

UNIVERSIDADE FEDERAL DE MATO GROSSO
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA AMBIENTAL

**PREENCHIMENTO DE FALHAS ALEATÓRIAS DE
SÉRIES TEMPORAIS MICROMETEOROLÓGICAS
PELA TÉCNICA DE MONTE CARLO**

FAHIM ELIAS COSTA RIHBANE

ORIENTADOR: PROF. DR DENILTON CARLOS GAIO
COORIENTADOR: PROF. DR CARLO RALPH DE MUSIS

Cuiabá, MT, Fevereiro de 2014

UNIVERSIDADE FEDERAL DE MATO GROSSO
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA AMBIENTAL

**PREENCHIMENTO DE FALHAS ALEATÓRIAS DE
SÉRIES TEMPORAIS MICROMETEOROLÓGICAS
PELA TÉCNICA DE MONTE CARLO**

FAHIM ELIAS COSTA RIHBANE

*Dissertação apresentado ao Programa de
Pós-graduação em Física Ambiental da
Universidade Federal de Mato Grosso,
como parte dos requisitos para obtenção
do Título de Mestre em Física Ambiental.*

**ORIENTADOR: PROF. DR DENILTON CARLOS GAIO
COORIENTADOR: PROF. DR CARLO RALPH DE MUSIS**

Cuiabá, MT, Fevereiro de 2014

Dados Internacionais de Catalogação na Fonte.

C837p Costa Rihbane, Fahim Elias.
Preenchimento de Falhas Aleatórias de Séries Temporais Micrometeorológicas
pela Técnica de Monte Carlo / Fahim Elias Costa Rihbane. -- 2014
48 f. : il. color. ; 30 cm.

Orientador: Denilton Carlos Gaio.

Co-orientador: Carlo Ralph De Musis.

Dissertação (mestrado) - Universidade Federal de Mato Grosso, Instituto de Física, Programa de Pós-Graduação em Física Ambiental, Cuiabá, 2014.

Inclui bibliografia.

1. Monte Carlo. 2. Preenchimento de falhas. 3. Séries Temporais. I. Título.

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

Permitida a reprodução parcial ou total, desde que citada a fonte.

UNIVERSIDADE FEDERAL DE MATO GROSSO
INSTITUTO DE FÍSICA
Programa de Pós-Graduação em Física Ambiental

FOLHA DE APROVAÇÃO

**TÍTULO: PREENCHIMENTO DE FALHAS ALEATÓRIAS DE
SÉRIES TEMPORAIS MICROMETEOROLÓGICAS PELA
TÉCNICA DE MONTE CARLO**

AUTOR: FAHIM ELIAS COSTA RIBBANE

Dissertação de Mestrado defendida e aprovada em 25 de fevereiro de 2014, pela
comissão julgadora:



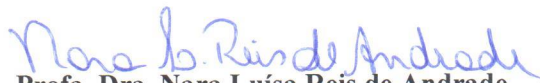
Prof. Dr. Denilton Carlos Gaio
Orientador
Instituto de Física - UFMT



Prof. Dr. Carlo Ralph De Musis
Coorientador
Universidade de Cuiabá - UNIC



Profa. Dra. Flávia Maria de Moura Santos
Examinadora Interna
Faculdade de Arquitetura, Engenharia e Tecnologia –
UFMT



Profa. Dra. Nara Luísa Reis de Andrade
Examinadora Externa
Departamento de Engenharia Ambiental/UNIR

DEDICATÓRIA

Ao meus pais, amigos, família e a todos aqueles que de alguma forma me ajudaram

AGRADECIMENTOS

Primeiramente agradeço a Deus por nos conceder o dom da vida e tudo que existe;

Ao meu pai Almir Ribane e minhas mães Cleocema Ferreira da Costa e Ana Maria Nunes pelo incentivo aos estudos me proporcionando tudo do bom e do melhor, pelo amor, carinho e ensinamentos que me foram passados;

Meu orientador, Prof. Dr. Denilton Carlos Gaio, pela paciência, compreensão, pela amizade e conhecimento compartilhado;

Ao meu coorientador Prof. Dr. Carlo Ralph De Musis, pela contribuição e compartilhamento de conhecimento;

Ao Amigo Mestre, Raphael de Souza pela amizade, paciência e dedicação em me ajudar;

Ao Amigo Mestre, Victor Hugo de Moraes Danelichen pela amizade e compartilhamento de conhecimento;

A prof^a. Dr.^a Luciana Sanches pelo incentivo ao ingresso do mestrado, contribuições nos trabalhos desenvolvidos, conhecimento compartilhados e amizade;

Natallia Sanches que idealizou a seguir em frente, fazer o mestrado na Física Ambiental e pelo companheirismo;

Ao coordenador prof. Dr. Jose de Souza Nogueira (Paraná), sempre disposto a ajudar, aconselhar e me incentivar rumo ao crescimento intelectual e pessoal;

A prof^a. Dr.^a Nara Luisa pelo aceite em participar da banca e pelas contribuições;

A prof^a. Dr.^a Marta Nogueira, pelo apoio e acolhimento de sempre;

A Soilce e ao Cesário, pela amizade e boa vontade para auxiliar nas questões burocráticas e amizade;

A Minha namorada Elizaura Alvez e Família, por estar ao meu lado, paciência e compreensão;

Aos Amigos e professores da família do Programa de Pós Graduação em Física Ambiental;

A CAPES, órgão financiador da minha bolsa de pesquisa;

E a todos amigos e familiares;

SUMÁRIO

LISTA DE FIGURAS	vii
LISTA DE ABREVIATURAS E SÍMBOLOS	ix
RESUMO	x
ABSTRACT	xi
1. INTRODUÇÃO	12
1.1. PROBLEMÁTICA	12
1.2. JUSTIFICATIVA	13
2. REVISÃO DA LITERATURA	14
2.1. SÉRIES TEMPORAIS	14
2.1.1. Série Temporal Discreta e Contínua	15
2.2. MODELOS DE SÉRIES TEMPORAIS	16
2.3. TIPOS DE MODELOS	16
2.3.1. Modelo Paramétrico	17
2.3.2. Modelo Não Paramétrico	17
2.4. CARACTERÍSTICAS DAS SÉRIES TEMPORAIS	17
2.4.1. Tendência	18
2.4.2. Sazonalidade & Ciclos	19
2.5. SÉRIES E FALHAS	19
2.6. ERRO DE MEDIÇÃO DO DATA LOGGER	21
2.7. MÉTODOS PARA PREENCHIMENTO DE FALHAS	21
2.8. MONTE CARLO	22
2.8.1. Aleatoriedade	24
2.8.2. Bootstrap	24

3. MATERIAL E MÉTODOS	26
3.1. DADOS DE ESTUDO	26
3.2. SISTEMA DE GERAÇÃO DE FALHAS PAREADAS	28
3.3. DISTRIBUIÇÃO DE FREQUÊNCIAS	28
3.4. BOOTSTRAP & MONTE CARLO	29
4. RESULTADOS E DISCUSSÕES	31
4.1. ANÁLISE DO PREENCHIMENTO DE FALHAS PARA SÉRIE TEMPORAL DA TEMPERATURA	31
4.2. ANÁLISE DO PREENCHIMENTO DE FALHAS PARA SÉRIE TEMPORAL DA UMIDADE RELATIVA DO AR	34
4.3. ANÁLISE DO PREENCHIMENTO DE FALHAS PARA SÉRIE TEMPORAL DE FLUXO DE CALOR NO SOLO.....	36
4.4. ANÁLISE ESTATÍSTICA DAS VARIÁVEIS	38
5. CONSIDERAÇÕES FINAIS.....	43
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	44

LISTA DE FIGURAS

Figura 1: (a) Totais mensais de passageiros em linhas aéreas internacionais nos EUA, (b) Número anual de lincos capturados em armadilhas no Canadá, (c) Medições anuais de vazões do Rio Nilo em Ashwan, (d) Consumo de gás no Reino Unido no primeiro trimestre de cada ano.(EHLERS,2009).	15
Figura 2: Localização da área de estudo. (SANCHES et al., 2006).....	26
Figura 3: Fluxograma representando os processos do preenchimento das falhas por etapa, círculo verde é o início do processo, círculo vermelha é o fim do processo, as caixas são cada processo individual, os processos em amarelo serão feitos um loop de acordo com o processo de cor preta.	30
Figura 4: Gráfico de dispersão com valores preenchidos e originais para a série temporal da temperatura do ar correspondente a quinta simulação, (a) 10% falhas, (b) 20% falhas, (c) 30% falhas, (d) 40% falhas, (e) 50% falhas, (f) 60% falhas e (g) 70% falhas.	32
Figura 5: Séries temporais da temperatura do ar referente a quinta simulação,original (em azul) e preenchida (em vermelho), (a) 10% falhas, (b) 20% falhas, (c) 30% falhas, (d) 40% falhas, (e) 50% falhas, (f) 60% falhas e (g) 70% falhas.	33
Figura 6: Valores preenchidos e originais para a série temporal da umidade relativa do ar correspondente a quinta simulação, (a) 10% falhas, (b) 20% falhas, (c) 30% falhas, (d) 40% falhas, (e) 50% falhas, (f) 60% falhas e (g) 70% falhas.	35
Figura 7: Séries temporais da umidade relativa do ar referente a quinta simulação,original (em azul) e preenchida (em vermelho), (a) 10% falhas, (b) 20% falhas, (c) 30% falhas, (d) 40% falhas, (e) 50% falhas, (f) 60% falhas e (g) 70% falhas.	36
Figura 8: Valores preenchidos e originais para a série temporal do fluxo de calor no solo correspondente a quinta simulação, (a) 10% falhas, (b) 20% falhas, (c) 30% falhas, (d) 40% falhas, (e) 50% falhas, (f) 60% falhas e (g) 70% falhas.	37
Figura 9: Séries temporais do fluxo calor no solo referente a quinta simulação,original (em azul) e preenchida (em vermelho), (a) 10% falhas, (b) 20%	

falhas, (c) 30% falhas, (d) 40% falhas, (e) 50% falhas, (f) 60% falhas e (g) 70% falhas.	38
Figura 10: Relação entre coeficiente de correlação (R^2) e porcentagem de falhas das nove simulações para variável temperatura do ar.	39
Figura 11: Relação entre coeficiente de correlação (R^2) e porcentagem de falhas das nove simulações para variável umidade relativa do ar.	39
Figura 12: Relação entre coeficiente de correlação (R^2) e porcentagem de falhas das nove simulações para variável fluxo de calor no solo.	39
Figura 13: Mediana, limite superior e limite inferior com 95% de intervalo de confiança para variável de temperatura do ar para as nove simulações.	40
Figura 14: Mediana, limite superior e limite inferior com 95% de intervalo de confiança para variável de umidade relativa do ar para as nove simulações.	40
Figura 15: Mediana, limite superior e limite inferior com 95% de intervalo de confiança para variável de fluxo de calor no solo para as nove simulações.	41
Figura 16: Teste-f para a variável de temperatura do ar para as nove simulações em que círculos com cores diferentes representam simulações diferentes.	41
Figura 17: Teste-f para variável de umidade relativa do ar para as nove simulações em que círculos com cores diferentes representam simulações diferentes.	42
Figura 18: Teste-f para variável de fluxo de calor no solo para as nove simulações em que círculos com cores diferentes representam simulações diferentes.	42

LISTA DE ABREVIATURAS E SÍMBOLOS

PPGFA	Programa de Pós-Graduação em Física Ambiental
UFMT	Universidade Federal de Mato Grosso
LBA	Experimento de Grande Escala da Biosfera - Atmosfera na Amazônia
Z	Temperatura
t	Tempo
ARIMA	Modelo Autorregressivo integrado de Média Moveis
Y	Valor da variável
T	Valor da componente tendência
C	Valor da componente t
S	Valor da componente Sazonal
E	Erro ou variação aleatória
K	Número de classes
N	Tamanho da série
CO ₂	Gás Carbonico
H ₂ O	Água
Δ	Delta

RESUMO

RIHBANE, F, E, C. ***Preenchimento de Falhas Aleatórias de Séries Temporais Micrometeorológicas pela Técnica de Monte Carlo***. Cuiabá, 2014. 48p. Dissertação (Mestrado) – Programa de Pós-graduação em Física Ambiental, Universidade Federal de Mato Grosso. Cuiabá. 2014

O uso de técnicas para preenchimento de falhas tem se desenvolvido fortemente nos últimos anos na área das ciências ambientais como apoio às técnicas de modelagem, com objetivo principal de conhecer a dinâmica da interação atmosfera-biosfera. Neste trabalho foi desenvolvida uma técnica que objetiva preservar a variabilidade das séries temporais micrometeorológicas em seu preenchimento. Essa é uma característica inovadora, dado que a literatura consultada não registra o uso de técnica que se preocupe em preservar o grau de variabilidade, isto é, em técnicas anteriores a variabilidade é "desprezada". O preenchimento é automatizado por meio de um programa computacional ancorado em métodos estatísticos, que empregam a técnica de frequência, Monte Carlo e *Bootstrap*. Testes de correlação, com coeficiente de regressão da ordem de 0.95, mostram que a série artificialmente falhada e preenchida por essa técnica preserva suas características até a presença de 70% de falhas aleatórias.

Palavras-chave: Monte Carlo, Preenchimento de falhas, Séries temporais.

ABSTRACT

RIHBANE, F, E, C. Gap filling random time séries micrometeorological technique Monte Carlo. Cuiabá, 2014. 48p. Dissertation (Master's in Environment Physics)-Institute of Physics, Federal University of Mato Grosso Cuiabá. 2014.

The use of techniques for gap filling has strongly developed in recent years in the environmental sciences abreast of modeling techniques with the main objective to know the dynamics of the atmosphere-biosphere interactions. In this work we developed a technique that aims to preserve the variability of micrometeorological time series in its filling. This is a novel feature, since there is no record in the literature, that worry in to preserve the variability degree, namely the variability in the prior art is "discarded." The filling is automated by a computer program anchored on statistical methods, which employs the use of frequency, Monte Carlo and Bootstrap techniques. Correlation tests, with a regression coefficient of around 0.95, show that artificially failed series and completed by this technique preserves its characteristics to the presence of 70% of random failures.

Keywords: Monte Carlo, Gap filling, Time series.

1. INTRODUÇÃO

1.1. PROBLEMÁTICA

O estudo de séries temporais é importante para se compreender a dinâmica das variáveis micrometeorológicas que descrevem ecossistemas. Entretanto, o monitoramento e a coleta de dados micrometeorológicos são realizados com equipamentos eletrônicos, data loggers, sensores, instalados em diferentes tipos de ambientes. Isso pode ocasionar falhas na coleta de dados por curtos ou longos períodos de tempo, devido a falhas do equipamento ou intempéries da natureza.

Os dados coletados por estações micrometeorológicas são utilizados, geralmente, em estudos em que as falhas de dados na série temporal, dependendo do objetivo, impossibilitam sua manipulação. Então, adota-se alguma técnica para preencher os dados faltosos, para que posteriormente possam ser trabalhados.

Existem diferentes técnicas para executar os preenchimentos das falhas, como uso de interpolação, geoestatística, redes neurais, algoritmos genéticos, regressão linear, entre outras. A escolha depende da necessidade e do tipo dos dados a serem trabalhados, porém essas técnicas têm uma característica comum: realizam o preenchimento com seus valores de tendência central, mediana ou média, por exemplo, de modo que a maioria dos métodos despreza a variabilidade das séries temporais.

Deste modo, há a necessidade do desenvolvimento de técnica para preenchimento de falhas de maneira a preservar a variabilidade de séries temporais. A fonte da variabilidade dos dados pode ser intrínseca à variável ou relacionada às técnicas de medida. Preencher um dado, sem considerar a aleatoriedade, leva à redução das medidas de variabilidade. Uma possibilidade é introduzir a variabilidade aplicando-se técnicas estatísticas combinadas: de Monte Carlo, Bootstrap e da Frequência.

1.2. JUSTIFICATIVA

A obtenção de dados meteorológicos é de muita importância para os estudos climáticos. Com base neles, têm-se as informações necessárias, para se realizar pesquisas.

Especificamente, no Programa de Pós-Graduação em Física Ambiental da Universidade Federal de Mato Grosso, (PPGFA-UFMT) muitas pesquisas que já foram realizadas e outras, que estão em andamento, utilizam dados micrometeorológicos que são coletados em torres instaladas com equipamentos eletrônicos dotados de data loggers, sensores, etc. Muitas vezes, devido a motivos diversos, esses dados que estão sendo coletados falham e, para a maioria das técnicas de estudo dos pesquisadores do programa, se faz necessário preencher a série. Neste sentido, este estudo se insere no projeto “Experimento de Grande Escala da Biosfera - Atmosfera na Amazônia (LBA)”.

Técnicas de modelagem muitas vezes exigem séries não falhadas, no entanto, as técnicas de preenchimento atualmente desenvolvidas, reduzem as medidas de variabilidade e, portanto, introduzem falsas informações, nessas condições esse trabalho tem como objetivo principal, desenvolver metodologia de preenchimento de falhas em séries temporais utilizando as técnicas de frequência, bootstrap e Monte Carlo, afim de manter a mesma média e variância.

A introdução de tendências aos dados no preenchimento de falhas tem levado muitos pesquisados a preferir descartar trechos de séries que contenham falhas, reduzindo a amostra, desprezando assim informações importantes do comportamento desses ecossistemas. As técnicas aqui utilizadas buscam minimizar a introdução de tendências, visando preservar a medida de variabilidade. Foram realizadas análises de correlação e teste f entre a série original e a submetida ao preenchimento de falhas artificialmente provocadas, houve a necessidade de implementar um sistema utilizando a plataforma Java que operacionalize o processo de preenchimento de falhas. Essas análises mostram bom ajuste entre as duas séries original e preenchida.

2. REVISÃO DA LITERATURA

2.1. SÉRIES TEMPORAIS

Séries temporais de dados registram informações de vários campos do conhecimento como Economia (preços diários de ações, taxa mensal de desemprego, produção industrial), Medicina (eletrocardiograma, eletroencefalograma), Epidemiologia (número mensal de novos casos de meningite), Meteorologia (precipitação pluviométrica, temperatura diária, velocidade do vento), etc.

Segundo Ehlers (2009), uma série temporal é uma coleção de observações feitas sequencialmente ao longo do tempo. Nesse sentido, Savi (2006) define uma série temporal como sendo um conjunto discreto numerável de valores de uma variável de estado de um sistema dinâmico. Pode-se entender então que série temporal é um termo genérico; ela descreve comportamento dos dados de características e áreas de estudo diferentes, por exemplo, dados diários de temperatura, umidade, radiação solar, bolsa de valores, entre outros. Morettin & Toloi (2006) afirmam que obtida a série temporal $Z(t_1)$, onde Z é temperatura e t é tempo, tem-se $Z(t_1), \dots, Z(t_n)$, observada nos instantes $t_1, \dots, t_n$. Os objetivos da análise de séries temporais são:

- a) Investigar o mecanismo gerador da série temporal;
- b) Fazer previsões de valores futuros, a curto ou longo prazo, com base em valores passados. Aqui, assume-se que o futuro envolve incerteza, ou seja, as previsões não são perfeitas. Porém, deve-se tentar reduzir os erros de previsão;
- c) Descrever apenas o comportamento da série; o padrão de tendência, existência de variação sazonal ou cíclica, observações discrepantes (*outliers*);
- d) Procurar periodicidade relevante nos dados, análise espectral.

A Figura 1 é um exemplo de séries temporais de diferentes tipos de variáveis.

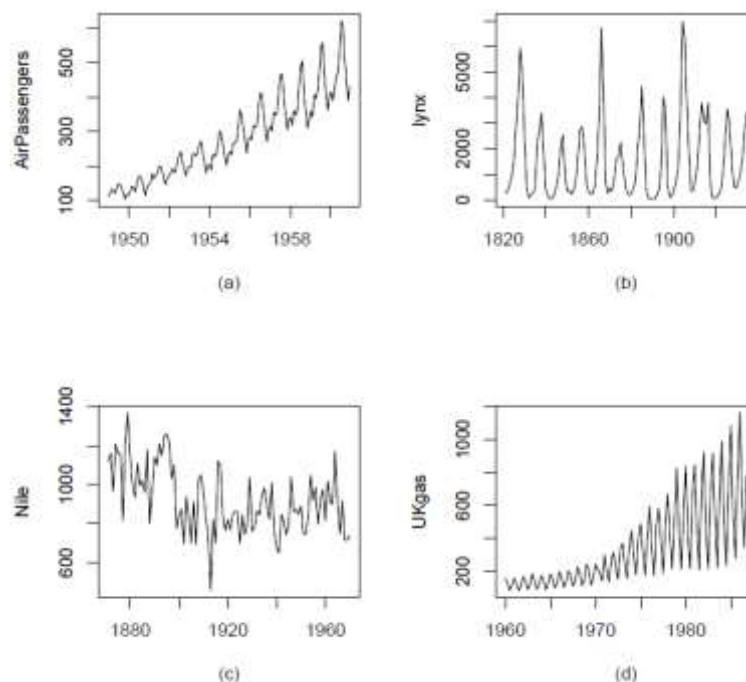


Figura 1: (a) Totais mensais de passageiros em linhas aéreas internacionais nos EUA, (b) Número anual de linces capturados em armadilhas no Canadá, (c) Medições anuais de vazões do Rio Nilo em Ashwan, (d) Consumo de gás no Reino Unido no primeiro trimestre de cada ano.(EHLERS,2009).

Em séries temporais, há existência de dependência do valor em um instante t qualquer com as observações vizinhas, em modelos de regressão, diferentemente, a ordem dos dados não tem importância para a análise.

As séries temporais podem ser classificadas basicamente em três tipos: discreta, contínua e multivariada (EHLERS, 2009).

2.1.1. Série Temporal Discreta e Contínua

Uma série temporal é dita discreta quando as observações são feitas em instantes de tempos específicos, usualmente igualmente espaçados (Ex: temperatura média mensal) (EHLERS, 2009).

Definindo o conjunto $T = \{t_1, \dots, t_n\}$ a série temporal será denotada por $\{X_t: t \in T\}$. Por simplicidade pode-se fazer $T = \{1, 2, \dots, n\}$.

Uma série temporal é dita ser contínua quando as observações são feitas continuamente no tempo. Exemplo: medidas de temperatura a partir de um termógrafo (EHLERS, 2009).

As séries temporais discretas podem surgir de várias formas, por exemplo, médias de valores diários de dados de temperatura do ar, umidade relativa do ar. Se

os valores são dados de tempo em tempo, a série é discreta. Já a informação que pode ser obtida a qualquer instante pode formar uma série contínua (por exemplo, uma linha em um termógrafo marca os valores da temperatura de modo contínuo).

Portanto, somente equipamentos analógicos permitem obter séries contínuas de dados. Séries de valores agregados ou acumulados em intervalos de tempo, por exemplo, quantidade de chuva medida diariamente.

2.2. MODELOS DE SÉRIES TEMPORAIS

Segundo Morettin & Toloí (2006), as séries temporais podem ser caracterizadas ou descritas de acordo com os modelos. O processo de medida controlado por leis probabilísticas são chamados de estocásticos. Tem-se várias classificações dos processos estocásticos, resultando em grande quantidade de modelos diferentes que descrevem uma série temporal específica. Os modelos podem ser construídos em função de vários fatores tais como comportamento do fenômeno ou conhecimento empírico da natureza e do objeto de análise conforme descrito no item 2.1 (série temporal). A descrição desses modelos pode contar com o auxílio de métodos de estimação e da disponibilidade de software adequado.

Segundo Albuquerque et al (2008), é importante notar que o processo considera situações em que são feitas observações quanto a um período de tempo, situações essas influenciadas por efeitos aleatórios ou de azar, não só em um único instante, mas por todo o intervalo de tempo ou sequência de tempos que se está a considerar. Em termos gerais, processo estocástico é o fenômeno que varia em algum grau, de forma imprevisível, à medida que o tempo passa (CLARKE & DISNEY (1979).

2.3. TIPOS DE MODELOS

Os modelos são utilizados para descrever e caracterizar as séries temporais. Morettin & Toloí (2006) afirmam que existem modelos para séries temporais classificados em paramétricos e não paramétricos, segundo o número de parâmetros envolvidos.

2.3.1. Modelo Paramétrico

Segundo Bezerra (2006), existem dois enfoques utilizados na análise de séries temporais. Em ambos, o objetivo é construir modelos para essas séries. No primeiro enfoque, a análise é feita no domínio temporal e os modelos propostos são modelos paramétricos. Morettin & Toloí (2006), os definem como sendo um número finito de parâmetros. Dentre os modelos paramétricos tem-se, por exemplo, o Modelo Autorregressivo integrado de Médias Móveis (ARIMA). É utilizado para realizar modelagem e previsões de séries temporais. Os valores da variável estudada devem ter distribuição normal, outra suposição é apresentar homogeneidade de variância o que normalmente não ocorre (JACQUES, 2003).

2.3.2. Modelo Não Paramétrico

Segundo Bezerra (2006), no domínio de frequências tem-se a análise espectral, com inúmeras aplicações em ciências físicas e engenharia, e que consiste em decompor a série dada em componentes de frequências. Morettin & Toloí (2006) os definem como séries contendo um número infinito de parâmetros, que como normalmente não se tem distribuição normal e homogeneidade, se encaixam nesse modelo (JACQUES, 2003).

2.4. CARACTERÍSTICAS DAS SÉRIES TEMPORAIS

De acordo com Oliveira & Favero (2002) são componentes periódicos das séries temporais: tendência, sazonalidade e ciclos. Segundo Morettin & Toloí (2006), para formar uma série temporal, somam-se os componentes periódicos isolados e a variação aleatória, de acordo com a equação 1:

$$Y_t = T_t + C_t + S_t + E_t \quad \text{eq. (1)}$$

em que Y_t é o valor da variável em um instante de tempo dado, T_t é o valor da componente de tendência em um tempo dado, C_t é o valor da componente de ciclo, S_t é o valor da componente sazonal e E_t é o erro ou a variação aleatória.

A diferença entre o valor teórico e os valor medido é a camada de erro da medida, podendo ser causada por diversos motivos, tais como fatores ambientais, calibração, picos de luzes, equipamento com defeito, dentre outros (SOUZA & SOBRINHO, 2005).

Erros determinados por calibração interferem na definição do resultado final. Erros podem ter diversas origens: histerese, falta de repetibilidade, falta de linearidade, entre outras. Designa-se por exatidão, a qualidade que caracteriza a capacidade de um sensor fornecer resultados próximos do valor real da quantidade medida. Um sensor será tão mais exato quanto menor for a diferença entre o valor medido e o verdadeiro. (SOUZA & SOBRINHO, 2005).

2.4.1. Tendência

De acordo com os autores Morettin & Tolo (2006) pode-se conceituar tendência como a descrição de um movimento suave dos dados, em longo prazo, crescente, decrescente ou estável, pelo qual identifica-se a velocidade com que essas mudanças acontecem. As tendências podem estar relacionadas ao crescimento econômico de um país, ao aumento das temperaturas devido ao efeito do aquecimento global ou local, entre outros.

Morettin & Tolo (2006) verificaram que as tendências, independentemente do tipo de variável, podem ser identificadas com testes estatísticos como médias móveis, utilizadas para analisar dados calculando o valor médio em um intervalo de tempo ou período; teste de sequências (Wald-Wolfowitz), para amostras independentes; teste do sinal (Cox-Stuart), para amostras emparelhadas, etc.

Marengo & Alves (2005) realizaram estudos para verificar a tendência da vazão do rio da Paraíba do Sul utilizando o método Mann-Kendall, para testar a hipótese: as observações são independentes e identicamente distribuída ou possuem tendência. Os autores verificaram que, de modo geral, houve uma tendência negativa, isto é, com o passar dos anos a vazão do rio vem diminuindo e isso pode ter sido causado por influências climáticas ou antrópicas.

Um outro exemplo: Blain et al (2009), ao analisarem séries anuais da temperatura mínima do ar no Estado de São Paulo, observaram tendência de elevação dessa variável climática. Os autores apontam como causa para essa tendência o fato

de fatores de escala local parecerem se sobrepor a possíveis fatores de escala global, como principais forçantes radioativas no aumento médio dos valores dessa variável meteorológica.

2.4.2. Sazonalidade & Ciclos

Segundo os autores Morettin & Toloi (2006), os ciclos são variações periódicas em torno de uma linha de tendência. Um ciclo com período anual é chamado de ciclo sazonal. É exemplo de ciclo com período maior que um ano o ciclo de atividade solar responsável pelas manchas solares, de aproximadamente 11 anos.

A grande maioria das séries temporais de variáveis meteorológicas exhibe variações com período anual, que são exemplos de sazonalidade. A temperatura é usualmente maior no verão que no inverno. Essas características são encontradas na região do cerrado, e, em algumas localidades, a precipitação possui um ciclo sazonal bem definido. O efeito sazonal pode ser aditivo ou multiplicativo.

Os fenômenos sazonais estão associados às estações do ano, como exemplo de eventos sazonais pode-se citar a variação da temperatura ao longo do ano, o consumo de artigos de estação, tais como sorvetes e ovos de páscoa, entre outros.

2.5. SÉRIES E FALHAS

A PPGFA possui inúmeros trabalhos que utilizam séries temporais de diversas variáveis, entretanto apenas em alguns desses trabalhos as falhas são tratadas. Isso faz com que o pesquisador escolha medidas alternativas quanto ao tratamento dos dados, serão abordados alguns destes trabalhos que utilizaram série temporal.

Capistrano (2007) analisou séries temporais de variáveis microclimatológicas utilizando a teoria da complexidade em uma floresta de transição cerrado e floresta tropical úmida, com dados coletados de 2001 a 2006, três meses de dados pertencente às estações seca ou chuvosa. O objetivo foi a caracterização do atrator das séries temporais a partir das variáveis de fluxo de calor latente e sensível, fluxo de CO₂, temperatura, concentração de CO₂ e de H₂O, velocidade do vento e

velocidade de fricção. Os dados retirados foram substituídos por dados repetidos anteriores. Nota-se, que no estudo da teoria da complexidade, essa adequação é permitida devido não sofrer alteração no estudo da complexidade das variáveis.

Mariano (2008), com o trabalho "Análise Espectral de Séries Temporais de Variáveis Micro - Climatológicas em uma Área de Ecótono Entre os Biomas Amazônia e Cerrado no Norte de Mato Grosso", realizou estudo na região entre os biomas tropical úmido e cerrado. Os dados coletados por correlação de vórtices turbulentos (*Eddy Covariance*) e analisados foram os de temperatura, fluxo de calor sensível e fluxo de calor latente no período entre os anos 1999 e 2006, tendo então seis séries para estação úmida e seis séries para estação seca. As séries foram organizadas com o máximo de dados sem ocorrência de falhas, as condições de campo e de manutenção geram falhas na medição ou armazenamento de horas, dias, limitando a quantidade de dados, o que ocorreu nos anos de 2000 e 2004. Assim descartando as informações, para processar os dados, não poderiam existir falhas.

Gallon (2005), em sua dissertação com tema "Um Estudo Sobre a Dinâmica de Sistemas Complexos a Partir de Séries Temporais de Dados Microclimatológicos para uma Floresta de Transição no Noroeste de Mato Grosso", trabalhou com série de dados do LBA de 1999 a 2003, analisando fluxo de CO₂, calor latente e calor sensível, estes dados apresentaram falhas com período superior a um dia. Quando ocorreu falhas com período de ordem de mês, a autora optou por descartar os dados. As análises foram dificultadas para os meses com grande número de dados ausentes.

Palú (2008) com tema de dissertação "Determinação do Tempo de Defasagem mais Adequado para Análise de Séries Temporais de Variáveis Microclimatológicas Medidas numa Floresta de Transição no Norte de Mato Grosso" também utilizando dados da floresta de transição entre o cerrado e tropical úmida coletados no ano de 2002 a 2006, analisou somente os dados referente a estações seca e chuvosa, das variáveis de fluxo de calor latente, fluxo de carbono, velocidade do vento e direção do vento, visto que nos anos de coleta de dados houve muitas falhas em diversos meses, os dados foram ignorados devido a "grande importância o número de dados da série e não o tempo de sua distribuição" (PALÚ, 2008, p. 16). Mas ao trabalhar com os dados verificou também que as falhas ocorridas e ignoradas podem ter alterado os resultados obtidos.

2.6. ERRO DE MEDIÇÃO DO DATA LOGGER

Segundo Instrutemp (2007), o data logger é um dispositivo eletrônico que registra sinais emitidos pelos sensores portáteis. É normalmente pequeno, funciona a pilhas, baterias ou diretamente a uma fonte de energia e, equipado com um microprocessador, uma memória interna para armazenar os dados.

A principal vantagem do data logger é a capacidade de coletar automaticamente e armazenar os dados. Por essa facilidade, esses dispositivos são muito usados em diferentes modalidades como em indústrias; para coletar e armazenar dados para a empresa como produtividade, gastos, temperatura de caldeiras, etc, assim facilitando o trabalho que se torna inviável manualmente, ou até mesmo impossível, dependendo das condições de trabalho.

Também são utilizados em tratores e colheitadeiras mantendo uma base de dados com as informações importantes sobre o equipamento, como produtividade, permite detectar, analisando os dados, algum problema mecânico, e assim corrigi-lo. Portanto, o data logger tem inúmeras utilidades e vantagens com relação ao tempo, economia e eficiência. Atualmente, a automação em geral vem cada vez mais sendo adotada independentemente da aplicação.

Os data loggers são utilizados com sensores, que monitoram processos e são transdutores, pois realizam a conversão de grandezas físicas em grandezas elétricas, utilizados em diversas áreas como já mencionado.

Sensores são dispositivos que mudam seu comportamento sob a ação de uma grandeza física, podendo fornecer diretamente ou indiretamente um sinal que indica o valor dessa grandeza. Atualmente existe uma grande variabilidade de sensores que medem diferentes tipos de grandezas físicas. Os tipos de sensores mais comuns são de pressão, temperatura, luz, velocidade, vazão, posição, entre outros tipos de sensores.

2.7. MÉTODOS PARA PREENCHIMENTO DE FALHAS

Atualmente existe um grande número de métodos de preenchimento de falhas. Muitas vezes em uma série de dados não é necessário utilizar preenchimento de falhas, mas isso varia de acordo com o objetivo do trabalho realizado, tipo de

dados, e tamanho das falhas. Entretanto, quando se faz necessário preencher as falhas, é devido ao tipo de variável ou quantidade de dados falhos e o volume de dados a serem analisados.

Segundo Rihbane et al. (2012) de um modo geral, os modelos necessitam de longas séries temporais sem falhas. Faz-se necessário, portanto, o preenchimento dos dados para que sejam processados e utilizados em pesquisas científicas. Ao se preencher as séries temporais, pode se realizar estimativas e análises de variáveis como carbono líquido, fluxo de calor latente, fluxo de calor sensível e outras que nos permitem conhecer a dinâmica da atmosfera ou de um ecossistema, em diferentes escalas espaciais e temporais. Séries falhadas são, de modo geral, fragmentadas nas técnicas de modelagem, o que obriga a redução de escala nas análises.

Wheelwright (1985) apud Mueller (1996), descreve a possibilidade de extrapolação futura dos dados a partir da identificação das componentes individuais presentes no padrão básico da série temporal de dados (tendência, ciclo, sazonalidade e aleatoriedade). De maneira análoga, há um grande número de técnicas em desenvolvimento para preenchimento de falhas: por meio de métodos de interpolação que utilizam krigagem ordinária (DEUS et al., 2010), métodos geoestatísticos, que realizam preenchimento com regras pré-estabelecidas de acordo com o tempo de falha, combinando interpolação linear e médias (GAIO et al., 2008), métodos de correlação (FALGE et al., 2001), entre outros. Observa-se, no entanto, nesses trabalhos, que a variabilidade dos dados não foi tratada. A fonte da variabilidade dos dados pode ser intrínseca à variável ou relacionada às técnicas de medida. Preencher um dado sem considerar a aleatoriedade leva à redução das medidas de variabilidade como mostrado em Gaio et al. (2007).

2.8. MONTE CARLO

Durante a Segunda Guerra Mundial, cientistas e pesquisadores trabalhavam intensamente em um projeto científico que posteriormente desencadeou um grande avanço tecnológico. Graças a ele tem-se o conhecimento, facilidade do desenvolvimento de tecnologias e equipamentos eletrônicos. Estes pesquisadores trabalhavam no projeto ENIAC que era o primeiro computador eletrônico a ser

desenvolvido. Durante sua pesquisa e desenvolvimento, John Von Neumann descobriu que os circuitos podiam contar, fazer operações matemáticas, e resolver equações e cálculos em uma velocidade muito alta.

John Von Neumann era um consultor de matemática que, ao se deparar com a descoberta, se interessou pela área termonuclear que veio a desencadear o desenvolvimento da bomba atômica.

De acordo com Metropolis & Ulam (1949), denominou-se Método Monte Carlo por fazer referência à famosa cidade de Mônaco, conhecida mundialmente como a capital dos jogos de azar, pelo fato de Ulam e Von Neumann frequentarem e aplicarem matemática nos jogos de azar, observando aleatoriedade empírica. O método baseia-se em eventos que ocorrem de forma aleatória. Segundo Yoriyaz, (2009), pode ser descrito como método estatístico no qual utiliza-se uma sequência de números aleatórios para realização de simulações.

Segundo Metropolis (1987), é um método antigo com forte embasamento teórico e prático, que tem como função gerar números aleatórios. Esses números aleatórios darão suporte à inserção da variabilidade dos dados calculados para serem preenchidos.

Segundo Metropolis (1987), o método permite simular qualquer processo que dependa de aleatoriedade e também problemas que não tenham relação com a aleatoriedade, e criar modelos probabilísticos para resolvê-los.

A técnica de Monte Carlo basicamente é a geração de valores aleatórios, com o objetivo de produzir "n" quantidades de cenários, em que "a distribuição dos valores calculados deve refletir a probabilidade de ocorrência dos mesmos". (SOUZA, 2006. p. 67).

De acordo com Souza (2006), existem algumas vantagens em utilizar a simulação de Monte Carlo, tais como:

- a) Não precisam ser aproximadas as distribuições das variáveis do modelo;
- b) Podem ser modeladas as correlações e outras interdependências;
- c) O computador é responsável por realizar a geração dos valores aleatórios;
- d) Aumentando o número de interações é possível aumentar o nível de precisão da simulação;

- e) Monte Carlo por ser uma simulação bastante reconhecida, tem uma aceitação considerável;
- f) O modelo pode ser alterado facilmente, em que gera novos resultados podendo ser comparados com os anteriores.

2.8.1. Aleatoriedade

Segundo Brinder & Heermann (1997), o método de Monte Carlo é responsável pela aleatoriedade nos modelos de simulação. Consegue gerar dados novos que se assemelham aos dados da natureza. Permitindo simular modelo, por mais tempo do que o período usado na coleta de dados.

Para gerar números aleatórios é necessário inicialmente que sejam geradas sequências uniformemente distribuídas no intervalo $[0,1]$. Os números "aleatórios" podem ter três tipos de classificações:

- a) - **Números Aleatórios:** a principal característica desses números é o fato de serem completamente imprevisíveis. Esses só podem ser gerados por algum processo físico natural.
- b) - **Números Pseudoaleatórios:** São os mais utilizados, sendo gerados em computador por meio de algum algoritmo simples. Com isso, não são extremamente números aleatórios, no sentido estrito do termo.
- c) - **Números Quasi-aleatórios:** também são gerados por algum algoritmo numérico simples. Contudo, eles são produzidos de maneira que sejam distribuídos o mais uniformemente possível.

2.8.2. Bootstrap

De acordo com Barros (2005), em 1979 foi publicado o artigo *Annals of Statistics*, proposto por Bradley Efron, que apresentava o método de simulação Bootstrap. A técnica se resume basicamente em construção de simulações por reamostragem a partir de uma amostra inicial. É bastante comum aplicar o método para procedimentos estatísticos em séries temporais. O termo *Bootstrap* refere-se às

simulações Monte Carlo que tratam a amostra original como a pseudo-população. Então reamostragens são feitas a partir da amostra original.

O método Bootstrap pode ser construído a partir de dois pressupostos: Bootstrap paramétrico, que simula amostras a partir de uma distribuição conhecida, podendo conter um ou mais parâmetros desconhecidos, e Bootstrap não paramétrico, amostras obtidas com reposição dos dados originais repetindo o processo n vezes. Nesse último caso, os dados não são gerados a partir da distribuição de probabilidade como no método paramétrico (Barros 2005). Segundo Costa (2006, p. 42), "o verdadeiro poder da reamostragem vem de amostragem com reposição".

3. MATERIAL E MÉTODOS

Este trabalho utiliza algoritmos computacionais no desenvolvimento de metodologia de preenchimento de falhas utilizando os conceitos, Bootstrap, Monte Carlo e cálculo de frequência.

3.1. DADOS DE ESTUDO

Os dados utilizados para este estudo foram coletados em uma estação micrometeorológica localizada em Sinop - MT. O local está a 423 m acima do nível do mar nas coordenadas geográficas $11^{\circ}24'43,4''\text{S}$ e $55^{\circ}19'25,7''\text{O}$, Figura 2.



Figura 2: Localização da área de estudo. (SANCHES et al., 2006).

Utilizou-se de séries temporais de temperatura, da umidade relativa do ar e do fluxo de calor no solo, provenientes de uma torre micrometeorológica de 42 m de altura, instalada a 50 km do município de Sinop, Mato Grosso, Brasil. Foram

fornecidos dados de umidade relativa do ar, temperatura do ar, radiação fotossinteticamente ativa, radiação solar global, fluxo de calor no solo, fluxo de calor sensível e latente, velocidade do vento, entre outras variáveis. Como a maior parte dos procedimentos estatísticos foi desenvolvida para analisar observações independentes, o estudo de séries temporais requer o uso de técnicas específicas.

Neste trabalho foram utilizadas as variáveis temperatura, umidade relativa do ar e fluxo de calor no solo, essas três variáveis ocorreram falhas nos dados reais de forma pareadas, e as outras variáveis existentes essa característica não ocorreu. Os dados desta estação originaram diversos trabalhos de pesquisas como teses, dissertações e artigos no Programa de Pós-Graduação em Física Ambiental da UFMT.

Os dados utilizados correspondem a um ano, de janeiro a dezembro de 2007, os valores utilizados são médias de 30 minutos gravados em data loggers CR10X (Campbell Scientific, Logan, UT, USA). Entretanto, havia dados com falhas, aproximadamente 57% e utilizados somente os dados existentes para realizar os testes, aproximadamente 43% que equivale aproximadamente 5 meses de dados. Os testes estatísticos foram aplicados a quinta simulação, esta simulação foi escolhida aleatoriamente, tendo em vista que as nove simulações apresentam a mesma variação do percentil.

3.2. SISTEMA DE GERAÇÃO DE FALHAS PAREADAS

Muitos dados podem possuir falhas, porém, é inviável preencher as séries a fim de verificar se o método é consistente, pois não há como saber qual era o valor original para compará-los com os dados preenchidos pelo método desenvolvido, quaisquer que sejam. De modo a estimar a qualidade do preenchimento, foram utilizadas séries originais sem falhas, as quais foram artificialmente falhadas de forma pareadas por meio de algoritmo desenvolvido em linguagem Java. O motivo da escolha dessa linguagem foi o fato de que é de alto nível, multiplataforma e tem muitos recursos de programação disponíveis.

Para todos os processos, foi utilizado um notebook com as seguintes configurações:

Intel(R) Core(TM) i7-2670QM, CPU 2.20GHz, com 8 GB de memória RAM (7.90 GB utilizável), GPU (placa de vídeo) GeForce GT 540M (2 GB), Foram geradas falhas artificiais aleatórias de 1% a 70% com passo de 1%, para essa base de dados de forma pareadas.

3.3. DISTRIBUIÇÃO DE FREQUÊNCIAS

A aplicação computacional desenvolvida para o preenchimento das falhas adotou um cálculo estatístico chamado distribuição de frequência para organizar os dados em classes de modo a fornecer a quantidade e/ou porcentagem de dados de cada classe.

Primeiramente, os dados foram separados e organizados de acordo com as respectivas horas, por exemplo, seleciona-se todos os dados das 10h00 dos dias Juliano 1 ao 365. Realiza-se esse processo para todas as 24 horas do dia. O algoritmo de preenchimento de falhas em seguida contabiliza quantos dados existem em determinada hora e verifica os valores máximo e mínimo para cada hora contabilizada.

À variação da grandeza medida entre os intervalos de tempo deu-se o nome de delta (Δ). O delta representa a diferença entre duas medidas subsequentes, por

exemplo, diferença entre a temperatura das 10:30 e das 10:00. O algoritmo calcula cada delta para todos os intervalos de tempo.

Para determinar a quantidade de classes foi utilizada a regra de Sturges (regra do logaritmo) (FALCO, 2008), aplicado a identificação e contagem dos delta, representada pela equação 2:

$$K = 1 + 3,3 * \log(N) \quad \text{eq. (2)}$$

em que K é número de classes e N é o tamanho da série, assim, aplicando a equação 2, para N= 365 obtém-se K= 9 classes.

Definida a quantidade de classes, realizou-se a distribuição dos dados para cada hora determinada de acordo com as classes. Determina-se então as frequências de cada classe. O algoritmo de preenchimento de falhas, identifica todas as falhas nos arquivos ao considerar o tipo do dado, hora e dia. Assim aplica-se o cálculo da distribuição de frequência para as respectivas horas que correspondem às falhas. A tabela 1 representa um exemplo da distribuição de frequências para uma determinada hora.

3.4. BOOTSTRAP & MONTE CARLO

Com os dados de frequências, realiza-se um sorteio, para cada classe, em que 100 valores aleatórios são sorteados, com reposição. Em seguida, desses 100 números sorteados em cada classe, são selecionados aleatoriamente um número de valores de acordo com a porcentagem da classe. Tem-se agora um único conjunto de 100 números com as mesmas frequências das classes originais. Faz-se a média desses números e armazena-se esse valor gerado que vamos chamar de n1. Repete-se esse procedimento 100 vezes até obter cem valores, nomeados: n1, n2, ..., n99, n100. Aplica-se então a média novamente de n1 a n100, obtendo-se um único valor, o qual é o dado que preenche a falha. A Figura 3 mostra o fluxograma das etapas desse processo. O ícone representado pela círculo verde representa o início, o círculo vermelho significa fim do processo. A caixa cinza representa a repetição (100 vezes) dos processos das caixas amarelas.

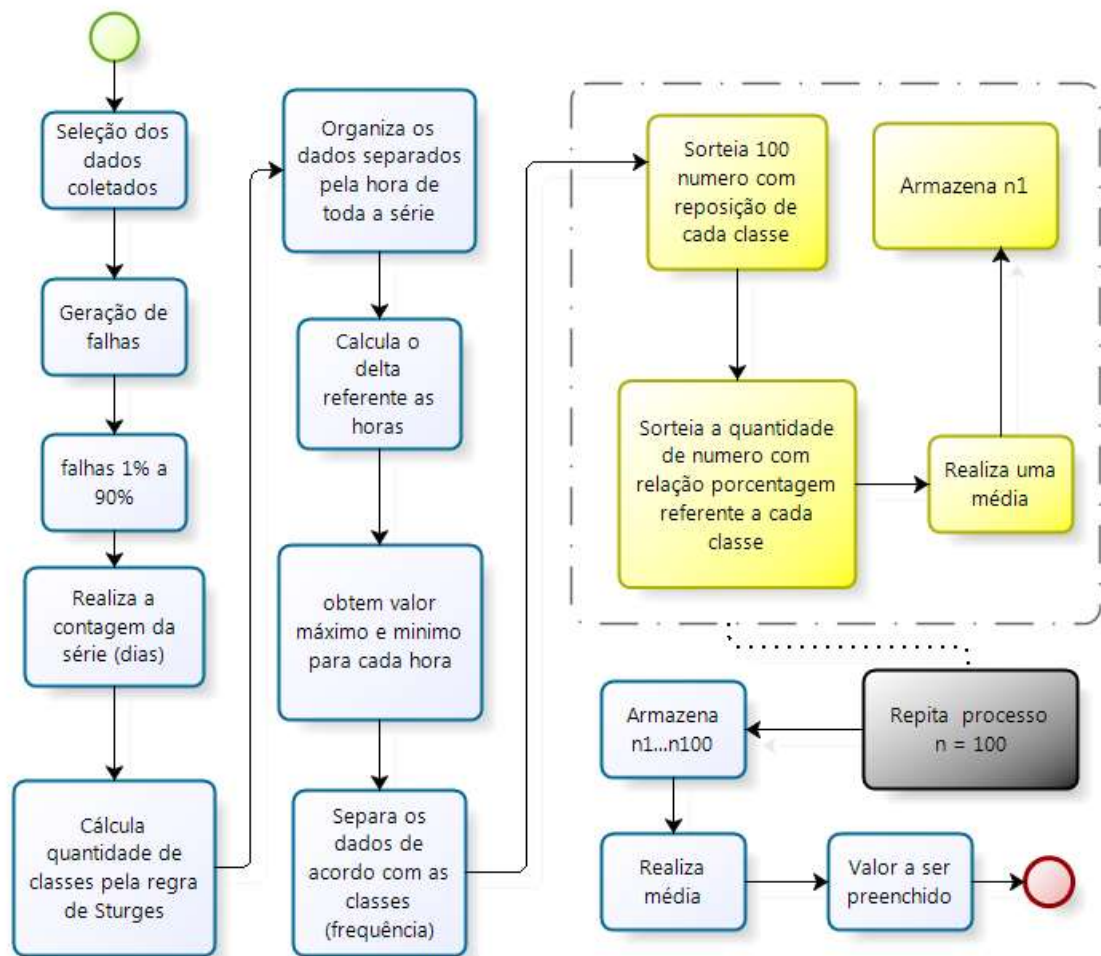


Figura 3: Fluxograma representando os processos do preenchimento das falhas por etapa, círculo verde é o início do processo, círculo vermelha é o fim do processo, as caixas são cada processo individual, os processos em amarelo serão feitos um loop de acordo com o processo de cor preta.

4. RESULTADOS E DISCUSSÕES

O desempenho do método de preenchimento de falhas foi analisado para nove simulações falhadas aleatoriamente, de maneira igual para as variáveis de temperatura do ar, umidade relativa do ar e fluxo de calor no solo. Nas sessões a seguir serão detalhados os resultados apresentados com suas respectivas falhas e suas correlações. São visualizados e analisados os gráficos de 10% a 70% de falhas; com passo de 10%.

Para gerar as falhas aleatórias de 1% a 70%, leva-se em média 37 segundos de processamento. Para todas as variáveis contidas na planilha de dados, o gerador de falhas pode ser alterado de acordo com sua necessidade. Entretanto, para preencher de 1% a 70% de falhas para todas as 9 simulações, gasta-se em média 300 minutos de processamento para preencher todos os dados faltosos da planilha para todas as variáveis.

4.1. ANÁLISE DO PREENCHIMENTO DE FALHAS PARA SÉRIE TEMPORAL DA TEMPERATURA

Nas figuras 4 e 5, com 10% a 70% de falhas, pode-se visualizar os dados originais e os dados preenchidos de temperatura do ar. Na Figura 5 a linha vermelha no gráfico refere-se aos dados preenchidos e a linha azul refere-se aos dados originais. Todos os gráficos de linha contêm uma quantidade reduzida de dados de aproximadamente mil dados, facilitando a visualização e comparação do comportamento das duas séries, entretanto o gráfico de dispersão é referente a todos os dados da série (Figura 4).

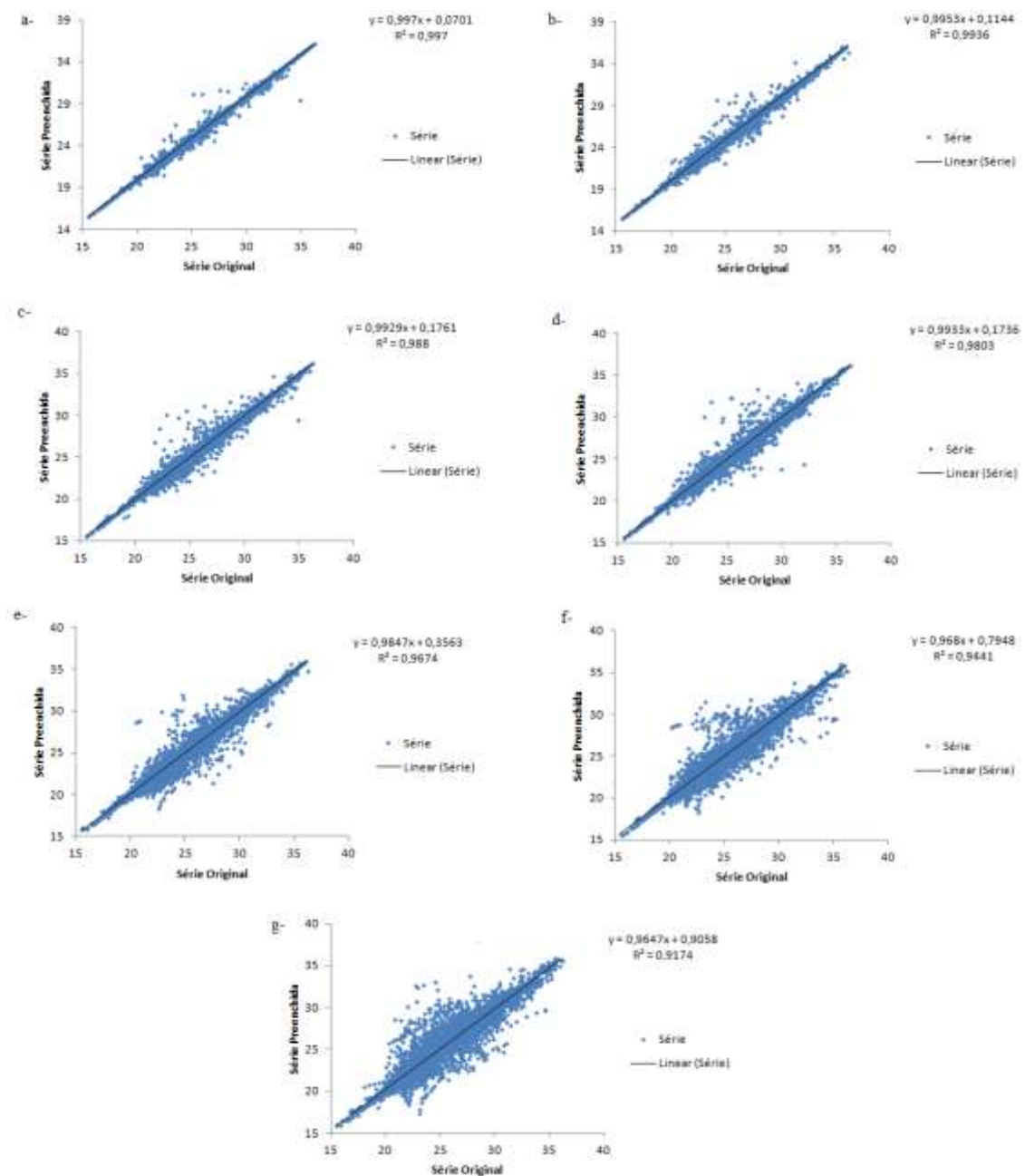


Figura 4: Gráfico de dispersão com valores preenchidos e originais para a série temporal da temperatura do ar correspondente a quinta simulação, (a) 10% falhas, (b) 20% falhas, (c) 30% falhas, (d) 40% falhas, (e) 50% falhas, (f) 60% falhas e (g) 70% falhas.

Para 1% à 30% de falhas aleatórias obteve-se aproximadamente um coeficiente de correlação 0,99. Nota-se que ao aumentar a porcentagem de falhas, a correlação diminui. Entretanto, para a variável temperatura a variação dos coeficientes de correlação não foi muito acentuada. Para 40% de falhas tem-se 0,98, 50% tem-se 0,96, 60% tem-se 0,94 e 70% tem-se 0,91 para o coeficiente de correlação, Figura 4.

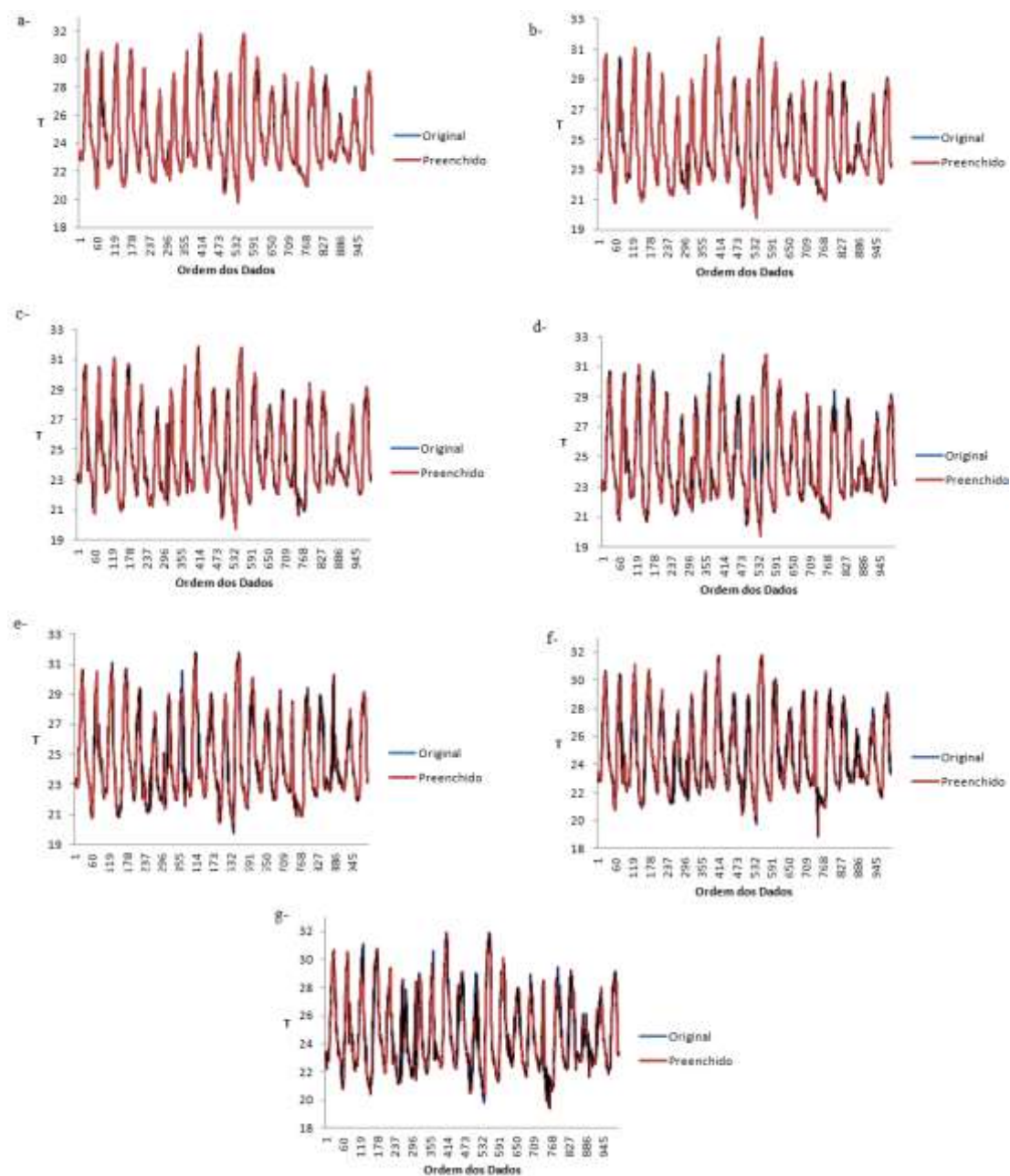


Figura 5: Séries temporais da temperatura do ar referente a quinta simulação, original (em azul) e preenchida (em vermelho), (a) 10% falhas, (b) 20% falhas, (c) 30% falhas, (d) 40% falhas, (e) 50% falhas, (f) 60% falhas e (g) 70% falhas.

Ventura (2012) desenvolveu um programa para preenchimento de falhas baseado em inteligência artificial e obteve os seguintes resultados para variável de temperatura do ar: a simulação com 5% teve correlação de 0,97, a simulação com 10% teve um resultado de aproximadamente 0,95, as simulações de 20% e 30% tiveram um resultado semelhante a 0,93, e a de 40%, com a menor correlação, 0,92,

porém, relativamente próxima às anteriores. Do ponto de vista da correlação, portanto, observam-se resultados semelhantes

4.2. ANÁLISE DO PREENCHIMENTO DE FALHAS PARA SÉRIE TEMPORAL DA UMIDADE RELATIVA DO AR

A figura 6 representa a regressão referente aos dados preenchidos e originais de 10% a 70% de falhas para a variável umidade relativa do ar.

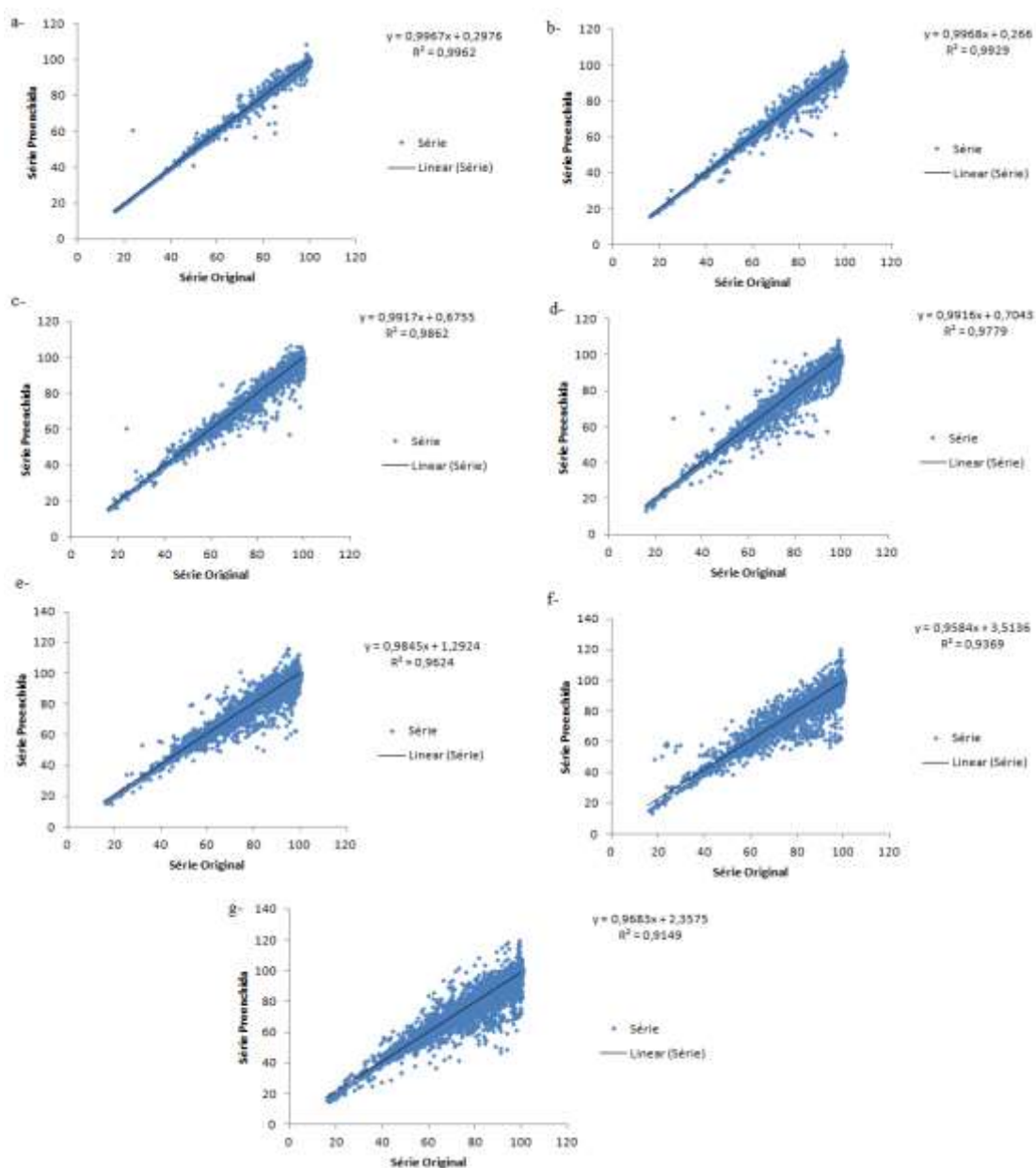


Figura 6: Valores preenchidos e originais para a série temporal da umidade relativa do ar correspondente a quinta simulação, (a) 10% falhas, (b) 20% falhas, (c) 30% falhas, (d) 40% falhas, (e) 50% falhas, (f) 60% falhas e (g) 70% falhas.

Ao analisar a série temporal para umidade relativa do ar (Figura 6), percebe-se que os coeficientes de correlação de acordo com a porcentagem de falha são bons, mas um pouco menores quando comparados aos resultados obtidos com a série de dados de temperatura do ar. Porém, possuem valores próximos: 0,99 para 10% e 20% de falhas, 0,98 para 30% de falhas, 40% tem-se 0,97, 50% tem-se 0,96, 60% tem-se 0,93 e para 90% de falhas tem-se 0,91.

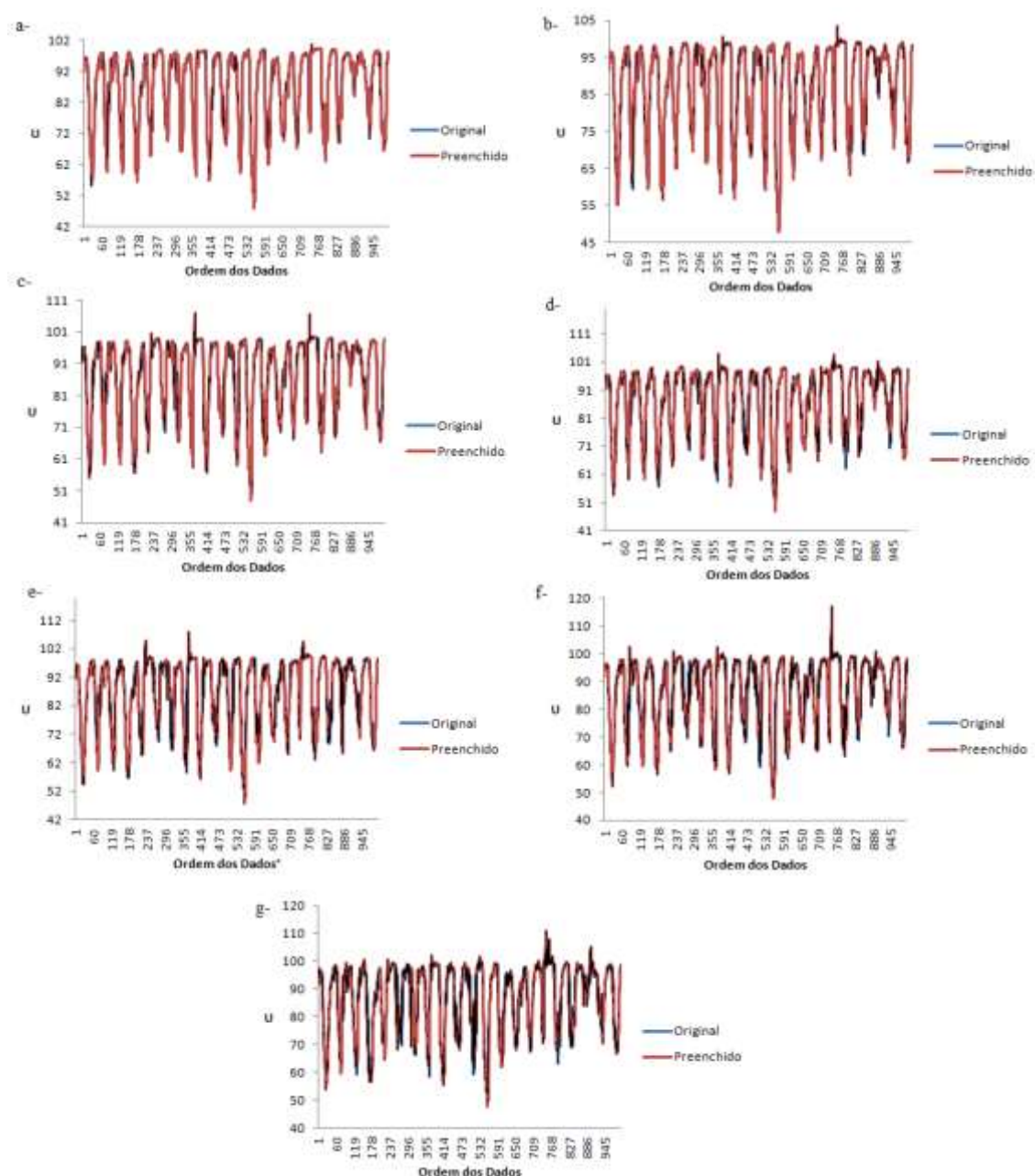
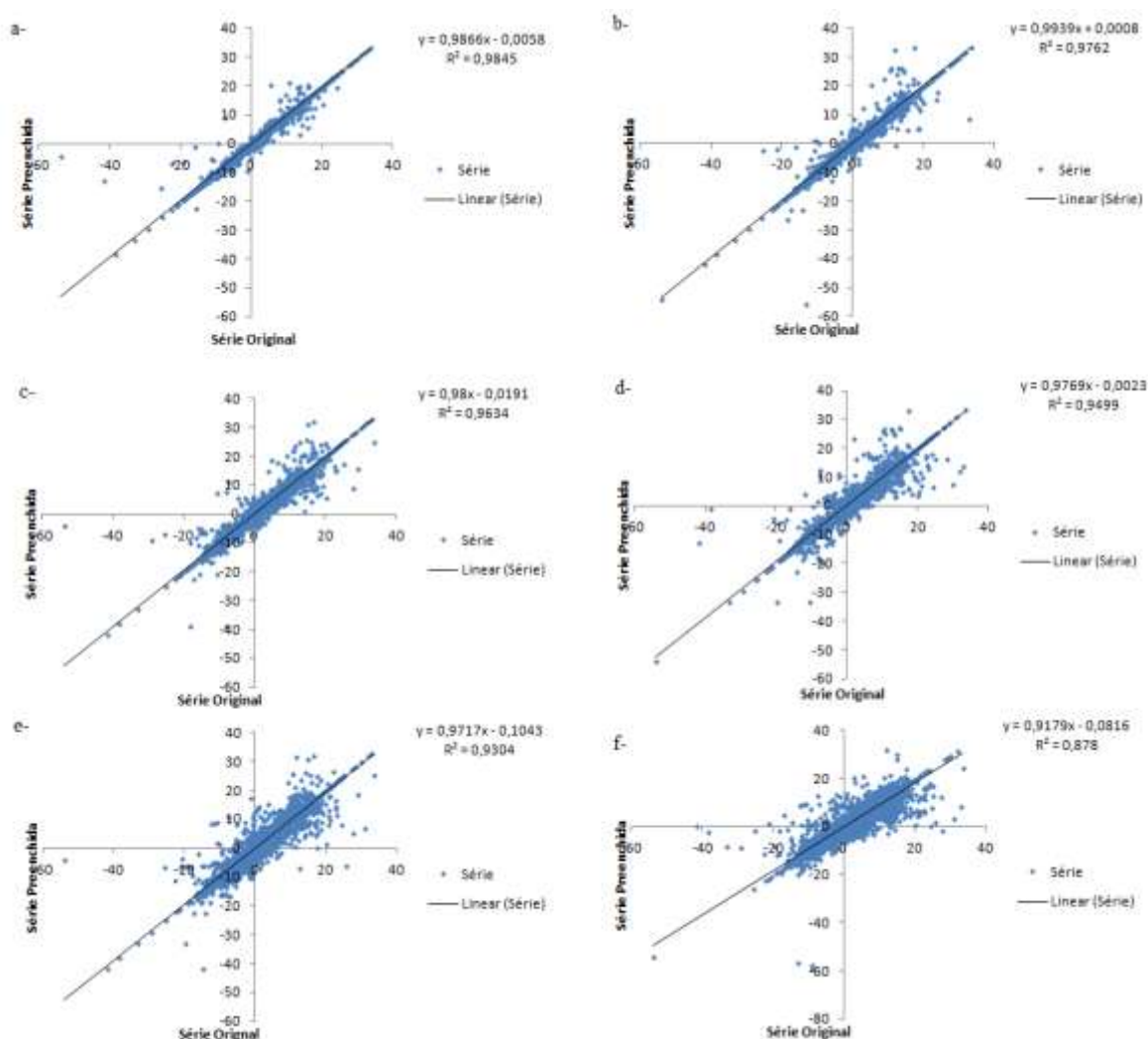


Figura 7: Séries temporais da umidade relativa do ar referente a quinta simulação, original (em azul) e preenchida (em vermelho), (a) 10% falhas, (b) 20% falhas, (c) 30% falhas, (d) 40% falhas, (e) 50% falhas, (f) 60% falhas e (g) 70% falhas.

Comparando novamente com o trabalho de Ventura (2012), observam-se valores semelhantes, esse autor, ao aplicar técnicas de inteligência artificial obteve como resultado para a variável umidade relativa do ar, o coeficiente de correlação de 0,91 para 5% de falhas, 0,93 para 10% de falhas, 0,95 para 20% de falhas, 0,93 e 0,92 para 30% e 40% de falhas respectivamente.

4.3. ANÁLISE DO PREENCHIMENTO DE FALHAS PARA SÉRIE TEMPORAL DE FLUXO DE CALOR NO SOLO

A figura 8 apresenta a regressão referente aos dados preenchidos e originais de 10% a 70% de falhas para a variável fluxo de calor no solo.



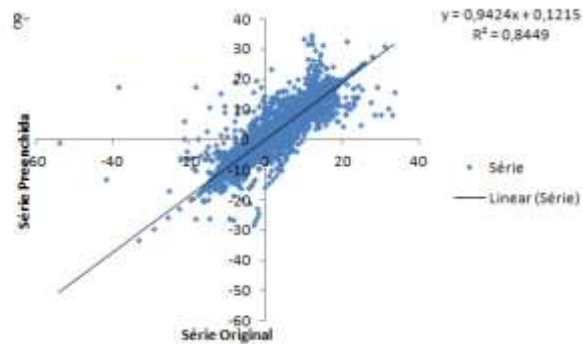
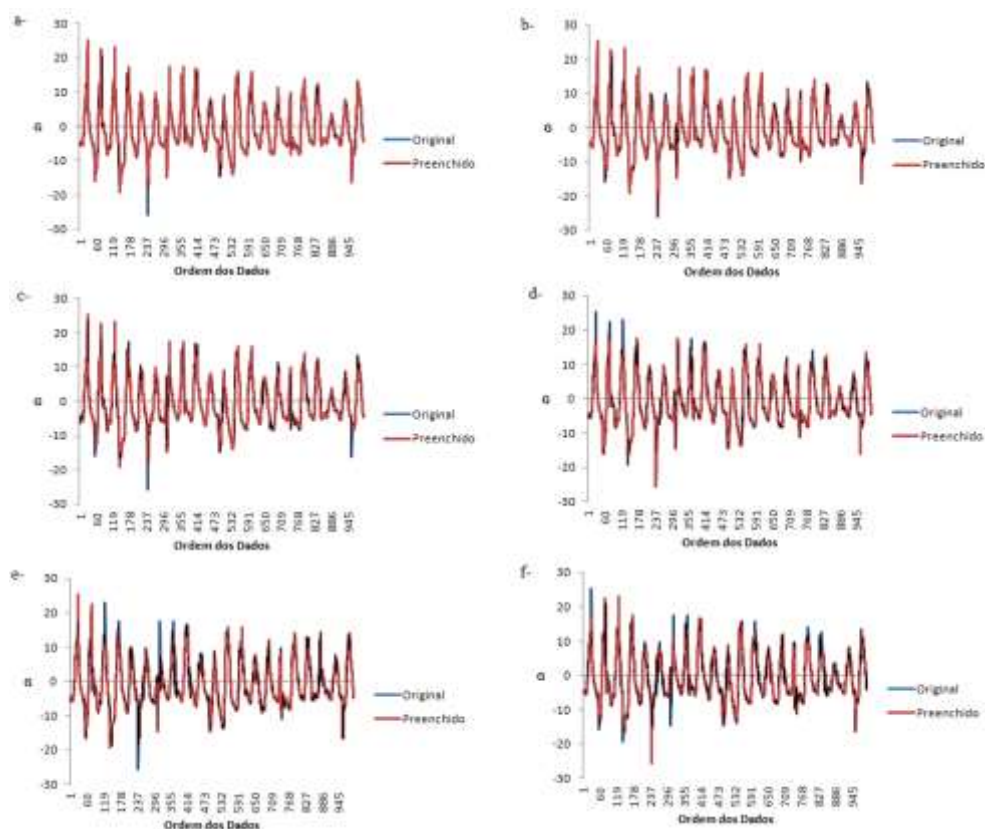


Figura 8: Valores preenchidos e originais para a série temporal do fluxo de calor no solo correspondente a quinta simulação, (a) 10% falhas, (b) 20% falhas, (c) 30% falhas, (d) 40% falhas, (e) 50% falhas, (f) 60% falhas e (g) 70% falhas.

Os resultados obtidos para fluxo de calor no solo foram os que tiveram correlação mais baixa quando comparados às variáveis de temperatura do ar e umidade relativa do ar. Entretanto, conseguiu-se um bom desempenho apesar de ser uma série com maior variabilidade. Obteve-se coeficiente de correlação de 0,98 para 10% e 20% de falhas, 0,96 para 30% de falhas, 0,95 para 40% de falhas, 0,91 para 50% de falhas, 0,88 para 60% de falhas e 0,84 para 70% de falhas, Figura 8.



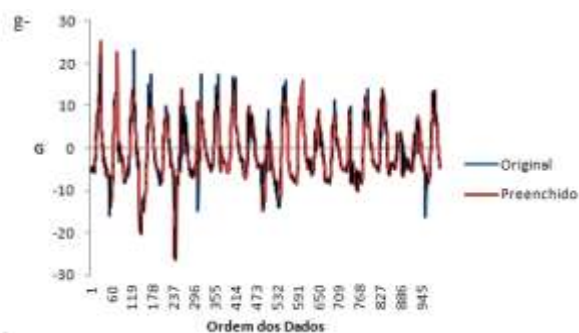


Figura 9: Séries temporais do fluxo calor no solo referente a quinta simulação, original (em azul) e preenchida (em vermelho), (a) 10% falhas, (b) 20% falhas, (c) 30% falhas, (d) 40% falhas, (e) 50% falhas, (f) 60% falhas e (g) 70% falhas.

4.4. ANÁLISE ESTATÍSTICA DAS VARIÁVEIS

O desempenho do sistema foi analisado para nove simulações, com os dados de temperatura do ar, umidade relativa do ar e fluxo de calor no solo. Foram provocadas falhas aleatórias de 1% a 70% de acordo com a quantidade de dados existentes originais. Nove simulações com os mesmos dados foram realizadas visando a repetibilidade da informação para averiguar se, independente da realização de um preenchimento ou mais, a precisão, a média e variância se mantêm.

Os gráficos de dispersão do coeficiente de correlação com relação ao percentual de falhas, das Figuras 10 a 12, indicam o declínio exponencial do coeficiente de correlação para temperatura do ar, umidade relativa do ar e fluxo de calor no solo, respectivamente. Em média, entre as nove simulações, para aproximadamente 60% de falhas aleatórias obteve-se um coeficiente de correlação de 0,95 para a variável de temperatura do ar, para 56% de falhas obteve-se um coeficiente de correlação de 0,95 para umidade relativa do ar e, para 40% de falhas, obteve-se um coeficiente de correlação de 0,95 para fluxo de calor no solo.

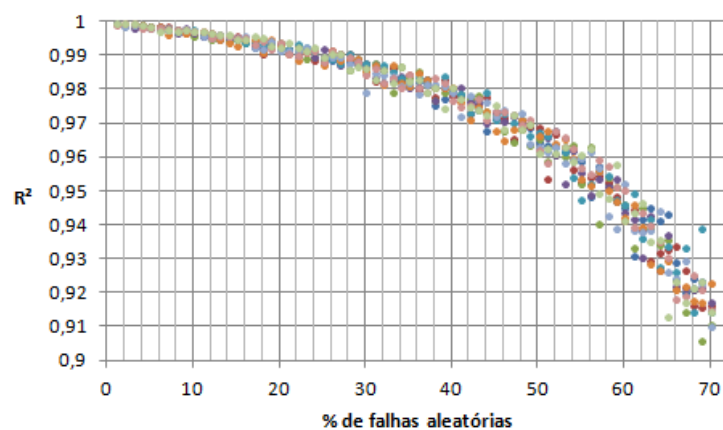


Figura 10: Relação entre coeficiente de correlação (R^2) e porcentagem de falhas das nove simulações para variável temperatura do ar.

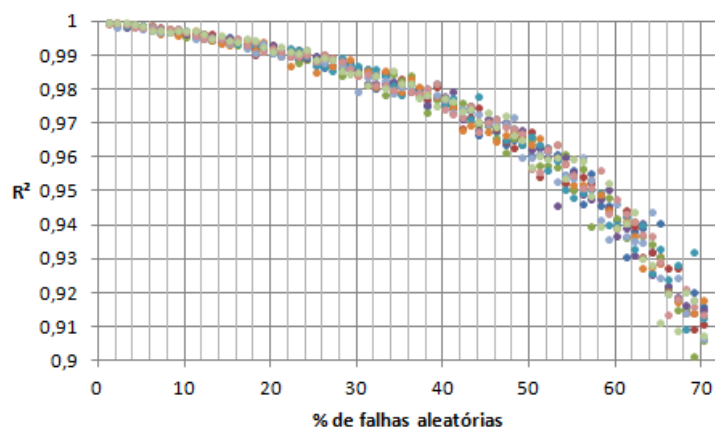


Figura 11: Relação entre coeficiente de correlação (R^2) e porcentagem de falhas das nove simulações para variável umidade relativa do ar.

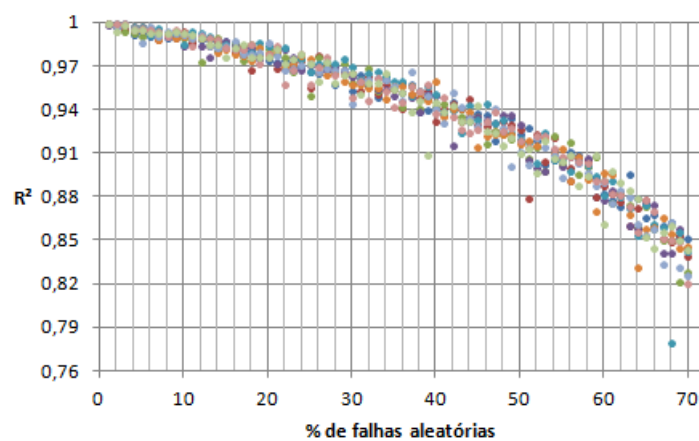


Figura 12: Relação entre coeficiente de correlação (R^2) e porcentagem de falhas das nove simulações para variável fluxo de calor no solo.

Após determinação do coeficiente de regressão em função do percentual de falhas, foi calculada a mediana e os percentis para suas respectivas variáveis, conforme apresentado nas Figuras 13 a 15, em que tem-se a delimitação do intervalo de confiança, por Bootstrap não paramétrico com intervalo de confiança 95% para as variáveis temperatura do ar, umidade relativa do ar e fluxo de calor no solo.

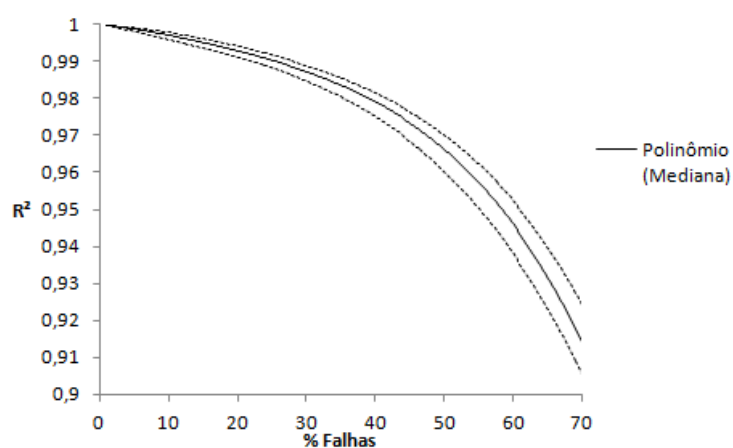


Figura 13: Mediana, limite superior e limite inferior com 95% de intervalo de confiança para variável de temperatura do ar para as nove simulações.

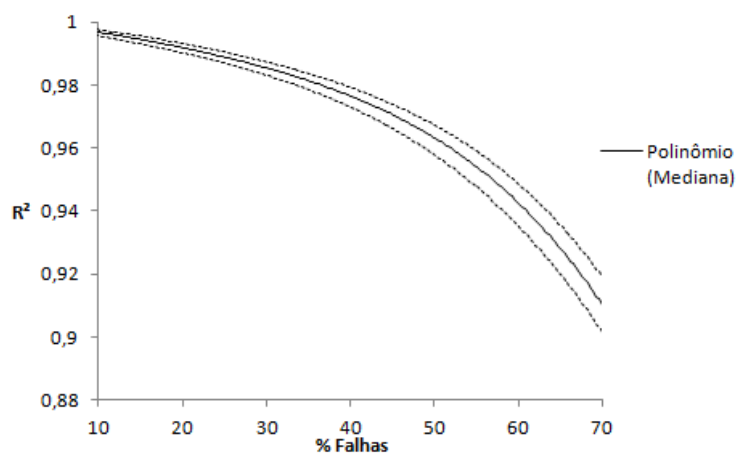


Figura 14: Mediana, limite superior e limite inferior com 95% de intervalo de confiança para variável de umidade relativa do ar para as nove simulações.

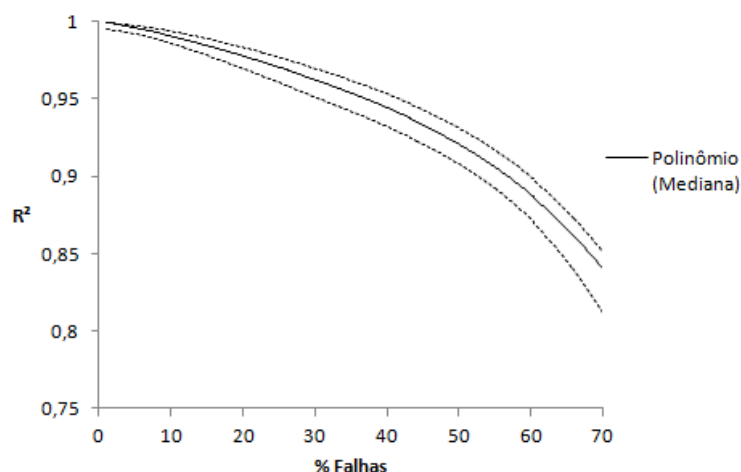


Figura 15: Mediana, limite superior e limite inferior com 95% de intervalo de confiança para variável de fluxo de calor no solo para as nove simulações.

Para verificar a variância entre os dados, foi realizado um teste f para as nove simulações, para cada porcentagem de falhas. Em quase todas as análises, encontrou-se p-valor (é a probabilidade de se obter uma estatística de teste igual ou mais extrema que aquela observada em uma amostra), acima de 0,05, indicando homocedasticidade entre as séries gerada e preenchida. E para a variável temperatura do ar apenas um valor referente a quarta simulação teve como p-valor 0,015 referente a 68% de falha (Figura 16).

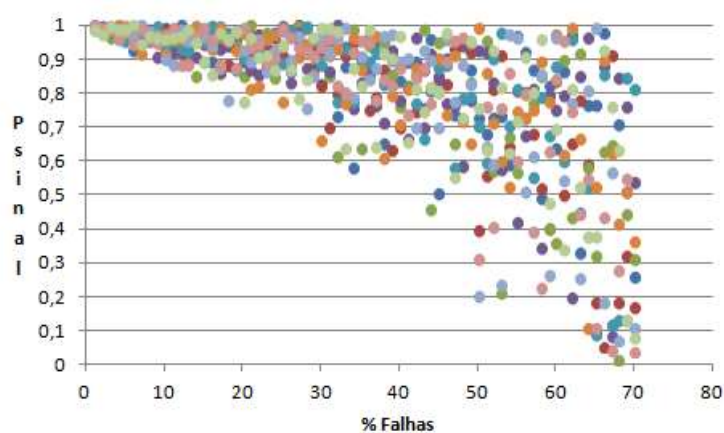


Figura 16: Teste-f para a variável de temperatura do ar para as nove simulações em que círculos com cores diferentes representam simulações diferentes.

Para simulação 8 teve apenas um valor com p-valor baixo, para 70%, 0,02 e na simulação 9 dois valores com diferença significativa, 69%, 0,01 e 70%, 0,02 (Figura 17).

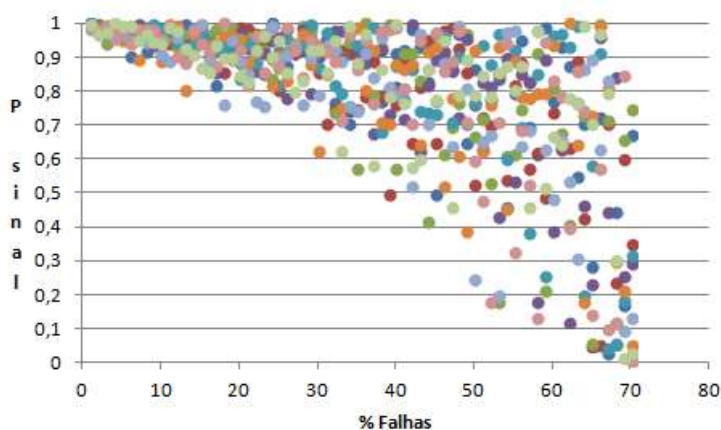


Figura 17: Teste-f para variável de umidade relativa do ar para as nove simulações em que círculos com cores diferentes representam simulações diferentes.

O teste f para a variável fluxo de calor no solo apresenta uma maior dispersão comparado com umidade relativa do ar e temperatura do ar. Em quatro porcentagens referentes a 4^o simulações de fluxo de calor no solo da série gerada (63%, 67%, 69% e 70%), no entanto, encontrou-se p-valor abaixo de 0,05, o que significa que nestas séries há diferença significativa entre as variâncias (Figura 18). A partir de 50% de falhas começam a ter diferença significativa em alguns valores conforme aumenta a porcentagem de falhas tem-se maior frequência de p-valor baixo (Figura 18).

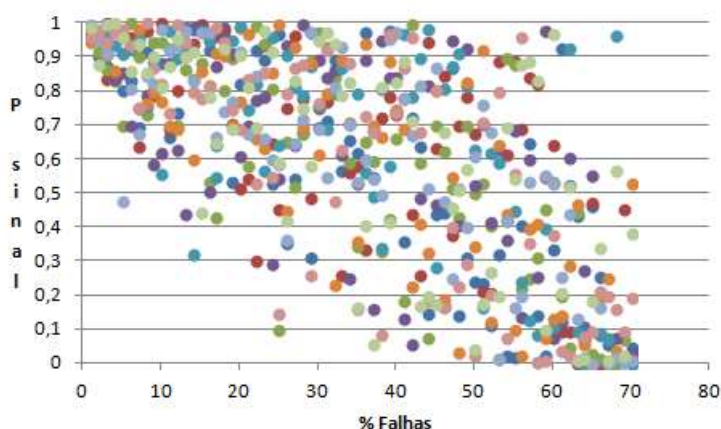


Figura 18: Teste-f para variável de fluxo de calor no solo para as nove simulações em que círculos com cores diferentes representam simulações diferentes.

5. CONSIDERAÇÕES FINAIS

A proposta deste trabalho consistiu em desenvolver uma metodologia em que houve a junção de técnicas estatísticas e automatização computacional para realizar preenchimento de falhas aleatórias de dados micrometeorológicos. Os valores do preenchimento de falhas são estimados com base nos dados existentes da própria série temporal.

Para testar a eficácia da metodologia proposta foram realizadas nove simulações gerando falhas e as preenchendo de 1% a 70% em todo o dado, para as variáveis temperatura, umidade do ar e fluxo de calor no solo.

Os testes foram realizados para três variáveis, entretanto a metodologia desenvolvida tem como objetivo preencher dados faltantes para qualquer tipo de variável e para qualquer região, pois a técnica não necessita de outras variáveis pareadas; o preenchimento é realizado com base somente nos dados da própria série temporal.

Os resultados apresentaram altos coeficientes de correlação entre a série preenchida e a original, preservando a variabilidade da série original. Apresentaram-se algumas exceções, o que implica em aumento do controle do número de repetições, em trabalhos futuros.

Ao se realizar o estudo com falhas aleatórias, observou-se também que se pode melhorar a qualidade do preenchimento, ao se considerar características, tais como, as falhas de longo período (inclusive para mais de um dia completo); falhas se encontrarem em horários de inversão do comportamento da variável; comportamento da série em função da periodicidade de suas componentes (ciclos, sazonalidade).

Esse estudo implicará na necessidade de se desenvolver métodos de classificação das falhas, de combinação de preenchimento da esquerda para a direita e da direita para a esquerda; considerações sobre a natureza sazonal das variáveis micrometeorológicas e características dos sensores, dentre outras implementações que se prevê para os trabalhos futuros.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ALBUQUERQUE, J. P. de A.; FORTES, J. M. P.; FINAMORE, W. A. **Probabilidade, Variáveis Aleatórias e Processos Estocásticos**, editora interferência. Rio de Janeiro. 2008.

BARROS, E. A. C, **APLICAÇÕES DE SIMULAÇÃO MONTE CARLO E BOOTSTRAP**. Monografia - Centro de Ciências Exatas Departamento de Estatística, Universidade Estadual de Maringá. 2005.

Bezerra, M. I. S. **Apostila de Análise de Séries Temporais**.UNESP. 2006.

BLAIN. G. C.; PICOLI. M. C. A.; LULU. J. **ANÁLISES ESTATÍSTICAS DAS TENDÊNCIAS DE ELEVAÇÃO NAS SÉRIES ANUAIS DE TEMPERATURA MÍNIMA DO AR NO ESTADO DE SÃO PAULO**.

v.68, n.3, p.807-815, 2009

BRINDER, K.; HEERMANN, D. W. **Monte Carlo Simulation in Statistical Physics**. 3 ed. New York: Springer, 1997. 150p.

CAPISTRANO. V. B. **ANÁLISE DE SÉRIES TEMPORAIS DE VARIÁVEIS MICROCLIMATOLÓGICAS MEDIDAS EM SINOP-MT UTILIZANDO A TEORIA DA COMPLEXIDADE**. 2007. 48 f. Dissertação(Mestrado em física Ambiental) - Instituto de Física, Universidade Federal de Mato Grosso, Cuiabá, 2007.

Costa. G. G. de O. **UM PROCEDIMENTO INFERENCIAL PARA ANÁLISE FATORIAL UTILIZANDO AS TÉCNICAS BOOTSTRAP E JACKKNIFE: CONSTRUÇÃO DE INTERVALOS DE CONFIANÇA E TESTES DE HIPÓTESES**. 2006. 189 f. Tese (Doutorado em Engenharia Elétrica) - Engenharia Elétrica, Pontifícia Universidade Católica, Rio de Janeiro, 2006.

CLARKE, A. B.; DISNEY, R. L. **Probabilidade e Processos Estocásticos**. Livros técnicos e científicos editora. Rio de Janeiro. 1979.

DEUS, B. V de; ZEILHOFER, P.; ARAUJO, G. C.; SANTOS, A. S. L. . Interpolação pluviométrica na bacia do alto e médio rio teles pires: uma análise de séries históricas e interpoladores. in: **III Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação** 27-30 (2010) p.001-00.

EHLERS, R. S. **Análise de Séries Temporais**. Curso de Séries Temporais. Agosto de 2009 e 2011.

FALCO, J, G. **Estatística Aplicada**. Cuiabá: ED UFMT, 2008.

FALGE, E. et al. Agricultural and Forest Meteorology 107 (2001) 43–69.

GAIO, D. C.; SANCHES, L.; NOGUEIRA, J.S. ; COSTA, M.H. ; ANDRADE, N. ; FRAGA,C. Gap Filling for Forest Ecosystem Modelling: A Critical Assessment of Accuracy. In: **The 6th European Conference on Ecological Modelling**. TRIESTE. 2007.

GAIO, D.C. ; MUSIS, C. R. ; PAULO, S. R. ; SANCHES, L. . Utilização de geoestatística para análise e preenchimento de falhas em séries temporais longas. In: **Conferência Científica Internacional Amazônia em Perspectiva - Ciência Integrada para um Futuro Sustentável**, 2008, Manaus. Conference Abstract: Internationa Scientific Conference Amazon in Perspective. Manaus.

GALLON. M. M. P. **UM ESTUDO SOBRE A DINÂMICA DE SISTEMAS COMPLEXOS A PARTIR DE SÉRIES TEMPORAIS DE DADOS MICROCLIMATOLÓGICOS PARA UMA FLORESTA DE TRANSIÇÃO NO NOROESTE DE MATO GROSSO**. 2005. 107 f. Dissertação(Mestrado em física Ambiental) - Instituto de Física, Universidade Federal de Mato Grosso, Cuiabá, 2005.

ISNTRUTEMP, **INSTRUMENTOS DE MEDIÇÃO**,2007. Disponível em: <<http://medicao.instrutemp.com.br/data-logger.php>>. Acesso em: 10 set. 2013.

JACQUES, S, M, C. Bio Estatística Princípios e Aplicações. Porto Alegre. Artmed. 2003.

MARENGO, J. A, ALVES, L. M. **Tendências Hidrológicas da Bacia do Rio Paraíba do Sul**. Revista Brasileira de Meteorologia, **v.20, n.2**, 215-226, 2005.

MARIANO. R. T. G. **ANÁLISE ESPECTRAL DE SÉRIES TEMPORAIS DE VARIÁVEIS MICRO-CLIMATOLÓGICAS EM UMA ÁREA DE ECÓTONO ENTRE OS BIOMAS AMAZÔNIA E CERRADO NO NORTE DE MATO GROSSO**. 2008. 79 f. Dissertação(Mestrado em física Ambiental) - Instituto de Física, Universidade Federal de Mato Grosso, Cuiabá, 2008.

METROPOLIS. N, ULAM S. **The Monte Carlo method**. J Am Stat Assoc. 1949;44 (247): 335-41.

MORETTIN, P. A.; TOLOI, C. M. **Análise de Séries Temporais**. 2ed. São Paulo: Egard Blucher. 2006.

OLIVERA, M. A. & FAVERO, L.P.L. **Uma breve descrição de algumas técnicas para análise de séries temporais: Séries de Fourier, Wavelets, Arima, Modelos Estruturais para séries de tempos e redes neurais**. V I SEMEAD Ensaio mqi.USP. São Paulo. 13p, 2002.

PALÚ. A. E. R. **DETERMINAÇÃO DO TEMPO DE DEFASAGEM MAIS ADEQUADO PARA ANÁLISE DE SÉRIES TEMPORAIS DE VARIÁVEIS MICROCLIMATOLÓGICAS MEDIDAS NUMA FLORESTA DE TRANSIÇÃO NO NORTE DE MATO GROSSO**. 2008. 35 f. Dissertação(Mestrado em física Ambiental) - Instituto de Física, Universidade Federal de Mato Grosso, Cuiabá, 2008.

RIHBANE, F. E. C, GAIO, D. C, Sanches, L. artigo apresentado, **XVII Congresso Brasileiro de Meteorologia. Estudo da variabilidade em séries históricas para preenchimento de falhas.** Gramado, RS. 2012

SANCHES, L.; SULI, G. S.; PRIANTE, N. F.; VOURLITIS, G. L.; NOGUEIRA J. S. Índice de área foliar em floresta semidecídua. **Revista Ciência Natura**, v. 1, p. 37-40, 2006.

SAVI, M, A. **Dinâmica Não linear e Caos.** Editora e-papers. Rio de Janeiro. 2006.

SOUZA, G. C. U. I de. **AVALIAÇÃO DE TÍTULOS CONVERSÍVEIS COM OPÇÕES DE COMPRA E VENDA IMPLÍCITAS EM CONTRATO.** 2006. 174 f. Tese (Doutorado em Engenharia de Produção: Finanças e Análise de Investimentos) – Engenharia Industrial, Pontifícia Universidade Católica, Rio de Janeiro, 2006

SOUZA, de K. P.; SOBRINHO, D. M, Instrumentos de medida e sistemas de instrumentação, 2005. Disponível em:

<<http://www.shs.eesc.usp.br/downloads/instrumentacao/Instrumentos.pdf>>

VENTURA, T. M. **PREENCHIMENTO DE FALHAS DE DADOS MICROMETEOROLÓGICOS UTILIZANDO TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL.** 2012. 73 f. Dissertação (Mestrado em Física Ambiental) – Instituto de Física, Universidade Federal de Mato Grosso, Cuiabá, 2012.

WHEELWRIGHT, Steven C.; MAKRIDAKIS, Spyros. **Forecasting Methods for Management.** 4th edition. New York : John Wiley & Sons Inc, 1985, apud MUELLER, A. **Uma Aplicação de Redes Neurais Artificiais na Previsão do Mercado Acionário.** Dissertação Programa de Pós-graduação em Engenharia de Produção FLORIANÓPOLIS, JULHO DE 1996.

Disponível em:

<<http://www.eps.ufsc.br/disserta96/mueller/index/index.htm#sumario>>

YORIYAZ. H. Monte Carlo Method: principles and applications in Medical Physics. Revista Brasileira de Física Médica. 2009;3(1):141-9.