



**Universidade do Estado do Rio de Janeiro**  
Centro de Tecnologia e Ciências  
Instituto Politécnico

Nely Grillo Guzmán

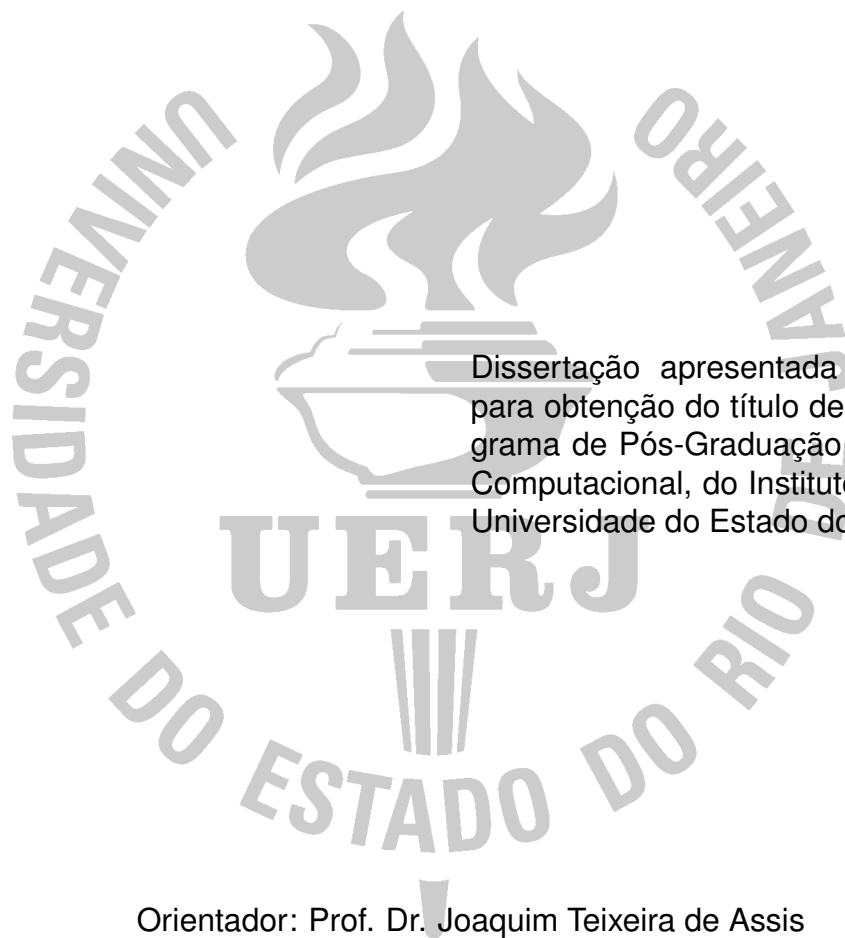
**Modelagem para estimativa de dados faltantes em série de dados  
meteorológicos**

Nova Friburgo

2018

Nely Grillo Guzmán

**Modelagem para estimativa de dados faltantes em série de dados  
meteorológicos**



Dissertação apresentada como requisito para obtenção do título de Mestre, ao Programa de Pós-Graduação em Modelagem Computacional, do Instituto Politécnico, da Universidade do Estado do Rio de Janeiro.

Orientador: Prof. Dr. Joaquim Teixeira de Assis

Nova Friburgo

2018

CATALOGAÇÃO NA FONTE  
UERJ / REDE SIRIUS / BIBLIOTECA CTC/E

G993 Guzmán, Nely Grillo.  
Modelagem para estimativa de dados faltantes em série de  
dados meteorológicos / Nely Grillo Guzmán. - 2018.  
69 f. : il.

Orientador: Joaquim Teixeira de Assis.  
Dissertação (Mestrado) – Universidade do Estado do Rio de  
Janeiro, Instituto Politécnico.

1. Meteorologia - Metodos de simulação – Teses. 2. Controle  
meteorológico - Metodos de simulação – Teses. 3. Previsão  
numérica do tempo (Meteorologia) – Teses. I. Assis, Joaquim  
Teixeira de . II. Universidade do Estado do Rio de Janeiro. Instituto  
Politécnico. III. Título.

CDU 551.51:519.872

Bibliotecária: Sandra Mueller CRB7/3633

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta  
dissertação.

---

Assinatura

---

Data

Nely Grillo Guzmán

**Modelagem para estimativa de dados faltantes em série de dados meteorológicos**

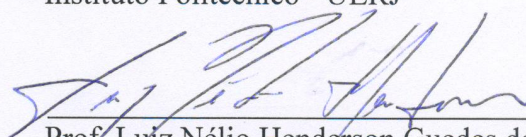
Dissertação apresentada como requisito parcial para obtenção do título de Mestre ao Programa de Pós-Graduação em Modelagem Computacional do Instituto Politécnico, da Universidade do Estado do Rio de Janeiro. Linha de Pesquisa: Matemática Aplicada e Computação Científica.

Aprovada em 05 de abril de 2018.

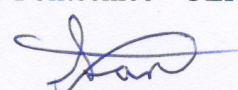
Banca examinadora:

  
\_\_\_\_\_  
Prof. Joaquim Teixeira de Assis, D.Sc. (Orientador)

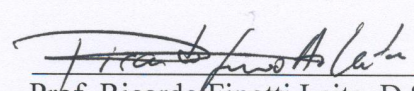
Instituto Politécnico - UERJ

  
\_\_\_\_\_  
Prof. Luiz Nélio Henderson Guedes de Oliveira, D.Sc.

Instituto Politécnico - UERJ

  
\_\_\_\_\_  
Prof. Joel Sanchez Dominguez, D.Sc.

Instituto Politécnico - UERJ

  
\_\_\_\_\_  
Prof. Ricardo Finotti Leite, D.Sc.

Universidade Estácio de Sá

Nova Friburgo

2018

## RESUMO

GUZMÁN, Nely Grillo. Modelagem para estimativa de dados faltantes em série de dados meteorológicos. 2018. 69f. Dissertação (Mestrado em Modelagem Computacional) - Instituto Politécnico, Universidade do Estado do Rio de Janeiro, Nova Friburgo, 2018.

Neste trabalho é apresentada a implementação de um método preditivo auto regressivo para a estimativa de dados faltantes em série de dados meteorológicos, usando somente a informação da mesma série. Foi usado o método de Burg para o cálculo dos coeficientes auto regressivos, o qual diminui o erro gerado no cálculo desses coeficientes. Para a implementação do método foi usado o framework multi-plataforma Qt através da linguagem de programação C++. Com o Qt conseguiu se obter uma interface gráfica amigável que facilita o uso do modelo. A validação do modelo foi feita com os dados das variáveis: temperatura do ar, velocidade e direção do vento, pressão atmosférica, umidade relativa e precipitação da Estação Meteorológica localizada no IPRJ/UERJ. Para a precipitação, não foram obtidos resultados favoráveis com o uso do modelo, mas o resultado obtido para as outras variáveis estão de acordo com a tolerância reportada, particularmente com as ordens 1 e 2 do modelo e a probabilidade de acerto nas estimativas não depende da quantidade de dados da série. Os dados estimados conservam as características estatísticas da série original.

Palavras-chave: Estimativa de dados faltantes. Método preditivo auto regressivo. Método de Burg.

## ABSTRACT

GUZMÁN, Nely Grillo. Modeling to estimate missing values in meteorological data series. 2018. 69f. Dissertação (Mestrado em Modelagem Computacional) - Instituto Politécnico, Universidade do Estado do Rio de Janeiro, Nova Friburgo, 2018.

In this work is presented the implementation of an autoregressive-predictive method for the estimation of missing data in meteorological data series, using just the information from the same serie. The Burg method was used for the calculation of the autoregressive coefficients, which reduces the error generated in the calculation of these coefficients. For the implementation of the method was used the multiplatform framework Qt through the programming language C++. With Qt we were able to obtain a friendly graphical interface that facilitates the use of the model. The validation of the model was done with data of the variables: air temperature, wind speed and direction, atmospheric pressure, relative humidity and precipitation of the Meteorological Station located at the IPRJ/UERJ. For precipitation no favorable results were obtained with the use of the model, nonetheless the results for the other variables are in agreement with the reported tolerance, particularly with orders 1 and 2 of the model and the probability of success in the estimates does not depend on the amount of data of the series. The estimated data retains the statistical characteristics of the original series.

Keywords: Estimation of missing data. Auto regressive predictive method. Burg Method.

## LISTA DE FIGURAS

Figura 1 - Localização da estação meteorológica no IPRJ/UERJ . . . . .	14
Figura 2 - Interface gráfica do modelo . . . . .	46
Figura 3 - Erro Quadrático Médio por ordens . . . . .	51
Figura 4 - Diferenças máximas entre os valores reais da série e os estimados pelo modelo para cada ordem . . . . .	54
Figura 5 - Diferenças máximas entre os valores reais da série e os estimados pelo modelo para cada posição . . . . .	55
Figura 6 - Diferenças máximas entre os valores reais e os estimados pelo modelo para a ordem igual a 1 nas diferentes posições dos dados faltantes na série . . . . .	56
Figura 7 - Diferença máxima entre os valores reais da série de chuva e os estimados pelo modelo com diferentes quantidades de dados . . . . .	57
Figura 8 - Comportamento mensal da precipitação. . . . .	66
Figura 9 - Comportamento mensal da temperatura. . . . .	67
Figura 10 - Comportamento mensal da pressão atmosférica. . . . .	68
Figura 11 - Comportamento mensal da umidade relativa. . . . .	69

## LISTA DE TABELAS

Tabela 1 - Máxima diferença entre os valores estimados e originais da série de temperatura do ar. . . . .	53
Tabela 2 - Diferença máxima permissível entre os valores estimados e originais da série para cada variável. . . . .	53
Tabela 3 - Máxima diferença entre os valores estimados e originais da série de chuva. . . . .	58
Tabela 4 - Estrutura do arquivo de dados a usar pelo modelo. . . . .	65



## SUMÁRIO

	<b>INTRODUÇÃO</b> . . . . .	9
1	<b>FUNDAMENTAÇÃO TEÓRICA</b> . . . . .	13
1.1	<b>A Estação Meteorológica de estudo</b> . . . . .	14
1.1.1	<u>Precipitação</u> . . . . .	15
1.1.2	<u>Temperatura do Ar</u> . . . . .	15
1.1.3	<u>Vento</u> . . . . .	16
1.1.4	<u>Pressão Atmosférica</u> . . . . .	16
1.1.5	<u>Umidade Relativa</u> . . . . .	17
1.2	<b>Modelos de previsão do tempo</b> . . . . .	17
1.2.1	<u>Componentes dos modelos numéricos de previsão do tempo</u> . . . . .	19
1.2.2	<u>Problemas da previsão numérica</u> . . . . .	21
1.2.3	<u>Tipos de Modelos Meteorológicos</u> . . . . .	22
1.2.4	<u>Modelos mais usados para a previsão do tempo no Brasil e no mundo</u> . . . . .	24
2	<b>REVISÃO BIBLIOGRÁFICA</b> . . . . .	28
2.1	<b>Pesquisas para estimar dados faltantes em séries temporais que usam os dados de mais de uma estação meteorológica</b> . . . . .	29
2.2	<b>Pesquisas para estimar dados faltantes em séries temporais que usam os dados da mesma estação meteorológica</b> . . . . .	32
2.3	<b>Pesquisas para estimar dados faltantes em séries temporais que usam o método preditivo auto regressivo</b> . . . . .	34
3	<b>MATERIAIS E MÉTODOS</b> . . . . .	37
3.1	<b>Estação de estudo</b> . . . . .	37
3.2	<b>O <i>Framework</i> Multiplataforma Qt</b> . . . . .	37
3.3	<b>O modelo preditivo auto regressivo</b> . . . . .	38
3.3.1	<u>O Método de Burg</u> . . . . .	40
3.4	<b>Método para a estimativa de dados faltantes em séries de dados meteorológicos</b> . . . . .	42
3.4.1	<u>Algoritmo para a modelagem da estimativa de dados faltantes</u> . . . . .	42
3.4.2	<u>Algoritmo para a modelagem do método preditivo auto regressivo</u> . . . . .	42
3.4.3	<u>Algoritmo para o cálculo dos coeficientes do método</u> . . . . .	43

3.5	<b>A ordem do modelo <math>AR(p)</math></b> . . . . .	45
4	<b>RESULTADOS E DISCUSSÃO</b> . . . . .	46
4.1	<b>Interface gráfica</b> . . . . .	47
4.2	<b>Os dados meteorológicos</b> . . . . .	48
4.3	<b>O modelo Preditivo Auto regressivo de ordem <math>p</math>, <math>AR(p)</math></b> . . . . .	49
4.3.1	<u>A ordem do método</u> . . . . .	50
4.3.2	<u>Validação do modelo</u> . . . . .	50
	<b>CONCLUSÕES E RECOMENDAÇÕES</b> . . . . .	60
	<b>REFERÊNCIAS</b> . . . . .	62
	<b>APÊNDICE A</b> -Estrutura do arquivo de dados a usar pelo modelo . . . . .	65
	<b>APÊNDICE B</b> -Comportamento mensal da Precipitação na estação meteorológica do IPRJ/UERJ nos anos desde 2009 até 2016 . . . . .	66
	<b>APÊNDICE C</b> -Comportamento mensal da Temperatura na estação meteorológica do IPRJ/UERJ nos anos desde 2009 até 2016 . . . . .	67
	<b>APÊNDICE D</b> -Comportamento mensal da Pressão Atmosférica na estação meteorológica do IPRJ/UERJ nos anos desde 2009 até 2016 . . . . .	68
	<b>APÊNDICE E</b> -Comportamento mensal da Umidade Relativa na estação meteorológica do IPRJ/UERJ nos anos desde 2009 até 2016 . . . . .	69

## INTRODUÇÃO

A sofisticação dos problemas com que a humanidade tem se deparado nas últimas décadas nos exige respostas mais sofisticadas para eles. Nas últimas décadas, o incremento do desempenho dos computadores, as interfaces gráficas cada vez mais amigáveis e o aprimoramento das técnicas de desenvolvimento de modelos matemáticos têm possibilitado um vertiginoso desenvolvimento das técnicas de modelagem computacional resolvendo cada vez problemas mais complexos nos diferentes ramos da ciência e da tecnologia. Assim, tem sido possível obter respostas mais sofisticadas com relação às que tivemos necessidade de gerar no passado. (LESA; DOMÍNGUEZ, 2007)

A Meteorologia é uma ciência interdisciplinar que estuda a atmosfera terrestre e tem como foco o estudo dos processos atmosféricos e a previsão do tempo. Os estudos neste campo da ciência foram iniciados há mais de dois milênios, mas só a partir do século XVII é que a Meteorologia progrediu significativamente. A partir do século XIX seu desenvolvimento ganhou um ímpeto ainda mais significativo com o progresso das redes de intercâmbio de dados entre os países, que junto ao desenvolvimento de modelos meteorológicos, fizeram que as primeiras previsões numéricas do tempo se tornassem possíveis. A invenção do computador e o desenvolvimento da Internet tornou mais rápido e ainda mais eficaz o processamento e o intercâmbio dos dados meteorológicos, proporcionando assim um maior entendimento dos eventos e variáveis meteorológicas e, conseqüentemente, tornou possível uma maior precisão na previsão do tempo.

Atualmente, na hora de fazer uma pesquisa que permita conhecer os elementos meteorológicos, variabilidade, frequência e probabilidade de valores críticos, é preciso que os dados históricos sejam contínuos e coerentes, diminuindo assim os riscos de resultados errados. Também é necessário ter dados precisos para avaliar e prever o comportamento futuro das variáveis meteorológicas. Porém, a estimativa de valores faltantes nas séries de dados é essencial para conseguir uma futura previsão do tempo mais acertada.

Muito frequentemente é preciso conhecer os valores faltantes numa série de dados meteorológicos de uma estação que encontra-se isolada ou numa localização

com características físico-geográficas particulares (diferentes das estações que ficam próximas). Nesta situação poderíamos usar os dados da mesma estação para fazer a estimativa daqueles que estiverem faltando.

O objetivo geral deste trabalho é resolver o problema da estimativa dos dados faltantes nas séries de dados meteorológicos, usando os valores da mesma série. Os objetivos específicos são:

- desenvolver um modelo capaz de estimar os dados faltantes nas séries de dados meteorológicos;
- criar uma interface gráfica através da qual o uso do modelo seja mais fácil;
- desenvolver o modelo e a interface num software livre para expandir seu uso sem custo algum;
- criar o modelo com código aberto para que outros cientistas interessados em usá-lo possam incluir as mudanças precisas para adequá-lo às suas necessidades.

Para resolver este problema, será usado um modelo preditivo auto regressivo, conhecido como  $AR(p)$ , o qual é um modelo linear que usa os valores de  $p$  tempos da amostragem anteriores e posteriores para estimar o valor da série num tempo dado. (ULRICH; CLAYTON, 1976; ULRICH; BISHOP, 1975)

Esse método permite preencher os dados faltantes usando a informação da série conhecida, característica que o faz muito útil nos casos em que não existe uma outra estação próxima com características similares à estação que tem dados em falta. (ALFARO; SOLEY, 2009)

Para a implementação do método foi usado um *framework* multiplataforma para desenvolvimento de interfaces gráficas em C++ criado pela Empresa norueguesa Trolltech, o “Qt”. O Qt é uma estratégia tecnológica que permite criar, desenhar, desenvolver e manter um software rapidamente e com um baixo custo, além de fornecer iguais experiências em todos os dispositivos do usuário, já que os aplicativos desenvolvidos nesta plataforma podem ser executados no Windows, Linux e Mac OS X, sem precisar mudar o código fonte. (QT COMPANY, 2018)

Para o teste e implementação deste modelo, foram usados os dados da Estação Meteorológica do Centro de Tecnologia em Meio Ambiente (CETEMA) que está localizada no bairro Vila Amélia, no Campus Regional da UERJ em Nova Friburgo à 870 metros de altitude, sob as coordenadas de 22° 28'68"S e 42° 54'31"W. (MAPS AND DIRECTIONS, 2018)

### **Estrutura da dissertação:**

#### Capítulo 1. Marco Teórico:

Nesse capítulo são descritas as definições e conceitos precisos para um melhor entendimento do tema que vai ser abordado no resto da dissertação. É explicada a importância dos dados meteorológicos e sua utilização. Também é mostrada uma descrição da estação meteorológica da qual foram usados os dados, os sensores com que são feitas as medições e as variáveis meteorológicas.

#### Capítulo 2. Revisão Bibliográfica:

Nesse capítulo são expostas as pesquisas feitas anteriormente com os principais resultados obtidos sobre a estimativa de dados faltantes nas séries de dados temporais e meteorológicos. Para um melhor entendimento dos antecedentes da pesquisa, esse capítulo foi dividido em três seções: pesquisas para estimar dados faltantes em séries temporais que usam os dados de mais de uma estação meteorológica, da mesma estação meteorológica e modelos que usam o método preditivo auto regressivo.

#### Capítulo 3. Materiais e Métodos:

Nesse capítulo são descritos os dados usados na pesquisa, assim como o *framework* multiplataforma Qt usado para a programação do método e obtenção da interface gráfica resultado desta pesquisa. Também foi descrito o método preditivo auto regressivo de ordem  $p$ , usado no desenvolvimento do modelo, mostrando os al-

goritmos precisos para o desenvolvimento do método.

#### Capítulo 4. Resultados e Discussão:

Nesse capítulo são descritos os principais resultados obtidos através desta pesquisa. É mostrada uma descrição da interface gráfica obtida e como pode ser usada pelo usuário. São apresentados os testes feitos para a validação do modelo, assim como os resultados obtidos com cada um deles.

#### Conclusões e Recomendações:

Nesse capítulo são apresentados as principais conclusões obtidas no desenvolvimento da pesquisa assim como as recomendações propostas para continuar com o estudo e melhoria do modelo.

## 1 FUNDAMENTAÇÃO TEÓRICA

As atividades climáticas, em particular as pesquisas climáticas, são baseadas nas observações do estado da atmosfera e do tempo. A necessidade de contar com informação mais precisa e oportuna está aumentando rapidamente, dado que a variedade dos usuários tem aumentado consideravelmente, assim como as aplicações deste tipo de informações e pesquisas. A climatologia é aplicada em vários setores da sociedade, economia e o meio ambiente como na agricultura, energia, indústria, produção e distribuição de bens de consumo, construção, transporte, turismo, gestão de recursos hídricos, pesca, e o desenvolvimento de zonas costeiras, entre outros. É por isso que cada país deve usar práticas coerentes ao fazer observações climáticas, gerenciar registros climáticos e manter a qualidade e a utilidade necessária dos serviços prestados.

A principal utilidade dos dados meteorológicos e as previsões climáticas é planejar a atenuação de desastres e desenvolvimento sustentável para enfrentar as consequências das mudanças climáticas, objetivo este que foi firmemente estabelecido em várias convenções, como a Convenção Geral das Nações Unidas sobre Mudanças Climáticas. (ORGANIZACIÓN METEOROLÓGICA MUNDIAL, 2011)

Todos os produtos e serviços climáticos estão baseados na coleta dos dados primários através das observações e o registro amplo e sistemático de uma série de variáveis principais que permitem caracterizar o clima em uma ampla gama de escalas temporais.

Na hora de fazer o cálculo estatístico para analisar as séries de dados meteorológicos, se essas séries estão incompletas, os resultados podem ser distorcidos. Como geralmente existe uma correlação muito grande entre os dados climáticos, a falta consecutiva de dados tem uma maior repercussão nas análises estatísticas que a falta, da mesma quantidade de dados, espalhados aleatoriamente no período de dados todo. Por isso é recomendado que as análises estatísticas sejam feitas quando, no mínimo, o 80% dos dados estiverem disponíveis. (ORGANIZACIÓN METEOROLÓGICA MUNDIAL, 2011)

No caso de algumas variáveis, as falhas nas séries de dados são mais significativas do que em outras. Por exemplo, os dados de chuva mensal total podem estar

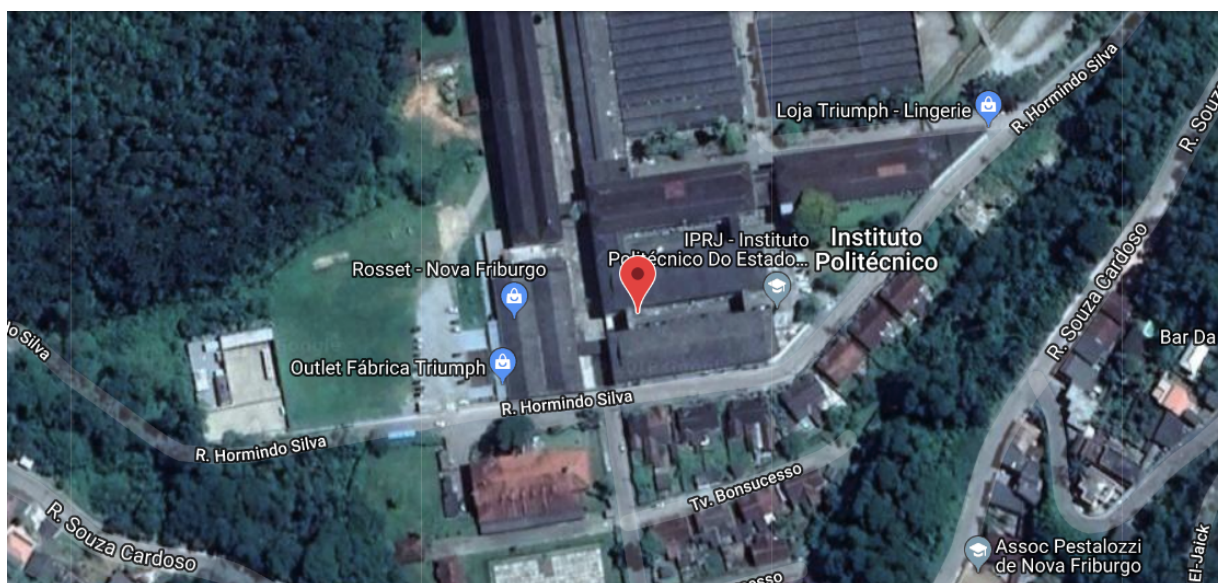
muito afetados se faltam dados na série, principalmente se aconteceu algum fenômeno gerador de chuva nos dias da falha. Por outro lado, os dados de temperatura média mensal podem ser menos afetados pela falta de dados, se durante o mês com falta de dados a temperatura se manteve estável.

Então, para algumas aplicações é preciso que a série de dados esteja completa para poder obter resultados confiáveis.

### 1.1 A Estação Meteorológica de estudo

A Estação Meteorológica do Centro de Tecnologia em Meio Ambiente (CETEMA), usada para este trabalho, está localizada no bairro Vila Amélia, no Campus Regional de UERJ em Nova Friburgo (IPRJ) (Fig. 1) à 870 metros de altitude, sob as coordenadas de 22° 28'68"S e 42° 54'31"W. (MAPS AND DIRECTIONS, 2018)

Figura 1 – Localização da estação meteorológica no IPRJ/UERJ



Fonte: CETEMA (2018).

Essa Estação Meteorológica foi projetada para complementar uma ampla gama de aplicações meteorológicas, climatológicas, hidrológicas e de pesquisas ambientais a serem executadas na região centro-norte fluminense. Trata-se de uma plataforma automática capaz de controlar todo o processo de coleta, armazenamento e comunicação de dados meteorológicos por ela coletados através da transmissão remota



utilizando uma placa de modem. Por essa razão pode ser utilizada como uma unidade independente e/ou integrada em uma rede de estações interligadas a uma única unidade receptora central. (CETEMA, 2018)

Nessa estação são coletadas as variáveis: Precipitação, Temperatura do Ar, Direção e Velocidade do Vento, Pressão Atmosférica e Umidade Relativa.

#### 1.1.1 Precipitação

A precipitação, em meteorologia, se refere a qualquer tipo de queda de água do céu, em qualquer um dos seus estados (neve, chuva e chuva de granizos). Todos os valores de precipitação são expressos em milímetros (mm) ou polegadas (“), e podem ser medidos através de pluviômetros ou baseados em imagens de satélite.

A estação opera com um sensor de precipitação denominado pluviômetro da marca Squitter, modelo S2163.

O pluviômetro mede a quantidade de precipitação (chuva) acumulada num certo intervalo de tempo. Consiste de um funil coletor de alumínio com borda em faca que direciona a água para um mecanismo de balança de equilíbrio instável. Um ímã é acoplado à balança e aciona uma chave magnética a cada mudança de posição do mecanismo, produzindo um fechamento momentâneo da chave. Cada pulso gerado corresponde a um acúmulo de 0,25 mm de chuva. (CETEMA, 2018)

#### 1.1.2 Temperatura do Ar

A temperatura do ar é a medida que se refere ao grau de calor específico do ar em um determinado local e hora. O aquecimento da atmosfera acontece principalmente por causa da incidência dos raios do sol, os quais chegam até a terra em forma de onda curta que são absorvidas pela terra e emitidas em forma de calor, o que faz que as maiores temperaturas do ar se encontrem perto da superfície terrestre.

A estação opera com um sensor de umidade relativa e temperatura do ar acoplados da marca Squitter, modelo S2095.

O elemento, sensor de temperatura do ar, é constituído de um resistor de platina que possui uma bem calibrada e estável relação entre resistência elétrica e tempe-

ratura. As variações de resistência são medidas por um circuito eletrônico que apresenta em sua saída uma tensão contínua com 1,0 Volt representando +60° C e 0,0 Volts representando -40° C. (CETEMA, 2018)

### 1.1.3 Vento

O vento não é mais do que o ar em movimento, e pode ser gerado pelo aquecimento (dilatação) o resfriamento (contração) de massas de ar à mesoescala, ou pelo efeito de sistemas meteorológicos de maior escala como sistemas sinóticos ou de escala global.

A estação opera com sensor de ventos acoplado (velocidade e direção) da marca Squitter, modelo S2014.

A velocidade do vento é medida usando-se um *reed-switch* selado, que gera uma série de pulsos a uma taxa proporcional à velocidade de rotação do sensor, numa faixa operacional de 0 a 50 m/s.

A direção do vento é medida usando-se um cata-vento diretamente acoplado a um potenciômetro de precisão. Variações na direção do vento produzem uma correspondente mudança de tensão no potenciômetro. (CETEMA, 2018)

### 1.1.4 Pressão Atmosférica

A pressão atmosférica num ponto determinado não é outra coisa que o peso da coluna de ar que está acima desse ponto. As variações nas medidas estão ligadas inversamente à temperatura do ar. Uma vez que a temperatura aumenta, o ar se expande e diminui o peso da coluna de ar que fica acima do ponto de medição, fazendo com que diminua a pressão atmosférica. Quando a temperatura diminui acontece o processo contrário, o ar se contrai e aumenta o peso da coluna de ar acima do ponto de medição fazendo a pressão atmosférica aumentar.

A estação opera com o sensor de pressão atmosférica denominado barômetro, da marca Squitter, modelo S2021.

O sensor capacitivo de cerâmica segue um projeto mecânico único. Uma cápsula cerâmica simétrica se deforma proporcionalmente à pressão aplicada. (CETEMA,

2018)

### 1.1.5 Umidade Relativa

A umidade relativa é a relação da quantidade de água que há no ar e a quantidade máxima que poderia haver com essa mesma temperatura, expressa em percentuais. Os valores desta variável estão compreendidos no intervalo 0 a 100%, sendo igual a 100% quando o ar não consegue mais absorver água, o ar está saturado com água nessa temperatura.

A estação opera com um sensor de umidade relativa e temperatura do ar acoplados da marca Squitter, modelo S2095.

O elemento sensor de umidade relativa do ar é um filme higroscópico posicionado entre dois eletrodos, constituindo um capacitor. A capacitância é proporcional à umidade relativa do ar e depende da umidade absorvida pelo filme higroscópico (o dielétrico do capacitor). A capacitância medida é convertida numa tensão contínua com correção automática para a temperatura. A saída de tensão é 0,0 Volts para 0% de umidade relativa e 1,0 Volt para 100% de umidade. (CETEMA, 2018)

## 1.2 **Modelos de previsão do tempo**

Para o estudo da atmosfera e dos fenômenos meteorológicos, são usadas diferentes ferramentas, tais como as observações e medições das variáveis meteorológicas, que têm representado por séculos a base no estudo de tempo da climatologia e da meteorologia. A partir da década de 1950 começa a se desenvolver outro tipo de ferramentas que são essenciais para a compreensão e previsão da atmosfera: os Modelos Meteorológicos.

O Modelo Meteorológico é o desenvolvimento computacional das equações físico-matemáticas que descrevem os processos atmosféricos. Um modelo meteorológico permite, a partir de um estado inicial da atmosfera, simular a evolução atmosférica traduzindo, através de métodos numéricos, as leis da mecânica dos fluidos e da termodinâmica.

Existem vários modelos de acordo com a escala temporal e espacial dos fenô-

menos a estudar. Os mais utilizados são os modelos de Circulação Geral Hidrostática (MGC). São modelos que trabalham com baixa resolução espacial, à 200 km, e estudam os fenômenos que têm períodos de anos, é dito estar trabalhando na macro escala. São modelos que tem como prioridade a simulação do fluxo de circulação geral da atmosfera e, por conseguinte, deve trabalhar em uma malha de integração que cobre a terra toda. Eles são usados principalmente para a previsão a prazo médio (até 6 dias), sazonais (estações do ano) e climáticos (mais de 1 ano). Na outra extremidade são colocados os modelos de alta resolução, conhecidos como modelos não-hidrostáticos de nuvem. Eles trabalham com resoluções espaciais de centenas de metros ou menos, descrevendo eventos desde horas até minutos de duração, trabalham na microescala. (GARCÍA J., 2002)

Modelos regionais e de mesoescala são os que descrevem principalmente os fenômenos entre a macroescala e a microescala. Não há uma divisão clara entre estes dois tipos, mas pode ser considerado que um modelo regional trabalha com resoluções de 50-150 km e os de mesoescala, de 1-50 km. (GARCÍA J., 2002)

Os modelos regionais, modelos de mesoescala e de nuvem, também são conhecidos como modelos LAM (acrônimo em inglês de Modelos de Área Limitada), porque se concentram em regiões específicas e têm limites laterais bem estabelecidos, ao contrário dos modelos globais. (JORBA; BALDASANO, 2005)

Um modelo numérico é um conjunto de equações matemáticas cuja solução precisa de métodos numéricos. As equações básicas de um modelo numérico de tempo são aquelas que, por exemplo, regulam a circulação do ar (horizontal e vertical), conservação de massa e energia, transformações termodinâmicas, os processos de formação e desenvolvimento de nuvens, etc. Os métodos numéricos mais comuns utilizados para resolver o sistema de equações diferenciais em derivadas parciais (modelo numérico do tempo) são: diferença finita, métodos espectrais e elementos finitos.

As leis da física atmosférica são desenvolvidas em uma série de expressões matemáticas que são transformados para implementá-las em um ambiente de computação. Devido à incapacidade de ser resolvido analiticamente, deve ser resolvido a partir da aplicação de métodos numéricos, tais como:

- diferenças finitas;
- técnicas espectrais;

- técnicas pseudo-espectrais;
- elementos finitos;
- esquemas de interpolação.

De todas essas técnicas, as mais utilizados são: diferenças finitas, elementos finitos e esquemas de interpolação. Embora as técnicas espectrais tenham provado ser muito precisas, o seu uso não foi estendido devido à complexidade das expressões com as quais têm de trabalhar. (FOX; DEARDROFF, 1972)

O conjunto de leis que descrevem os processos físicos e dinâmicos na atmosfera têm uma natureza altamente não-linear. Esta não linearidade está dada pela interação entre os diferentes processos, que causam que a evolução da atmosfera seja sensível às condições iniciais, sendo um sistema de dissipação de energia. (JORBA; BALDASANO, 2005)

### 1.2.1 Componentes dos modelos numéricos de previsão do tempo

A estrutura básica de um modelo numérico de previsão do tempo é composta de um grupo de componentes que interagem com o objetivo de produzir uma previsão objetiva. Estes componentes básicos são: a assimilação de dados, parametrizações, a dinâmica do modelo (dinâmica da atmosfera) e, finalmente, a previsão feita pelo modelo considerado. (LÓPEZ L., 2004)

O primeiro passo é a **assimilação de dados**, que é usualmente realizada sequencialmente. O modelo organiza e difunde informações a partir de observações anteriores. A informação de novas observações são usadas para modificar o estado do modelo e torná-lo o mais consistente possível.

Em seguida, vem a **parametrização** dos diferentes processos físicos da atmosfera, tais como a radiação, convecção e interação com o exterior, nas bordas ou fronteiras do modelo. Estes processos precisam ser representados numericamente; alguns deles ocorrem em uma escala que é pequena demais para serem resolvidos de acordo com a resolução do modelo.

Os esquemas de parametrização podem fazer suposições de acordo com as restrições computacionais envolvidas no processo.

Alguns processos parametrizados são (JORBA; BALDASANO, 2005):

- *radiação*, que é dada pela emissão do sol e sua representação, é essencial em um modelo de previsão numérica. A radiação na atmosfera é dividida em ondas curtas e ondas longas. Os fluxos de ondas curtas dependem principalmente do ângulo zenital do sol (varia de acordo com a latitude, estação e hora do dia), tipo de nuvem e albedo do solo. Os fluxos de ondas longas dependem da quantidade de emissão média e a emissividade da terra;
- os *processos superficiais e sub-superficiais*, onde é determinado, por exemplo, a influência do tipo de solo e da vegetação;
- os *processos dentro de nuvens de grande escala e sua precipitação*, o qual é considerado através dos valores explícitos de quantidade de céu nublado, vapor de água e relação de mistura em nuvens de gelo;
- a *precipitação convectiva*, que é iniciada através de um teste de instabilidade convectiva. Se a temperatura potencial de algum nível é superior ao nível imediatamente por cima, começa a convecção. Isto irá continuar até que o ar quente continue a subir, devido à sua densidade mais baixa;
- a *orografia*: O esquema de orografia parametriza os efeitos na camada limite da atmosfera e sua topografia em escalas de 10 km ou menos. Esta parametrização é calculada em termos de um coeficiente de arrasto constante e depende linearmente da forma do terreno, e é considerado um coeficiente de rugosidade que depende do tipo de vegetação. Esta fórmula é baseada na teoria *Monin-Obukov*, utilizada para calcular os fluxos de superfície.

Em seguida, é considerada a **dinâmica do modelo numérico**, que é condicionado pelas equações que governam a atmosfera.

Essas equações são escritas em um sistema de coordenadas esféricas e discretizadas horizontalmente em uma grade regular em latitude e longitude. Na configuração global, a grade de polos coincide com os polos geográficos.

Na vertical, é usado um sistema de coordenadas com base na pressão. A coordenada é a “*terrain*”, que segue um comportamento  $\sigma$  na superfície, normalizado para a pressão superficial ( $\sigma = 1$ ), e um comportamento de pressão nas camadas superiores, que tendem a ser mais horizontais.

Entre a superfície  $\sigma$ , e a camada superior, os níveis de coordenada diminuem gradualmente sua espessura. O sistema de coordenada “*terrain*” tem a vantagem de não ter interseção com as superfícies orográficas e atende às condições da camada superficial baixa (a velocidade vertical no sistema  $\sigma$  é zero sobre a superfície e no topo do modelo).

Finalmente, são considerados os campos de superfície, cuja representação descreve o comportamento da terra, para determinar as situações tais como a topografia, o tipo de solo e da vegetação, a temperatura do mar, umidade do solo e quantidade de neve acumulada, etc.,.

### 1.2.2 Problemas da previsão numérica

A limitação no acerto das previsões dos modelos dinâmicos para a previsão numérica do tempo não é associada somente ao tratamento inadequado dos processos físicos da atmosfera. Os problemas nas previsões numéricas também são fornecidos por:

– *Resolução espacial dos modelos:*

Por causa das limitações no *hardware* dos computadores que são usados, é impossível ter um modelo que consiga descrever todos os fenômenos interessantes e importantes da atmosfera. Porém, há casos de previsões ruins devido ao fato de que o fenômeno de interesse para a previsão não pode ser simulado pelo modelo.

– *Inicialização dos modelos:*

Em muitos locais do mundo, a previsão meteorológica de até três dias depende muito do estado inicial da atmosfera ao redor da área de interesse, e da quantidade de observações na superfície e na altura. Consequentemente, os problemas na inicialização dos modelos são responsáveis por

mais de 50% dos erros na previsão meteorológica.

– *Parametrização dos processos físicos sub-grade:*

Todos os fenômenos físicos que acontecem numa resolução menor do que a resolução do modelo não podem ser representados diretamente nas equações dos modelos. Os mais importantes são representados de uma forma simples que é chamado de “parametrização”. Deficiências nas parametrizações físicas são muitas vezes responsáveis por previsões ruins quando há um fenômeno muito importante que se desenvolve na pequena escala. Um exemplo típico é a convecção, que é muito importante em baixa latitude, e na Europa no verão.

– *Acoplamento com a superfície:*

Um modelo de previsão numérica depende de alguns campos constantes que descrevem: orografia, as condições de superfície (albedo, vegetação ...) e as condições de contorno (para um Modelos de Área Limitada).

A orografia é muito importante para a previsão de certos fenômenos. O problema da representação da orografia é muito próximo ao problema da resolução espacial dos modelos, porque as montanhas não podem ser representadas com uma resolução melhor que a resolução do modelo.

– *Capacidade computacional:*

Modelos com uma maior resolução vão precisar de maior tempo de computação, ou de computadores mais potentes para poder obter resultados com tempo suficiente para fazer a previsão.

### 1.2.3 Tipos de Modelos Meteorológicos

Apesar da evolução da previsão numérica nos últimos quarenta anos, ainda não existe uma solução única para resolver o problema da previsão do tempo. Quando consideramos a concepção de um modelo numérico devemos começar colocando quais são nossas limitações e nossas prioridades. A limitação mais importante é o tempo máximo entre o início e o fim da previsão que nós queremos fazer. Depen-



dendo da velocidade de cálculo do computador que temos, seria limitado o número de pontos da grade que pode incluir a nossa área de integração, portanto, a resolução espacial do nosso modelo.

Em resposta a estas prioridades e limitações, alguns dos tipos de modelos meteorológicos são (LIGHEZZOLO R A., 2014):

– *Modelos de Circulação Geral ou Global:*

Nesses modelos a prioridade é a simulação do fluxo de circulação geral da atmosfera, já que a sua integração na rede deverá cobrir a Terra toda. São utilizados principalmente para previsões a médio prazo, sazonais e climáticos, porque ele não precisa de dados externos (exceto as condições iniciais) para fazer as previsões. Isto é uma grande vantagem, mas a sua principal desvantagem é que eles precisam de uma grande infraestrutura de manutenção. Para ter uma resolução horizontal adequada (entre 50 e 80 Km) o número de nós da integração na rede é tão grande que são necessários computadores poderosos para integrá-los. Além disso, dado que a cobertura é global, precisa de observações do mundo inteiro para a assimilação, como a cobertura de observações é muito desigual, os resultados desses modelos dependem muito da área do globo que é considerada.

– *Modelos de Área Limitada (LAM):*

Esse tipo de modelo deve ser usado quando a área de interesse de previsão ou estudos que devem ser feitos com o modelo é muito específico e de curto alcance. Pode economizar muito tempo de cálculo operando o modelo numérico em uma grade plana cobrindo apenas essa área, é por isso que são chamados de “modelos de área limitada” . Eles precisam das previsões de um modelo que inclui a grade de integração para usá-la como condições de fronteira e têm limitações no tempo máximo das previsões que podem ser feitas com eles. Considera-se que as estruturas que entram nos contornos do modelo devem ser menores que a área de maior interesse na duração da previsão, isto é, se considerarmos a velocidade típica de uma perturbação em latitudes médias, o período

de previsão obtido com o modelo deve ser menor do que o tempo gasto pela perturbação para viajar a partir da entrada na área até a área de interesse.

Normalmente, esses modelos são usados para previsões de curto prazo (até 48 horas) e tem como vantagem uma maior resolução horizontal do que os modelos globais típicos.

– *Modelos de Mesoescala:*

Na medida em que aumenta a resolução horizontal dos modelos, as previsões obtidas com eles melhoram. No entanto, esta melhoria tem o seu limite nos 5 km, onde começam a ser inválidos alguns dos pressupostos básicos com os quais se construíram os modelos numéricos de curto e médio prazo. Esses modelos concebidos para funcionar em resoluções acima de 5 km, são conhecidos pelo nome de “modelos de mesoescala”. Estes são modelos com uns esquemas de parametrizações físicas sofisticados, mas cuja contrapartida é que exigem uma grande velocidade de cálculo.

Dado que os seus resultados operacionais ainda não estão à altura dos recursos de computação que necessitam, esses modelos são utilizados principalmente como uma ferramenta de pesquisa e não para o trabalho diário.

#### 1.2.4 Modelos mais usados para a previsão do tempo no Brasil e no mundo

Entre os principais modelos meteorológicos de previsão do tempo que são usados no Brasil e no mundo, podem ser citados os seguintes:

**GFS** (*Global Forecast System*) foi desenvolvido pelo Centro Nacional de Previsão Ambiental (NCEP, pelo acrônimo do nome em inglês) EUA, é um sistema global de previsão numérica do tempo que é executado quatro vezes por dia e produz previsões de até 16 dias.

Basicamente, o modelo é executado em duas partes. O primeiro tem alta resolução e prevê 192 horas (8 dias), enquanto que a segunda parte é

executada a partir de 192 a 384 horas (16 dias) em baixa resolução. A resolução do modelo varia em cada parte do modelo: horizontalmente a resolução varia de 35 a 70 km, verticalmente divide a atmosfera em 64 camadas e temporariamente produz previsões a cada 3 horas durante as primeiras 192 horas e a cada 12 horas até 384 horas.

O GFS é um modelo acoplado que consiste em quatro modelos separados: modelo da atmosfera, modelo do oceano, modelo de terra/solo, e um modelo do gelo do mar. Estes quatro modelos trabalham juntos para fornecer uma representação mais precisa das condições climatológicas.

**WRF** (*Weather Research and Forecasting*) é projetado para ser usado bem em investigação como para a implementação de previsões operacionais. Ele permite uma ampla gama de aplicações meteorológicas em faixas com escalas que vão desde poucos a milhares de quilômetros.

O modelo WRF permite aos pesquisadores produzir simulações considerando dados reais ou em condições atmosféricas idealizadas, enquanto que ao mesmo tempo fornece uma previsão operacional flexível numa plataforma robusta com base na contribuição dos mais recentes avanços na física, modelagem numérica e técnicas de assimilação de dados desenvolvidos por pesquisadores.

No Brasil existem trabalhos sendo desenvolvidos com o WRF por Instituições acadêmicas, entre elas o Instituto Nacional de Pesquisas Espaciais (INPE), a Universidade Federal do Rio de Janeiro (UFRJ) e a Universidade de São Paulo (USP). Na Força Aérea Brasileira, estudos com o WRF foram iniciados em 2011 no Centro Nacional de Meteorologia Aeronáutica (CNMA) e em 2012 o Grupo de Trabalho (GT) definiu o modelo WRF como o próximo modelo de previsão numérica a ser implementado no Instituto de Controle do Espaço Aéreo (ICEA) dando continuidade aos trabalhos realizados com o MM5.

**MM5** é um LAM de simulação numérica desenvolvido no final da década de 70 pela *Penn State University*, em conjunto com o *National Center for Atmospheric Research (NCAR)*. Atualmente, se encontra na quinta geração

e ao longo dos anos foram sendo incorporadas diversas modificações, como a capacidade de múltiplos aninhamentos de grade com interação bidirecional, dinâmica não hidrostática e assimilação de dados em 4 dimensões, além de várias parametrizações físicas e portabilidade em diversas plataformas computacionais. (GOMEZ C. ET AL., 2014)

Este modelo utiliza um sistema de coordenadas que segue a topografia e resolve a equação da continuidade, as equações de *Navier-Stokes* em três dimensões, a Primeira Lei da Termodinâmica e a equação de transferência radiativa.

Como todo LAM, são necessárias condições iniciais e de contorno. No MM5, as quatro fronteiras possuem campos de ventos horizontais, temperatura e umidade especificados. Estes dados podem vir de análises de modelos de grande escala, simulações prévias do próprio MM5 ou de outro modelo de previsão. O processamento dos dados e obtenção das previsões é feito pelo módulo MM5 e o pós-processamento, ou seja, a visualização dos resultados obtidos, pode ser feita por vários programas, como por exemplo o GRADS. (LIRA; CATALDI, 2016)

**BRAMS** (*Brazilian Regional Atmospheric Modelling System*) é uma adaptação às condições tropicais do modelo RAMS (*Regional Atmospheric Modeling System*). O modelo resolve as equações da dinâmica de convecção e possui uma série de submodelos que representam processos como a interação solo-vegetação-atmosfera, troca de fluxos turbulentos, transferências radiativas, microfísica de nuvens, entre outros.

O modelo numérico BRAMS é inicializado diariamente, sendo obtidas previsões das condições atmosféricas para 1, 2, 3 e 4 dias, ou seja, 24, 48, 72 e 96 horas. As inicializações do BRAMS são realizadas a partir das análises do modelo global T254L64 do NCEP/NOAA (*National Centers of Environmental Predictions/National Oceanic and Atmospheric Administration*).

O modelo assimila dados iniciais de umidade do solo e dados de temperatura da superfície do mar (TSM) média semanal, topografia e vege-

tação. (FREITAS J. ET AL., 2011)

## 2 REVISÃO BIBLIOGRÁFICA

Todo cientista que trabalha com séries de dados meteorológicos é confrontado com o problema da falta dos mesmos nas séries em muitas ocasiões. Alguns métodos para a análise desses dados podem ser adequados a essa situação, mas a maioria deles exige que a série esteja completa.

Um dos fornecimentos essenciais ao estudo e conhecimento de um fenômeno meteorológico são os dados observados. A partir do seu tratamento e interpretação, é possível planejar processos, formular e criar alternativas de solução, conseguindo finalmente, diminuir o grau de incerteza na tomada de decisões. Na medida em que os dados possam fornecer registros periódicos que dão continuidade às observações, vai ser possível obter resultados mais próximos da realidade do fenômeno que está sendo estudado. Porém, existem situações que causam a falta de continuidade nas bases de dados disponíveis, que podem afetar a qualidade das variáveis em estudo, obtendo séries de dados descontínuos que limitam grandemente os resultados que vão ser usados. Isso é o que acontece com as estações meteorológicas, que devido a situações como: equipamentos não calibrados, deterioração em unidades de medida, medidas inesperadas ou deficiências no procedimento de observação e medição, entre outros, causam a ausência de dados nas séries.(RODRIGO; FERNÁNDEZ, 2014)

Para um melhor entendimento das pesquisas realizadas acerca do preenchimento de dados faltantes em séries de dados meteorológicos, a análise foi dividida em três grupos:

- pesquisas para estimar dados faltantes em séries temporais que usam os dados de mais de uma estação meteorológica;
- pesquisas para estimar dados faltantes em séries temporais que usam os dados da mesma estação meteorológica;
- pesquisas para estimar dados faltantes em séries temporais que usam o método preditivo auto regressivo.

## 2.1 Pesquisas para estimar dados faltantes em séries temporais que usam os dados de mais de uma estação meteorológica

Foi proposta por Infante; Ortega; Cedeño (2008), a aplicação de dois métodos complementares para a reconstrução de séries de dados de precipitações, para aplicá-lo nas estações meteorológicas da Venezuela com o objetivo de corrigir o problema da falha nas séries de dados meteorológicos.

O primeiro reconstrói a dinâmica e o tempo de atraso do sistema dinâmico da série temporal, e o segundo usa um modelo de redes neurais para estimar os dados faltantes. Os dados usados para testar a metodologia foram os das séries mensais de chuva desde 1971 até 2000 de 36 estações do serviço meteorológico da Força Aérea de Venezuela. O primeiro método vem da teoria dos sistemas dinâmicos, e procura compreender o comportamento do sistema gerado pelas séries de dados de chuva, através da reconstrução dos estados do sistema. Com o segundo método se propõe usar um modelo de redes neurais multicamadas, que são capazes de imitar o comportamento da série temporal através de algoritmos recursivos.

Com essa pesquisa conseguiu-se utilizar modelos de redes neurais para estimar os dados ausentes em séries de tempo irregulares, obtendo-se uma reconstrução robusta, levando em consideração a relação espaço-tempo da série de precipitação mensal. Além disso, na maioria das estações foi usado entre 2 e 10 neurônios para obter o menor erro de estimativa. O erro quadrático médio foi estimado para validar o modelo, obtendo resultados favoráveis.

Usando a “*Transformada de Wavelets*” (TW), Carbajal M. et al. (2010) descreveram a aplicação de um método para o preenchimento de dados faltantes de precipitação diária. Foram usadas séries completas de dados de precipitação diária para os períodos 1999-2000, 2003-2004, 2004-2005 de 6 estações meteorológicas da Montanha peruana.

O processo consistiu na decomposição e reconstrução do sinal em dois níveis, utilizando o *Wavelet Haar*. O programa desenvolvido usa o sinal que resulta do filtro de baixa frequência, vindo da estação a ser corrigida e o sinal de alta frequência proveniente de uma estação vizinha, que fornece o componente aleatório que é usado no preenchimento dos dados faltantes.

A seleção da estação vizinha foi baseada na semelhança da distribuição cumulativa de ambos os sinais previamente padronizados.

O procedimento foi validado, gerando lacunas aleatórias, preenchendo-as com a *Transformada Wavelets* ou o gerador de dados climáticos ClimGen e comparando os sinais reconstruídos com o sinal original sem alteração.

Eles são provados em todos os casos e o método baseado no TW produziu melhores coeficientes de determinação do que o ClimGen. Os estimadores estatísticos das distribuições descobriram que os sinais reconstruídos e originais são semelhantes.

Depois da análise feita por Carbajal M. et al. (2010), foi concluído que o método baseado na TW, é o mais apropriado para a estimativa dos dados diários de precipitação faltantes, pois além de gerar dados muito mais próximos dos reais, a geração dos parâmetros para a reconstrução precisa de poucos dados em comparação a outros métodos convencionais. Porém, esse modelo é uma alternativa eficiente para estimar dados faltantes só quando há informação do mesmo período na mesma escala temporal e com probabilidades semelhantes de ocorrência, provenientes de outra estação próxima.

Foi apresentada, por Ventura T. M. (2012) uma abordagem computacional que utiliza a técnica de redes neurais, combinada com algoritmos genéticos, aplicada a dados reais, com o objetivo de preencher falhas (valores ausentes) em séries de dados de estações micrometeorológicas em uma região de cerrado no estado de Mato Grosso. Os valores utilizados no preenchimento das falhas são estimados por outros sensores obtidos no mesmo local e momento da falha detectada baseando-se nos valores de outras séries de dados mensurados.

Para avaliar os resultados, foram analisados: o coeficiente de correlação entre os dados reais e os dados estimados, o erro médio absoluto e o desvio padrão dos dados.

Os resultados mostram que a proposta apresentou bons resultados para corrigir séries de dados com pequeno número de erros, mas também foi eficiente quando a série de dados tinha uma grande parte de falhas. Uma das vantagens desta metodologia é que ela pode ter uma boa precisão nos resultados mesmo que as falhas estejam em sequência, ou seja, as falhas nas séries de dados ocorram durante horas, dias ou



até mesmo meses.

Fi feita por Toro A. et al. (2015), uma pesquisa com o objetivo de determinar a confiabilidade dos seguintes métodos de preenchimento de séries de dados meteorológicos:

- *U.S National Weather Service (WS)*;
- dedutivo racional (RD);
- a regressão múltipla (RM);
- a regressão linear (RL).

Determinaram qual deles era o melhor para preencher os dados faltantes nas séries de dados de precipitação, temperatura máxima e mínima no período 2006-2009 das estações na zona norte do eixo bananeiro de Urabá Antioqueño. A estimativa do erro foi feito da seguinte forma: raiz quadrada do quadrado médio do erro (RCCME), coeficiente de determinação (R2), erro médio absoluto (MAE), erro relativo (RE) e índice de concordância de Willmott (d). Os métodos RM e RL apresentaram RCCME e MAE semelhantes aos de WS, o que levaria a erros semelhantes. Dado que o R2 desses métodos foram menores que 0,8 e o índice (d) para WS foi maior ou semelhante ao resto dos métodos, determinaram que o WS fosse usado para o preenchimento das séries.

Com a pesquisa feita por Fante; Sant'anna (2016), foi possível comparar e constatar a eficiência na utilização de técnicas geoestatísticas para a estimativa de dados de temperatura faltantes. Por meio de testes feitos pelas técnicas: regressão linear, ponderação regional, técnica de Alfonsi; Pinto; Pedro JR (1974), e técnica de Pinto e Alfonsi com compensação de dados de estação apoio Fante K. (2014), o artigo buscou avaliar quais os métodos que apresentaram os melhores resultados e que simularam os valores mais próximos do real. Foi utilizado para esta análise o conjunto de dados mensais de temperatura máxima e mínima da estação meteorológica de Piracicaba/SP (ESALQ/USP), entre os anos de 2001 e 2011. Os melhores resultados foram alcançados pela ponderação regional, tanto para as temperaturas máximas como para as mínimas, demonstrando a eficiência e confiabilidade na estimativa de dados climatológicos faltantes por técnicas estatísticas. Em seguida, ressalta-se a técnica consagrada por Pinto e Alfonsi com compensação de dados de estação apoio, a

técnica de ALfonsi; Pinto; Pedro JR (1974) e por último, a técnica de regressão linear que apresentou os resultados menos confiáveis, os mais discrepantes do real.

A teoria afirma que, nos métodos para estimar dados faltantes que usam estações de referência, elas devem estar o mais próximo possível, porém, esse critério não se aplica para as estações que estão em regiões de montanha, pois a proximidade das estações não implica a mesma condição topoclimática.(MEDINA; MONTOYA; JARAMILLO, 2008)

Então, às vezes não há estação meteorológica que possa ser usada como referência para fazer a estimativa de dados faltantes na estação que está sendo estudada. Nesses casos, só podem ser usados os dados da mesma estação.

## **2.2 Pesquisas para estimar dados faltantes em séries temporais que usam os dados da mesma estação meteorológica**

Para estimar os dados inexistentes em uma série de dados meteorológicos, usando os dados da mesma série, podem ser usados métodos univariados. Esses métodos levam em consideração a possível correlação entre os dados da série (autocorrelação) ou o comportamento físico da variável (modelagem física).

Segundo Rodrigo; Fernández (2014), uma vez que a autocorrelação da série é estudada, os dados inexistentes podem ser substituídos basicamente por alguns métodos simples:

**critério dos dados anteriores e posteriores:** consiste em substituir o valor faltante pelo registro anterior ou posterior, dependendo da escala de tempo. Foi concluído nesse estudo que este método só pode ser aplicado a umidade relativa.

**critério do valor médio:** consiste em substituir o valor faltante pelo valor médio da série e, só é útil usá-lo em séries com desvios típicos muito menores a seu valor médio. Foi concluído que este método só pode ser usado para o preenchimento de séries de dados de pressão e umidade relativa a escala mensal.

**critério de diferenças:** consiste em substituir o valor mensal faltante pelo valor do

mês anterior, adicionando o valor médio das diferenças entre o mês anterior e o próprio mês. Foi concluído que este método pode ser apropriado para o preenchimento de séries de dados mensais da umidade relativa.

Para estimar estatisticamente os valores faltantes em séries de dados de chuva, Medina; Montoya; Jaramillo (2008), propuseram um método de fácil aplicação, fundamentado e estatisticamente validado, baseado em séries de dados de precipitação acumulada por ciclos em períodos mensais, agrupados de acordo com a condição do El Niño, La Niña ou Neutro. Com o método proposto, o erro de estimativa relativa foi menor de 30% quando o usuário tem até quatro dados faltantes, independentemente da condição. Além disso, a estimativa dos dados se ajusta às características de cada estação e sua estimativa só depende da sua história.

Em geral, as etapas do método, propostas para estimar a falta de dados de precipitação, são as seguintes:

1. Identificar o grupo ao qual pertence o mês com o dia onde está faltando o dado: El Niño, La Niña ou Neutro.
2. Obter os acumulados de chuva (X) sem levar em conta o dado faltante.
3. Se a probabilidade de chuva no dia é menor ou igual a 20%, então a estimativa do dado faltante é “zero”.
4. Se a probabilidade de chuva é menor de 20%, e X é descritivamente maior que o limite superior do intervalo do acumulado histórico do mês e do grupo, respectivamente, então a estimativa do dado faltante é “zero”.
5. Se a probabilidade de chuva é maior de 20%, e X não é descritivamente maior que o limite superior do intervalo do acumulado histórico do mês e do grupo, respectivamente, então a estimativa do dado faltante é “zero”.

Foi concluído que o erro não depende do número de dados faltantes na série e o erro relativo médio de estimativa é menor de 30%.

### 2.3 Pesquisas para estimar dados faltantes em séries temporais que usam o método preditivo auto regressivo

Para a estimativa de dados faltantes nas séries de dados meteorológicos, segundo Ulrich; Clayton (1976) e Ulrich; Bishop (1975), pode ser usado um modelo preditivo auto regressivo conhecido como “ $AR(p)$ ”, o qual é um modelo linear que usa os valores de  $p$  tempos anteriores e posteriores de amostragem para estimar o valor em um determinado momento. Esse método pode recuperar o sinal estacional e aqueles outros sinais cuja persistência no tempo sejam compatíveis com o tempo da amostra.

Além disso, estes filtros possuem a propriedade, pelo princípio de Máxima Entropia, que os valores calculados são consistentes com as propriedades estatísticas da série, sem incluir suposições externas aos dados, ainda que a informação ausente tenha se perdido, os valores estimados são consistentes estatisticamente com o resto da série.

A estimativa  $AR$  foi originalmente desenvolvida para o processamento de dados geofísicos, onde foi denominado Método de Entropia Máxima (MEM). Foi utilizado para aplicações em radares, imagens, radioastronomia, biomedicina, oceanografia e sistemas ecológicos. (STEVEN; MARPLE, 1981)

Foram apresentadas por Alfaro; Soley (2009), duas metodologias para o preenchimento de dados ausentes nas séries de dados geofísicos:

- a primeira é baseada na decomposição de componentes principais da matriz de correlação de dados da mesma variável entre estações próximas e em períodos de tempo comuns. Este método multivariável permite incorporar os fenômenos de maior escala nos valores preenchidos a partir das informações das estações próximas;
- o segundo método pode ser usado quando não houver estações próximas e o preenchimento deve ser feito com as informações da mesma estação. Consiste em ajustar um modelo auto regressivo de ordem  $p$  ou  $AR(p)$  às séries temporais e usar esse modelo para estimar os dados faltantes. Dois algoritmos foram avaliados para calcular os coeficientes auto regressivos: o estimador de Burg Burg R. (1967) e Burg R. (1968)

e o proposto por Ulrich; Clayton (1976).

O primeiro é apropriado para processos estocásticos e o segundo, para séries determinísticas.

As duas metodologias descritas nesse artigo são recursivas: é feita uma primeira estimativa dos dados ausentes executando o algoritmo, ignorando os dados que faltam, se possível, ou aproximando-os rudemente. Em seguida, o algoritmo é executado com os novos valores que substituem os valores estimados na execução anterior. Esse ciclo termina quando a diferença máxima dos valores estimados entre duas execuções sucessivas é inferior a um valor estabelecido com antecedência pelo usuário.

Em ambos métodos, a estimativa dos dados perdidos não alterou as propriedades estatísticas das séries temporais.

Baseados na pesquisa feita por Alfaro; Soley (2009), foi proposta também uma combinação dos dois métodos desenvolvidos por Ureña; Alfaro; Soley (2016). Foi usado o método auto regressivo como uma primeira aproximação da estimativa dos dados faltantes e depois é usada de forma iterativa pelo método das componentes principais. Foi feita uma comparação deste método integrado com o método originalmente proposto por Alfaro; Soley (2009) através do cálculo do erro quadrático médio, o erro absoluto médio e a variância. A análise desses resultados permitiu concluir que a rotina integrada gera menores erros e economiza tempo de trabalho do usuário.

Para o cálculo dos coeficientes gerados pelo modelo auto regressivo, os métodos mais usados são o método de Yule-Walker (YW) e o Método de Burg (RODRÍGUEZ G., 1995). A diferença principal destes métodos é a forma de estimar a sequência de autocorrelação da série temporal estudada. No método de Yule-Walker, a sequência de auto correlação é obtida de forma independente, usando o estimador distorcido, mas o uso do mesmo implica em admitir que os dados fora do intervalo de observação são zero, pressuposto que viola o princípio da entropia máxima.

Com o objetivo de resolver este problema, Burg (BURG R., 1967, 1968) propõe diminuir diretamente o erro de estimativa. O erro de estimativa é considerado como a soma das estimativas dos erros diretos e inversos. A diminuição desses erros juntos, leva à obtenção de um operador de previsão de erro de atraso mínimo e, implicitamente, à estimativa de uma sequência de autocorrelação semidefinida positiva (SMYLE; CLARKE; ULRYCH, 1973). Assim, não é preciso conhecer a função de au-

to correlação com antecedência.

A aplicação dos modelos auto regressivos à geofísica teve considerável sucesso. O método foi originalmente proposto por Burg (BURG R., 1967, 1968), que apresentou a formulação resultante da aplicação da consideração da entropia às determinações espectrais, mas também um método para calcular os coeficientes de erro de previsão requeridos. O modelo *AR*, usando o método de Burg, foi aplicado por Ulrich T. J. (1972), que mostrou as propriedades de resolução notáveis desta abordagem. (ULRICH; BISHOP, 1975; ANDERSEN N., 1974)

Foi demonstrado numericamente por Ulrich; Bishop (1975) que, para séries temporais curtas, a variância da autocovariância estimada por Burg é maior do que a estimada por YW, e não diminui para zero com o aumento do atraso. As propriedades de resolução do método de Burg são consideravelmente superiores às da técnica de YW.

Um resumo com algumas técnicas de análise de espectro de séries de tempo discreto foi desenvolvido por Steven; Marple (1981). Algumas técnicas discutidas foram a autoregressiva (entropia máxima), média móvel e média móvel auto regressiva. Eles apresentam uma tabela de resumo com uma visão geral de todos os métodos, incluindo as referências principais e equações apropriadas para o cálculo de cada estimativa espectral.

### 3 MATERIAIS E MÉTODOS

Neste trabalho será feita uma estimativa dos dados faltantes nas séries de dados meteorológicos. Para isso vai ser implementado um modelo preditivo auto regressivo de ordem  $p$ ,  $AR(p)$ , junto com o método de Burg para o cálculo dos coeficientes gerados no  $AR(p)$ .

#### 3.1 Estação de estudo

A Estação Meteorológica usada para este trabalho coleta, a cada 10 minutos, os dados das variáveis: chuva acumulada [ $mm$ ], temperatura do ar [ $^{\circ}C$ ], direção do vento [ $graus$ ], velocidade do vento [ $m/s$ ], pressão atmosférica [ $hPa$ ] e umidade relativa [% $UR$ ]. Foi usada a informação no período desde 2009 até 2016 (informação disponível desde a instalação da mesma até a hora de começar o estudo para a geração do modelo). Estes dados não estão disponíveis para serem usados por qualquer pessoa, devem ser solicitados ao pessoal responsável com a justificativa para a qual serão usados. Mas os dados da estação podem ser visualizado no site do IPRJ/UERJ em tempo real no endereço: <http://www.clima.iprj.uerj.br/>.

#### 3.2 O Framework Multiplataforma Qt

O *framework* multiplataforma Qt é uma tecnologia em expansão que nos fornece um conjunto de ferramentas e elementos gráficos para a criação de interfaces e aplicativos. Atualmente, tem um grande sucesso e uma ótima implementação em diferentes áreas, desde aplicativos de desktop até sistemas eletrônicos industriais e integrados. (GUTIÉRREZ D., 2009)

O Qt esteve disponível pela primeira vez para a população em 1995. Foi desenvolvido por dois engenheiros noruegueses, Haavard Nord e Eirik Chanble-Eng, em resposta à necessidade de ter uma GUI (acrônimo em inglês de Interface Gráfica do Utilizador) para um aplicativo C++ multiplataforma orientado a objetos. A empresa foi fundada em 1994 como *Quasar Technologies*, e evoluiu para *Trolltech*. Esta empresa começou a oferecer o Qt com a licença GPL (acrônimo em inglês de Licença Pública

Geral) a partir do ano 2000, sendo gratuito para o desenvolvimento de *software* livre, mas de pagamento para o desenvolvimento de *software* privado. Depois da compra da *Trolltech* pela *Nokia* no início de 2009, a Qt começou a ser oferecida com a licença LGPL (acrônimo em inglês de Licença Pública Geral Menor). O objetivo disso foi popularizar e ampliar o Qt baseado em permitir sua utilização em programas proprietários, pois não será necessário pagar licença nenhuma para sua utilização. (GUTIÉRREZ D., 2009)

Vantagens de programar com C++ em Qt: (MENEZES A., 2009)

- desenvolvimento Multiplataforma;
- programação C++ mais amigável com Qt;
- implemente uma vez;
- compilação em qualquer lugar;
- aplicações KDE são feitas com Qt;
- criação de interfaces gráficas elegantes e amigáveis;
- utilização de uma API rica e útil;
- licenças: Comercial, LGPL e GPL.

### 3.3 O modelo preditivo auto regressivo

Foi escolhido o modelo preditivo auto regressivo porque não se contava com os dados de nenhuma estação perto da estação em estudo nas datas que faltava informação. Uma propriedade dos filtros preditivos é que podem recolher sinais nas quais a persistência é comparável à dimensão do filtro. Também, pelo princípio de Máxima Entropia, os valores calculados são consistentes com as propriedades estatísticas da série, sem incluir pressupostos externos aos dados; ou seja, ainda que a informação em falta já não exista, os valores estimados são estatisticamente consistentes com o resto da série. (ALFARO; SOLEY, 2009)

Teoricamente, o método deve convergir para as séries de tempo meteorológicas, mas na prática, pode divergir por razões externas como: a acumulação de



erros de truncamento, erros na coleta dos dados, o uso de um número de coeficientes auto regressivos incongruente com a persistência da série, ou um alto percentual de dados em falta.

A primeira aproximação usada para estimar os dados em falta foi a média dos valores existentes na série de dados.

Este método poderia adicionar efeitos de pequena escala e seria útil só para períodos de tempo pequenos. (ALFARO; SOLEY, 2009)

O modelo auto regressivo de ordem  $p$ ,  $AR(p)$ , obedece à Equação 1:

$$y[t] = \sum_{i=1}^p a_i y[t - i] + \varphi[t] \quad (1)$$

Onde a saída no tempo  $t$  depende dos  $p$  valores anteriores mais um valor  $\varphi$ , o qual corresponde ao ruído introduzido nos cálculos. Então o filtro preditivo correspondente é mostrado na Equação 2:

$$y_b[t] = \sum_{i=1}^p a_i y[t - i] \quad (2)$$

O método também executa o filtro em tempo reverso, então agora o dado no tempo  $t$  se estima com os  $p$  valores futuros da série (Eq. 3):

$$y_f[t] = \sum_{i=1}^p a_i y[t + i] \quad (3)$$

Note que ambos filtros preditivos são executados dentro dos dados, sem sair dos limites da série. Então, o filtro que usa os  $p$  valores anteriores não produz saída para os  $p$  primeiros valores da série, e o filtro que usa os  $p$  valores posteriores não produz saída para os  $p$  últimos valores da série.

Porém, para os valores dos extremos, é usada a única saída disponível, e para os valores do meio, é usada a média das duas saídas, como mostrado nas Equações

4, 5 e 6:

$$y_{est}[t] = y_f[t] \quad \text{para} \quad 0 \leq t \leq p \quad (4)$$

$$y_{est}[t] = y_f[t] + y_b[t] \quad \text{para} \quad p < t \leq N - p \quad (5)$$

$$y_{est}[t] = y_b[t] \quad \text{para} \quad N - p < t \leq N \quad (6)$$

Onde:

**N:** quantidade de dados da série

**p:** ordem do modelo

$y_{est}$ : série estimada

$y_b$ : filtro preditivo que usa os  $p$  valores anteriores da série

$y_f$ : filtro preditivo que usa os  $p$  valores posteriores da série

### 3.3.1 O Método de Burg

O cálculo dos coeficientes ( $a_i$ ) de um modelo  $AR(p)$  pode ser feito através de vários métodos. Os mais usados segundo Rodríguez G. (1995) são: o método de Yule-Walker e o método de Burg.

Na nossa pesquisa decidiu-se usar o Método de Burg, o qual é usado para o tratamento de séries determinísticas.

Foram descritas por Rodríguez G. (1995) as principais vantagens e desvantagens do método de Burg em comparação com os métodos convencionais.

**Vantagens:**

- não precisa do uso de janelas que modifiquem a série original, dado que a extrapolação da função de auto correlação maximizando a entropia evita as discontinuidades nos extremos;
- não são requeridos procedimentos de suavização da série para diminuir a variação das estimativas;
- a extrapolação de entropia máxima da função de autocorrelação dá origem a uma função de densidade espectral que é uma função contínua, proporcionando assim uma melhoria significativa na resolução de frequência.

**Desvantagens:**

- o principal problema deste método é a seleção da ordem do modelo  $AR(p)$  que precisa se ajustar, problema que é muito frequente nas técnicas de análise espectral;
- o método exibe uma divisão de linhas espectrais para altas relações sinal-ruído;
- para ordens do modelo muito altos, o método também pode introduzir picos falsos;
- para sinais sinusoidais com alto nível de ruído, o método de Burg é sensível para a fase inicial de uma sinusóide, especialmente em registros de dados curtos.

Para o cálculo dos coeficientes do modelo através do Método de Burg, foi usado o algoritmo desenvolvido por Andersen N. (1974), onde pode ser encontrado um desenvolvimento matemático rigoroso do método. Neste trabalho será apresentado, somente o pseudocódigo usado na implementação do método dentro do modelo.

### 3.4 Método para a estimativa de dados faltantes em séries de dados meteorológicos

A estimativa de dados faltantes em séries de dados meteorológicos é um dos problemas que enfrenta a meteorologia hoje, pois qualquer estudo climatológico precisa de uma série de dados completa e estável. Com esta pesquisa nos propomos estimar os valores faltantes nas séries de dados meteorológicos usando os dados da mesma estação, mas sempre deve ser levado em consideração que os valores obtidos não são mais do que uma estimativa, e para serem usados devem ser analisados por um especialista que valide que estão dentro da faixa de valores estabelecidos para a variável que está sendo estudada.

#### 3.4.1 Algoritmo para a modelagem da estimativa de dados faltantes

Serám mostrados os pasos seguidos para o desenvolvimento do modelo:

1. é preciso ter a série de dados num arquivo de texto que possa ser lido pelo programa (\*.txt) (Estrutura do arquivo mostrado no Apêndice A);
2. detectar onde faltam dados e substituí-los pelo código '9999';
3. fazer a primeira aproximação do método, que em nossa pesquisa foi feita com o valor médio dos valores da série toda;
4. aplicar o método preditivo auto regressivo;
5. imprimir os resultados obtidos num arquivo de texto para a sua futura utilização (\*.txt ou \*.csv).

#### 3.4.2 Algoritmo para a modelagem do método preditivo auto regressivo

Fazer enquanto a diferença máxima entre duas iterações consecutivas seja maior do que um valor definido com antecedência pelo usuário, ou chegar a uma quantidade de iterações também definida pelo usuário com antecedência.

1. calcular os coeficientes do método;
2. calcular os filtros preditivos anterior e posterior (Equações 2 e 3);
3. definir o vetor de valores estimados (Equações 4, 5 e 6);
4. redefinir o vetor de valores estimados. Este passo foi feito substituindo os valores estimados somente nas posições dos dados faltantes na série com os dados meteorológicos originais;
5. calcular a diferença máxima entre a iteração atual e a anterior.

### 3.4.3 Algoritmo para o cálculo dos coeficientes do método

1. Inicializar os vetores auxiliares  $b_1$  (Eq. 7) e  $b_2$  (Eq. 8). Estes vetores vão ter um tamanho de  $N - 1$ , onde  $N$  é a quantidade de valores da série meteorológica onde estão sendo estimados os dados faltantes.

$$b_1[i] = y[i] \quad (7)$$

$$b_2[i] = y[i - 1] \quad (8)$$

#### ***Fazer enquanto m seja menor que a ordem do modelo***

2. Calcular o coeficiente correspondente à ordem da iteração através da Equação 9:

$$a_m = \frac{\sum_{i=1}^{N-m} b_1[i]b_2[i]}{\sum_{i=1}^{N-m} (b_1^2[i] + b_2^2[i])} \quad (9)$$

3. Se:  $m > 1$

Recalcular os coeficientes anteriores aos da presente iteração através da Equação 10:

$$a[i] = a_{anterior}[i] - a_m[i]a_{anterior}[m - 1] \quad (10)$$

4. Aumentar o iterador  $m$

5. Fazer o vetor dos coeficientes anterior igual ao vetor dos coeficientes na iteração:

$$a_{anterior}[i] = a_m[i] \quad (11)$$

6. Recalcular os vetores auxiliares  $b1$  (Eq. 12) e  $b2$  (Eq. 13):

$$b1[i] = b1[i] - a_{anterior}[m - 1]b2[i] \quad (12)$$

$$b2[i] = b2[i + 1] - a_{anterior}[m - 1]b1[i + 1] \quad (13)$$

### 3.5 A ordem do modelo $AR(p)$

Selecionar a ordem do modelo é uma parte muito importante no processo da modelagem do método preditivo auto regressivo. Se o sinal estudado é parte de um processo  $AR(p)$  puro, todos os coeficientes  $a(i)$  para  $i > p$  são maiores que zero, existindo uma só solução para a ordem do modelo certo. Por outro lado, a maioria dos processos reais que ocorrem na natureza, não são processos  $AR(p)$  puros, então para cada processo é preciso selecionar a ordem certa do modelo.

Na prática trabalhamos com séries de dados finitas, então selecionar a ordem do modelo muito grande (próximo ao tamanho da série) leva a um aumento do erro de estimativa dos parâmetros do modelo. (RODRÍGUEZ G., 1995)

Existem alguns critérios para estimar a ordem do modelo. Entre os mais usados estão: Critério do Erro de Predição Final, Critério de Informação do Akaike, Critério da Função de Transferência Auto regressiva e Critério dos Coeficientes de Correlação Parcial. Segundo Rodríguez G. (1995). Os resultados experimentais obtidos por diferentes autores indicam que a seleção da ordem, usando estes critérios, não oferece resultados confiáveis. Ainda que esses métodos sejam uma ajuda para a seleção da ordem, o valor de  $p$  é geralmente determinado empiricamente.

Na nossa pesquisa, a ordem do modelo foi determinada empiricamente. Para determinar quais valores da ordem foram as que obtiveram melhores estimativas, foi calculados o Erro Quadrático Médio (Eq. 14) e a diferença máxima entre os dados estimados e os dados originais da série.

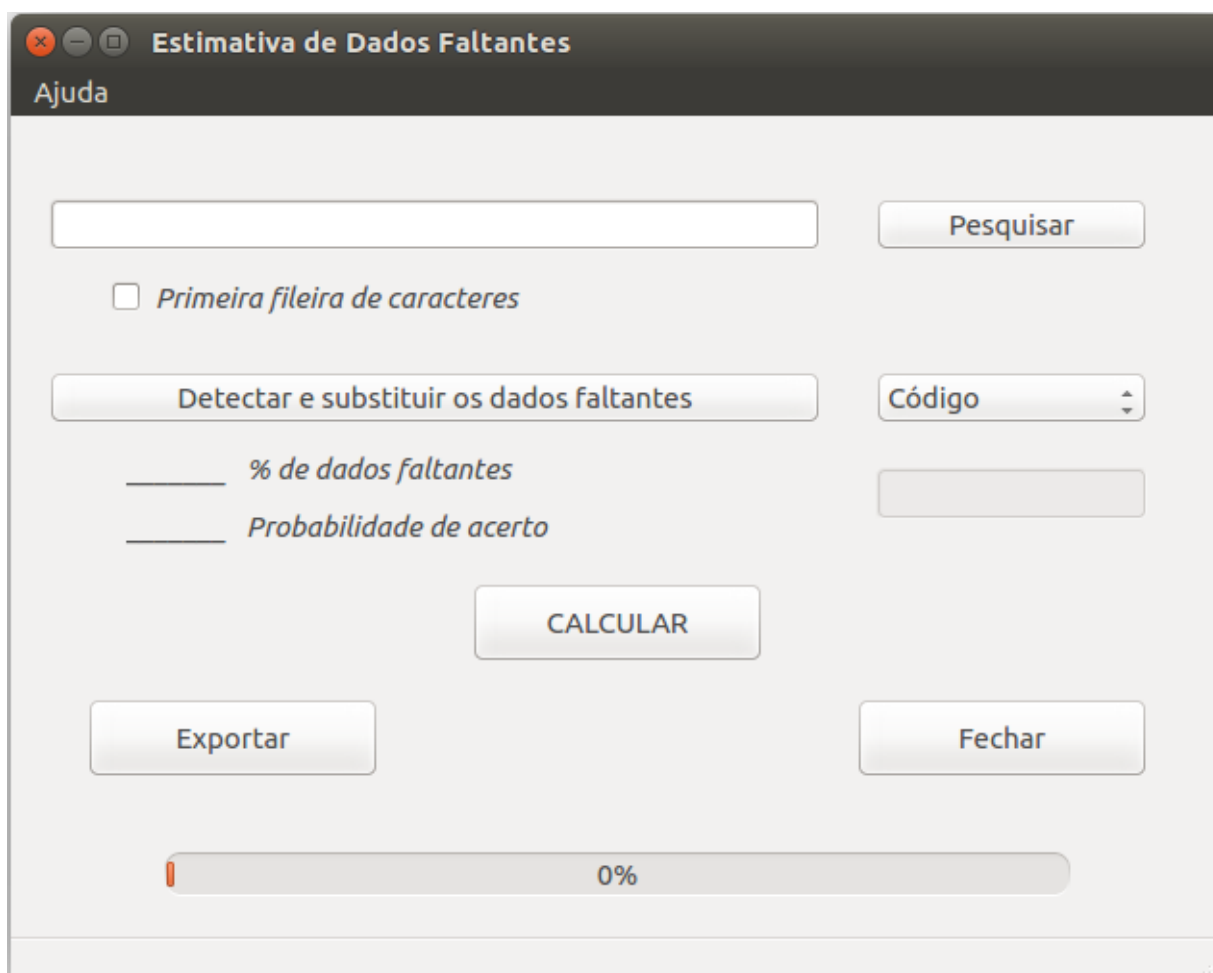
$$EQM = \frac{1}{N} \sum_{i=1}^N (y_{est}[i] - y[i])^2 \quad (14)$$

Na modelagem do método, este erro foi calculado com todas as ordens possíveis, para ser devolvido o resultado com o menor valor. Este erro foi calculado entre os valores estimados e os correspondentes valores reais que temos.

## 4 RESULTADOS E DISCUSSÃO

Como resultado deste trabalho foi criado um modelo para estimar os valores faltantes nas séries de dados meteorológicos, usando somente os dados da mesma estação. Para desenvolver este modelo foi usada a linguagem de programação C++, por meio do *framework* multiplataforma Qt, o qual permitiu fazer a interface gráfica da Fig. 2 para facilitar seu uso. Desta forma o modelo pode ser usado por pessoas que não necessariamente precisam ter um amplo conhecimento científico.

Figura 2 – Interface gráfica do modelo



Fonte: O autor, 2018.



## 4.1 Interface gráfica

**Ajuda:** Para o uso do modelo é preciso um arquivo de dados com a informação meteorológica, e esse arquivo precisa ter entre as primeiras colunas: a data (dividida por colunas em: Ano, Mês e Dia) e a hora das observações (dividida em horas e minutos). Também é preciso que seja conhecida a quantidade de colunas com dados meteorológicos e a informação que haverá em cada uma delas.

Os números devem usar o ponto como separador decimal, e não devem ter caracteres especiais, só números. A primeira fileira pode ser definida como de caracteres para não ser usada nos cálculos do modelo, marcando a caixa embaixo do quadro onde será pesquisado o arquivo com a série de dados.

No Menu “Ajuda”, o usuário poderá encontrar tudo o que precisa saber para o uso do modelo.

**Pesquisar:** Através desse botão é possível procurar o arquivo dos dados meteorológicos onde precisam ser procurados e estimados os dados faltantes. Ao dar click no botão vai sair uma janela na pasta principal do ordenador, desde onde pode ser procurado e selecionado o arquivo.

**Detectar e substituir os dados faltantes:** Ao dar click nesse botão vai ser executado o programa que detecta e substitui os valores faltantes na série de dados meteorológicos. Os valores faltantes neste passo vão ser substituídos pelo código definido pelo usuário embaixo desse botão. O código pode ser “9999”, “Nan” ou o usuário poderia definir um novo código na seção “Agregar Novo Código”. Ao seleccionar a opção “Agregar Novo Código”, vai ser habilitado um quadro de texto embaixo dos códigos onde o usuário poderá escrever o novo código a ser usado para preencher as falhas nas séries de dados meteorológicos.

Logo depois de clicar, aparecerão à direita do botão, o percentual de dados faltantes e a probabilidade de acerto na previsão que vai ser feita.

A probabilidade de acerto na previsão diminui na medida que aumenta o percentual de dados faltantes.

**Calcular:** Com esse botão será executado o modelo com o método que permite estimar os valores faltantes nas séries. Na medida que se esteja fazendo o cálculo, o progresso será mostrado numa barra de progresso no final da janela. O modelo foi explicado no capítulo de Materiais e Métodos.

**Guardar:** Ao clicar nesse botão, poderão ser guardados os valores obtidos na modelagem do método, para arquivos de texto tipo \*.txt e \*.csv . Para isto surgirá uma nova janela onde o usuário poderá indicar o endereço onde quer guardar o arquivo gerado com os dados estimados.

**Fechar:** Através desse botão poderá ser fechada a janela. Antes de fechar, se não foram guardados os resultados do modelo, aparecerá uma outra janela perguntando se o usuário quer guardar o resultado dos cálculos feitos. Nesse momento o usuário poderá optar por “Aceitar” ou “Cancelar” a operação.

## 4.2 Os dados meteorológicos

A análise da informação da Estação Meteorológica localizada no IPRJ concluiu que os valores de chuva acumulada mensal variam de algumas dezenas até várias centenas de milímetros entre o período chuvoso (de janeiro até março e de outubro até dezembro) e o pouco chuvoso (de abril até setembro), com os máximos acumulados nos meses de janeiro, fevereiro e dezembro (Apêndice B).

A temperatura do ar tem uma média entre 14 e 15 °C nos meses de inverno e entre 20 e 21 °C nos meses de verão (Apêndice C), obtendo-se o menor valor (3,91 °C) no dia 4 de junho de 2009 às 6:40 horas e o maior valor (35,55 °C) no dia 25 de setembro de 2015 às 13:30 horas.

A pressão atmosférica teve uma mudança no mês de outubro de 2012 dado que a estação meteorológica mudou de lugar ao serem trasladadas as instalações do IPRJ para uma alocação diferente. Antes desta mudança, a variável tinha uma média entre 900 e 910 hPa, e depois mudou para valores entre 915 e 925 hPa (Apêndice D).

A Umidade Relativa média teve valores entre 80 e 100 %, sendo maior nos meses de abril a agosto, coincidindo com o período mais frio do ano (Apêndice E).

A análise da quantidade de dados faltantes na estação meteorológica concluiu que os anos com maior quantidade de dados faltantes foram 2016 e 2013, com 72,25% e 64,33% dos dados ausentes respectivamente. O ano 2011 foi o único em que não faltou dado nenhum, e depois dele, o ano com menor quantidade de dados ausentes foi o 2014 com o 3,17% de falhas.

Dado que o ano 2011 não teve dados faltantes e também não foi preciso fazer ajuste nenhum nos dados, é dizer, não tinha nenhum dado com erro que fosse preciso tirar da série, foi decidido usar os dados deste ano para fazer os testes do modelo, tirando inicialmente alguns dados da série para serem estimados pelo modelo e poder comparar as estimativas feitas com os valores reais das variáveis.

Não foram usados os dados do mês de Janeiro dado que nos dias 11 e 12 desse mês se registrou uma anomalia nos dados de precipitação, causada por grandes acumulados de chuva. As chuvas na Região Serrana do Rio de Janeiro ocorreram devido ao desenvolvimento de profundas nuvens convectivas em uma atmosfera úmida e muito instável, associada à forte convergência de umidade gerada pelo escoamento na troposfera baixa. (ANDRADE; MOURA; PINHEIRO, 2011)

Para os testes do modelo foram usadas as variáveis “Chuva”, “Temperatura do Ar”, “Pressão Atmosférica” e “Umidade Relativa”, dado que são estas as variáveis que a Estação Meteorológica coleta.

#### 4.3 O modelo Preditivo Auto regressivo de ordem $p$ , $AR(p)$

Através da linguagem de programação C++ foi desenvolvido o modelo usando o filtro preditivo auto regressivo de ordem  $p$ ,  $AR(p)$ .

Se conseguiu gerar uma interface gráfica amigável através da qual se podem estimar os dados faltantes nas séries de dados meteorológicos. O código que gerou essa interface e o desenvolvimento do método foram disponibilizados no *GitHub*, para estar disponível a todo aquele que quiser usar, modificar ou adequar o método para atender às suas necessidades.

O *GitHub* é uma plataforma de hospedagem de código-fonte com controle de

versão usando o *Git*. Ele permite que programadores, utilitários ou qualquer usuário cadastrado na plataforma contribuam em projetos privados e/ou de código aberto de qualquer lugar do mundo.

Este método é usado quando não existe informação meteorológica perto da estação de estudo. É usada a informação da mesma estação para estimar os dados faltantes.

Tem como desvantagem principal que não estima variações por causa de eventos meteorológicos extremos ou de escala sinóptica.

#### 4.3.1 A ordem do método

Para determinar a ordem ótima do modelo, foi executado com as ordens desde 1 até a quantidade de dados da série que está sendo estudada. Ao fazer este cálculo se chegou à conclusão que quando a ordem é maior do que a metade da quantidade de dados, o método diverge, então esses valores não devem ser usados no modelo, e as melhores estimativas são obtidas com as ordens 1 e 2 do método.

Para obter o valor final da estimativa dos dados, o método é executado com todas as ordens que poderiam ser usadas no método. É calculado o Erro Quadrático Médio entre os dados obtidos pela estimativa do método e os dados originais da série (calculando apenas com a parte dos dados que o usuário possui).

Finalmente é devolvido como estimativa final, aquela onde é obtido o menor Erro Quadrático Médio.

#### 4.3.2 Validação do modelo

A validação do método foi feita com os dados da Estação Meteorológica localizada no campus do IPRJ/UERJ em Nova Friburgo.

Para fazer esta validação foram usados os dados do ano 2011 dado que este é o único ano que tem todos os dados completos. Dessa forma o modelo pode ser validado a través dos seguintes passos:

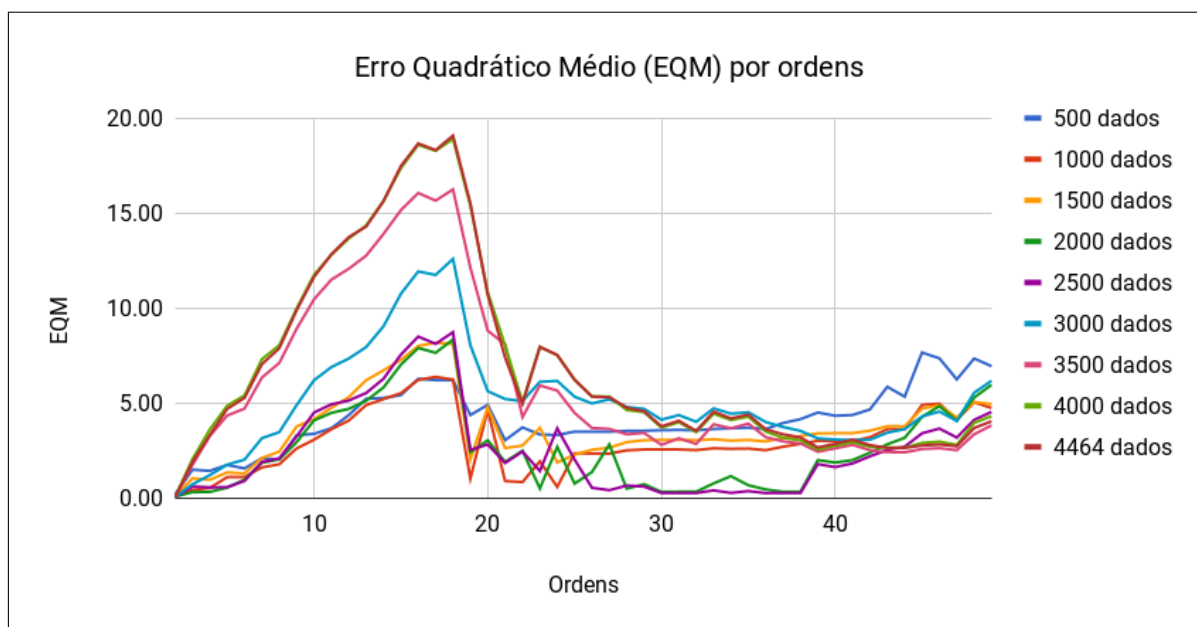
- apagar valores aleatoriamente da série;

- estimar os valores apagados usando o modelo;
- comparar a série de dados estimados pelo modelo com a série de dados originais.

Foi executado o modelo com valores entre 4464 (quantidade máxima de dados) e 500 dados com o mesmo percentual de valores faltantes, diminuindo de 500 em 500 dados.

Com isto se conseguiu concluir que a probabilidade de acerto na estimativa dos dados faltantes não depende da quantidade de dados, pois foram obtidos valores similares em cada execução para os mesmos valores da ordem do modelo. As melhores estimativas foram obtidas com as ordens 1 e 2, e as piores estimativas com as ordens entre 10 e 20 (Figura 3). No gráfico só foram representados os resultados para as ordens até 50, dado que para ordens maiores aparecem os picos falsos gerados porque o método preditivo auto regressivo é baseado na Transformada de Fourier.

Figura 3 – Erro Quadrático Médio por ordens



Fonte: O autor, 2018.

Foram usados 1000 dados de temperatura da série de dados meteorológicos, e foram gerados erros para testar o funcionamento do modelo.

Primeiro foram introduzidos valores em falta alternados com os valores reais, é dizer, uma posição sim e a outra não (50% de dados faltantes), primeiro começando

com o valor real e depois com o valor faltante. Foi executado o modelo com esses dados e, segundo os resultados obtidos, a melhor estimativa foi gerada com a ordem do modelo igual a 1, obtendo uma diferença máxima entre os valores estimados e os originais de 1,92°C para a série de dados que começa com o dado faltante, e 1,75°C para a série de dados que começa com o valor original da série. Para as restantes ordens, a diferença entre os dados estimados e os originais é maior do que 40°C.

O modelo foi executado com as ordens desde 1 até 500, faltando 500 dados consecutivos, começando pela primeira posição da série (50% de dados faltantes consecutivos). A melhor estimativa obtida pelo modelo com esta série de dados foi gerada com a ordem igual a 1 e foi obtida uma diferença máxima de 2,75°C entre os dados estimados e os dados originais. O erro quadrático médio obtido para essa estimativa foi 0,85. Com estes resultados conseguiu se concluir que com o 50% dos dados faltantes não é obtida uma boa estimativa do método.

Então a estimativa feita com o 50% de dados faltantes consecutivos, tem um erro muito alto e não é recomendado o uso do modelo para cálculos com esse percentual de dados faltantes consecutivos, começando pela primeira posição.

Depois, o percentual de dados faltantes consecutivos foi diminuído (Tabela 1) para chegar ao ponto onde o Erro Quadrático Médio (EQM) e a máxima diferença entre os valores estimados e os valores originais da série, tivessem um valor permissível para o uso do modelo. A quantidade de dados foi diminuída até só 1% dos valores faltantes começando na primeira posição da série. Com esses dados, a máxima diferença obtida foi de 1,82°C e o EQM foi 0,095. Ainda que este último valor tenha sido baixo, a máxima diferença entre os valores estimados e os valores originais da série, teve um valor não permissível (Tabela 2) para ser usado como substituição dos valores faltantes.

Pode-se concluir nesta parte, que para as estimativas feitas em séries de dados nas quais tem dados faltantes na primeira posição, o modelo não obtém estimativas ótimas para serem usadas pelos especialistas.

O modelo foi executado com valores faltantes consecutivos, começando na posição 3 da série, desde 1 valor faltante até 39, onde foi alcançada uma diferença máxima de 0,5 ° C entre os valores estimados e os valores originais da série, valor permissível para esta variável (Tabela 2).

É necessário fazer o esclarecimento de que este valor vai depender da variável que se esteja estimando. Neste caso está sendo estudada a temperatura do ar, e nesta pesquisa foi considerado que uma diferença máxima entre os valores estimados e os valores originais da série de 0,5 ° C é um valor permissível para usar os dados obtidos pela estimativa feita pelo modelo.

Tabela 1 – Máxima diferença entre os valores estimados e originais da série de temperatura do ar.

	50 %	40 %	30 %	20 %	10 %	5 %	4 %	3 %	2 %	1 %
1	2,7	2,6	2,3	2,2	1,9	1,9	1,9	1,8	1,8	1,8
2	17,7	17,7	17,7	17,7	17,7	17,7	17,7	17,6	17,6	17,4
3	45,8	45	43,7	44,7	42,7	27,9	17,1	16,9	16,4	16,5
4	30,5	29,1	26,9	29,2	25,8	17,3	17,0	16,7	18,6	277,4
5	20,7	21,3	20,5	20,8	19,6	17,3	17,9	17,9	18,6	574,7
6	21,1	21,2	20,7	19,7	16,6	18,7	18,3	18,1	17,2	212,4
7	21,4	21,3	21,1	20,5	15,2	18,6	20,3	19,8	17,2	491,7
8	21,4	20,8	20,3	20,7	16,7	18,5	22,9	21,3	678,9	627,4
9	22,4	21,6	21,7	22,6	18,1	16,5	19,0	21,1	10220,6	686,0
10	22,0	22,6	21,4	22,4	20,3	19,5	17,9	523,6	646,8	798,9

Nota: Dados para diferentes percentuais de valores faltantes consecutivos começando na primeira posição da série e para diferentes ordens do modelo

Fonte: O autor, 2018

Tabela 2 – Diferença máxima permissível entre os valores estimados e originais da série para cada variável.

Variável	Diferença máxima permissível
Chuva	0,01 mm
Temperatura do ar	0,5 ° C
Direção do vento	1,0 °
Velocidade do Vento	0,2 m/s
Pressão atmosférica	2,0 hPa
Umidade Relativa	2,0 %

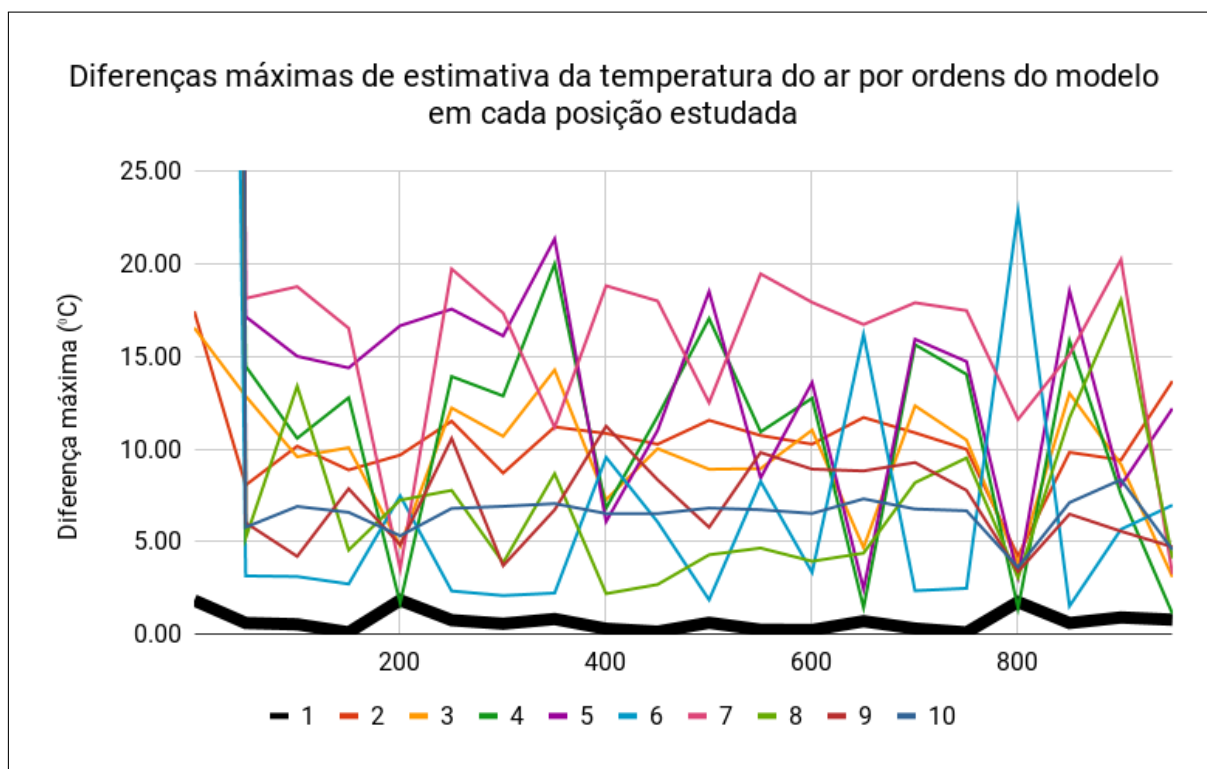
Fonte: O autor, 2018

Foi executado o modelo com a série de dados de temperatura faltando 10 dados consecutivos (Fig. 4 e 5), com as ordens de 1 até 500 em diferentes posições da série de temperatura do ar (nas posições 1, 51, 101, 151, 201, 251, 301, 351, 401, 451, 501, 551, 601, 651, 701, 751, 801, 851, 901 e 951). Na Figura 4 são mostrados

os resultados só com as ordens de 1 até 10 para um melhor entendimento do comportamento dos valores da diferença máxima entre os valores estimados pelo modelo e os valores reais da série.

Na Figura 4, cada série representa o comportamento das diferenças máximas para cada ordem do modelo (1 - 10) nas diferentes posições testadas. No eixo  $x$  se encontram as posições onde foram testados os dados, no eixo  $y$  estão as diferenças entre os dados estimados pelo modelo e os dados reais da série, e cada linha representa o comportamento das diferenças máximas para cada ordem com que foi executado o modelo.

Figura 4 – Diferenças máximas entre os valores reais da série e os estimados pelo modelo para cada ordem



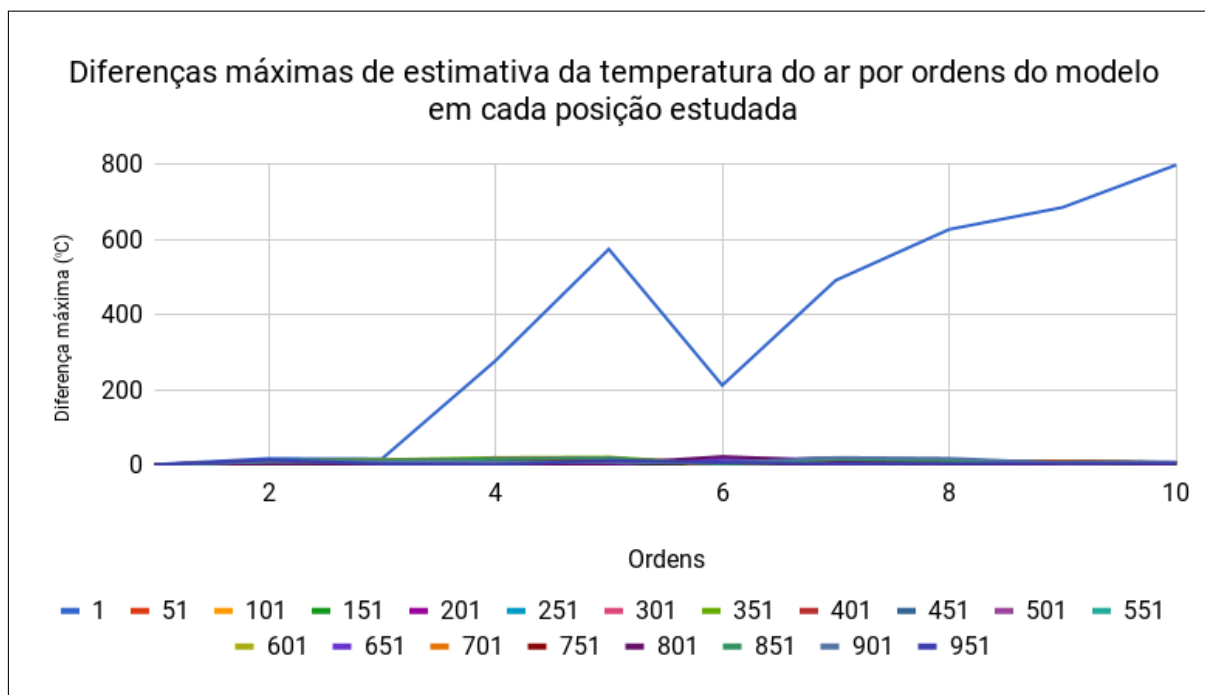
Fonte: O autor, 2018.

Na Figura 5, cada série representa o comportamento das diferenças máximas para cada posição onde foi testado o modelo nas diferentes ordens testadas. Foi representado na figura com as ordens de 1 até 10, dado que o comportamento dos dados além desses valores tem um comportamento similar aos mostrados. No eixo  $x$  se encontram as ordens do modelo, no eixo  $y$  estão as diferenças entre os dados estimados pelo modelo e os dados reais da série, e cada linha representa o com-



portamento das diferenças máximas para cada posição onde começaram os dados faltantes usadas para a execução do modelo.

Figura 5 – Diferenças máximas entre os valores reais da série e os estimados pelo modelo para cada posição



Fonte: O autor, 2018.

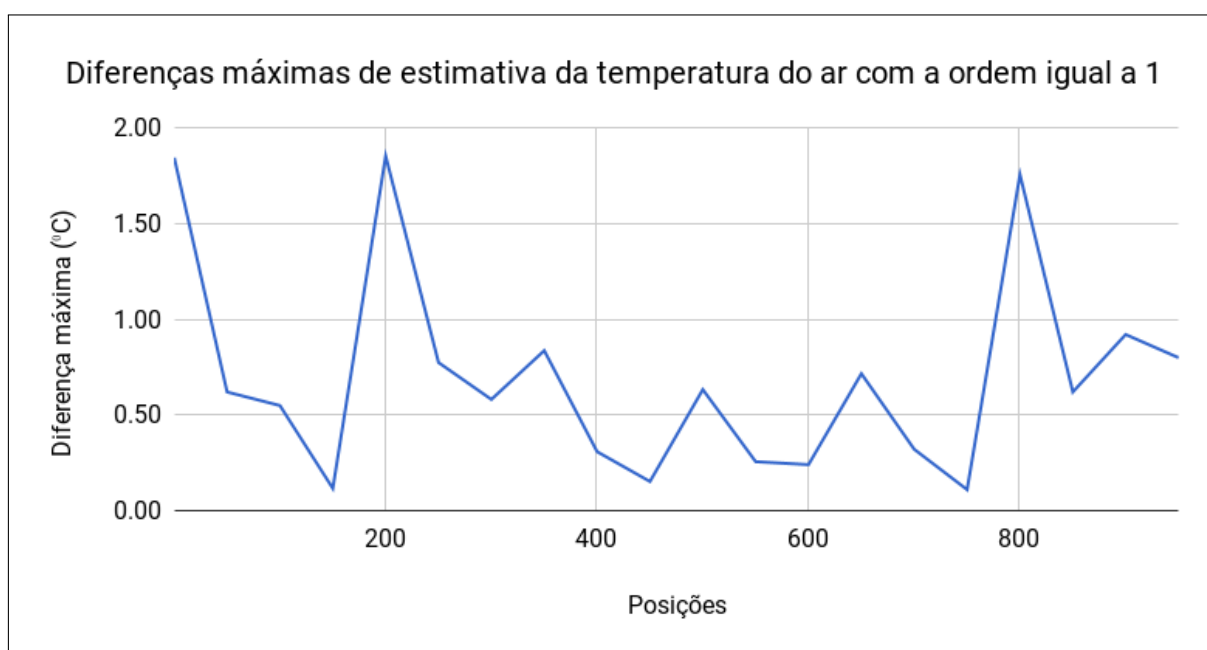
As estimativas feitas com os dados faltantes começando na primeira posição, foram as que obtiveram os maiores erros, como pode se ver na Figura 5, onde os valores dessa série são superiores ao das restantes séries de forma que não conseguem se diferenciar os valores delas, só das estimativas feitas com os dados faltantes na primeira posição. Isso acontece porque o modelo faz as estimativas com os dados anteriores e posteriores ao dado que está sendo estimado, e no caso da primeira posição, o modelo pode usar só o dado posterior, pois o anterior não existe. Os restantes dados são calculados usando as estimativas feitas anteriormente, o que introduz um erro maior do que quando os dados faltantes consecutivos estão em outras posições.

As melhores estimativas foram obtidas com a ordem igual a 1, como pode se ver na Figura 4, onde a série 1 (estimativas feitas com a ordem igual a 1) encontra-se por baixo do resto das séries geradas pelas estimativas feitas com as restantes ordens. Com esta ordem do modelo, as melhores estimativas obtiveram-se nas posições 151 e

751 (Fig. 6), ainda que nas posições 401, 451, 551, 601 e 701 também foram obtidas boas estimativas pois as diferenças máximas foram menores do que  $0,5^{\circ}\text{C}$ .

As estimativas menos confiáveis obtidas com a ordem igual a 1 foram geradas com os dados faltantes consecutivos começando nas posições 1 (explicado com antecedência), 201 e 801. Para estas duas últimas posições não conseguiu se achar uma explicação para o comportamento do modelo.

Figura 6 – Diferenças máximas entre os valores reais e os estimados pelo modelo para a ordem igual a 1 nas diferentes posições dos dados faltantes na série



Fonte: O autor, 2018.

Os mesmos testes foram feitos com as variáveis: velocidade e direção do vento, pressão atmosférica e umidade relativa, e foram obtidos resultados similares. Isso aconteceu porque estas são as variáveis que têm um ciclo diurno bem definido, e o modelo preditivo auto regressivo faz boas estimativas para séries de dados com certa periodicidade.

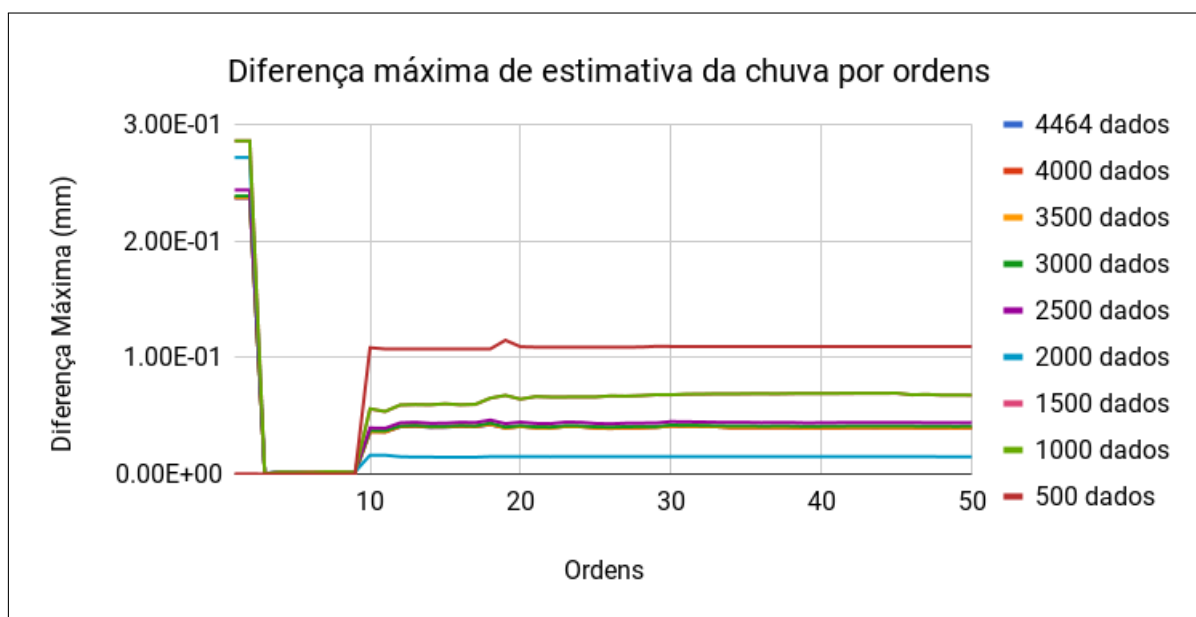
Foram feitos com os valores de chuva, os mesmos procedimentos que com os dados de temperatura do ar, só que na hora de fazer o análise com o EQM para as diferentes quantidades de dados e o mesmo percentual de dados faltantes, foi decidido analisar a diferença máxima entre os dados estimados e os dados reais da série, pois dado que os dados são muito pertos de zero, o EQM tem muita pouca variação e não se consegue ver com clareza qual é a melhor estimativa e qual a pior.

Então, foi executado o modelo com valores entre 4464 (quantidade máxima de dados) e 500 dados com o mesmo percentual de valores faltantes, diminuindo de 500 em 500 dados.

Como é mostrado na Figura 7, as melhores estimativas da precipitação feitas pelo modelo preditivo auto regressivo, foram obtidas com as ordens desde 3 até 9 para todas as quantidades de dados estudadas. Na Figura 7 são mostrados os valores das ordens de 1 até 50, pois nesta faixa de dados encontram-se os resultados mais significativos, e assim é mais fácil a compreensão do gráfico.

Também pode se concluir desta figura que, como nas séries de dados das outras variáveis, com as séries de dados de chuva, as estimativas do modelo não dependem da quantidade de dados que tem a série.

Figura 7 – Diferença máxima entre os valores reais da série de chuva e os estimados pelo modelo com diferentes quantidades de dados



Fonte: O autor, 2018.

Assim como com os dados de temperatura do ar, foram usados 1000 dados de chuva da série de dados meteorológicos, e foram gerados erros para testar o funcionamento do modelo.

Faltando 50% dos dados, o modelo foi executado primeiro com os dados faltantes alternados com os dados reais, é dizer, um sim e um no, primeiro começando com o dado faltante e depois começando com o dado real da série. Para ambos testes

o resultado foi que o modelo não gera estimativas adequadas para serem usadas em substituição dos dados perdidos, sendo que a diferença máxima entre os dados estimados pelo modelo e os dados reais da série são maiores que 0,4 mm no primeiro caso e maiores que 0,8 mm no segundo, valores esses que são muito grandes considerando que a média da série é 0,1 mm e a diferença máxima permissível para a chuva é 0,01 mm (Tabela 2).

O modelo foi executado também, faltando o 50% dos dados consecutivos, começando pela primeira posição da série de dados de chuva, mas os resultados não foram muito diferentes do teste anterior. Depois foi-se diminuindo o percentual de dados faltantes consecutivos (Tabela 3). A quantidade de dados foi diminuída até só 1% dos valores faltantes a partir da primeira posição da série, mas a diferença máxima obtida entre os valores estimados pelo modelo e os valores reais da série foi muito alta para esta variável (0,5 mm).

Tabela 3 – Máxima diferença entre os valores estimados e originais da série de chuva.

	50%	40%	30%	20%	10%	1%
1	0,5	0,4999	0,4999	0,4998	0,4999	0,4999
2	0,5	0,5	0,4999	0,4999	0,4999	0,4999
3	0,4987	0,4984	0,4983	0,4981	0,4978	0,4975
4	3,7832	0,4876	0,488	0,489	0,4893	0,4884
5	2,6469	3,2865	0,4846	0,4869	0,4873	0,4875
6	2,7982	2,8781	2,9181	2,8955	2,5133	0,436
7	1,2267	3,1156	0,4933	0,4936	0,4894	0,5147
8	2,5996	0,496	0,4952	0,4945	0,4939	0,4931
9	2,4736	2,585	2,7128	2,7138	2,1089	0,4674
10	2,3041	2,368	2,449	2,544	2,0819	0,4368

Nota: Dados para diferentes percentuais de dados faltantes consecutivos começando na primeira posição da série e para diferentes ordens do modelo (1-10).

Fonte: O autor, 2018.

O modelo foi executado aumentando a quantidade de dados faltantes consecutivos com o objetivo de encontrar a quantidade máxima de dados faltantes consecutivos que se conseguiria estimar com um erro permissível. Também foi executado colocando a mesma quantidade de dados faltantes em diferentes posições (1, 51, 101, 151, 201, 251, 301, 351, 401, 451, 501, 551, 601, 651, 701, 751, 801, 851, 901 e 951).

Como resultado destes testes não foi possível encontrar um padrão no comportamento das estimativas, pois a variável não tem um comportamento cíclico, as estimativas variam muito ao mudar a posição dos dados faltantes. Por exemplo, nos casos onde os dados faltantes estavam em posições onde não tinha acontecido chuva, a estimativa do modelo foi aceitável (diferença máxima entre os dados estimados e os dados reais da série menor de 0,01 mm), mas nos casos onde faltavam dados nas posições onde tinha acontecido o fenômeno, as estimativas sempre foram menores que os dados reais em mais de 0,5 mm.

## CONCLUSÕES E RECOMENDAÇÕES

Foi atingido o objetivo de estimar os dados faltantes em séries de dados meteorológicos através do método preditivo auto regressivo usando somente os dados da mesma estação. Para isto foi usado o *framework* multiplataforma “Qt” através da linguagem de programação C++, o qual permitiu fazer uma interface gráfica amigável que facilita seu uso.

Os códigos através dos quais foram gerados o modelo e a interface gráfica, foram disponibilizados no GitHub, para que pudesse estar disponível para todo aquele que quiser usar, modificar ou adaptá-lo às suas necessidades.

A análise da quantidade de dados faltantes na estação meteorológica do IPRJ concluiu que o melhor ano para usar nos testes da validação do modelo foi o 2011, dado que não tinha dados faltantes e também não foi preciso fazer ajuste nenhum, pois não tinha valores espúrios.

Ao fazer os testes no modelo a partir das séries de dados de temperatura do ar, velocidade e direção do vento, pressão atmosférica e umidade relativa, foram obtidas as seguintes conclusões:

- quando a ordem é maior do que a metade da quantidade de dados, o método diverge, estão esses valores não devem ser usados no modelo;
- as melhores estimativas são obtidas com as ordens 1 e 2 do método;
- a probabilidade de acerto na estimativa dos dados faltantes não depende da quantidade de dados;
- nas estimativas feitas a partir de séries de dados com falhas alternadas foram obtidos altos valores de diferença máxima entre os valores estimados pelo modelo e os valores reais da série, e então não é recomendado o uso do modelo nesses casos;
- as estimativas feitas em séries de dados nos quais tem dados faltantes consecutivos começando na primeira posição, o modelo não obtém estimativas boas para serem usadas pelos especialistas;

- a partir de 40 dados faltantes consecutivos, dos 1000 dados que foram usados para os testes, sem começar na primeira posição da série, as estimativas deixam de ser boas para seu uso pelos especialistas.

Ao fazer os testes no modelo a partir das séries de dados de chuva, foram obtidas algumas diferenças em comparação com os resultados a partir das outras séries estudadas:

- as melhores estimativas feitas pelo modelo preditivo auto regressivo, foram obtidas com as ordens desde 3 até 9 para todas as quantidades de dados estudadas;
- não foi possível encontrar um padrão no comportamento das estimativas do modelo ao serem usadas a mesma quantidade de dados faltantes em diferentes posições, pois a variável não tem um comportamento cíclico e as estimativas variam muito ao mudar a posição dos dados faltantes.

Para fazer a estimativa dos valores faltantes nas séries de dados meteorológicos através do método preditivo auto regressivo, é recomendado que:

- sejam feitos outros testes que permitam determinar a probabilidade de acerto nas estimativas do modelo;
- o modelo seja executado com outras variáveis com o objetivo de saber se são obtidos os mesmos resultados;
- seja testado o método com dados diários das variáveis em séries de tempo maiores, por exemplo, séries anuais;
- seria muito interessante fazer um estudo mais a fundo do problema para conseguir explicar por quê as estimativas nas posições 201 e 801 são muito ruins;
- se continue estudando o método com o objetivo de fazê-lo mais eficiente;
- seja divulgado seu uso e cada vez mais usuários possam modificá-los para que seja aprimorado o modelo.

## REFERÊNCIAS

- ALFARO, E. J.; SOLEY, F. J. Descripción de dos métodos de rellenado de datos ausentes en series de tiempo meteorológicas. Revista de Matemática: Teoría y Aplicaciones, v. 16, n. 1, p. 60–75, 2009.
- ALFONSI, R. R.; PINTO, H. S.; PEDRO JUNIOR, M. J. Estimativas das normais de temperatura média mensal e anual do estado de Goiás (BR) em função de altitude e latitude. Caderno de Ciências da Terra, v. 45, p. 1-27, 1974.
- ANDERSEN, N. Short Notes: On the calculation of filters coefficients for maximum entropy spectral analysis. Geophysics, v. 39, n. 1, p. 69-72, 1974.
- BARRERA, Antonio. Métodos univariados. (Técnicas de completado de series mensuales y aplicación al estudio de la influencia de la NAO en la distribución de la precipitación en España). 2004. Tese (Doctorado de Astronomía y Meteorología)-Universidad de Barcelona, Barcelona. 2004.
- BURG, R. Máximum Entropy spectral analysis. In: MEETING OF THE SOCIETY OF EXPLORATION GEOPHYSICISTS, 37., 1967, Oklahoma City. Proceedings of 37th Meeting, Society of Exploration Geophysicists, Oklahoma City. Oklahoma City: SEG, 1967.
- BURG, R. A new analysis technique for time series analysis. In: NATO ADVANCED STUDY INST. ON SIGNAL PROCESSING WITH EMPHASIS ON UNDERWATER ACOUSTICS, 1968, Enschede, The Netherlands. Proceedings of the NATO advanced study... [s.l.]: NATO, 1968.
- CARBAJAL, Mariela et al. Reconstrucción de datos faltantes de precipitación pluvial diaria mediante la Transformada Wavelet. Revista peruana Geo-Atmosférica RPGA, v. 2, p. 76-88, 2010.
- CENTRO DE TECNOLOGIA EM MEIO AMBIENTE do IPRJ/UERJ. Estação Meteorológica. Disponível em: <http://www.clima.iprj.uerj.br/>. Acesso em: 15 jan. 2018.
- FANTE, K. P. Variabilidade da temperatura em áreas urbanas não metropolitanas do Estado de São Paulo: Brasil no período de 1961 a 2011. 2014 Dissertação (Mestrado) - Universidade Estadual Paulista, Presidente Prudente, SP, 2014.
- FANTE K. P.; SANT'ANNA J. L. Técnicas estatísticas para a homogeneização de dados de temperatura em séries temporais climatológicas. Revista Brasileira de Climatologia, v. 18, n. 12, 2016.
- FOX D G.; DEARDORFF J. W. Computer Methods for Simulation of Multidimensional, Nonlinear, Subsonic, Incompressible Flow. Journal of Heat Transfer, Nov. 1972.
- FREITAS, Janice et al. Análise da acurácia das simulações do modelo BRAMS na Amazônia Ocidental. In: CONGRESSO BRASILEIRO DE AGROMETEOROLOGIA, 17., 2011, Guarapari, ES. Anais do XVII Congresso Brasileiro de Agrometeorologia. SESC Centro de Turismo de Guarapari, Guarapari-ES, 2011.



GUTIÉRREZ, David G. Tutorial de Qt4 Designer y QDevelop. Proyecto de Fin de Carrera FIB - UPC 2008/09 Q2. Disponível em: <http://upcommons.upc.edu/bitstream/handle/2099.1/7656/memoriafinalPFC.pdf>. Acesso em: 24 mai. 2017.

GOMES, Luiz Claudio et al. Influência dos parâmetros de configuração do modelo CALMET sobre a simulação da circulação atmosférica na região metropolitana do Rio de Janeiro. Revista Brasileira de Meteorologia, v 29, n 4, p 579-596, 2014.

GARCÍA, José A. Los modelos numéricos de predicción del tiempo. TETHYS, Revista de Meteorología, n. 2, 2002.

INFANTE S.; ORTEGA J.; CEDEÑO F. Estimación de datos faltantes en estaciones meteorológicas de Venezuela vía un modelo de redes neuronales Revista de Climatología, v. 8, p. 51-70, 2008.

JORBA C, O.; BALDASANO R, J. M. Simulación de los campos de viento de la Península Ibérica y el área geográfica de Catalunya con alta resolución espacial para distintas situaciones meteorológicas típicas. Capítulo 6: Modelo de pronóstico meteorológico mesoescalar. Institute Universitat Politècnica de Catalunya. 2005. Tese (Doutorado em Engenharia Ambiental) - Universitat Politècnica de Catalunya, 2005.

LESA, F. L.; DOMÍNGUEZ, D. S. Modelagem Computacional do Problema de Desintegração Radioativa. In: CONGRESSO DE MATEMÁTICA APLICADA E COMPUTACIONAL, 30., 2007. Anais do XXX CNMAC. Santa Catarina. 2007.

LIGHEZZOLO, Rafael Andrés. Integración de modelos numéricos de predicción meteorológica en sistemas de alerta temprana a emergencias. Dissertação (Maestría) - Universidad Nacional de Córdoba, 2014.

LIRA, C; CATALDI, M. Avaliação do ensemble de parametrizações físicas do modelo MM5 no evento de precipitação intensa ocorrido entre os dias 05 e 06 de abril de 2010 no município do Rio de Janeiro. Ciência e Natura [en linea], n. 38, 2016.

LÓPEZ, Luis Alfonso. La predicción del tiempo a partir de modelos numéricos. Revista del aficionado a la Meteorología, n. 22, Ago. 2004.

MAPS AND DIRECTIONS. -22.2869770; -42.5418978. Disponível em: <https://www.mapsdirections.info>. Acesso em: 15 mar. 2018.

MEDINA R.; MONTOYA R.; JARAMILLO R. Estimación estadística de valores faltantes en series históricas de lluvia. Dissertação (Maestría) - Universidad Tecnológica de Pereira, Colombia. Cenicafé, v. 59, n. 3, p. 260-273, 2008.

MENEZES, Antonio M. Introdução a Programação C++ com Qt 4. In: FORUM DE TECNOLOGIA EM SOFTWARE LIVRE, 2., 2009, Porto Alegre. Anais do II Forum de Tecnologia em Software Livre. Porto Alegre: SERPRO, 2009. Apresentação oral.

ORGANIZACIÓN METEOROLÓGICA MUNDIAL. OMM-No 100, Guía de prácticas climatológicas. Edición de 2011. Disponível em: [http://www.wmo.int/pages/prog/wcp/ccl/guide/guide\\_climat\\_practices.php#updates](http://www.wmo.int/pages/prog/wcp/ccl/guide/guide_climat_practices.php#updates). Acesso em: 15 jun. 2017.

RODRIGO, M.; FERNÁNDEZ, M. E. Estimación de datos faltantes de precipitación diaria para las distintas ecorregiones de la República Argentina. In: ENCUENTRO DE INVESTIGADORES EN FORMACIÓN EN RECURSOS HÍDRICOS/IFRH, 2014. Anais do Encontro de investigadores em formação em Recursos Hídricos/IFRH Ezeiza-Argentina: Instituto Nacional del Agua, 2014. Apresentação Oral.

RODRÍGUEZ, Germán. Métodos de análisis espectral del oleaje. Estudio Comparativo. Universidad de Las Palmas de Gran Canaria. Departamento de Física, 1995.

SMYLIE, D. E.; CLARKE, G. K. C.; ULRYCH, T. J. Analysis of irregularities in the earth's rotation. Methods in Computational Physics, v. 13, p. 391-430, 1973.

STEVEN, M. K.; MARPLE, S. L. Spectrum Analysis-A Modern Perspective. Proceedings of the IEEE. v. 69, n. 11, 1981.

QT COMPANY. Disponível em: <https://www.qt.io/>. Acesso em: 10 dic. 2018.

TORO, Ana María et al. Relleno de series diarias de precipitación, temperatura mínima, máxima de la región norte del Urabá Antioqueño. Revista Mexicana de Ciencias Agrícolas. v. 6, n. 2, p. 577-588, 2015.

ULRYCH, T. J. Maximum entropy power spectrum of truncated sinusoids. J. Geophys. Res, n. 77, p. 1396-1400, 1972.

ULRYCH, T. J.; CLAYTON, R. W. Time Series Modelling and Maximum Entropy. Physics of the Earth and Planetary Interiors, n. 12, p. 188-200, 1976.

ULRYCH, Tad J.; BISHOP, THOMAS N. Maximum Entropy Spectral Analysis and Autoregressive Decomposition. Reviews of Geophysics and Space Physics, v. 13, n. 1, p. 183-200, 1975.

UREÑA, P.; ALFARO, E. J.; SOLEY, J. Propuestas metodológicas para el relleno de datos ausentes en series de tiempo geofísicas. Guía Práctica de uso. Disponível em: <http://kerwa.ucr.ac.cr/> e <http://kimuk.conare.ac.cr/>. Universidad de Costa Rica, 2016.

VENTURA, T. M. Preenchimento de falhas de dados micrometeorológicos utilizando técnicas de inteligência artificial. 2012 Dissertação (Mestrado em Física Ambiental) - Instituto de Física, Universidade Federal de Mato Grosso, Cuiabá, 2012.

ANDRADE, K.; MOURA, C.; PINHEIRO, H. A maior catástrofe climática do Brasil so a visão operacional do CPTEC/INPE. Centro de Previsão do Tempo e Estudos Climáticos - CPTEC/INPE, São Paulo-SP. Brasil. 2011. Disponível em: <http://www.sbmet.org.br/sic2011/arq/65702891341666570289134.pdf>. Acesso em: 22 fev. 2017.

**APÊNDICE A** -Estrutura do arquivo de dados a usar pelo modelo

Tabela 4 – Estrutura do arquivo de dados a usar pelo modelo.

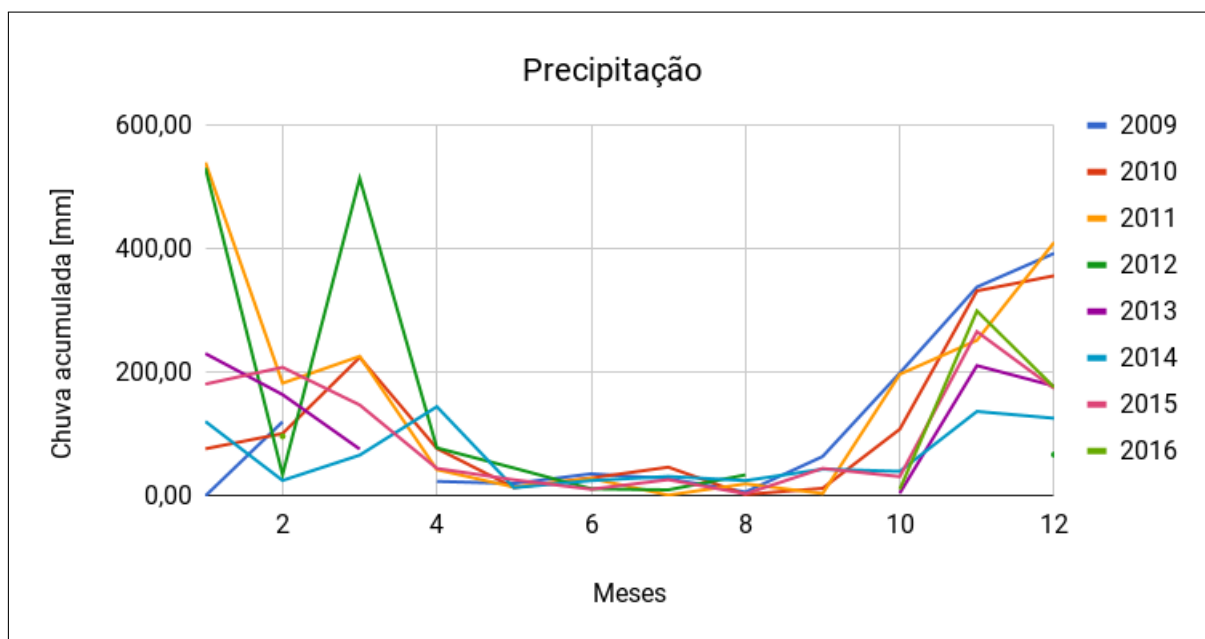
---

Ano	Mês	Dia	Hora	Minuto	Variável
2011	01	01	01	01	x
⋮	⋮	⋮	⋮	⋮	⋮

Fonte: O autor, 2018.

**APÊNDICE B** -Comportamento mensal da Precipitação na estação meteorológica do IPRJ/UERJ nos anos desde 2009 até 2016

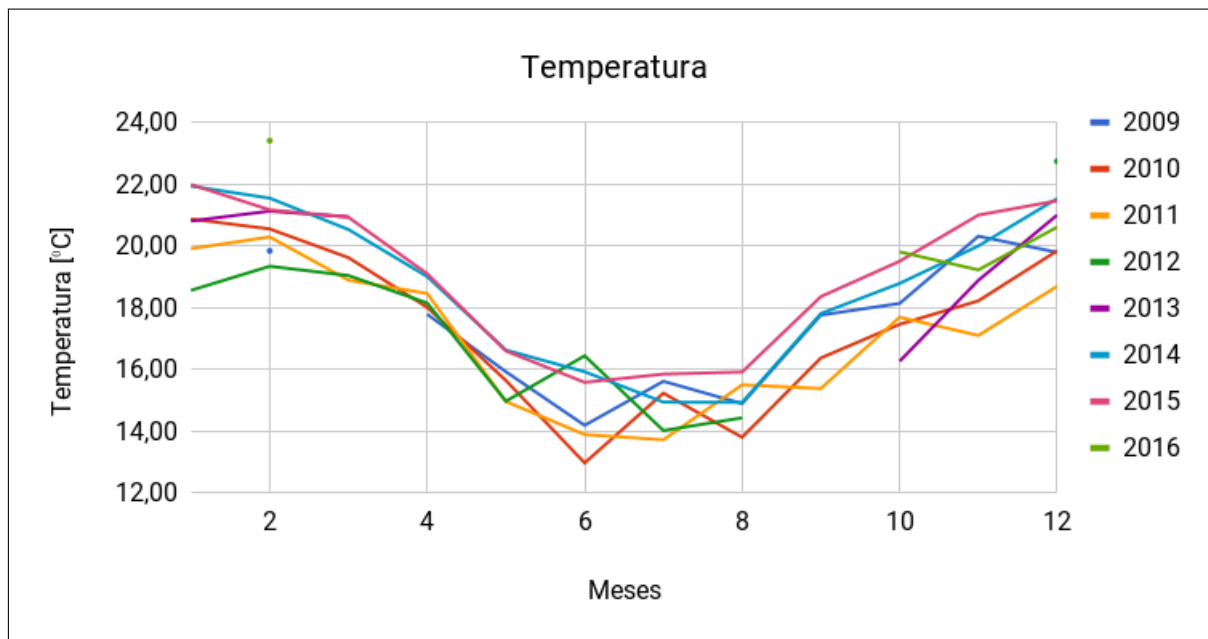
Figura 8 – Comportamento mensal da precipitação.



Fonte: O autor, 2018.

**APÊNDICE C** -Comportamento mensal da Temperatura na estação meteorológica do IPRJ/UERJ nos anos desde 2009 até 2016

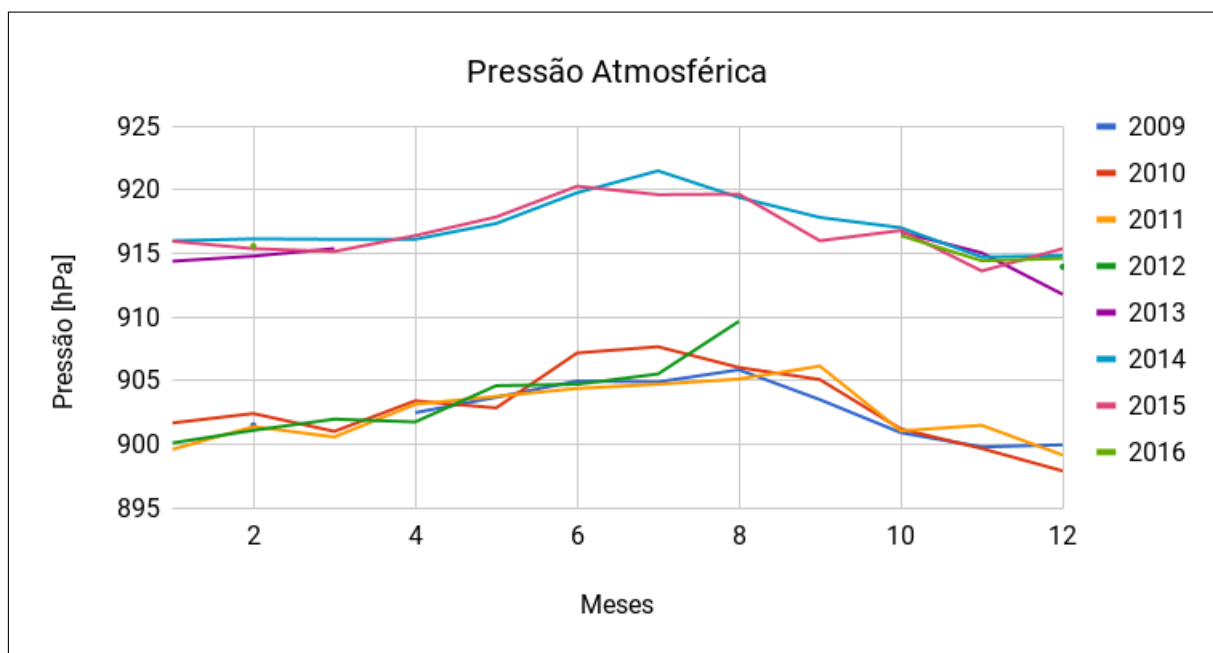
Figura 9 – Comportamento mensal da temperatura.



Fonte: O autor, 2018.

**APÊNDICE D** -Comportamento mensal da Pressão Atmosférica na estação meteorológica do IPRJ/UERJ nos anos desde 2009 até 2016

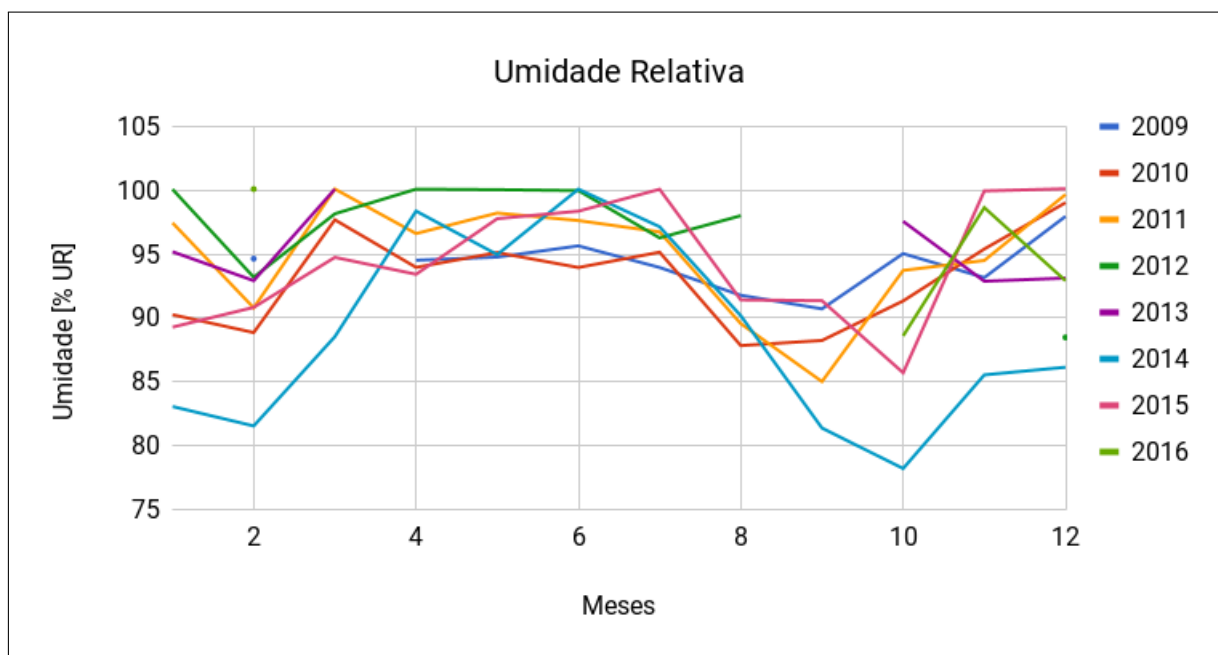
Figura 10 – Comportamento mensal da pressão atmosférica.



Fonte: O autor, 2018.

**APÊNDICE E** -Comportamento mensal da Umidade Relativa na estação meteorológica do IPRJ/UERJ nos anos desde 2009 até 2016

Figura 11 – Comportamento mensal da umidade relativa.



Fonte: O autor, 2018.