot direto de apenas uma variável

##### MORTANDADE

#### Imputação direta

ame2 <- read.csv("mortandade.csv", h=T); ame2
ame2$Bacia <- as.character(ame2$Bacia)
ame2$time <- 1:2788

vaz_imp_mortandade <- ame2[,c(2,3,5)]
bds2 <- matrix(c(2, 0, 150), nrow = 1, ncol = 3)
out_vaz2 <- amelia(vaz_imp_mortandade, m = 1, ts = "time", cs = "Bacia", p2s = 2,
    intercs = TRUE, polytime = 3, bounds=bds2, max.resample=1000)

tscsPlot(out_vaz2, cs = "Nativa", var = "Vazao", ylab="Valores imputados de vazão
    especifica", xlab="Observações", main = "", ylim=c(0,500))
compare.density(out_vaz2, var = 2, main="Valores imputados e observados de vazão
    especifica para microbacia Mortandade") # para plot direto de apenas uma variável

##### Separando as Falhas em intervalo com 200 observações antes e depois de falhas

ame_i1c2 <- ame2[1:(12+200),]
vaz_imp1c2 <- ame_i1c2[,c(2,3,5)]
out_vaz1c2 <- amelia(vaz_imp1c2, m = 1, ts = "time", cs = "Bacia", p2s = 2,
    intercs = TRUE, polytime = 3, bounds=bds2, max.resample=1000)
tscsPlot(out_vaz1c2, cs = "Nativa", var = "Vazao", ylab="Valores imputados de vazão
    especifica", xlab="Observações", main = "", ylim=c(0,500))

IMP1c2 <- cbind(out_vaz1c2$imputations$imp1[,2],out_vaz1c2$missMatrix[,2])

ame_i2c2 <- ame2[(581-200):(809),]
vaz_imp2c2 <- ame_i2c2[,c(2,3,5)]
out_vaz2c2 <- amelia(vaz_imp2c2, m = 1, ts = "time", cs = "Bacia", p2s = 2,
    intercs = TRUE, polytime = 3, bounds=bds2, max.resample=1000)
tscsPlot(out_vaz2c2, cs = "Nativa", var = "Vazao", ylab="Valores imputados de vazão
    especifica", xlab="Observações", main = "", ylim=c(0,500))

IMP2c2 <- cbind(out_vaz2c2$imputations$imp1[,2],out_vaz2c2$missMatrix[,2])

ame_i3c2 <- ame2[630:863,]
vaz_imp3c2 <- ame_i3c2[,c(2,3,5)]
out_vaz3c2 <- amelia(vaz_imp3c2, m = 1, ts = "time", cs = "Bacia", p2s = 2,
    intercs = TRUE, polytime = 3, bounds=bds2, max.resample=1000)
tscsPlot(out_vaz3c2, cs = "Nativa", var = "Vazao", ylab="Valores imputados de vazão
    especifica", xlab="Observações", main = "", ylim=c(0,500))

IMP3c2 <- cbind(out_vaz3c2$imputations$imp1[,2],out_vaz3c2$missMatrix[,2])

ame_i4c2 <- ame2[850:(883+200),]
vaz_imp4c2 <- ame_i4c2[,c(2,3,5)]
out_vaz4c2 <- amelia(vaz_imp4c2, m = 1, ts = "time", cs = "Bacia", p2s = 2,
    intercs = TRUE, polytime = 3, bounds=bds2, max.resample=1000)

```

```

tsctsPlot(out_vaz4c2, cs = "Nativa", var = "Vazao", ylab="Valores imputados de vazão
específica", xlab="Observações", main = "", ylim=c(0,500))

IMP4c2 <- cbind(out_vaz4c2$imputations$imp1[,2],out_vaz4c2$missMatrix[,2])

ame_i5c2 <- ame2[(1667-200):(1667+200),]
vaz_imp5c2 <- ame_i5c2[,c(2,3,5)]
out_vaz5c2 <- amelia(vaz_imp5c2, m = 1, ts = "time", cs = "Bacia", p2s = 2,
intercs = TRUE, polytime = 3, bounds=bds2, max.resample=1000)
tsctsPlot(out_vaz5c2, cs = "Nativa", var = "Vazao", ylab="Valores imputados de vazão
específica", xlab="Observações", main = "", ylim=c(0,500))

IMP5c2 <- cbind(out_vaz5c2$imputations$imp1[,2],out_vaz5c2$missMatrix[,2])

ame_i6c2 <- ame2[(2102-200):2145,]
vaz_imp6c2 <- ame_i6c2[,c(2,3,5)]
out_vaz6c2 <- amelia(vaz_imp6c2, m = 1, ts = "time", cs = "Bacia", p2s = 2,
intercs = TRUE, polytime = 3, bounds=bds2, max.resample=1000)
tsctsPlot(out_vaz6c2, cs = "Nativa", var = "Vazao", ylab="Valores imputados de vazão
específica", xlab="Observações", main = "", ylim=c(0,500))

IMP6c2 <- cbind(out_vaz6c2$imputations$imp1[,2],out_vaz6c2$missMatrix[,2])

ame_i7c2 <- ame2[2120:2201,]
vaz_imp7c2 <- ame_i7c2[,c(2,3,5)]
out_vaz7c2 <- amelia(vaz_imp7c2, m = 1, ts = "time", cs = "Bacia", p2s = 2,
intercs = TRUE, polytime = 3, bounds=bds2, max.resample=1000)
tsctsPlot(out_vaz7c2, cs = "Nativa", var = "Vazao", ylab="Valores imputados de vazão
específica", xlab="Observações", main = "", ylim=c(0,500))

IMP7c2 <- cbind(out_vaz7c2$imputations$imp1[,2],out_vaz7c2$missMatrix[,2])

ame_i8c2 <- ame2[2158:(2212+200),]
vaz_imp8c2 <- ame_i8c2[,c(2,3,5)]
out_vaz8c2 <- amelia(vaz_imp8c2, m = 1, ts = "time", cs = "Bacia", p2s = 2,
intercs = TRUE, polytime = 3, bounds=bds2, max.resample=1000)
tsctsPlot(out_vaz8c2, cs = "Nativa", var = "Vazao", ylab="Valores imputados de vazão
específica", xlab="Observações", main = "", ylim=c(0,500))

IMP8c2 <- cbind(out_vaz8c2$imputations$imp1[,2],out_vaz8c2$missMatrix[,2])

ame_i9c2 <- ame2[(2667-200):2729,]
vaz_imp9c2 <- ame_i9c2[,c(2,3,5)]
out_vaz9c2 <- amelia(vaz_imp9c2, m = 1, ts = "time", cs = "Bacia", p2s = 2,
intercs = TRUE, polytime = 3, bounds=bds2, max.resample=1000)
tsctsPlot(out_vaz9c2, cs = "Nativa", var = "Vazao", ylab="Valores imputados de vazão
específica", xlab="Observações", main = "", ylim=c(0,500))

IMP9c2 <- cbind(out_vaz9c2$imputations$imp1[,2],out_vaz9c2$missMatrix[,2])

ame_i10c2 <- ame2[2677:2785,]
vaz_imp10c2 <- ame_i10c2[,c(2,3,5)]
out_vaz10c2 <- amelia(vaz_imp10c2, m = 1, ts = "time", cs = "Bacia", p2s = 2,
intercs = TRUE, polytime = 3, bounds=bds2, max.resample=1000)
tsctsPlot(out_vaz10c2, cs = "Nativa", var = "Vazao", ylab="Valores imputados de vazão
específica", xlab="Observações", main = "", ylim=c(0,500))

IMP10c2 <- cbind(out_vaz10c2$imputations$imp1[,2],out_vaz10c2$missMatrix[,2])

ame_i11c2 <- ame2[2761:2788,]
vaz_imp11c2 <- ame_i11c2[,c(2,3,5)]
out_vaz11c2 <- amelia(vaz_imp11c2, m = 1, ts = "time", cs = "Bacia", p2s = 2,
intercs = TRUE, polytime = 3, bounds=bds2, max.resample=1000)
tsctsPlot(out_vaz11c2, cs = "Nativa", var = "Vazao", ylab="Valores imputados de vazã
o específica", xlab="Observações", main = "", ylim=c(0,500))

IMP11c2 <- cbind(out_vaz11c2$imputations$imp1[,2],out_vaz11c2$missMatrix[,2])

IMP2c <- c(IMP1c2[IMP1c2[,2]==1,][,1],IMP2c2[IMP2c2[,2]==1,][,1],IMP3c2[IMP3c2[,2]==1,]
[,1],IMP4c2[IMP4c2[,2]==1,][,1],IMP5c2[IMP5c2[,2]==1,][1],
IMP6c2[IMP6c2[,2]==1,][,1],IMP7c2[IMP7c2[,2]==1,][,1],IMP8c2[IMP8c2[,2]==1,]
[,1],IMP9c2[IMP9c2[,2]==1,][,1],
IMP10c2[IMP10c2[,2]==1,][,1],IMP11c2[IMP11c2[,2]==1,][,1])

```

```

IMP2 <- cbind(out_vaz2$imputations$imp1[,2],out_vaz2$missMatrix[,2])
IMP2[IMP2[,2]==1,][,1] <- IMP2c

out_vaz2$imputations$imp1[,2] <- IMP2[,1]

tscsPlot(out_vaz2, cs = "Nativa", var = "Vazao", ylab="Valores imputados de vazão
específica", xlab="Observações", main = "", ylim=c(0,500))
compare.density(out_vaz2, var = 2, main="Valores imputados e observados de vazão
específica para microbacia Mortandade") # para plot direto de apenas uma variável

#####

# Criar os objetos com as séries completas

#####

sxv <- out_vaz$imputations$imp1[1:2788,2];
ts.plot(sxv, ylab="Vazão Específica (L/s.Km)", xlab="Observações (Sem Falha)", main = "Série de
vazão específica para microbacia Colônia")

syv <- out_vaz$imputations$imp1[2789:5576,2];
ts.plot(syv, ylab="Vazão Específica (L/s.Km)", xlab="Observações (Sem Falha)", main = "Série de
vazão específica para microbacia Mortandade")

quadro <- data.frame(sxv, syv)
write.csv(quadro, "serie_imputada.csv")

dados <- read.csv("serie_imputada.csv", h=T); dados
sxv <- dados[,3]
syv <- dados[,4]

#####

# Avaliação Visual para heterogeneidade de variâncias das Séries Imputadas

#####

vazpin2 <- ts(sxv, start=2005, freq=365)
vaux2 <- rep(1, each = length(vazpin2))
summary(vazpin2)

vaznat2 <- ts(syv, start=2005, freq=365)
vaux2 <- rep(1, each = length(vaznat2))
summary(vaznat2)

### Pinus

vaux <- rep(1:557, each = 5)
M<- tapply(sxv[1:2785],vaux,mean)
#SD<- tapply(sxv[1:2785],vaux,sd)
A<- tapply(sxv[1:2785],vaux,max)-tapply(sxv[1:2785],vaux,min)
# par(mfrow=c(1,2))
plot(M,A, xlab="Média", ylab="Amplitude")
#plot(M,SD, , xlab="Média", ylab="Desvio Padrão")

### Nativa

vaux <- rep(1:557, each = 5)
M<- tapply(syv[1:2785],vaux,mean)
#SD<- tapply(syv[1:2785],vaux,sd)
A<- tapply(syv[1:2785],vaux,max)-tapply(syv[1:2785],vaux,min)
#par(mfrow=c(1,2))
plot(M,A, xlab="Média", ylab="Amplitude")
#plot(M,SD, , xlab="Média", ylab="Desvio Padrão")

##### Presença de Heterocedasticidade - Aplica-se o Logaritmo Natural

### Pinus
lsxv <- log(sxv)

```



```

vaux <- rep(1:557, each = 5)
M<- tapply(lsxv[1:2785],vaux,mean)
A<- tapply(lsxv[1:2785],vaux,max)-tapply(lsxv[1:2785],vaux,min)
plot(M,A, xlab="Média", ylab="Amplitude")

ts.plot(lsxv, ylab="log(Vazão Específica)", xlab="Observações", main = "Logaritmo natural da
série de vazão específica para microbacia Colônia")

### Nativa
lsyv <- log(syv)

vaux <- rep(1:557, each = 5)
M<- tapply(lsyv[1:2785],vaux,mean)
A<- tapply(lsyv[1:2785],vaux,max)-tapply(lsyv[1:2785],vaux,min)
plot(M,A, xlab="Média", ylab="Amplitude")

ts.plot(lsyv, ylab="log(Vazão Específica)", xlab="Observações", main = "Logaritmo natural da
série de vazão específica para microbacia Mortandade")

#####

# Teste de Cox-Stuart - verificação de Tendência

#####

cox.stuart <- function(series, trend=c("+","-")) {
  trend <- match.arg(trend)
  N <- length(series)
  p1 <- series[1:(N/2)]
  p2 <- series[(N/2+1):N]
  sinais <- sign(p2-p1)
  if(trend=="-") T2 <- sum(sinais<0) else T2 <- sum(sinais>0)
  n <- length(sinais[sinais!=0])
  tcrit <- qbinom(.95, n, .5)
  pv <- 1-pbinom(T2, n, .5)
  if(round(pv, 4)==0) pv <- "<0.0001" else pv <- round(pv, 4)
  data.frame("N"=N, "n"=n, "estatistica"=T2, "valor.critico"=tcrit, "tendencia"=trend, "p-valor"=pv)
}

##### Colônia - Pinus

cox.stuart(lsxv, trend="+") # Série transformada PINUS apresenta tendência positiva - Toma-se a série diff
cox.stuart(diff(lsxv), trend="+") # Série diferenciada PINUS não apresenta tendência

dlsxv <- diff(lsxv)
ts.plot(dlsxv, ylab="diferença[log(Vazão Específica)]", xlab="Observações", main = "Diferença da série
transformada de vazão específica para microbacia Colônia")
acf(dlsxv, main = "Função de autocorrelação da série estacionária para microbacia Colônia")

##### Mortandade - Nativa

cox.stuart(lsyv, trend="+") # Série transformada NATIVA apresenta tendência positiva - Toma-se a série diff
cox.stuart(diff(lsyv), trend="+") # Série diferenciada NATIVA não apresenta tendência

dlsyv <- diff(lsyv)
ts.plot(dlsyv, ylab="diferença[log(Vazão Específica)]", xlab="Observações", main = "Diferença da série
transformada de vazão específica para microbacia Mortandade")
acf(dlsyv, main = "Função de autocorrelação da série estacionária para microbacia Mortandade")

pacf(dlsxv, main = "Função de autocorrelação parcial da série estacionária para microbacia Colônia")

pacf(dlsyv, main = "Função de autocorrelação parcial da série estacionária para microbacia Mortandade")

#####

# Teste de Fisher - Sazonalidade

#####

dlsxv <- diff(log(sxv))

pgram <- spectrum(dlsxv, fast=FALSE, detrend=FALSE, log="no", plot=T, main = "Periodograma da série

```

```

    estacionária para microbacia Colônia")$spec

pvalue.g <- function(g.obs, n) n*(1-g.obs)^(n-1)

n <- length(pgram)
g.obs <- max(pgram)/sum(pgram)
pvalue.g(g.obs, n) # > 0.05, portanto não existe sazonalidade no período p

pgram1 <- spectrum(dlsyv, fast=FALSE, detrend=FALSE, log="no", plot=T, main = "Periodograma da série
    estacionária para microbacia Mortandade")$spec

n1 <- length(pgram1)
g.obs <- max(pgram1)/sum(pgram1)
pvalue.g(g.obs, n1) # > 0.05, portanto não existe sazonalidade no período p

# qual é o período p?
pgram2 <- data.frame(pgram1, 1:n1, p=pgram1==max(pgram1))
pgram2[pgram2$p==T,]

#####

require(tseries)

adf.test(dlsxv)
adf.test(dlsyv)

#####

# Teste de Box-Pierce - Autocorrelações dos resíduos estimados (iid)

#####

Box.test(dlsxv, type = "Box-Pierce")
Box.test(dlsyv, type = "Box-Pierce")

#####

# Comparação de Séries Temporais - Teste da Igualdade das Funções de Autocorrelação (Quenouille)

#####

quenouille.test <- function(sx, sy, j=20, P=j, p, plot=F) {
  n1 <- length(sx)
  n2 <- length(sy)
  acfx <- acf(sx, j, type="correlation", plot=F)$acf
  acfy <- acf(sy, j, type="correlation", plot=F)$acf
  acfcom <- ((n1*acfx)+(n2*acy))/(n1+n2)
  acf <- acfcom
  zeros <- rep(0,j)
  acf <- c(zeros, acf)
  phi <- numeric(j)
  phi[1] <- acf[j+2]
  vetmat <- numeric(j)
  vetmat[1] <- list(1)
  for (m1 in 2:j) {
    mat1 <- numeric(m1*m1)
    mat1 <- matrix(mat1, m1, m1)
    for (m2 in 1:m1) {
      lim1 <- j+2-m2
      lim2 <- j+m1+1-m2
      fila <- acf[lim1:lim2]
      mat1[m2,] <- fila
    }
  }
  diagsup <- mat1
  matden <- diagsup+t(diagsup)
  matden <- matden-diag(m1)
  vetmat[m1] <- list(matden)
  denom <- determinant(matden, logarithm=F)
  denom <- (denom$modulus)*(denom$sign)
  lim3 <- j+2
  lim4 <- j+m1+1
  col <- acf[lim3:lim4]
  col <- matrix(col,m1,1)

```

```

mat2 <- matden[,-c(m1)]
mat3 <- cbind(mat2,col)
num <- determinant(mat3, logarithm=F)
num <- (num$modulus)*(num$sign)
phi[m1] <- num/denom
}
facp <- c(phi[1:j])
facp <- abs(facp)
maximo <- max(n1,n2)
inter <- 2*(1/sqrt(maximo))
vetorlag <- numeric(j)
for (m in 1:j){
  if (facp[m]>inter){vetorlag[m]<-m} else {vetorlag[m]<-1}
}
vetorlag <- c(vetorlag[1:j])
if(missing(p)) p <- max(vetorlag)
if(plot) {
  plot(1:j, facp, type="h", xlab="lag", ylab="PACF Conjunta")
  abline(h=0)
  abline(h=c(-inter,inter), lty=2, col=4)
  cat("Digite o valor de p")
  p <- scan(n=1)
}
lim1 <- j+2
lim2 <- j+1+p
r <- acf[lim1:lim2]
r <- matrix(r,p,1)
R <- vetmat[[p]]
R <- matrix(R,p,p)
hipar <- solve(R)%*%r
hipar <- matrix(hipar,p,1)
p1 <- p+1
np1 <- n1-p
np2 <- n2-p
x <- sx[c(p1:n1)]
x <- matrix(x,np1,1)
y <- sy[c(p1:n2)]
y <- matrix(y,np2,1)
dim1 <- np1*p
dim2 <- np2*p
xmat <- numeric(dim1)
xmat <- matrix(xmat,np1,p)
ymat <- numeric(dim2)
ymat <- matrix(ymat,np2,p)
lim1 <- p1
lim2 <- n1
lim3 <- n2
for (l in 1:p){
  lim1 <- lim1-1
  lim2 <- lim2-1
  lim3 <- lim3-1
  colx <- sx[lim1:lim2]
  coly <- sy[lim1:lim3]
  xmat[,l] <- colx
  ymat[,l] <- coly
}
xmat <- xmat
ymat <- ymat
res1 <- x-xmat%%hipar
res2 <- y-ymat%%hipar
fac1 <- acf(res1, P, type="partial", plot=F, main = "Série Residual 1")
fac2 <- acf(res2, P, type="partial", plot=T, main = "Série Residual 2")
S <- fac1$lag
fac1 <- fac1$acf
fac2 <- fac2$acf
DEN <- (1/(n1-S))+(1/(n2-S))
NUM <- (fac1-fac2)^2
sq <- sum(NUM/DEN)
pv1 <- 1-pchisq(sq,P)
data.frame("estatistica"=sq,"gl"=P,"p-valor"=pv1)
}

```

```

quenouille.test(dlsxv, dlsyv, j=30, P=30, plot=T) # Funções de Autocorrelação não são iguais

```

```
#####

# Comparação de Séries Temporais - Comparação método Silva et al (2001)

#####

# 1 - Efetua-se a diferença entre as duas séries temporais em estudo

difer <- dlsyv - dlsxv; difer
ts.plot(difer, ylab="Valores da diferença", xlab="Observações", main = "Série da diferença
entre as séries estacionárias das microbacias Colônia e Mortandade")

acf(difer, main = "Função de autocorrelação para a série da diferença entre as microbacias
Colônia e Mortandade")
pacf(difer, main = "Função de autocorrelação parcial para a série da diferença entre as
microbacias Colônia e Mortandade")

# 2 - Na Série resultante do passo anterior aplica-se o teste de Cox-Stuart

cox.stuart <- function(series, trend=c("+","-")) {
  trend <- match.arg(trend)
  N <- length(series)
  p1 <- series[1:(N/2)]
  p2 <- series[(N/2+1):N]
  sinais <- sign(p2-p1)
  if(trend=="-") T2 <- sum(sinais<0) else T2 <- sum(sinais>0)
  n <- length(sinais[sinais!=0])
  if(trend=="+") tcrit <- qbinom(.95, n, .5) else tcrit <- qbinom(.05, n, .5)
  if(trend=="-") pv <- 1-pbinom(T2, n, .5) else pv <- pbinom(T2, n, .5)
  if(round(pv, 4)==0) pv <- "<0.0001" else pv <- round(pv, 4)
  data.frame("N"=N, "n"=n, "estatistica"=T2, "valor.critico"=tcrit, "tendencia"=trend,
    "p-valor"=pv)
}

cox.stuart(difer, trend = "+") # ----- pvalor = 0.2266 não tem tendência

# 3 - Aplica-se o teste de Fisher

pgram2 <- spectrum(difer, fast=FALSE, detrend=FALSE, log="no", plot=T, main = "Periodograma
para a série da diferença entre as microbacias Colônia e Mortandade")$spec

n2 <- length(pgram2)
g.obs <- max(pgram2)/sum(pgram2)
pvalue.g(g.obs, n2) # > 0.05, portanto não existe sazonalidade

# qual é o período p? ----- lag=1387?
pgram2 <- data.frame(pgram2, 1:n2, p=pgram2==max(pgram2))
pgram2[pgram2$p==T,]

# 4 - Aplica-se o teste de Box e Pierce ----- p-valor < 0.0001 não são ruído branco

Box.test(difer, lag = 15, type = "Ljung-Box") # Hipotese nula rejeitada H0 -
```