

UNIVERSIDADE ESTADUAL DE PONTA GROSSA
SETOR DE CIÊNCIAS AGRÁRIAS E DE TECNOLOGIA
CURSO DE MESTRADO EM COMPUTAÇÃO APLICADA

LUIZ RAFAEL SCHMITKE

PREENCHIMENTO DE FALHAS DE SÉRIES DE DADOS CLIMÁTICOS UTILIZANDO
REDES P2P

PONTA GROSSA
2012

LUIZ RAFAEL SCHMITKE

PREENCHIMENTO DE FALHAS DE SÉRIES DE DADOS CLIMÁTICOS UTILIZANDO
REDES P2P

Dissertação apresentada como requisito parcial
para obtenção do título de Mestre na
Universidade Estadual de Ponta Grossa, Área
de Computação Aplicada.

Orientador: Prof. Dr. Luciano José Senger

PONTA GROSSA
2012

Ficha Catalográfica Elaborada pelo Setor Tratamento da Informação Belém/UEPG

S355p Schmitke, Luiz Rafael
Preenchimento de falhas de séries de dados climáticos utilizando
redes P2P / Luiz Rafael Schmitke. Ponta Grossa, 2012.
65 f.

Dissertação (Mestrado em Computação Aplicada), Universidade
Estadual de Ponta Grossa.
Orientador: Prof. Dr. Luciano José Senger.

1. Redes P2P. 2. Correção. 3. Dados climáticos. I. Senger, Luciano
José. II. Universidade Estadual de Ponta Grossa. Mestrado em
Computação Aplicada. III. T.

CDD: 004.35

TERMO DE APROVAÇÃO

LUIZ RAFAEL SCHMITKE

"Preenchimento de Falhas de Séries de Dados Climáticos Utilizando redes P2P"

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Programa de Pós-Graduação em Computação Aplicada da Universidade Estadual de Ponta Grossa, pela seguinte banca examinadora.

Orientador: Luciano José Senger
UEPG

Maria Salete Marcon Gomes Vaz
UEPG

Jorim Souza das Virgens Filho
UEPG

Ponta Grossa, 30 de Julho de 2012.

*Dedico este trabalho à minha família que sempre me apoiou
e a minha esposa Kelly Cristina de Castro.*

AGRADECIMENTOS

Primeiramente, quero agradecer professores que encontramos durante a nossa caminhada da vida. O primeiro deles, Jesus, que além de um grande professor é também um exemplo de profissional que cumpre seus objetivos por mais obstáculos que apareçam.

Ao meu pai Luiz Carlos Schmitke, minha mãe Joclena do Rocio Schmitke e minha irmã Talâmita Maria Schmitke que como professores me ensinam uma disciplina que não pode ser encontrada em nenhum lugar, a disciplina da vida, e por último aos meus professores do mestrado que de uma forma ou outra contribuíram para meu crescimento profissional, em especial os professores Luciano José Senger, meu orientador neste trabalho, e Jorim Sousa das Virgens Filho que me ajudaram durante o desenvolvimento do trabalho.

Também, quero prestar meus agradecimentos a minha esposa Kelly Cristina de Castro que desde o início do trabalho me apoiou e me ajudou em algumas horas e a sua família que sempre esteve ao meu lado e também me apoiaram neste momento e a Fundação ABC e o Instituto Tecnológico SIMEPAR por fornecer os dados climáticos necessários para esta pesquisa.

Por fim agradeço a todos aqueles que de uma forma ou outra contribuíram para o meu desenvolvimento pessoal e profissional, durante o desenvolvimento do trabalho, e aos grandes colegas que jamais serão esquecidos.

Muito obrigado a todos vocês.

RESUMO

A agricultura é uma das atividades onde o clima tem mais impacto, influenciando as técnicas e os cultivos empregados. Grande parte da produtividade agrícola se deve as condições climáticas que são criadas por fatores naturais e não são passíveis de controle. Embora não seja possível controlar o clima, pode-se prevê-lo ou até simular suas condições para tentar minimizar seu impacto na agricultura. Para que seja possível realizar estas previsões e simulações são necessários dados coletados em estações climáticas que podem ser convencionais ou automáticas e que precisam estar sem dados anormais ou lacunas. Grande parte desses erros se deve a interferência no sinal, desconexão, oxidação de cabos e a variação espaço-temporal do clima que por consequência acabam gerando aqueles problemas nas bases climáticas. Desta forma, este trabalho de pesquisa tem como objetivo principal criar um modelo capaz de corrigir as lacunas existentes nas bases de dados climáticas, salientando-se que não visa à correção de observações anormais e nem a substituição dos métodos estatísticos para o mesmo fim. Para tanto foi criado um modelo de correção das lacunas em dados climáticos entre as estações utilizando a arquitetura P2P. Com este modelo, foi criada uma aplicação para testar seu desempenho em corrigir as lacunas encontradas. Também para a realização dos testes foram utilizadas bases das cidades de Ponta Grossa, Fernades Pinheiro e Telêmaco Borba, fornecidas pelo Instituto Tecnológico SIMEPAR, e bases das cidades de Castro, Carambeí, Tibagi e Pirai do Sul fornecidas pela Fundação ABC, sendo estes dados, diários e coletados em estações automáticas. Como resultados foi possível observar que o desempenho do modelo de correção P2P foi satisfatório quando comparado ao simulador utilizado nos testes, apresentando resultados inferiores somente no mês de fevereiro, que corresponde ao período de verão, para as estações de outono, inverno e primavera o modelo P2P foi melhor que o simulado. Ainda foi verificado que a quantidade de estações que participa da rede na hora da correção influencia os resultados, sendo que quanto maior ela for, melhores são os resultados obtidos com a correção.

Palavras-chave: Redes P2P, Correção, Dados Climáticos.

ABSTRACT

Agriculture is an activity where the weather has more impact, influencing techniques and crops employed. Much of the agricultural productivity is affected by climatic conditions that are created by natural factors and are not likely to control. Although you can't control the weather, we can predict it, or even simulate their conditions to try minimize its impact on agriculture. To be able to make these predictions and simulations are necessary data collected from weather stations that can be conventional or automatic and must be without gaps or abnormal data. Most of these errors are caused by signal interference, disconnection, oxidation of cables and spatio-temporal variation of climate which consequently end up generating those problems at the climates bases. Thus, this research work has as main objective to create a model capable of correcting gaps in climate databases, observing that not to correct abnormal observations or replace statistical methods for the same purpose. Therefore a model was created to correct the gaps in weather data between stations using the P2P architecture. With this model, an application was created to test its performance to correct the gaps. Also to perform the tests were used bases in the cities of Ponta Grossa, Fernandes Pinheiro and Telêmaco Borba provided by Instituto Tecnológico SIMEPAR, and bases of the cities of Castro, Carambeí, Pirai do Sul and Tibagi provided by Fundação ABC, which are collected daily on automatic stations. As a result it was observed that the performance of P2P correction model was satisfactory when compared to the simulator used in the tests, with lower results only in February, which corresponds to the period of summer, to the autumn, winter and spring the model P2P was better than simulated. Although it was found that the number of stations participating in the network at the time of correcting influences the results, and the higher it is, the better the results obtained with the correcting.

Keywords: P2P Networks, Correct, Climate Data.

LISTA DE ILUSTRAÇÕES

Figura 1: Classificação climática do Brasil segundo Köppen	15
Figura 2: Classificação climática do Paraná segundo Köppen.....	16
Figura 3 – Esquema de comunicação TCP	19
Figura 4 – Esquema de Comunicação UDP	19
Figura 5 – Arquitetura Cliente-Servidor.....	20
Figura 6 – Arquitetura <i>Peer-to-Peer</i> (P2P)	21
Figura 7 – Organização Híbrida P2P	22
Figura 8 – As três camadas do padrão JXTA	23
Figura 9 – Comunicação entre <i>peers</i> utilizando <i>pipes</i> e <i>endpoints</i>	24
Figura 10 – Estrutura de classes P2PComp	25
Figura 11 – Código de uma aplicação de exemplo utilizando o P2PComp	26
Figura 12 – Mapa com as estações utilizadas no trabalho.....	29
Figura 13 – Modelo de correção criado.....	31
Figura 14 – Dados observados, simulados e corrigidos no modelo P2P para o mês de Fevereiro em 3 localidades.	36
Figura 15 - Dados observados, simulados e corrigidos no modelo P2P para o mês de Maio em 3 localidades.	37
Figura 16 - Dados observados, simulados e corrigidos no modelo P2P para o mês de Agosto em 3 localidades.	37
Figura 17 - Dados observados, simulados e corrigidos no modelo P2P para o mês de Novembro em 3 localidades.	38
Figura 18 - Dados observados, simulados e corrigidos no modelo P2P para o mês de Fevereiro em todas as localidades.....	40
Figura 19 – Dados observados, simulados e corrigidos no modelo P2P para o mês de Maio em todas as localidades.	41
Figura 20 - Dados observados, simulados e corrigidos no modelo P2P para o mês de Agosto em todas as localidades.....	41
Figura 21 – Dados observados, simulados e corrigidos no modelo P2P para o mês de Novembro em todas as localidades.	42

LISTA DE QUADROS E TABELAS

Quadro 1 – Tipos de redes de computadores.....	18
Quadro 2 – Localização das bases climáticas utilizadas na pesquisa.....	28
Quadro 3 – Períodos onde foram geradas lacunas.....	29
Tabela 1 – Resultados da correção utilizando três estações para Fevereiro	34
Tabela 2 – Resultados da correção utilizando três estações para Maio.....	34
Tabela 3 – Resultados da correção utilizando três estações para Agosto.....	35
Tabela 4 - Resultados da correção utilizando três estações para Novembro.....	35
Tabela 5 - Resultados da correção utilizando todas as estações para Fevereiro.....	38
Tabela 6 – Resultados da correção utilizando todas as estações para Maio.....	39
Tabela 7 - Resultados da correção utilizando todas as estações para Agosto.	39
Tabela 8 - Resultados da correção utilizando todas as estações para Novembro.....	39
Tabela 9 – Comparação do desempenho do modelo P2P para três localidades e todas as localidades com base no <i>p-value</i> do teste t.....	42

LISTA DE ALGORITMOS

Algoritmo 1 – Cálculo de distância entre as estações	33
Algoritmo 2 – Processo de correção	33

LISTA DE ABREVIATURAS E SIGLAS

EMC – Estação Meteorológica Convencional
EMA – Estação Meteorológica Automática
SAD – Sistema de Apoio a Decisão
OMM – Organização Mundial de Meteorologia
SIMEPAR – Instituto Tecnológico SIMEPAR
FABC – Fundação ABC
LAN – *Local Area Network*
MAN – *Metropolitan Area Network*
WAN – *Wide Area Network*
WLAN – *Wireless Local Area Network*
WMAN – *Wireless Metropolitan Area Network*
WWAN – *Wireless Wide Area Network*
DSL – *Digital Subscriber Line*
GSM – *Global System for Mobile Communications*
WiMAX – *Worldwide Interoperability for Microwave Access*
TCP – *Transmission Control Protocol*
IP – *Internet Protocol*
PC – *Personal Computer*
P2P – *Peer-to-Peer*
ACK – *Acknowledge*
UDP – *User Datagram Protocol*
FTP – *File Transfer Protocol*
UEPG – Universidade Estadual de Ponta Grossa
AMD – *Advanced Micro Device*
GB – Gigabytes
RAM – *Random Access Memory*
PM – *Peer Main*
PA – *Peer Atuador*
UTM – Sistema Universal Transverso de Mercator

SUMÁRIO

INTRODUÇÃO	12
CAPITULO 1 - OBJETIVOS DA DISSERTAÇÃO	13
1.1 OBJETIVO GERAL.....	13
1.2 OBJETIVOS ESPECÍFICOS	13
CAPITULO 2 - FUNDAMENTAÇÃO TEÓRICA	14
2.1 CLIMA E TEMPO	14
2.2 CLASSIFICAÇÃO CLIMÁTICA DO BRASIL.....	14
2.3 DADOS CLIMÁTICOS	17
2.4 REDES DE COMPUTADORES	17
2.5 COMUNICAÇÃO CLIENTE-SERVIDOR.....	18
2.6 <i>PEER-TO-PEER (P2P)</i>	20
2.7 USO DE REDES DE COMPUTADORES PARA COMPARTILHAMENTO E DISTRIBUIÇÃO DE DADOS CLIMÁTICOS	22
2.8 O PADRÃO JXTA.....	23
2.9 P2PCOMP	25
2.10 CONCEITOS ESTATÍSTICOS	26
CAPITULO 3 - MATERIAIS E MÉTODOS	28
3.1 LOCAL DO DESENVOLVIMENTO DA PESQUISA.....	28
3.2 FERRAMENTAS.....	28
3.3 DADOS CLIMÁTICOS.....	28
3.4 TESTES E MEDIDAS ESTATÍSTICOS.....	30
CAPITULO 4 - RESULTADOS.....	31
4.1 MODELO DE CORREÇÃO.....	31
4.2 A APLICAÇÃO DE SINCRONIZAÇÃO	32
4.3 TESTES ESTATÍSTICOS DAS CORREÇÕES FEITAS UTILIZANDO TRÊS LOCALIDADES	34
4.4 TESTES ESTATÍSTICOS DAS CORREÇÕES FEITAS UTILIZANDO TODAS AS LOCALIDADES	38
4.5 CORREÇÃO COM TODAS AS LOCALIDADES X CORREÇÃO COM TRÊS LOCALIDADES	42
CAPITULO 5 - CONCLUSÕES	43
REFERÊNCIAS	44
APÊNDICE A – CÓDIGO FONTE DA CLASSE LCLIMA	47
APÊNDICE B – CÓDIGO FONTE DA BUSCA DE ESTAÇÕES E CÁLCULO DE DISTÂNCIA.....	52
APÊNDICE C – GRÁFICOS COMPARATIVOS DOS DADOS OBSERVADOS, GERADOS PELO SIMULADOR E PELO MODELO P2P UTILIZANDO TRÊS LOCALIDADES.....	53

APÊNDICE D – GRÁFICOS COMPARATIVOS DOS DADOS OBSERVADOS, GERADOS PELO SIMULADOR E PELO MODELO P2P UTILIZANDO TODAS AS LOCALIDADES.....	59
---	-----------

INTRODUÇÃO

A agricultura é uma atividade econômica onde o clima tem impacto direto, influenciando as técnicas de cultivo e os cultivares empregados. Segundo Hoogenboom (2000), grande parte da variação na produtividade agrícola deve-se as condições climáticas, e como os fatores que exercem essa influência são naturais, os agricultores não tem como controlá-los.

Entre esses fatores que compõe o clima, como radiação solar, temperatura do ar, umidade relativa do ar, pressão atmosférica, precipitação pluviométrica, velocidade e direção do vento (PEREIRA et al., 2002), os que mais afetam o desenvolvimento e a produtividade agrícola são a precipitação pluviométrica, temperatura do ar e a radiação solar (SENTELHAS & MONTEIRO, 2009, p.3). Embora não seja possível controlá-los, agricultores e pesquisadores utilizam ferramentas para prever e/ou simular o clima em datas futuras para minimizar o seu impacto nos cultivos e em outras atividades. Essas ferramentas se utilizam de dados climáticos coletados em estações meteorológicas que podem ser convencionais (EMC) ou automáticas (EMA) para fazer previsões ou simulações.

Segundo Almeida et al. (2008) e Strassburger et al.(2011), com o avanço tecnológico as EMA's estão substituindo, aos poucos, as EMC's devido a sua capacidade de coletar mais amostras e de poder ser operada em lugares de difícil acesso. Entretanto, as EMA's sofrem com problemas de interferência no sinal, desconexão e oxidação de cabos, entre outros problemas, gerando dados anormais ou lacunas na série de dados. Ainda, não somente esses problemas gerados na coleta, mas também a variação espaço-temporal do clima contribui para que a série de dados apresente dados anormais e lacunas, necessitando passar por modelos de geração e/ou correção antes de ser utilizada em aplicações agrícolas como, por exemplo, simuladores e Sistemas de Apoio a Decisão (SAD).

Entre as formas utilizadas, para gerar e/ou corrigir dados, encontram-se os métodos estatísticos, como, por exemplo, os modelos estocásticos (ZANETTI et al., 2006) que são capazes de gerar uma série de dados sintética, a partir de uma série real, combinando números aleatórios, mas para que isso seja possível os dados reais precisam estar corretos. Segundo Mateo e Leung (2008) e Hoogenboom (2000) a acuracidade dos dados reais é de grande importância para as pesquisas na agricultura, assim como o desenvolvimento de novos métodos para corrigir lacunas e/ou anormalidades também são.

CAPITULO 1 - OBJETIVOS DA DISSERTAÇÃO

1.1 OBJETIVO GERAL

O objetivo geral desta dissertação é criar um modelo capaz de corrigir as lacunas existentes nos dados climáticos, coletados de estações meteorológicas automáticas, utilizando para isso a Arquitetura P2P, definida no *framework* P2PComp.

1.2 OBJETIVOS ESPECÍFICOS

Os objetivos específicos desta dissertação são:

- Modelar o processo de correção de lacunas entre as estações meteorológicas;
- Implementar uma aplicação com base no modelo criado;
- Realizar testes com os dados corrigidos e com os dados gerados por um modelo estocástico.

CAPITULO 2 - FUNDAMENTAÇÃO TEÓRICA

2.1 CLIMA E TEMPO

O tempo e o clima são dois estados diferentes, embora sejam confundidos, o primeiro reflete a condição atual em uma determinada região, permitindo fazer previsões de um determinado dia ou semana, e o segundo é uma média dessas condições (PEREIRA et al., 2002; MORAES, 2003) caracterizada por uma série de 30 anos de dados definido pela Organização Mundial de Meteorologia (OMM), com base em métodos estatísticos de tendência do valor médio, e que serve como entrada para simulações que visam verificar os impactos no futuro. Comumente, ambos servem de subsídio para que agricultores e pesquisadores tomem decisões corretas, quando necessário.

Neste trabalho de pesquisa utilizar-se-á de dados sobre o clima, que tem em sua composição fatores como radiação solar, temperatura do ar, umidade relativa do ar, pressão atmosférica, precipitação pluviométrica, velocidade e direção do vento (PEREIRA et al., 2002), embora os que mais afetem o desenvolvimento e a produtividade agrícola são a precipitação pluviométrica, temperatura do ar e a radiação solar (SENTELHAS & MONTEIRO, 2009, p.3; HOOGENBOOM, 2000).

Segundo Hoogenboom (2000), na agricultura, a radiação solar fornece a energia necessária para a fotossíntese da planta, afetando assim o seu crescimento, já a temperatura do ar é responsável por vários processos nas plantas, como crescimento, transpiração, florescimento entre outros, e por isso é considerada uma das principais variáveis meteorológicas (PEREIRA et al, 2002; MAVI & TUPPER, 2004). E por fim, a precipitação pluviométrica que atua de forma indireta sobre o desenvolvimento da planta, quando em excesso afeta o solo e por consequência diminui a absorção de nutrientes pelas plantas e quando em falta afeta a fixação de CO₂ (HOOGENBOOM, 2000; SENTELHAS & MONTEIRO, 2009, p.4).

Também, cada região no Brasil apresenta características climáticas bem diferentes, o que deve ser levado em consideração na hora de executar uma simulação climática ou a previsão do tempo, pois o que pode aumentar a produtividade em um local, em outro pode reduzir.

2.2 CLASSIFICAÇÃO CLIMÁTICA DO BRASIL

Segundo Moraes (2003), no Brasil devido à sua extensão territorial, posição

geográfica, entre outros, não existe predomínio de somente um ambiente climático, como pode ser observado na Figura 1, onde estes são organizados segundo a classificação de Köppen.

Figura 1: Classificação climática do Brasil segundo Köppen



Fonte: Adaptado de PEREIRA, A. R; ANGELOCCI, L. R. SENTELHAS, P.C. Agrometeorologia: Fundamentos e Aplicações Práticas. Guaíba, RS: Editora Agropecuária, 2002.

Segundo Pereira (et al., 2002, p. 304), esta classificação é uma das mais conhecidas e utilizadas no meio científico por utilizar índices para determinar os ambientes climáticos e ainda definir grupos e subgrupos para detalhar cada ambiente, sendo que nesta classificação o Brasil apresenta nove regiões climáticas:

- Af: chuvas distribuídas ao longo do ano, sem estação seca;
- Am: pequena estação seca com influência de monções;
- Aw: inverno seco e chuvas máximas de verão;
- BSh: caracterizado por elevadas temperaturas e chuvas escassas;
- BWh: desértico quente de baixa latitude;
- Cfa: subtropical, sem estação seca e temperatura do mês mais quente maior que

22°C;

- Cfb: subtropical, sem estação seca e temperatura do mês mais quente menor que

22°C;

- Cwa: tropical de altitude, com inverno seco e temperatura do mês mais quente maior que 22°C;

- Cwb: tropical de altitude, com temperatura do mês mais quente menor que 22°C.

Comparando as séries climáticas obtidas junto ao Instituto Tecnológico SIMEPAR (SIMEPAR) das cidades de Ponta Grossa, Telêmaco Borba e Fernandes Pinheiro e as séries obtidas da Fundação ABC (FABC) das cidades de Castro, Piraí do Sul, Tibagi e Carambeí, observa-se que elas encontram-se dentro da zona Cfb, como mostra a Figura 2.

Figura 2: Classificação climática do Paraná segundo Köppen



Fonte: INSTITUTO AGRONÔMICO DO PARANÁ. Cartas Climáticas do Paraná. Disponível em: <
<http://www.iapar.br/modules/conteudo/conteudo.php?conteudo=597>> Acesso em: 24 ago. 2012.

Vale ressaltar que, cada ambiente possui suas próprias características climáticas e, portanto possui dados sobre o clima e o tempo específicos para a região, não podendo, por exemplo, utilizar dados climáticos da região sul para determinar o clima no nordeste, o que acarretaria em erros na simulação ou em previsões e gerariam incertezas quanto aos seus resultados.

2.3 DADOS CLIMÁTICOS

Os dados climáticos podem ser descritos como uma série temporal composta de variáveis meteorológicas, dos quais se destacam a temperatura do ar, radiação solar, precipitação pluviométrica, umidade relativa do ar, pressão atmosférica e vento.

Segundo Hoogenboom (2000), eles são importantes para a criação de aplicações agrícolas que envolvam simulação e ainda servem de subsídio na tomada de decisão dos agricultores (MATEO & LEUNG, 2008).

Para que seja possível a utilização dessas séries temporais os dados precisam ser coletados em estações que podem ser convencionais (EMC) ou automáticas (EMA) em períodos pré-determinados e não podem conter observações anormais e lacunas, embora seja difícil evitar esses problemas, pois as EMC's sofrem com a interferência humana e nas EMA's podem ocorrer problemas de conexão, no envio de dados, cabeamento rompido entre outros problemas. Schneider (2001), também cita o clima como um fator na ocorrência desses erros, devido a sua grande variação espaço-temporal.

Ainda, segundo Pedro Jr. et al. (1987), para uma coleta de dados eficiente devem-se tomar as seguintes precauções com a localização das estações: evitar condições extremas de relevo; área deve ser bem exposta; evitar elementos que possam projetar sombra; área plana e de fácil acesso e também a distância entre duas estações é definida pela OMM, e dista de no máximo 150 km. Para fins específicos esse valor é menor, como por exemplo, agrometeorologia (PEREIRA et al., 2002, p.456-457), onde quanto maior a quantidade de dados disponível para simulação e/ou previsão, melhores são os resultados.

Pereira et al. (2002) também cita que as estações de coleta de dados podem ser classificadas em três classes, conforme a quantidade de elementos monitorados, sendo que na primeira classe encontram-se aquelas que monitoram todos os elementos meteorológicos; na segunda classe não se mede a pressão do ar, ventos e radiação solar; e na terceira classe mede-se apenas temperatura do ar e precipitação pluviométrica

2.4 REDES DE COMPUTADORES

Segundo Comer (1998) e Coulouris et al. (2007), as redes de computadores estão por toda parte e são fundamentais na computação, podendo ser utilizadas para os mais variados fins, como agricultura, ambiente doméstico, telefonia móvel, entre outros.

Sua expansão foi rápida, fazendo com que em apenas alguns anos pessoas que não tinham acesso a qualquer tipo de rede passassem a utilizar suas facilidades e consumir seus mais variados produtos. Um exemplo dessa expansão é a Internet, onde Comer (2007) cita que no começo era apenas um projeto de pesquisa com poucos *sites*, mas atualmente é uma ferramenta de negócio bastante rentável e com o poder de alcançar milhões de pessoas.

Junto com essa explosão surgiram diversos tipos de redes como mostra o Quadro 1, sendo estas locais (LAN), metropolitanas (MAN) e de longa distância (WAN) e ainda recentemente variações destas sem fio, WLAN, WMAN e WWAN (COULOURIS et al., 2007; TANENBAUM, 2003).

Quadro 1 – Tipos de redes de computadores

Tipo	Exemplo	Alcance (km)	Banda (Mbps)	Latência (ms)
LAN	Ethernet	1-2	10-1000	1-10
WAN	Internet	Mundial	0,010-600	100-500
MAN	DSL	2-50	1-150	10
WLAN	WiFi	0,15-1,5	2-54	5-20
WWAN	GSM, 3G	Mundial	0,010-2	100-500
WMAN	WiMAX	5-50	1,5-20	5-20

Fonte: Adaptado de COULOURIS, G.; DOLLIMORE, J.; KINDBERG, T. *Sistemas Distribuídos: Conceitos e Projetos*. 4º Ed. Porto Alegre: Bookman, 2007.

Com essa variedade de tipos de rede faz-se necessária uma forma de comunicação entre elas, onde um dos seus protocolos mais conhecidos é o TCP/IP (COMER, 1998), formado por um conjunto de convenções para permitir a interconexão e a comunicação.

Segundo Comer e Stevens (1999), o TCP/IP é o protocolo mais utilizado para a comunicação inter-redes, embora existam outros, nenhum tem tanta interoperabilidade e abrange os mais diversos tipos de aplicativos.

Com a existência desses padrões de comunicação tornou-se possível a existência de aplicativos cliente-servidor e até mesmo o processamento distribuído de dados. Alguns exemplos dessas aplicações são o projeto SETI@home que utiliza computadores pessoais (PC) para buscar vida extraterrestre e a tecnologia *Peer-to-Peer* (P2P).

2.5 COMUNICAÇÃO CLIENTE-SERVIDOR

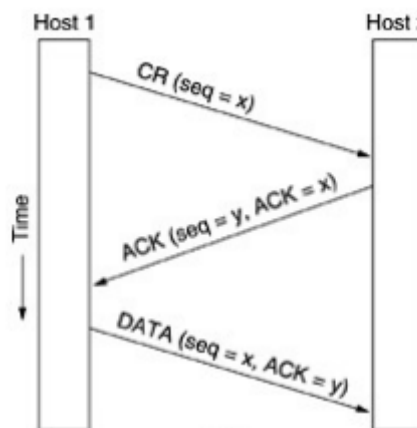
Segundo Coulouris et al. (2007) e Comer (1998), a comunicação cliente/servidor foi projetada para dar suporte a troca de mensagens entre cliente-servidor típicas, onde o servidor é um aplicativo qualquer, que fornece algum tipo de serviço, e que possa ser consumido por

um cliente.

Em sua forma mais comum de comunicação, o cliente e o servidor são síncronos, ou seja, o cliente envia uma requisição e fica bloqueado aguardando até que o servidor envie uma resposta (Figura 3).

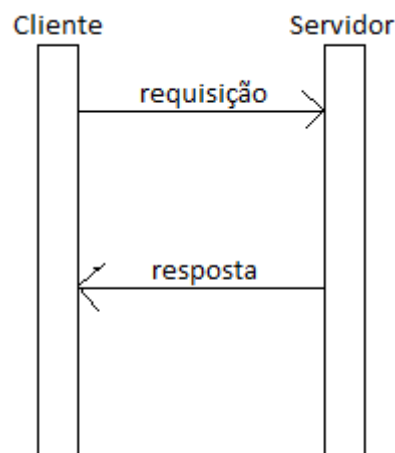
Coulouris et al. (2007, p.146) cita que o sincronismo é uma forma confiável de garantir que tanto cliente como servidor receberão a mensagem, pois são enviados pacotes de confirmação ACK (*acknowledge*) quando o cliente requisita uma conexão, no envio de dados e na desconexão, embora seja possível também a comunicação assíncrona, onde o cliente não fica bloqueado, como é o caso da utilização de datagramas UDP (*User Datagram Protocol*), exibido na Figura 4.

Figura 3 – Esquema de comunicação TCP



Fonte: TANENBAUM, Andrew S. Redes de Computadores. 4 ed. São Paulo: Campus, 2003.

Figura 4 – Esquema de Comunicação UDP



Fonte: Adaptado de COMER, Douglas E. Interligação em Rede com TCP/IP: Princípios, Protocolos e Arquitetura. 2º ed. Rio de Janeiro: Campus, 1998.

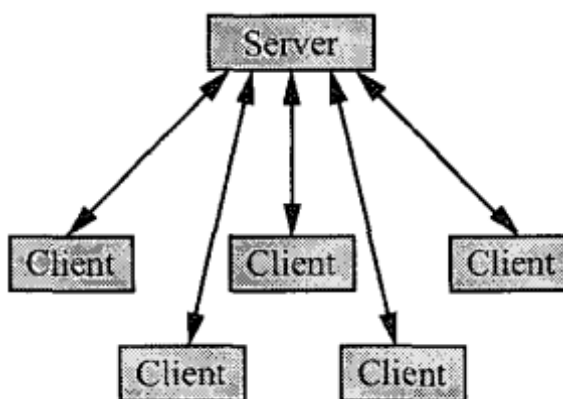
Quando se utiliza a comunicação cliente-servidor através de UDP, diferentemente do TCP, perde-se a confiabilidade na entrega das mensagens e como cita Tanenbaum (2003, p.559) não existe o controle de fluxo, controle de erros e retransmissão de pacotes perdidos, sendo estes controlados através da aplicação cliente. Um exemplo comum da aplicação de UDP é a Internet, onde se depois de feita uma requisição o cliente não obtém nenhuma resposta ele desconecta através de *timeout*.

Além dessas aplicações, onde cliente e servidor normalmente aparecem em computadores distintos, em alguns casos podem estar juntos, fazendo com que o computador adote o papel de cliente, quando precisa enviar uma requisição, e como servidor, quando é necessário algum tipo de processamento, como por exemplo, o P2P.

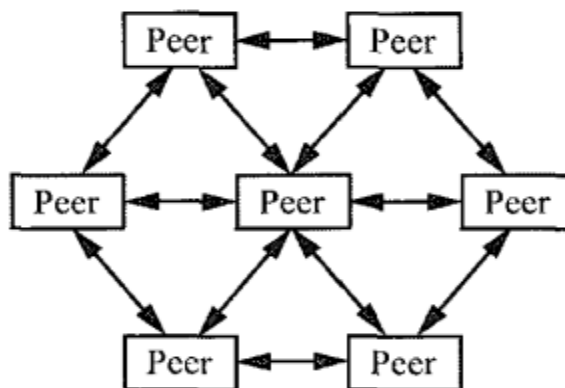
2.6 PEER-TO-PEER (P2P)

Segundo Endo et al. (2004) vários *softwares* de colaboração utilizam a arquitetura cliente-servidor (Figura 5), embora ela apresente algumas vantagens como por exemplo, manter os dados em um servidor e os clientes requisitarem esses dados quando necessário, quando o servidor está fora do ar nenhum cliente consegue compartilhar dados na rede.

Figura 5 – Arquitetura Cliente-Servidor



Fonte: ENDO, Satoru; MIYAMOTO, Toshiyuki; KUMAGAI, Sadatoshi; FUJII, Taku. A Data Synchronization Method for Peer-to-Peer Collaboration Systems. ISCIT. P.368-373, 2004.

Figura 6 – Arquitetura *Peer-to-Peer* (P2P)

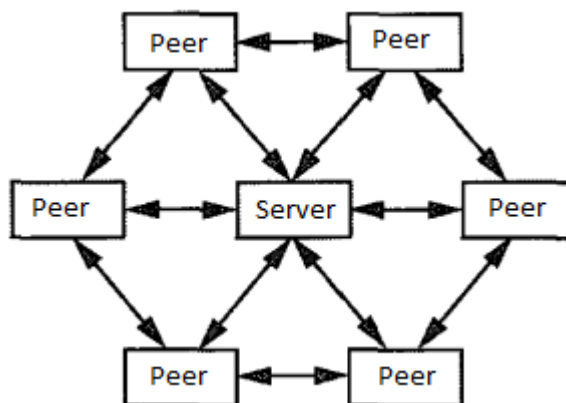
Fonte: ENDO, Satoru; MIYAMOTO, Toshiyuki; KUMAGAI, Sadatoshi; FUJII, Taku. A Data Synchronization Method for Peer-to-Peer Collaboration Systems. ISCIT. P.368-373, 2004.

Por outro lado, na Arquitetura P2P (Figura 6), que segundo Coulouris et al. (2007) surgiu como consequência do rápido crescimento da Internet, o objetivo principal é a colaboração de dados e recursos computacionais entre vários computadores, chamados *peers*, mesmo sem a presença de um gerenciador central. Nesta arquitetura estão presentes as seguintes características: cada *peer* tem as mesmas autoridades e responsabilidades; todos podem trocar dados com todos; um *peer* apresenta funções de servidor e de cliente; seu funcionamento não depende de gerenciadores centralizados (COULOURIS et al., 2007, p. 350; ENDO et al., 2004, p.368; MASUD et al., 2009, p.339).

Zhao et al. (2004) também cita que os recursos de todos os *peers* podem ser reunidos para formar um grande conjunto de informações, processamento e armazenamento e Schoder et al. (2005) complementa que estes podem ser escaláveis, possuir baixo custo e auto-organização. Assim, o P2P oferece maior flexibilidade, robustez e agrega mais recursos e capacidades computacionais que os tradicionais cliente-servidor (ZHAO et al., 2004; MOON e CHO, 2011).

Ainda no aspecto de organização, em P2P pode-se encontrar em duas formas: descentralizada e híbrida. Na forma descentralizada (Figura 6), a arquitetura trabalha sem servidores centrais e os *peers* são responsáveis pelas informações fornecidas na rede e por realizar as buscas quando precisam de algum recurso. Já atuando de forma híbrida (Figura 7), existe a presença de um servidor central, mas ele não é responsável pelo armazenamento dos dados, funcionando apenas como um índice, depois de estabelecida a conexão entre os *peers* o servidor para de atuar como um intermediário.

Figura 7 – Organização Híbrida P2P



Fonte: O Autor

Truelove et al. (2002) também cita que qualquer tipo de recurso pode ser um *peer*, já que as aplicações P2P utilizam técnicas de compartilhamento distribuído de recursos. Entre as aplicações já amadurecidas que utilizam P2P encontram-se o compartilhamento de arquivos, mensagens instantâneas e computação distribuída.

Neste trabalho de pesquisa a organização adotada é a híbrida, pois é a mais comumente utilizada por não sobrecarregar a rede com mensagens de busca de *peers*, e por que permite a localização direta de um *peer* recorrendo para isso à lista disponível no servidor.

2.7 USO DE REDES DE COMPUTADORES PARA COMPARTILHAMENTO E DISTRIBUIÇÃO DE DADOS CLIMÁTICOS

Trabalhos envolvendo o compartilhamento e/ou distribuição de dados climáticos dentro da área deste trabalho de pesquisa ainda são poucos, embora exista o processo de compartilhar de dados em outras áreas e/ou utilizando em conjunto com outras ferramentas e tecnologias.

Sangket et al.(2008), estudou uma forma de compartilhar dados de pesquisas biológicas entre vários centros de bioinformática e universidades, comparando para isso a eficiência do transporte de grande quantidade de dados utilizando *file transfer protocol* (FTP) e P2P. Com sua pesquisa foi possível verificar que o compartilhamento utilizando P2P se saiu melhor em termos de velocidade e quantidade de dados compartilhados que o FTP.

Masud e Kiringa (2009), apresentaram uma aplicação utilizando P2P, que realizava o processo de compartilhamento e também interagiu com um banco de dados relacional. No trabalho proposto quando um *peer* executa um processo de inserção, exclusão ou atualização

em sua base relacional ele propaga essa tarefa para que outros *peers* executem o mesmo processo em suas bases de dados.

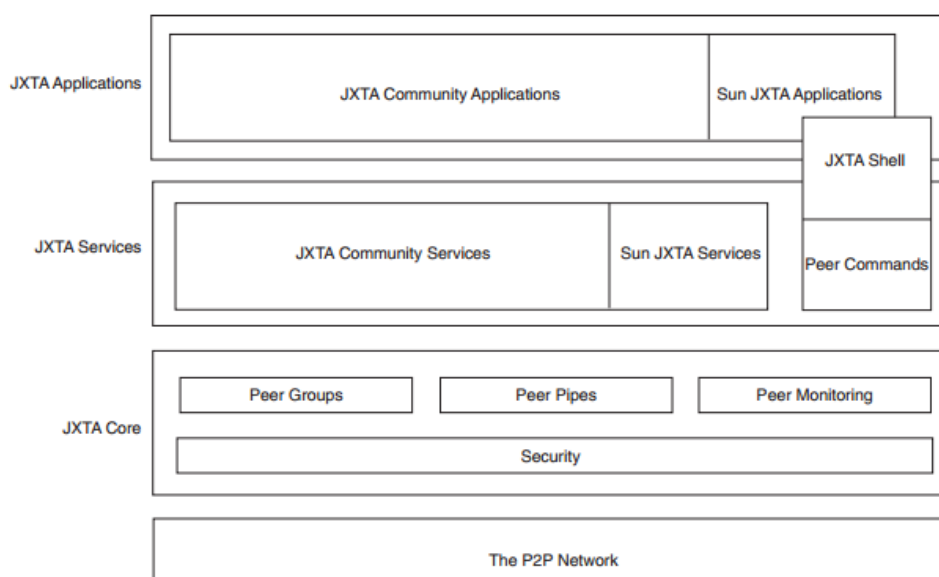
Já em seu trabalho Endo et al. (2004), apresentou um *software* colaborativo com P2P utilizado para compartilhar e gerenciar os dados através de permissões concedidas por *peers* administradores sem a presença de um servidor central, utilizando a Arquitetura P2P pura. Em seu trabalho quando um *peer* quer realizar a tarefa de compartilhamento, ele envia o pedido para um *peer* administrador que após aceitar o pedido propaga a tarefa para outros *peers* da rede.

2.8 O PADRÃO JXTA

Segundo Brookshier et al. (2002), Wilson (2002) e Verstrynge (2008), o padrão JXTA descreve um conjunto de protocolos projetados especialmente para a comunicação em redes P2P. Através desses protocolos, *peers* podem cooperar e formar grupos sem a necessidade de um gerenciador central. O padrão é organizado em camadas (Figura 8), núcleo, serviços e aplicações, onde este último utiliza o acesso provido pelos serviços aos protocolos e, este por sua vez acessa as funcionalidades do núcleo.

Ainda, segundo Oaks et al. (2002) e BrookShier et al. (2002), o JXTA apresenta alguns conceitos como *peer groups*, *pipes* e *endpoints* e também a comunicação e descoberta de *peers* através de anúncios que utilizam documentos do tipo *Extensible Markup Language* (XML).

Figura 8 – As três camadas do padrão JXTA



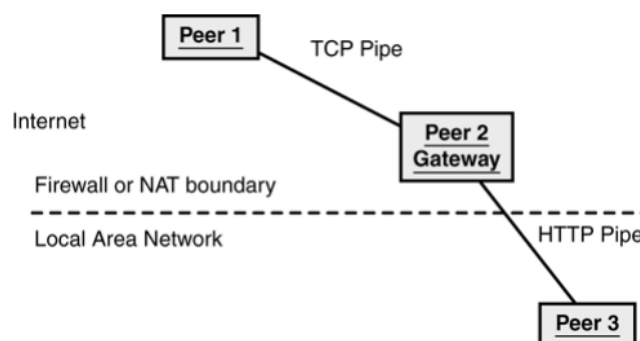
Fonte: WILSON, Brendon. JXTA. New Riders: Berkeley, CA, 2002.

Os *peergroups* são grupos de *peers* que compartilham uma mesma aplicação seguindo as políticas de segurança do grupo, sendo que os *peers* só podem se comunicar com outros *peers* do mesmo grupo, a menos que este faça parte de outros grupos, o que é permitido.

Os *pipes* provem uma conexão virtual entre os *peers*, permitindo sua comunicação e compartilhamento de informações nos mais diversos tipos de redes, passando por *firewalls* e outras possíveis barreiras, fazendo com que os *peers* se comuniquem mesmo sem saber onde estão localizados.

Segundo Brookshier et al. (2002), vários protocolos tem um endereço fixo, onde uma URL ou endereço IP é utilizado para se comunicar com os clientes, no padrão JXTA essa ideia de endereço cliente é abstraída e chamada de *endpoint*. Cada *peer* pode ter um ou vários *endpoints*, se comunicando por TCP, quando na rede local, e HTTP através de um *firewall*, como apresenta a Figura 9.

Figura 9 – Comunicação entre *peers* utilizando *pipes* e *endpoints*



Fonte: BROOKSHIER, Daniel; GOVONI, Darren; KRISHNAN, Navaneeth; SOTO, Juan Carlos. JXTA: Java P2P Programming. Indianapolis: Sams Publishing, 2002.

Ainda no padrão JXTA é possível encontrar dois tipos de *peers*, além do comum, *peers relay* e *peers rendezvous*. O primeiro é utilizado para descobrir os recursos da rede de forma dinâmica e permitir a comunicação entre *peers* através de HTTP (Figura 9), estes são conhecidos também como *gateways* e *routers*. Já o segundo funciona como uma lista estática contendo todos os *peers* presentes na rede e os recursos disponíveis, quando um *peer* busca por um recurso ele pergunta diretamente ao *peer rendezvous* se ele está disponível na rede para então acessá-lo.

Os anúncios são documentos XML utilizados para anunciar um recurso na rede P2P e procurar o que está disponível nela, esse recurso pode ser outro *peer*, *peergroups*, *pipes* entre outros.

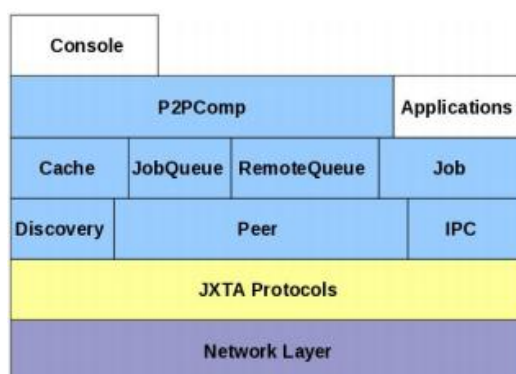
2.9 P2PCOMP

O P2PComp é um *framework* criado por Senger et al. (2010) e que utiliza os protocolos do padrão JXTA para seu funcionamento, o que permite sua utilização em redes onde existam *firewalls* ou outros tipos de bloqueios. Ele também utiliza a funcionalidade de anúncios do JXTA para criar um catálogo de elementos de processamento através da rede P2P, criando seu próprio anúncio personalizado contendo capacidade de processamento, carga média, número de processadores e capacidade de memória.

A comunicação entre os elementos, dentro do P2PComp, se dá basicamente através de matriz de custos, que contem o tempo de resposta para cada *peer* da rede. Através desta, um *peer* pode identificar e submeter uma aplicação para os *peers* com menores *delays* e minimizar a sobrecarga da rede, ainda segundo Senger et al. (2010) seu diferencial é que o *framework* é orientado ao modelo P2P puro e prove rotinas de comunicação ponto-a-ponto, o que permite a execução de programas paralelos do tipo *Single Program Multiple Data* (SPMD).

A estrutura do P2PComp é apresentada na Figura 10, onde segundo Senger et al. (2010), as principais funcionalidades relacionadas com a rede P2P estão implementadas nas classes *Peer*, *Discovery* e *IPC*. Na classe *Peer* estão os procedimentos necessários para a inicialização dos protocolos do padrão JXTA e para o controle das mensagens trocadas dentro do *framework*, a classe *Discovery* é responsável por fazer a publicação e buscar por anúncios de outros *peers* na rede *P2P* e a *IPC* é responsável pela comunicação entre os processos para implementação de aplicações paralelas.

Figura 10 – Estrutura de classes P2PComp



Fonte: SENGER, Luciano José; SOUZA, Márcio Augusto de; FOLTRAN JR, Dierone Cesar. Towards a peer-to-peer framework for parallel and distributed computing. 22nd International Symposium on Computer Architecture and High Performance Computing. p.127-134, 2010.

Todas as aplicações desenvolvidas, utilizando o *framework*, são colocadas sob o pacote *Applications* e com a ajuda do pacote *Console* é feito a criação e execução dos processos. Um exemplo simples de aplicação é apresentado na Figura 11, onde os processos filhos, com *rank* diferente de zero, enviam uma mensagem “*Hello World from*” para o processo mestre, *rank* zero, e este imprime na tela as mensagens recebidas.

Figura 11 – Código de uma aplicação de exemplo utilizando o P2PComp

```
public class HelloWorld extends Job{
    public void run(){
        if(getRank() == 0){
            // Master task
            size = getSize();
            i = 1;
            while(i < size){
                Message msg = recv(-1, -1);
                i++;
            }
            Vector results = new Vector();
            results.add(i);
            sendResults(results);
        }
        else{
            // A worker task
            message = new Vector();
            message.add(
                new String("Hello_world_from_"+
                    getRank())
            );
            send(0, 1, message);
        }
    }
}
```

Fonte: SENGER, Luciano José; SOUZA, Márcio Augusto de; FOLTRAN JR, Dierone Cesar. Towards a peer-to-peer framework for parallel and distributed computing. 22nd International Symposium on Computer Architecture and High Performance Computing. p.127-134, 2010.

2.10 CONCEITOS ESTATÍSTICOS

Segundo Bisquerra et al. (2007) costuma-se iniciar uma análise de dados sempre pela descrição das variáveis observadas, utilizando-se para isso medidas de tendência central, variabilidade, entre outros. Esses procedimentos iniciais permitem analisar os dados e verificar já no início possíveis valores fora do comum (*outliers*) e corrigi-los, pois se estiverem incorretos tudo o que for feito posteriormente ficará sem valor.

Uma das medidas de tendência central mais conhecida é a média aritmética, que representa o ponto médio de um conjunto de dados, embora existam outras como a mediana, a moda e outros tipos de média como a ponderada e a geométrica.

Já nas medidas de variabilidade, as principais são a variância e o desvio padrão, que expressam a dispersão dos dados descrevendo o quão longe se encontram da média, a segunda é a raiz quadrada da primeira e por ser menor sua interpretação é mais fácil e direta.

Além dessas análises também pode-se verificar a distribuição dos dados em relação a normal, que segundo Montgomery et al. (2004) é o modelo mais largamente utilizado para a distribuição de uma variável aleatório e que apresenta as seguintes características (BISQUERRA et al., 2007, p.52):

- é simétrica em torno do eixo que passa pela média;
- a ordenada máxima coincide com a média no eixo das abscissas;
- a média, a mediana e a moda coincidem;
- é assintótica em relação ao eixo das abscissas;
- o desvio padrão coincide com a distância entre o eixo de simetria e o ponto de inflexão da curva;

Conhecendo-se as medidas de tendência central e de variabilidade de um conjunto e se ele aproxima-se da distribuição normal pode-se fazer algumas inferências sobre esses resultados quando trabalha-se com mais de um conjunto de dados, como o teste F de Fisher e o t de Student. Segundo Bisquerra et al. (2007) e Montgomery et al. (2004) o teste F é utilizado para verificar a homogeneidade das variâncias, descrevendo se elas são homocedásticas, quando não existe diferença estatística significativa, ou heterocedásticas, quando essa diferença estatística é significativa.

Já para testar a média dos conjuntos de dados existe o teste t de Student, que descreve se as médias de dois conjuntos apresentam diferenças estatisticamente significantes ou se elas podem ser consideradas como procedentes de uma mesma população.

CAPITULO 3 - MATERIAIS E MÉTODOS

3.1 LOCAL DO DESENVOLVIMENTO DA PESQUISA

Todo o desenvolvimento da pesquisa e execução dos testes foram feitos no Laboratório de Computação de Alto Desempenho da Universidade Estadual de Ponta Grossa (UEPG), utilizando três computadores com Sistema Operacional Linux Slackware com processador AMD Athlon 64 X2 5200+ e 2 GB de Memória RAM.

3.2 FERRAMENTAS

Para a estruturação da rede P2P utilizada no modelo de correção foi utilizado o *framework* P2PComp proposto por Senger et al. (2010) e para a implementação em linguagem de programação foi utilizada a Linguagem Java SE 6u25.

A execução dos testes estatísticos, a simulação dos dados e os gráficos gerados para análise dos resultados foram feitos no Microsoft® Excel® 2007.

3.3 DADOS CLIMÁTICOS

Os dados climáticos utilizados nesta dissertação foram obtidos junto ao SIMEPAR e a FABC contendo dados diários das localidades apresentadas no Quadro 2 e foram coletados de estações automáticas.

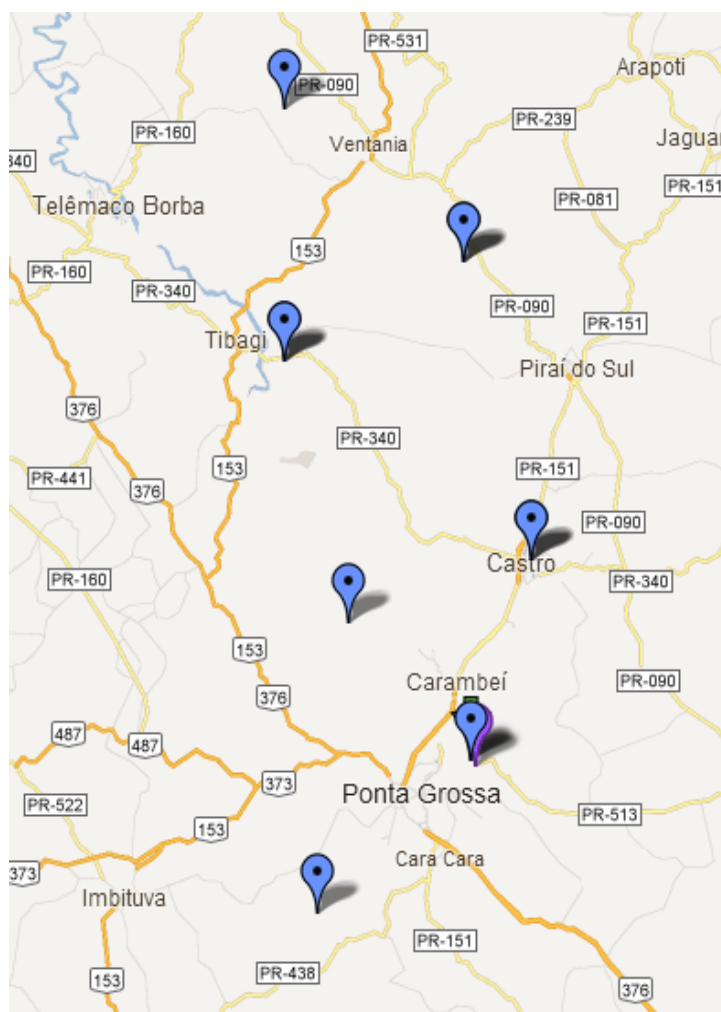
Quadro 2 – Localização das bases climáticas utilizadas na pesquisa

Cidade	Latitude	Longitude	Período	Fornecedor
Ponta Grossa	-25,05	-50,09	1997-2011	SIMEPAR
Telêmaco Borba	-24,20	-50,37	1997-2011	SIMEPAR
Fernades Pinheiro	-25,25	-50,32	1997-2011	SIMEPAR
Castro	-24,79	-50	2008-2012	FABC
Tibagi	-24,53	-50,37	2009-2012	FABC
Carambeí	-24,87	-50,22	2009-2012	FABC
Piraí do Sul	-24,40	-50,10	2009-2012	FABC

Fonte: O Autor

Foi utilizada a base de Carambeí para executar os testes com o modelo de correção criado por se apresentar mais central em relação as outras bases (Figura 12) e também nesta foram geradas lacunas no ano de 2011 nos períodos apresentados no Quadro 3.

Figura 12 – Mapa com as estações utilizadas no trabalho



Fonte: GOOGLE MAPS. Disponível em <<http://maps.google.com>> Acesso em 27 ago. 2012.

Quadro 3 – Períodos onde foram geradas lacunas

Mês	Período	Dias
Fevereiro	01 a 10	10
Fevereiro	11 a 20	10
Fevereiro	19 a 28	10
Fevereiro	01 a 20	20
Fevereiro	09 a 28	20
Fevereiro	01 a 28	28
Maio	01 a 10	10
Maio	11 a 20	10
Maio	22 a 31	10
Maio	01 a 20	20
Maio	12 a 31	20
Maio	01 a 31	31
Agosto	01 a 10	10
Agosto	11 a 20	10

Agosto	22 a 31	10
Agosto	01 a 20	20
Agosto	12 a 31	20
Agosto	01 a 31	31
Novembro	01 a 10	10
Novembro	11 a 20	10
Novembro	21 a 30	10
Novembro	01 a 20	20
Novembro	11 a 30	20
Novembro	01 a 30	30

Fonte: O Autor

Para a correção das lacunas geradas os testes foram divididos em duas baterias, uma utilizando três estações, Castro, Tibagi e Fernandes Pinheiro e outra utilizando todas as localidades, Castro, Tibagi, Fernandes Pinheiro, Ponta Grossa, Pirai do Sul e Telêmaco Borba.

A variável climática a ser utilizada na correção foi a temperatura por se apresentar mais próxima da normal e que segundo Steinnhaeuser et al. (2010), Sentelhas e Monteiro (2009, p.3) e Hoogenboom (2000) é a que mais apresenta significância por exemplo, para determinar regiões climáticas ou índices que atuam como preditores, embora a precipitação seja de igual importância, ela apresenta uma variação espaço-temporal muito grande que dificulta sua utilização.

3.4 TESTES E MEDIDAS ESTATÍSTICAS

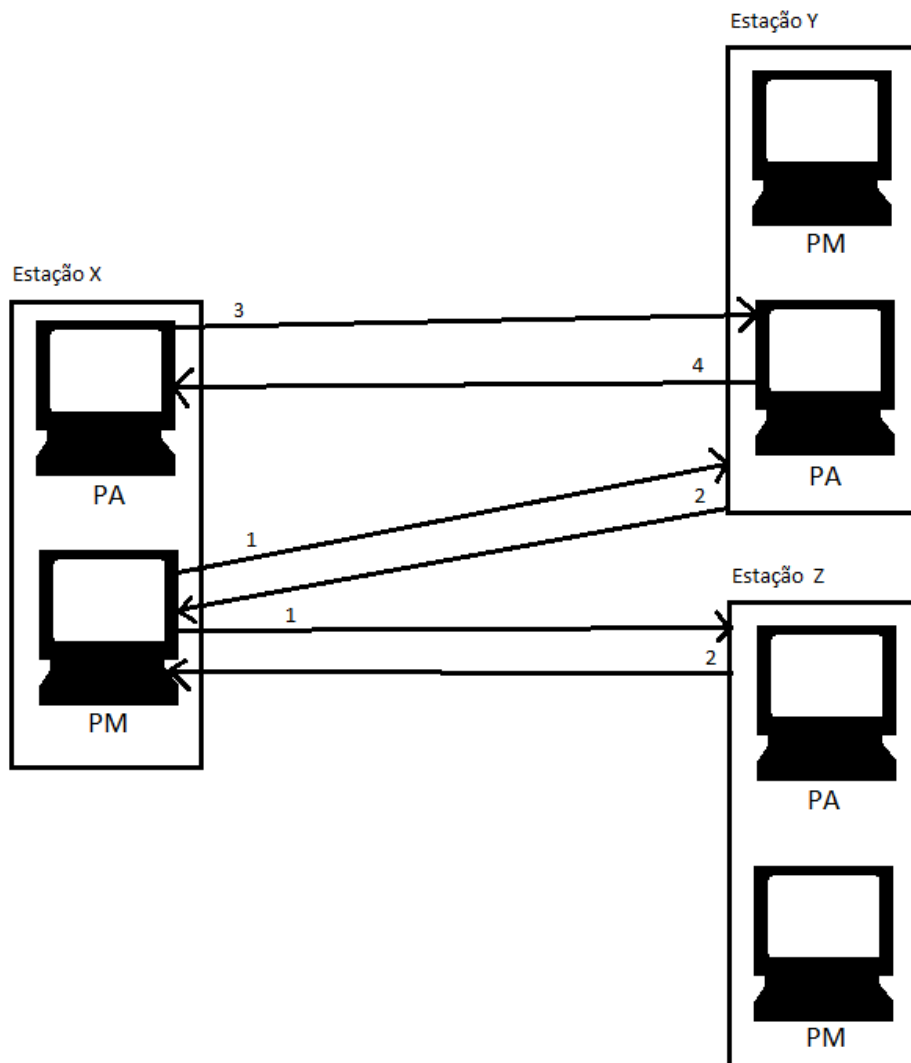
Os testes estatísticos foram utilizados após a correção das lacunas para verificar o desempenho do modelo criado em preencher as lacunas encontradas. Para tal foram utilizados o teste F de Fisher e o teste t de Student. As medidas utilizadas foram a média aritmética e o desvio padrão, para os dados observados, simulados no Microsoft® Excel® 2007 e gerados pelo modelo criado. Foram feitas comparações dos dados simulados no Microsoft® Excel® 2007 com os observados e também dos dados gerados pelo modelo criado com os dados observados.

CAPITULO 4 - RESULTADOS

4.1 MODELO DE CORREÇÃO

No modelo de correção criado o processo ocorre através de dois *peers* (Figura 11) para cada estação, um que funciona como gerenciador do processo de correção e que é responsável também por encontrar os *peers* que participaram da correção, chamado de *peer main* (PM), e outro que atua na rede P2P solicitando os períodos a serem corrigidos ou fornecendo os dados solicitados, chamado de *peer* atuador (PA).

Figura 13 – Modelo de correção criado



Fonte: O Autor

Dessa forma o modelo de correção funciona da seguinte maneira, o PM da estação X

propaga uma solicitação na rede P2P pedindo a localização, em coordenadas do Sistema Universal Transverso de Mercator (UTM), das estações presentes na rede P2P (Passo 1). Com a lista de localizações ele calcula a distância entre a sua localização e cada uma das outras estações relacionando as que estão mais próximas, sempre respeitando o limite de 150 Km como descrito em 2.3 (Passo 2).

Com a lista de estações criada o PM da estação X dispara o processo de correção na rede e o atuador PA da estação começa a procurar as lacunas nos dados climáticos, com as lacunas encontradas ele repassa os dias onde elas ocorreram para cada PA das outras estações presentes na lista gerada pelo PM (Passo 3)

Os PA's das outras estações buscam pelos dados climáticos dos dias solicitados, sempre utilizando dois dias antes (a e b) e dois dias depois do dia solicitado (c e d), realizando uma média ponderada com peso 2 para o dia solicitado e peso 1 para os dois dias anteriores e para os dois dias posteriores. O resultado desta média ponderada é retornado para o PA da estação que solicitou a correção (Passo 4).

$$média\ local = \frac{a + b + (x * 2) + c + d}{6}$$

Após receber todas as médias das estações elencadas na lista criada pelo PM o PA realiza uma média aritmética, onde n indica a quantidade de estações presente na lista criada por PM e Xn indica o resultado da média ponderada retornado de cada estação e assim é gerado o valor para o dia solicitado.

$$valor = \frac{\sum_{n=1}^{\infty} xn}{n}$$

4.2 A APLICAÇÃO DE SINCRONIZAÇÃO

Para a realização dos testes com o modelo de correção criado, foi implementada uma aplicação utilizando o *framework* e a linguagem de programação citados em 3.2. Para tanto foram criados dois métodos, seguindo o modelo criado em 4.1, um para localizar as estações da rede P2P e calcular a distância entre estas (Algoritmo 1) e outro para realizar a correção propriamente dita (Algoritmo 2).

Algoritmo 1 – Cálculo de distância entre as estações

```

lista_peers = buscar estações da rede
lista_peers_correcao = vazio;
enquanto existem peer em lista_peers faça
    temp_peer = peer
    se peer != peer_local e peer != peer main então
        distancia = raiz((temp_peer.UTM_E - peer_local.UTM_E)^2
+ (temp_peer.UTM_N - peer_local.UTM_N)^2)
        se distancia <= 150 então
            lista_peers_correcao.adicionar(temp_peer)
    fim se
fim se
fim enquanto

```

Fonte: O Autor

Algoritmo 2 – Processo de correção

```

i = 1
tam = lista_peers_correcao.tamanho()
se peer mestre então
    períodos_lacunas = buscar_períodos()
    enquanto i < tam faça
        enviar(i, períodos_lacunas)
        i++
    fim enquanto
    i = 1
resultados = vazio
    enquanto i < tam faça
        media_estacao_remota = receber(-1, -1)
        resultados.adicionar(media_estacao_remota)
        i++
    fim enquanto
    valor = calcula_media_aritmetica(resultados)
    imprime valor
senão
    períodos = receber(peer mestre, -1)
    pesos = {1, 1, 2, 1, 1}
    resultados = vazio
    enquanto existem períodos faça
        data = períodos.proximo()
        de = calcula_data(data, -2)
        ate = calcular_data(data, +2)
        valores = buscar_valores(de, ate)
        media = media_ponderada(valores, pesos)
        resultados.adicionar(media)
    fim enquanto
    enviar(peer mestre, resultados)
fim se

```

Fonte: O Autor

Dentro do *framework* P2PComp o Algoritmo 2 foi inserido no pacote *applications*, como mostra sua organização de classes na Figura 10, e no pacote *console* foi adicionado o Algoritmo 1 para localizar as estações presentes na rede P2P junto com suas coordenadas UTM.

Por se tratar de uma aplicação do tipo *console*, sua execução ocorre através de linha de comando, bem como suas saídas e outros tipos de informação.

4.3 TESTES ESTATÍSTICOS DAS CORREÇÕES FEITAS UTILIZANDO TRÊS LOCALIDADES

Com base nos testes e medidas estatísticos descritos em 3.4 obtiveram-se os seguintes resultados (Tabela 1, Tabela 2, Tabela 3 e Tabela 4) quando foi executado o procedimento de correção, no modelo P2P e na simulação através do Microsoft® Excel® 2007, para a localidade de Carambeí utilizando as estações de Tibagi, Fernandes Pinheiro e Castro como fornecedoras dos dados climáticos.

Tabela 1 – Resultados da correção utilizando três estações para Fevereiro

	10 início			10 meio			10 final		
	Real	Simulado	P2P	Real	Simulado	P2P	Real	Simulado	P2P
Média	20,9000	21,4200	21,4400	20,3600	21,0800	20,9944	21,2500	20,4000	21,5306
DP	0,4690	1,1989	0,1646	1,0977	1,5711	0,5916	1,1058	2,1782	0,6366
teste F	-	0,0100	0,0046	-	0,3003	0,0797	-	0,0560	0,1155
teste t	-	0,2177	0,0054	-	0,2503	0,1250	-	0,2857	0,4957

	20 início			20 final			inteiro		
	Real	Simulado	P2P	Real	Simulado	P2P	Real	Simulado	P2P
Média	20,6300	21,2500	21,2172	20,8700	20,8100	21,2661	20,8786	20,9571	21,3026
DP	0,8670	1,3713	0,4805	1,1379	1,8987	0,6582	0,9908	1,7006	0,5604
teste F	-	0,0524	0,0134	-	0,0310	0,0214	-	0,0066	0,0042
teste t	-	0,0956	0,0117	-	0,9042	0,1858	-	0,8737	0,0552

Fonte: O Autor

Tabela 2 – Resultados da correção utilizando três estações para Maio

	10 início			10 meio			10 final		
	Real	Simulado	P2P	Real	Simulado	P2P	Real	Simulado	P2P
Média	16,1800	14,3000	15,8011	14,3600	15,8500	14,6094	14,3300	15,6100	13,6422
DP	2,5059	2,8000	1,5415	2,9740	3,5837	2,2940	3,0793	1,8187	2,4064
teste F	-	0,7463	0,1639	-	0,5874	0,4511	-	0,1326	0,4740
teste t	-	0,1310	0,6886	-	0,3251	0,8360	-	0,2726	0,5847

	20 início			20 final			inteiro		
	Real	Simulado	P2P	Real	Simulado	P2P	Real	Simulado	P2P
Média	15,2700	15,0750	15,2053	14,1650	15,8400	13,9436	14,9548	15,2839	14,6720
DP	2,8348	3,2294	1,9980	2,7848	2,7398	2,1631	2,8514	2,7731	2,1939
teste F	-	0,5755	0,1362	-	0,9442	0,2796	-	0,8798	0,1568
teste t	-	0,8403	0,9339	-	0,0627	0,7804	-	0,6468	0,6632

Fonte: O Autor

Tabela 3 – Resultados da correção utilizando três estações para Agosto

	10 início			10 meio			10 final		
	Real	Simulado	P2P	Real	Simulado	P2P	Real	Simulado	P2P
Média	12,9800	17,1900	13,0078	16,4300	15,0700	15,7294	16,0800	14,5900	15,6122
DP	5,5196	4,1162	3,4783	2,6891	5,7579	1,8520	4,8803	6,1578	3,2660
teste F	-	0,3952	0,1851	-	0,0332	0,2818	-	0,4993	0,2472
teste t	-	0,0691	0,9894	-	0,5072	0,5061	-	0,5562	0,8040

	20 início			20 final			inteiro		
	Real	Simulado	P2P	Real	Simulado	P2P	Real	Simulado	P2P
Média	14,7050	16,1300	14,3686	15,8750	14,1250	15,3881	14,8871	15,4806	14,6272
DP	4,5814	4,9913	3,0504	4,3899	5,2755	2,8843	4,8191	5,3155	3,1922
teste F	-	0,7126	0,0842	-	0,4305	0,0748	-	0,5947	0,0273
teste t	-	0,3528	0,7861	-	0,2613	0,6808	-	0,6467	0,8032

Fonte: O Autor

Tabela 4 - Resultados da correção utilizando três estações para Novembro

	10 início			10 meio			10 final		
	Real	Simulado	P2P	Real	Simulado	P2P	Real	Simulado	P2P
Média	18,5600	18,5300	18,6572	16,8100	18,9800	18,0044	20,8700	18,7300	20,5889
DP	3,8768	3,1401	1,7055	1,8478	2,5015	0,9299	1,4499	1,6547	0,5896
teste F	-	0,5400	0,0225	-	0,3803	0,0531	-	0,7004	0,0132
teste t	-	0,9850	0,9429	-	0,0406	0,0845	-	0,0065	0,5771

	20 início			20 final			inteiro		
	Real	Simulado	P2P	Real	Simulado	P2P	Real	Simulado	P2P
Média	17,6850	18,7550	18,3308	18,8400	18,8550	19,2967	18,7467	18,7467	19,0835
DP	3,0891	2,7727	1,3782	2,6365	2,0682	1,5271	3,0391	2,4263	1,5889
teste F	-	0,6424	0,0009	-	0,2986	0,0217	-	0,2311	0,0008
teste t	-	0,2562	0,4009	-	0,9841	0,5067	-	1,0000	0,5933

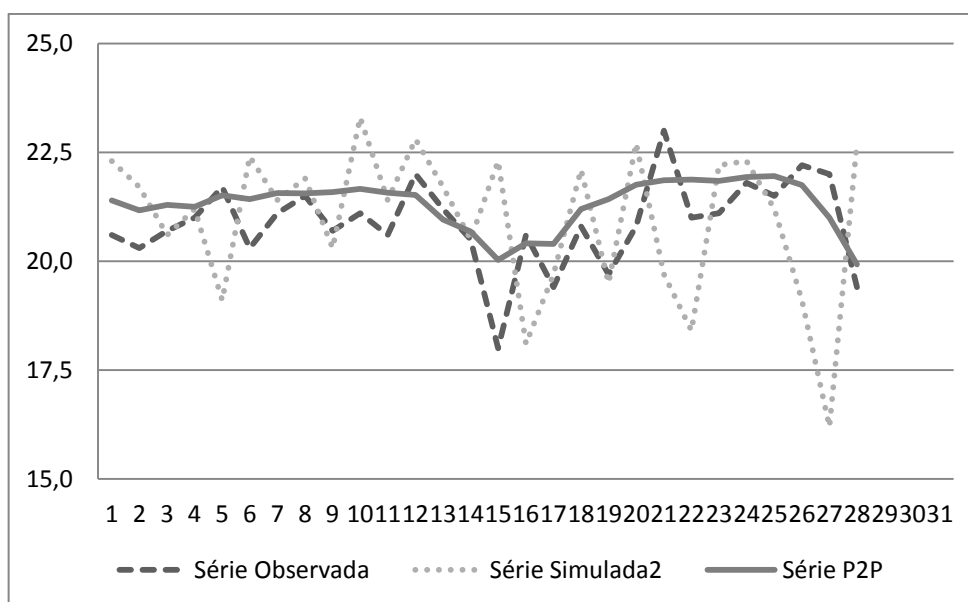
Fonte: O Autor

Analisando os resultados expostos nas Tabelas 1, 2, 3 e 4 é possível observar que no geral, utilizando três localidades, o modelo de correção através do P2P se saiu melhor que o

simulado, embora em Fevereiro, que é um mês de verão, o modelo simulado se saiu melhor sendo pior somente nos últimos 10 dias de Fevereiro.

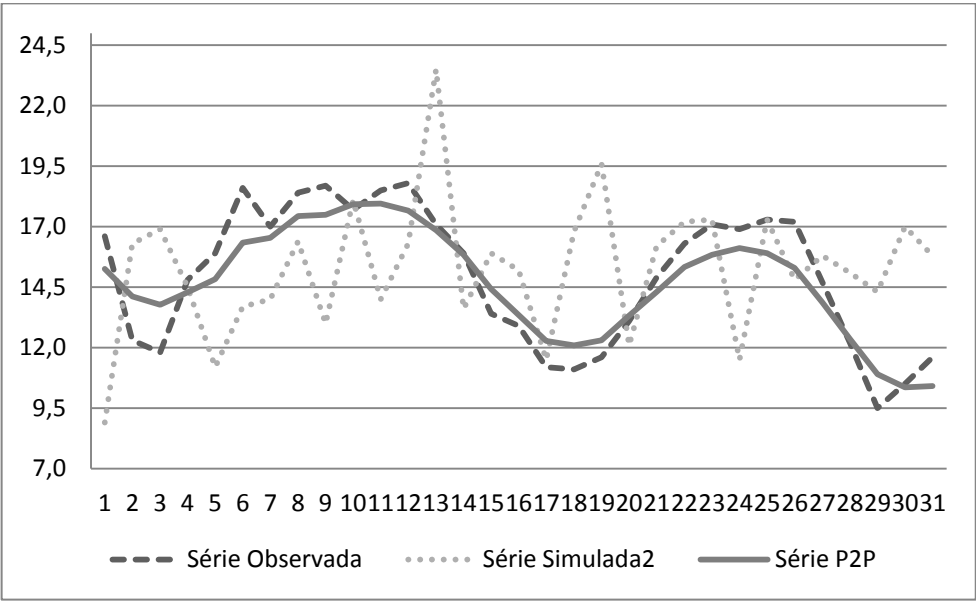
Observando os valores do teste t para o modelo P2P, pode-se dizer que ele conseguiu corrigidos os dados climáticos, que estatisticamente podem pertencer à mesma população, ou seja, os dados gerados refletem as mesmas informações de temperatura da base original. Também é possível verificar que o modelo P2P se saiu bem em corrigir dados para as estações de outono, inverno e primavera e para o verão obteve resultados inferiores as outras estações do ano (Figura 14, Figura 15, Figura 16 e Figura 17).

Figura 14 – Dados observados, simulados e corrigidos no modelo P2P para o mês de Fevereiro em 3 localidades.



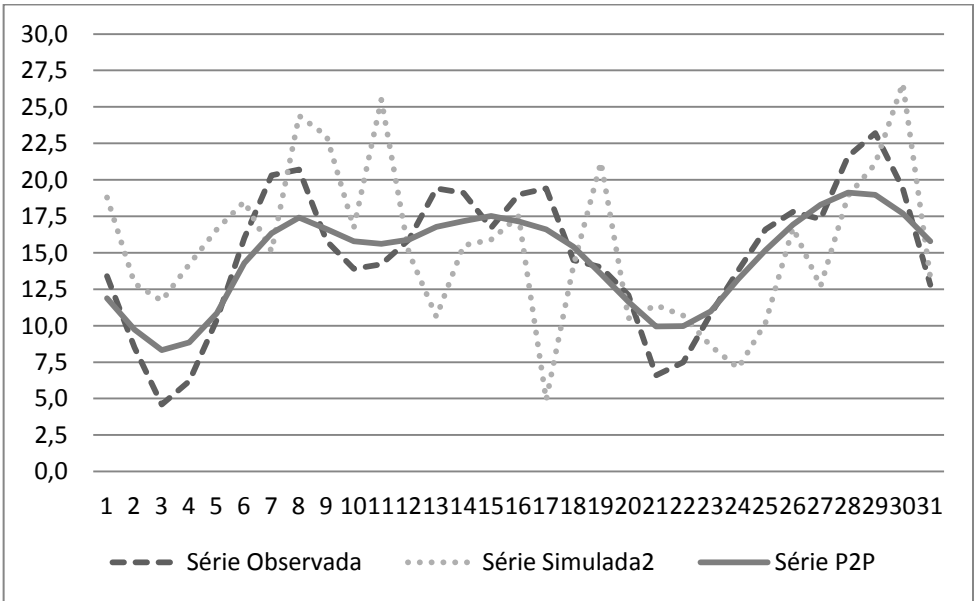
Fonte: O Autor

Figura 15 - Dados observados, simulados e corrigidos no modelo P2P para o mês de Maio em 3 localidades.



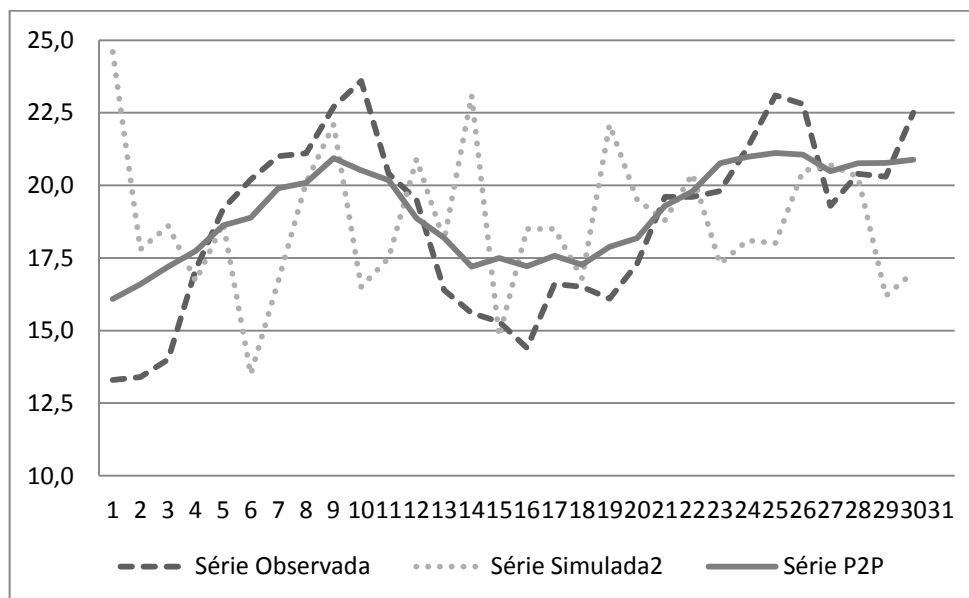
Fonte: O autor

Figura 16 - Dados observados, simulados e corrigidos no modelo P2P para o mês de Agosto em 3 localidades.



Fonte: O Autor

Figura 17 - Dados observados, simulados e corrigidos no modelo P2P para o mês de Novembro em 3 localidades.



Fonte: O Autor

4.4 TESTES ESTATÍSTICOS DAS CORREÇÕES FEITAS UTILIZANDO TODAS AS LOCALIDADES

Utilizando todas as localidades, Ponta Grossa, Tibagi, Castro, Pirai do Sul, Fernandes Pinheiro e Telêmaco Borba, para fornecer os dados para correção das lacunas na base de Carambeí, foram obtidos os seguintes resultados apresentados nas Tabelas 5, 6, 7 e 8.

Tabela 5 - Resultados da correção utilizando todas as estações para Fevereiro.

	10 início			10 meio			10 final		
	Real	Simulado	P2P	Real	Simulado	P2P	Real	Simulado	P2P
Média	20,9000	21,4200	21,4872	20,3600	21,0800	21,2081	21,2500	20,4000	21,8072
DP	0,4690	1,1989	0,1320	1,0977	1,5711	0,5436	1,1058	2,1782	0,6169
teste F	-	0,0100	0,0008	-	0,3003	0,0481	-	0,0560	0,0971
teste t	-	0,2177	0,0032	-	0,2503	0,0420	-	0,2857	0,1810

	20 início			20 final			inteiro		
	Real	Simulado	P2P	Real	Simulado	P2P	Real	Simulado	P2P
Média	20,6300	21,2500	21,3476	20,8700	20,8100	21,4747	20,8786	20,9571	21,4717
DP	0,8670	1,3713	0,4108	1,1379	1,8987	0,6279	0,9908	1,7006	0,5315
teste F	-	0,0524	0,0021	-	0,0310	0,0128	-	0,0066	0,0019
teste t	-	0,0956	0,0024	-	0,9042	0,0442	-	0,8337	0,0079

Fonte: O Autor

Tabela 6 – Resultados da correção utilizando todas as estações para Maio.

	10 início			10 meio			10 final		
	Real	Simulado	P2P	Real	Simulado	P2P	Real	Simulado	P2P
Média	16,1800	14,3000	15,9508	14,3600	15,8500	14,7722	14,3300	15,6100	13,8853
DP	2,5059	2,8000	1,6244	2,9740	3,5837	2,2629	3,0793	1,8187	2,3302
teste F	-	0,7463	0,2125	-	0,5874	0,4279	-	0,1326	0,4189
teste t	-	0,1310	0,8110	-	0,3251	0,7313	-	0,2726	0,7200

	20 início			20 final			inteiro		
	Real	Simulado	P2P	Real	Simulado	P2P	Real	Simulado	P2P
Média	15,2700	15,0750	15,3615	14,1650	15,8400	14,1463	14,9548	15,2839	14,8533
DP	2,8348	3,2294	2,0102	2,7848	2,7398	2,1101	2,8514	2,7731	2,1635
teste F	-	0,5755	0,1430	-	0,9442	0,2356	-	0,8798	0,1361
teste t	-	0,8403	0,9069	-	0,0627	0,9810	-	0,6468	0,8750

Fonte: O Autor

Tabela 7 - Resultados da correção utilizando todas as estações para Agosto.

	10 início			10 meio			10 final		
	Real	Simulado	P2P	Real	Simulado	P2P	Real	Simulado	P2P
Média	12,9800	17,1900	13,1750	16,4300	15,0700	15,9317	16,0800	14,5900	15,9511
DP	5,5196	4,1162	3,5668	2,6891	5,7579	1,7717	4,8803	6,1578	3,1222
teste F	-	0,3952	0,2094	-	0,0332	0,2298	-	0,4993	0,1994
teste t	-	0,0691	0,9263	-	0,5072	0,6305	-	0,5562	0,9447

	20 início			20 final			inteiro		
	Real	Simulado	P2P	Real	Simulado	P2P	Real	Simulado	P2P
Média	14,7050	16,1300	14,5533	15,8750	14,1250	15,6788	14,8871	15,4806	14,8723
DP	4,5814	4,9913	3,0843	4,3899	5,2755	2,7583	4,8191	5,3155	3,1706
teste F	-	0,7126	0,0928	-	0,4305	0,0494	-	0,5947	0,0249
teste t	-	0,3528	0,9029	-	0,2613	0,8665	-	0,6467	0,9887

Fonte: O Autor

Tabela 8 - Resultados da correção utilizando todas as estações para Novembro.

	10 início			10 meio			10 final		
	Real	Simulado	P2P	Real	Simulado	P2P	Real	Simulado	P2P
Média	18,5600	18,5300	18,4722	16,8100	18,9800	18,0739	20,8700	18,7300	20,3503
DP	3,8768	3,1401	1,1872	1,8478	2,5015	0,8411	1,4499	1,6547	0,5637
teste F	-	0,5400	0,0016	-	0,3803	0,0282	-	0,7004	0,0096
teste t	-	0,9850	0,9467	-	0,0406	0,0646	-	0,0065	0,3122

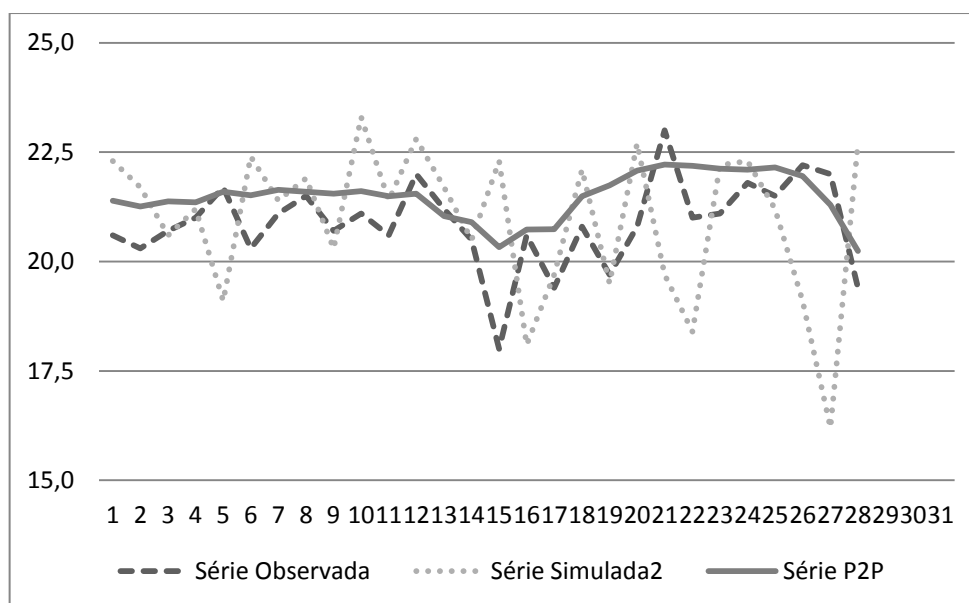
	20 início			20 final			inteiro		
	Real	Simulado	P2P	Real	Simulado	P2P	Real	Simulado	P2P

Média	17,6850	18,7550	18,2731	18,8400	18,8550	19,2121	18,7467	18,7467	18,9655
DP	3,0891	2,7727	1,0220	2,6365	2,0682	1,3599	3,0391	2,4263	1,3322
teste F	-	0,6424	0,0000	-	0,2986	0,0059	-	0,2311	0,0000
teste t	-	0,2562	0,4272	-	0,9841	0,5792	-	1,0000	0,7199

Fonte: O Autor

Observando os resultados obtidos nas Tabelas 5, 6, 7 e 8 pode-se dizer que no geral o modelo P2P se saiu melhor que o simulado, embora no mês de Fevereiro, assim como na correção com três localidades, o modelo simulado foi melhor que o P2P. É interessante notar também que com mais localidades aquele período de 10 dias do final de Fevereiro, utilizando três localidades, em que o modelo P2P se saiu melhor que o simulado desta vez não ocorreu, predominando o modelo simulado para este mês e/ou estação do ano (Figura 18).

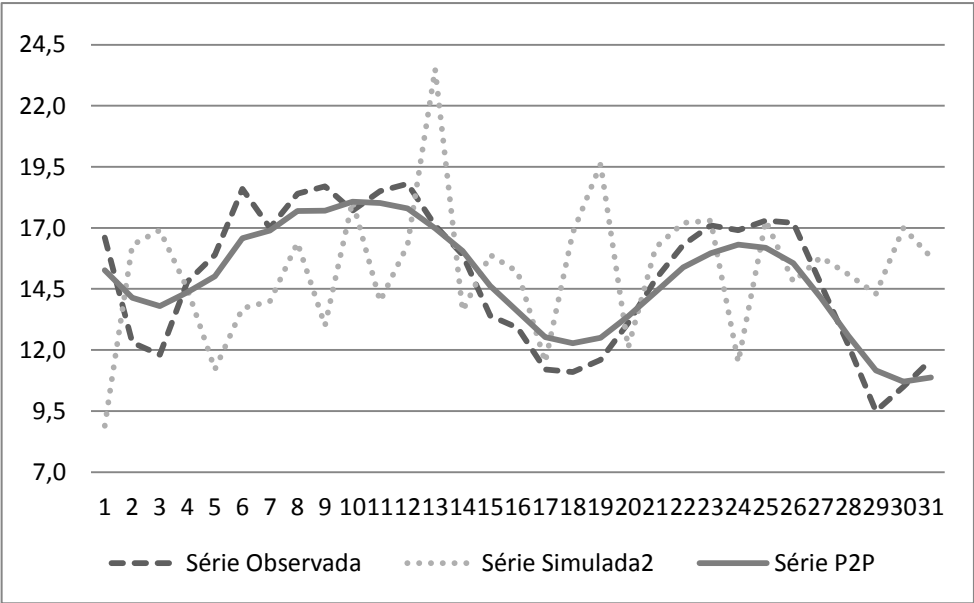
Figura 18 - Dados observados, simulados e corrigidos no modelo P2P para o mês de Fevereiro em todas as localidades.



Fonte: O Autor

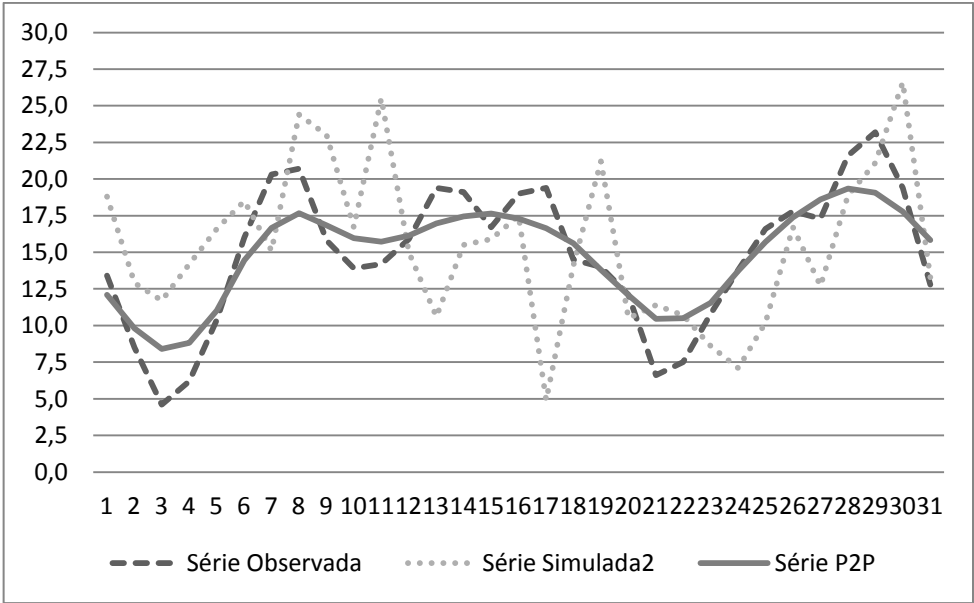
Nos demais períodos correspondentes as estações de outono, inverno e primavera (Figuras 19, 20 e 21) predominou o modelo de correção P2P, com resultados para o teste t indicando que com todas as localidades ele também foi capaz de corrigir dados climáticos, que podem ser considerados estatisticamente pertencentes a mesma população.

Figura 19 – Dados observados, simulados e corrigidos no modelo P2P para o mês de Maio em todas as localidades.



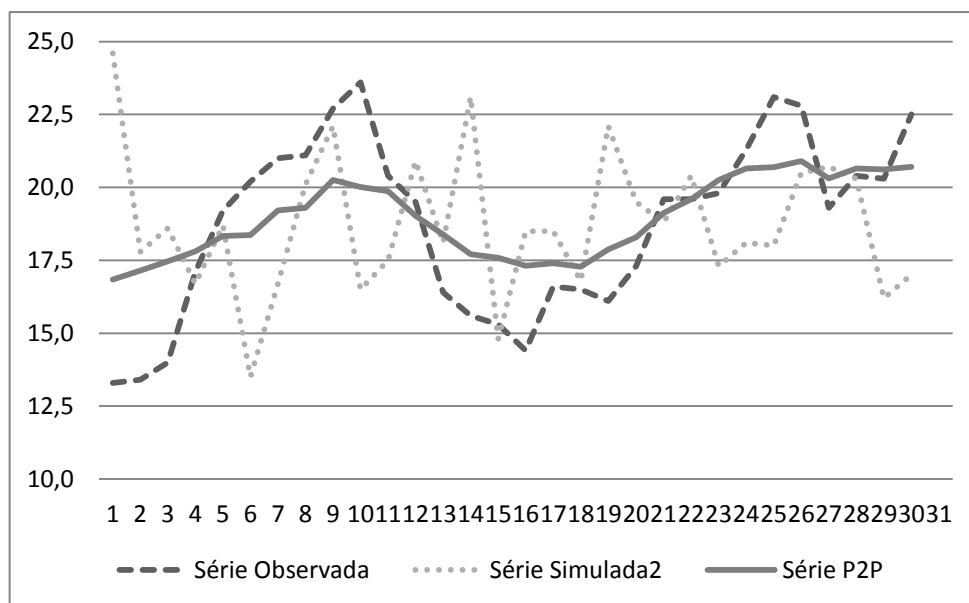
Fonte: O Autor

Figura 20 - Dados observados, simulados e corrigidos no modelo P2P para o mês de Agosto em todas as localidades.



Fonte: O Autor

Figura 21 – Dados observados, simulados e corrigidos no modelo P2P para o mês de Novembro em todas as localidades.



Fonte: O Autor

4.5 CORREÇÃO COM TODAS AS LOCALIDADES X CORREÇÃO COM TRÊS LOCALIDADES

Comparando o desempenho do modelo de correção P2P ao ser executado com três localidades e com todas as localidades, pode-se verificar que ao utilizar mais localidades para executar a correção o modelo foi melhor como mostra a Tabela 9.

Tabela 9 – Comparação do desempenho do modelo P2P para três localidades e todas as localidades com base no *p-value* do teste t.

		10 início	10 meio	10 final	20 início	20 final	inteiro
3 localidades	Fev	0,0054	0,1250	0,4957	0,0117	0,1858	0,0552
	Mai	0,6886	0,8360	0,5847	0,9339	0,7804	0,6632
	Ago	0,9894	0,5061	0,8040	0,7861	0,6808	0,8032
	Nov	0,9429	0,0845	0,5771	0,4009	0,5067	0,5933
todas	Fev	0,0032	0,0420	0,1810	0,0024	0,0442	0,0079
	Mai	0,8110	0,7313	0,7200	0,9069	0,9810	0,8750
	Ago	0,9263	0,6305	0,9447	0,9029	0,8665	0,9887
	Nov	0,9467	0,0646	0,3122	0,4272	0,5792	0,7199

Fonte: O Autor

Dos 24 períodos corrigidos, a execução da correção com três localidades foi melhor em 11 dos 24 pontos, já com todas as localidades a relação é de 13 em 24. Assim pode-se dizer que quanto maior o número de estações melhor é o resultado da correção.

CAPITULO 5 - CONCLUSÕES

Pode-se concluir que a utilização do modelo de correção criado nesta dissertação, para corrigir as lacunas presentes em bases climáticas, apresentou um desempenho melhor que os dados gerados através de um simulador, para os meses que compreendem as estações de outono, inverno e primavera, sendo menos eficiente durante o verão.

Também foi possível observar que quanto maior for a rede de estações participantes da correção, melhores são os resultados obtidos, embora com três estações já seja possível obter bons resultados no processo de correção.

A pesquisa também demonstrou que o modelo foi capaz, tanto ao utilizar somente três estações como utilizar todas, de corrigir dados climáticos que estatisticamente podem ser considerados pertencentes a uma mesma população, ou seja, continuaram refletindo as informações de temperatura apresentadas na base original.

Ainda pode-se destacar a estrutura da rede P2P como facilitadora para a correção, já que ela não necessita de um servidor central e permite que várias estações possam se inserir na rede para realizar a correção de seus dados.

Quanto à metodologia aplicada, pode-se destacar que a utilização de bases climáticas reais combinadas com testes estatísticos permitiu dar o apoio necessário durante a pesquisa para verificar se o modelo de correção criado realmente funcionou como esperado, verificando e corrigindo suas possíveis falhas.

Quanto ao modelo criado, foi possível implementá-lo na linguagem de programação descrita em 3.2 com a ajuda do *framework* P2PComp, embora nesta versão inicial ele não apresente comunicação direta com nenhum tipo de banco de dados, o que pode ser estudado para uma implementação futura.

REFERÊNCIAS

- ALMEIDA, H. A.; SOUZA, J. A.; ALCÂNTARA, H. M. Análise Comparativa de dados meteorológicos obtidos por estação convencional e automática no semi-árido paraibano. *Revista Brasileira de Agrometeorologia*, v.16, p.58-66, 2008.
- BROOKSHIER, D; GOVONI, D; KRISHNAN, N; SOTO, J. C. JXTA: Java P2P Programing. Indianapolis: Sams Publishing, 2002.
- BISQUERRA, R.; SARRIERA, J. C.; MARTINEZ, F. Introdução a Estatística: Enfoque informático com o pacote estatístico SPSS. São Paulo: Artmed, 2007.
- COMER, D. E. Redes de Computadores e Internet: Abrange transmissão de dados, ligações inter-redes, web e aplicações. Rio de Janeiro: Bookman, 2007.
- COMER, D. E; STEVENS, D. L. Interligação em Rede com TCP/IP: Projeto, implementação e detalhes internos. Rio de Janeiro: Campus, 1999.
- COMER, D. E. Interligação em Rede com TCP/IP: Princípios, protocolos e arquitetura. 2º ed. Rio de Janeiro: Campus, 1998.
- COULOURIS, G.; DOLLIMORE, J.; KINDBERG, T. Sistemas Distribuídos: Conceitos e Projetos. 4 ed. São Paulo: Bookman, 2007.
- ENDO, S; MIYAMOTO, T; KUMAGAI, S; FUJII, T. A Data Synchronization Method for Peer-to-Peer Collaboration Systems. *ISCIT*. p.368-373, 2004.
- GOOGLE MAPS. Disponível em <<http://maps.google.com>> Acesso em 27 ago. 2012.
- HOOGENBOOM, G. Contribution of agrometeorology to the simulation of crop production and its applications. *Agricultural and Forest Meteorology*, v.103, p.137-157, 2000.
- INSTITUTO AGRONÔMICO DO PARANÁ. Cartas Climáticas do Paraná. Disponível em: <<http://www.iapar.br/modules/conteudo/conteudo.php?conteudo=597>> Acesso em: 24 ago. 2012.
- JAVA. Disponível em <http://www.java.com/pt_BR/> Acesso em: 27 ago. 2012.
- MASUD, M. M; KIRINGA, I; URAL, H. Update Processing in Instance-Mapped P2P Data Sharing Systems. *IJCIS*. p.339-379, 2009.
- MATEO, M. A. F.; LEUNG, C. K. S. Design and Development of a Prototype System for Detecting Abnormal Weather Observations. *Proceedings of the 2008 C3S2E conference*, 2008.
- MAVI, H. S; TUPPER, G. J. Agrometeorology: Principles and application of climate studies in agriculture. New York: Food Products Press, 2004.
- MICROSOFT EXCEL. Disponível em <<http://office.microsoft.com/pt-br/excel/>> Acesso em: 27 ago. 2012.

MONTGOMERY, D. C.; RUNGER, G. C.; HUBELE, N. F. Estatística Aplicadas à Engenharia. 2º Ed. Rio de Janeiro: LTC, 2004.

MOON. J.; CHO Y. A Point-Based Incentive System to Prevent Free-Riding on P2P Network Environments. Computational Science and its Applications – ICCSA, p.462-471, 2011.

MORAES, P. R. Geografia: Geral e do Brasil. São Paulo: Editora Harbra, 2003.

OAKS, S; TRAVERSAT, B; GONG, L. JXTA in a Nutshell. Sebastopol: O'Reilly, 2002.

PAPPENBERGER, F.; THIELEN J.; MEDICO, M. D. The impact of weather forecast improvements on large scale hydrology: analysing a decade of forecasts of the European Flood Alert System. Hydrological Processes. v. 25, p.1091-1113, 2010.

PEDRO JR., M. J; CAMARGO, M. B. P; MACEDO, L. A. Guia para o observador dos postos agrometeorológicos do Instituto Agrônomo. Campinas, Boletim Técnico n.116, 1987.

PEREIRA, A. R; ANGELOCCI, L. R. SENTELHAS, P.C. Agrometeorologia: Fundamentos e Aplicações Práticas. Guaíba, RS: Editora Agropecuária, 2002.

SANGKET, U; PHONGDARA, A; CHOTIGEAT, W; NATHAN, D; KIM, W; BHAK, J; NGAMPHIW, C; TONGSIMA, S; KHAN, A. M; LIN, H; TAN, T. W. Automatic synchronization and distribution of biological databases and software over low-bandwidth networks among developing countries. Bioinformatics. v.24, p.299-301, 2008.

SCHNEIDER, T. Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. Journal of Climate, 2001.

SCHODER D.; FISCHBACH, K.; SCHMITT, C. Core Concepts in Peer-to-Peer Networking. Peer-to-Peer Computing: The evolution of a disruptive technology. Hershey: Idea Group Publishing, 2005.

SENGER, L. J.; SOUZA, M. A.; FOLTRAN, D. C. Towards a peer-to-peer framework for parallel and distributed computing. 22nd International Symposium on Computer Architecture and High Performance Computing. p.127-134, 2010.

SENTELHAS, P. C; MONTEIRO, J. E. B. de A. Agrometeorologia do Cultivos. Brasília, DF: INMET, 2009.

STEINNHAEUSER, K; CHAWLA, N. V; GANGULY, A. R. An Exploration of Climate Data Using Complex Networks. ACM SIGKDD Exploration Newsletter. v.12, p.25-32, 2010.

STRASSBURGER, A. S.; MENEZES, A. J. E. A de; PERLEBERG, T. D; EICHOLZ, E. D; MENDEZ, M. E. G; SCHÖFFEL, E.R. Comparação da temperatura do ar obtida por estação meteorológica convencional e automática. Revista Brasileira de Meteorologia, v.26, n.2, p.273-278, 2011.

TANENBAUM, A. S. Redes de Computadores. 4 ed. São Paulo: Campus, 2003.

TRUELOVE, K.; DOUGHERTY, D.; GONZE, L.; SHIRKY, C.; DORNFEST, R.; SCHNAPP, M. P2P Networking Overview: The Emergent P2P Platform of Presence, Identity, and Edge Resources. Sebastopol: O'Reilly & Associates. 2002.

VERSTRYNGE, J. Pratical JXTA: Cracking the P2P Puzzle. Lulu, 2008.

WILSON, B. JXTA. New Riders: Berkeley, CA, 2002.

ZANETTI, S. S.; OLIVEIRA, V. P. S.; PRUSKI, F. F. Validação do modelo ClimaBr em relação ao número de dias chuvosos e à precipitação total diária. Engenharia Agrícola, v. 26, p.96-102, 2006.

ZHAO, Y.; ZHOU, S.; ZHOU, A. E-Mail Services on Hybrid P2P Networks. Grid and Cooperative Computing Conference - GCC, p.610–61, 2004.

APÊNDICE A – CÓDIGO FONTE DA CLASSE LCLIMA

```
package br.uepg.p2pcomp.application;

import br.uepg.p2pcomp.ipc.IPCMessage;
import br.uepg.p2pcomp.kernel.Task;
import java.io.BufferedReader;
import java.io.FileReader;
import java.io.IOException;
import java.text.DateFormat;
import java.text.SimpleDateFormat;
import java.util.*;

public class LClima extends Task {

    public Double mediaPonderada(Vector<Double> valores,
Vector<Double> pesos){
        int tam = valores.size();
        Double temp = 0.0;
        Double temp2 = 0.0;

        for(int i = 0; i < tam; i++){
            temp += valores.elementAt(i) * pesos.elementAt(i);
            temp2 += pesos.elementAt(i);
        }

        return temp / temp2;
    }

    public String dataAnterior(String data, int dias) {
        String tmp[] = data.split("/");

        int ano = Integer.valueOf(tmp[2]);
        int mes = Integer.valueOf(tmp[1])-1;
        int dia = Integer.valueOf(tmp[0]);

        Calendar calen = Calendar.getInstance();
        calen.set(ano, mes, dia);
        calen.add(Calendar.DATE, dias);

        SimpleDateFormat formatador = new
SimpleDateFormat("dd/MM/yyyy");
        return formatador.format(calen.getTime());
    }

    public Vector<String> getDatas(String base){
        BufferedReader in = null;
        Vector<String> dados = new Vector();

        base = base.replace("p2pcomp-", "");
    }
}
```

```

        try {
            in = new BufferedReader(new FileReader("dados/" +
base + ".csv"));
            String linha;

            while ((linha = in.readLine()) != null) {
                String conteudo[] = linha.split(";");
                if(conteudo[1].compareToIgnoreCase("-99") ==
0){
                    dados.add(conteudo[0]);
                }
            }

        } catch(Exception ex) {
            ex.getMessage();
        } finally {
            try {
                in.close();
            } catch (IOException ex) {
                System.out.println("Nãfo foi possivel fechar
arquivo!");
            }
        }

        return dados;
    }

    public Vector<Double> getDados(String base, String de,
String ate) {
        BufferedReader in = null;
        Date _de = null;
        Date _ate = null;
        Vector<Double> dados = new Vector();

        base = base.replace("p2pcomp-", "");

        try {
            DateFormat formatador = new
SimpleDateFormat("dd/MM/yyyy");
            _de = formatador.parse(de);
            _ate = formatador.parse(ate);
            in = new BufferedReader(new FileReader("dados/" +
base + ".csv"));
            String linha;

            while ((linha = in.readLine()) != null) {

                if(!linha.trim().isEmpty()){
                    String conteudo[] = linha.split(";");

```

```

        Date atual =
formatador.parse(conteudo[0]);

        if((atual.getTime() >= _de.getTime()) &&
(atual.getTime() <= _ate.getTime())){

dados.add(Double.parseDouble(conteudo[1]));
        }
    }

    } catch(Exception ex) {
        ex.printStackTrace();
    } finally {
        try {
            in.close();
        } catch (IOException ex) {
            System.out.println("NÃO foi possível fechar
arquivo!");
        }
    }

    return dados;
}

@Override
public void run(){
    int i, tam;
    IPCMessage msg = null;

    if( this.getRank() == 0 ){

        tam = this.getSize();
        i = 1;

        Vector<String> datas =
this.getDatas(this.getHost().getName());
        Vector resultados = new Vector();

        while(i < tam){
            this.send(i, this.getRank(), datas);
            i++;
        }

        i = 1;

        while(i < tam){
            msg = this.recv(-1, -1);
            resultados.add((Vector<Double>)
msg.getPayload());
            i++;
        }
    }
}

```

```

    }

    tam = tam -1;

    int tam_datos = datas.size();
    int tam_res = resultados.size();
    Double temp = 0.0;
    Vector<Double> medias = new Vector<Double>();

    for(i = 0; i < tam_datos; i++){
        for(int j = 0; j < tam_res; j++){
            temp += ((Vector<Double>)
resultados.elementAt(j)).elementAt(i);
        }
        medias.add(temp / tam_res);
        temp = 0.0;
    }

    this.imprimirResultado(datas, medias);
    this.sendResults("FIM");

} else {
    msg = this.recv(0, -1);
    Vector<Double> result = new Vector();
    Vector<String> datas = (Vector<String>)
msg.getPayload();
    Iterator datasit = datas.iterator();
    Vector<Double> pesos = new Vector<Double>();
    pesos.add(1.0);
    pesos.add(1.0);
    pesos.add(2.0);
    pesos.add(1.0);
    pesos.add(1.0);

    while(datasit.hasNext()){
        String data = datasit.next().toString();
        String de = this.dataAnterior(data, -2);
        String ate = this.dataAnterior(data, 2);
        Vector<Double> valores =
this.getDados(this.getHost().getName(), de, ate);
        result.add(this.mediaPonderada(valores,
pesos));
    }

    this.send(0, this.getRank(), result);
}

}

public void imprimirResultado(Vector<String> datas,
Vector<Double> valores){

```

```
        int tam_dadas = datas.size();

        for(int i = 0; i < tam_dadas; i++){
            System.out.println(datas.elementAt(i) + ": " +
valores.elementAt(i));
        }
    }
}
```

APÊNDICE B – CÓDIGO FONTE DA BUSCA DE ESTAÇÕES E CÁLCULO DE DISTÂNCIA

```

if (line.startsWith("estprox")) {
    machines = Cache.getInstance().getPeers();
    peersEncontrados = 0;
    mprox = new Vector<Machine>();

    synchronized(machines) {

        Iterator teste = machines.iterator();

        it = machines.iterator();

        String peer_escravo =
p2pcomp.getMachineName().replace("main_", "");
        mprox.add(Cache.getInstance().findPeer(peer_escravo));

        System.out.println(peer_escravo);

        while(it.hasNext()){

            Machine temp = it.next();

            if(!temp.getName().equals(p2pcomp.getMachine().getName()) &&
!temp.getName().contains("main_") &&
!temp.getName().equals(peer_escravo)) {

                double dist =
p2pcomp.getMachine().calculaDistancia(temp.getUtm_N(),
temp.getUtm_E());

                if(dist <= 150.0){
                    mprox.add(temp);
                    System.out.println(temp.getName());
                }
            }

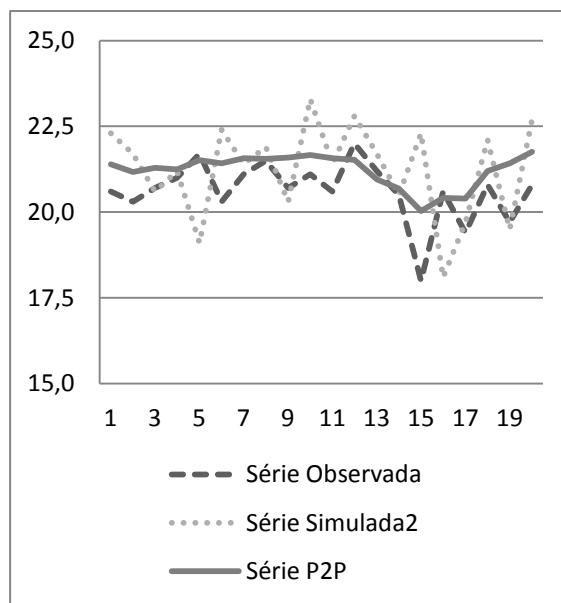
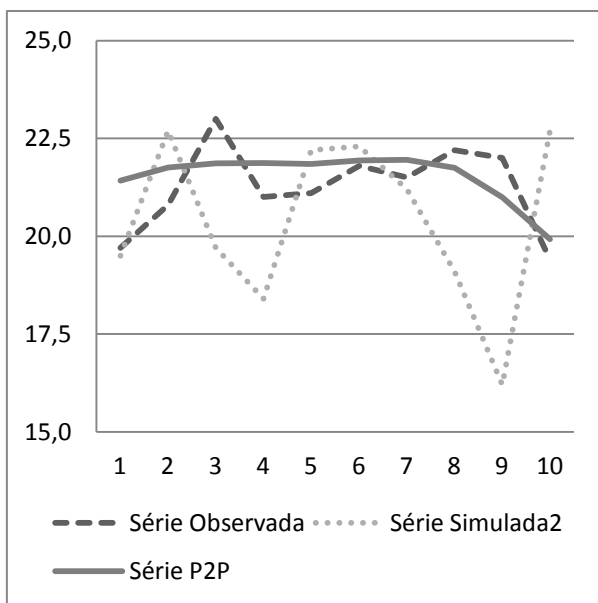
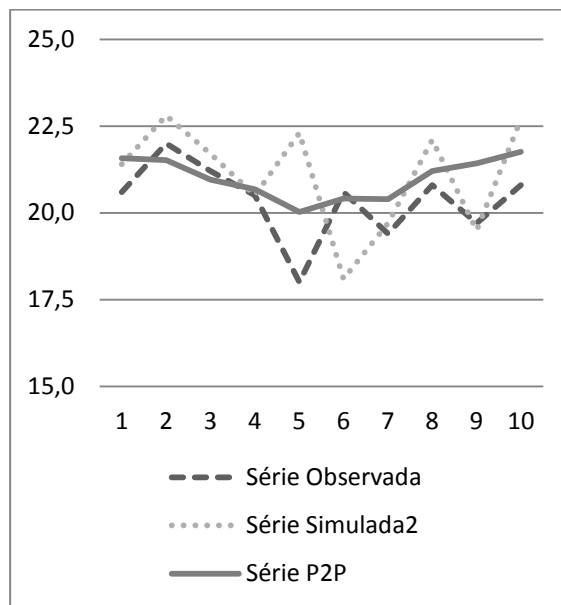
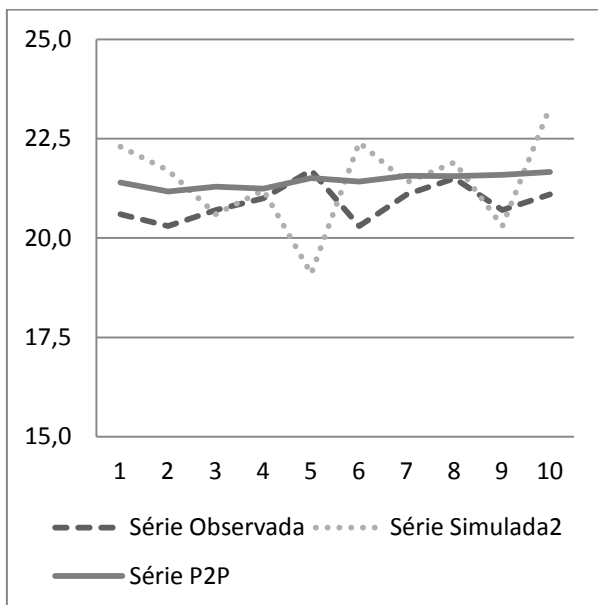
            peersEncontrados = mprox.size();
            System.out.println("Peers encontrados: " +
peersEncontrados);

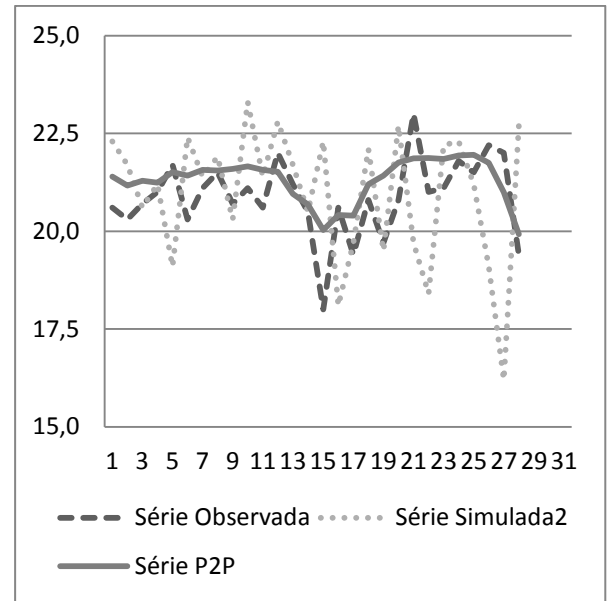
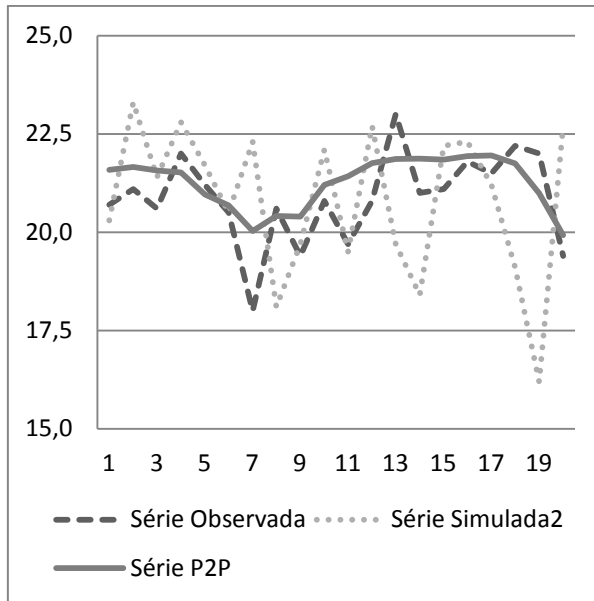
        }
    }
}

```

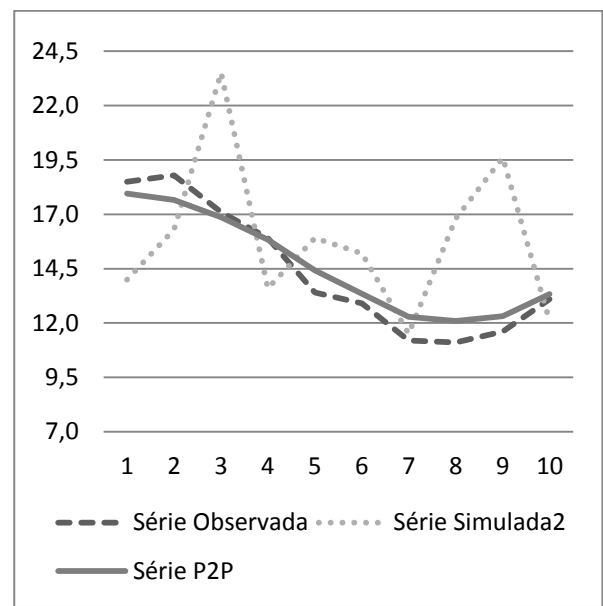
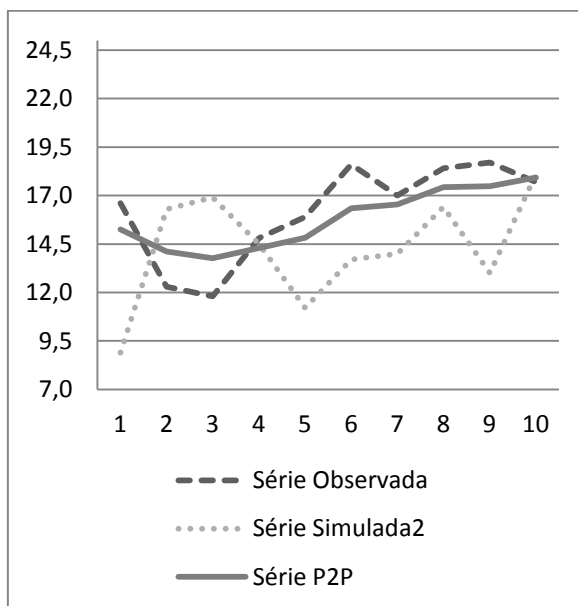
APÊNDICE C – GRÁFICOS COMPARATIVOS DOS DADOS OBSERVADOS, GERADOS PELO SIMULADOR E PELO MODELO P2P UTILIZANDO TRÊS LOCALIDADES.

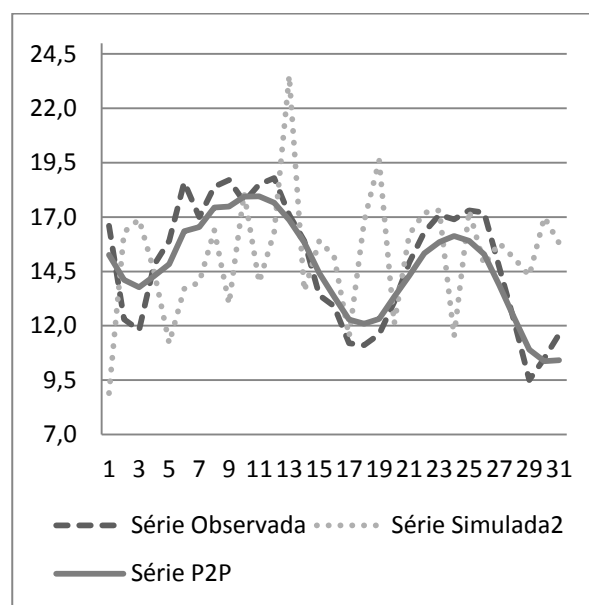
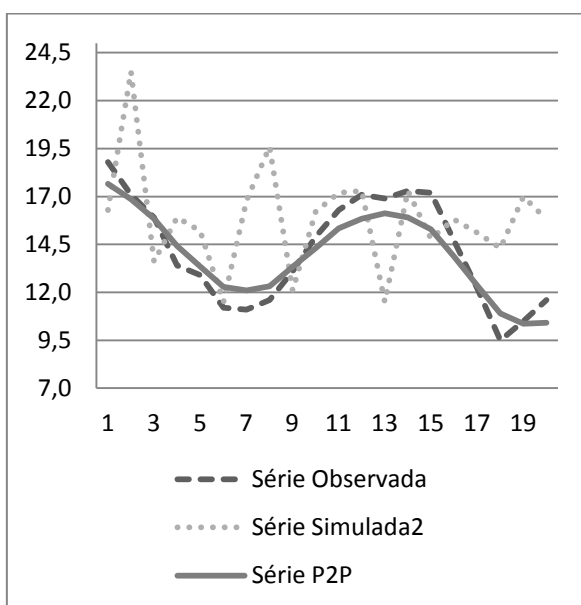
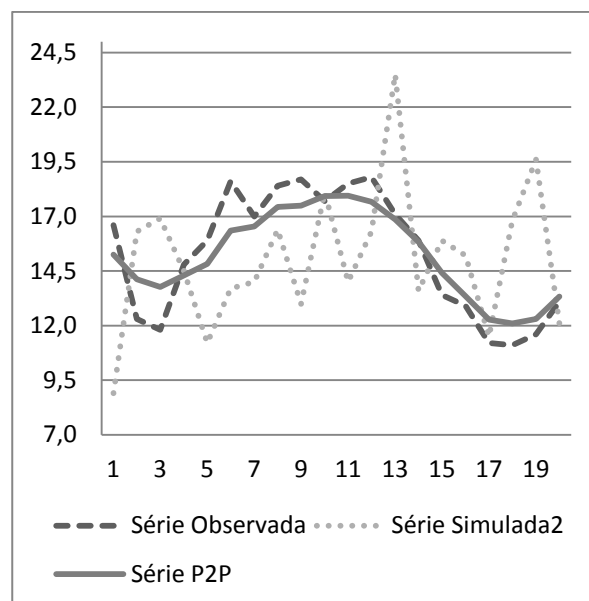
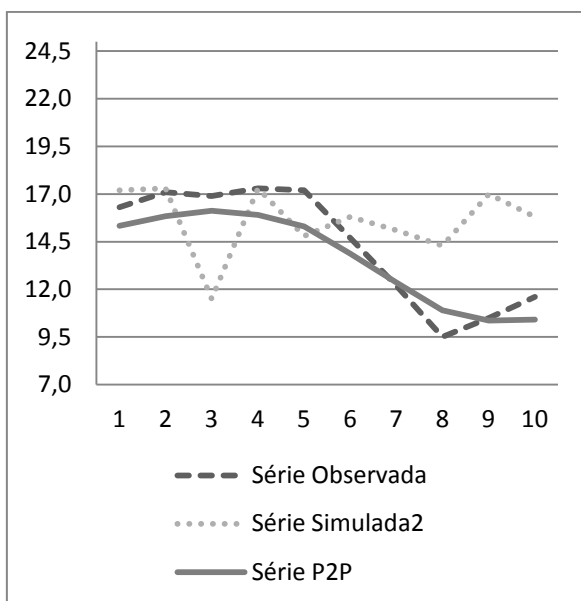
Gráficos do mês de Fevereiro para 10 dias iniciais, 10 dias do meio, 10 dias finais, 20 dias iniciais, 20 dias finais e do mês inteiros respectivamente.



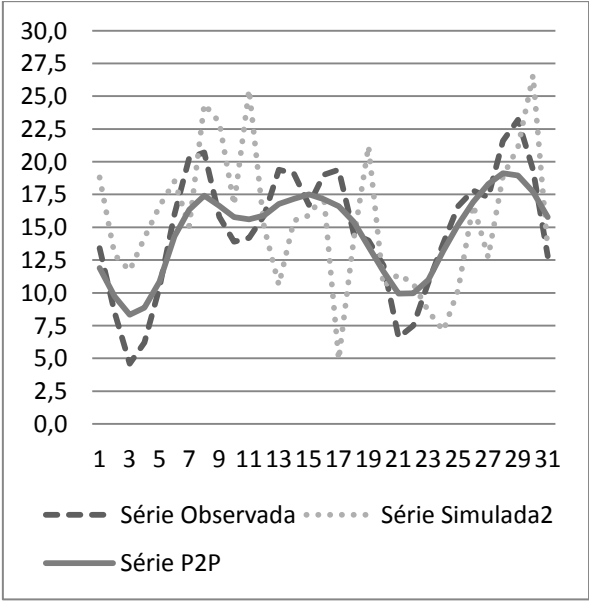
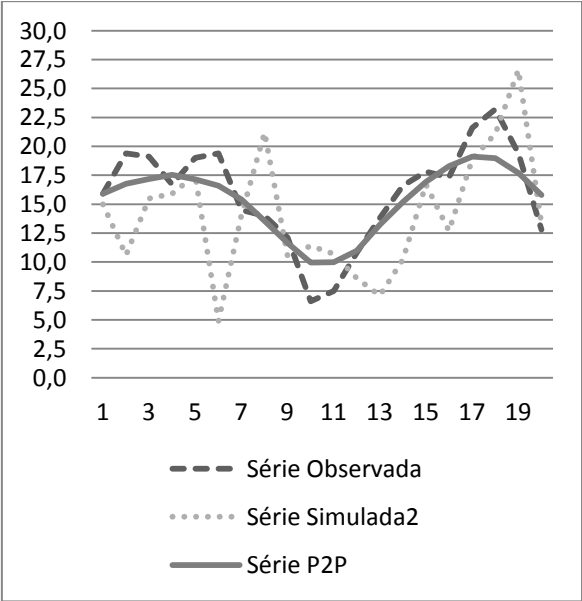
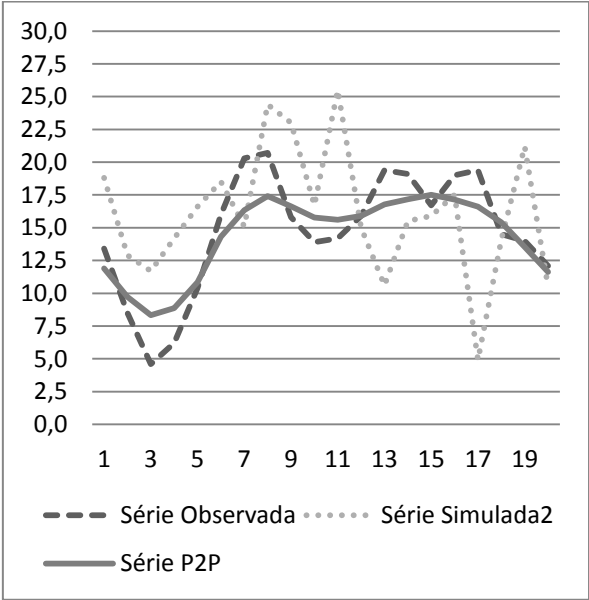
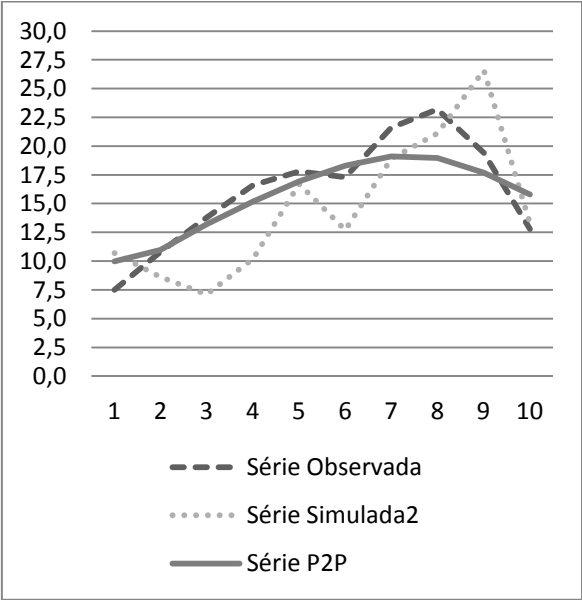
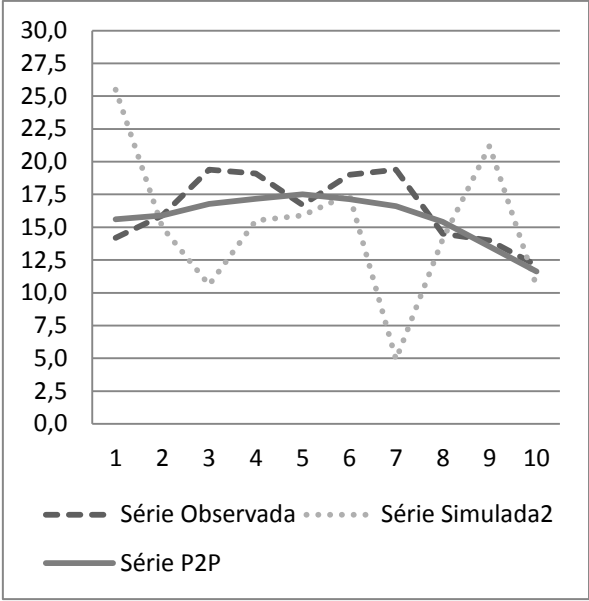
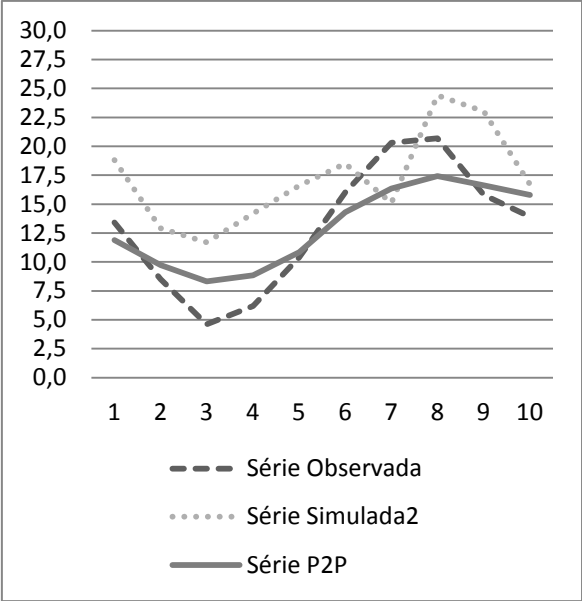


Gráficos do mês de Maio para 10 dias iniciais, 10 dias do meio, 10 dias finais, 20 dias iniciais, 20 dias finais e do mês inteiros respectivamente.

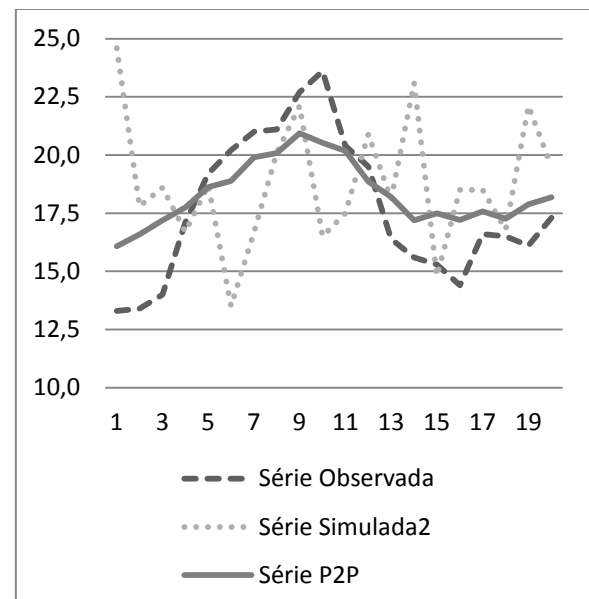
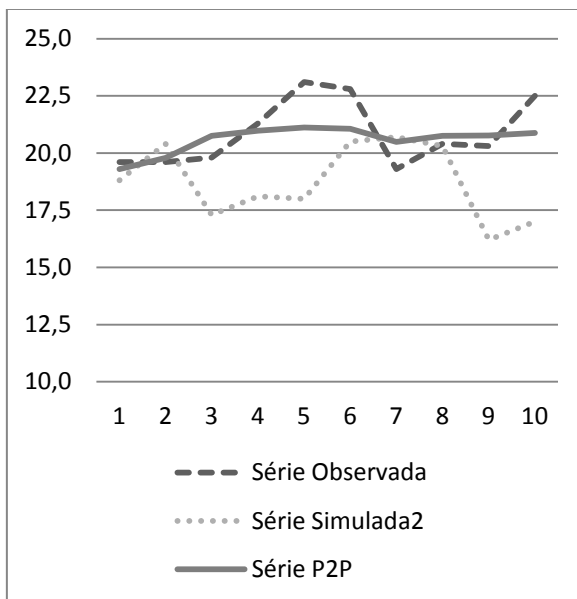
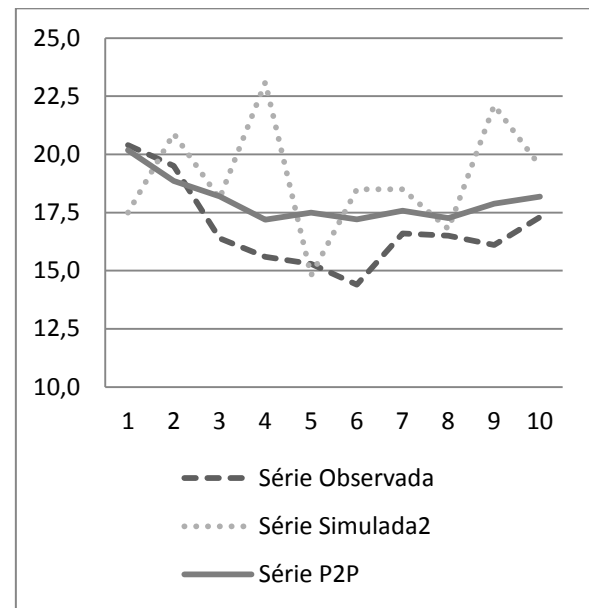
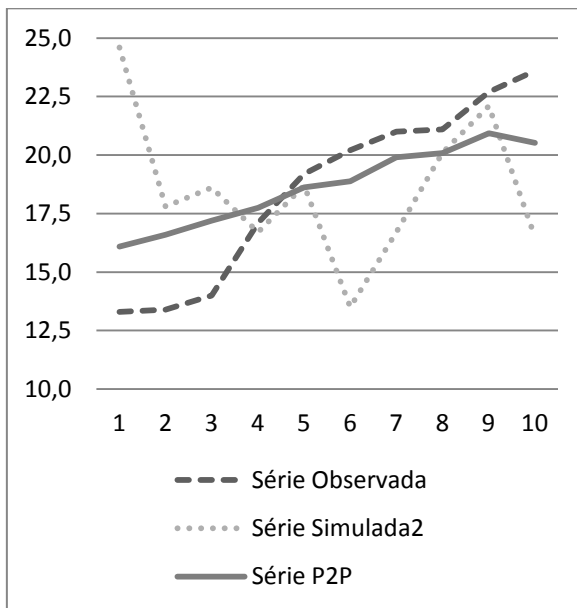


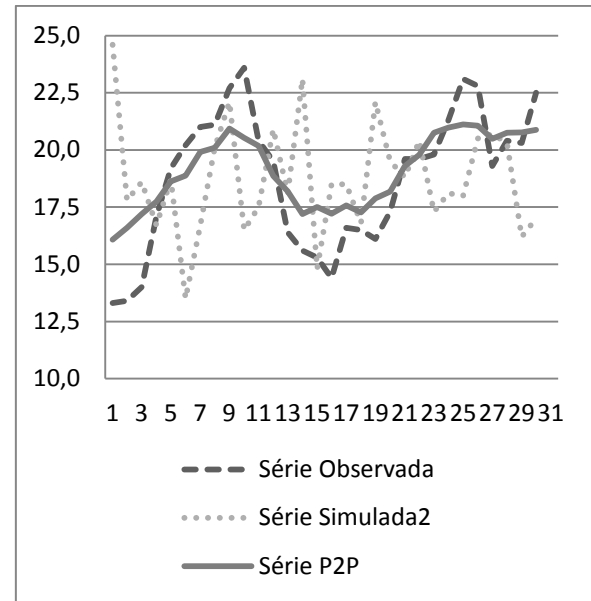
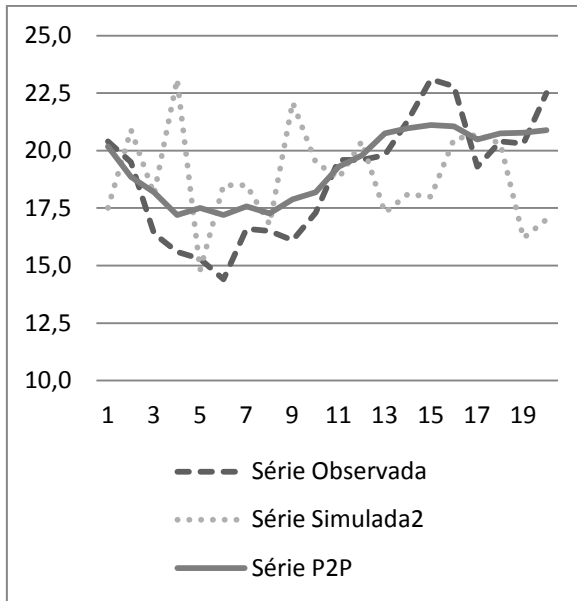


Gráficos do mês de Agosto para 10 dias iniciais, 10 dias do meio, 10 dias finais, 20 dias iniciais, 20 dias finais e do mês inteiros respectivamente.



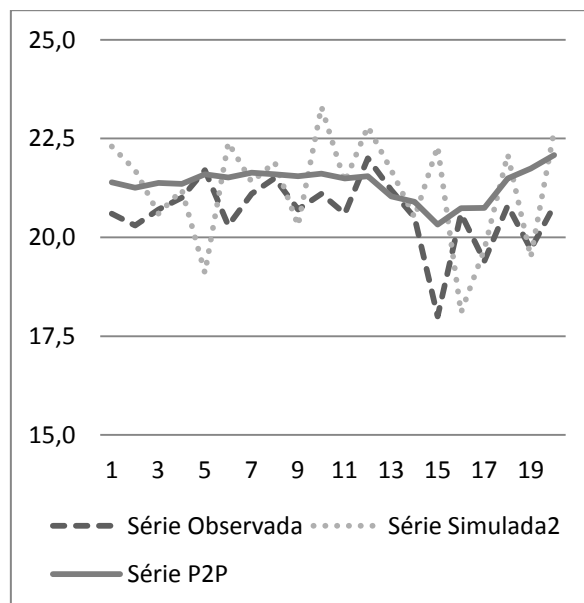
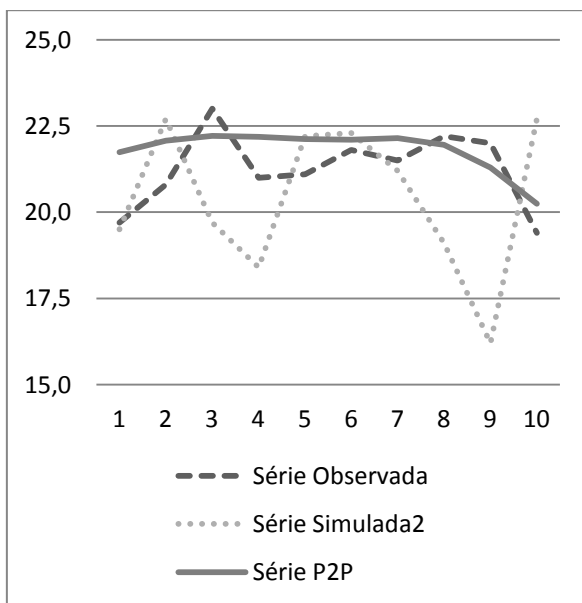
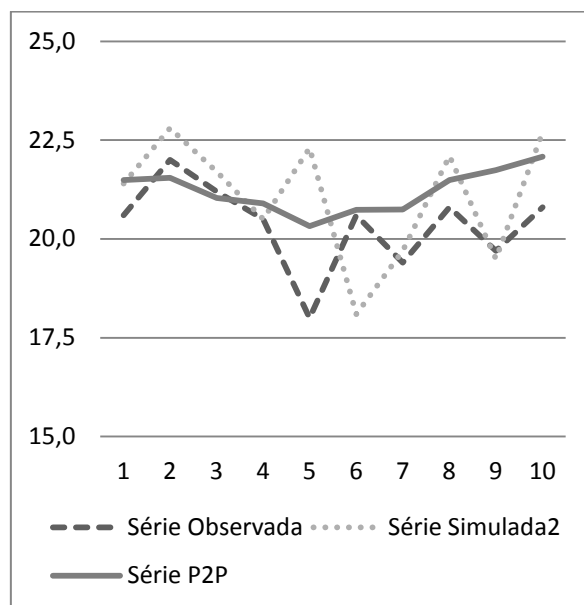
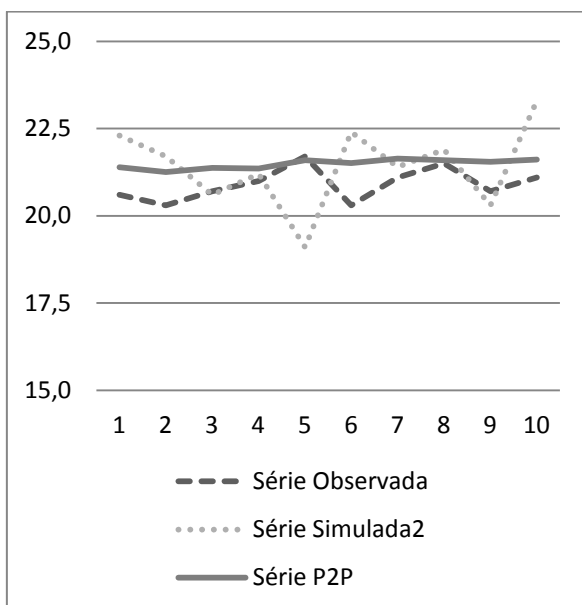
Gráficos do mês de Novembro para 10 dias iniciais, 10 dias do meio, 10 dias finais, 20 dias iniciais, 20 dias finais e do mês inteiros respectivamente.

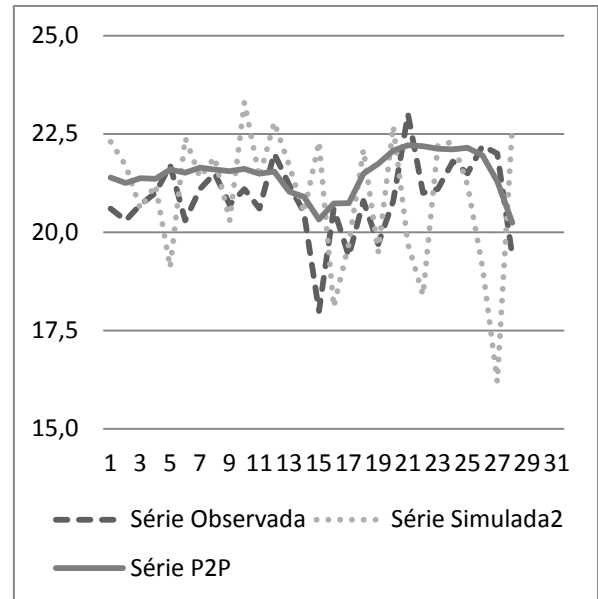
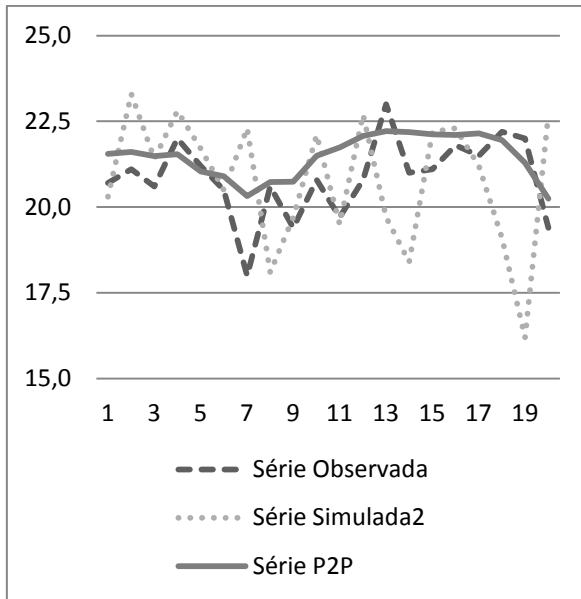




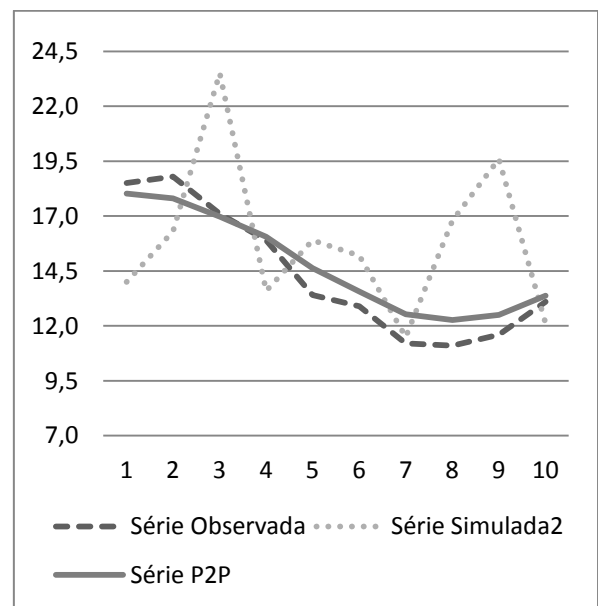
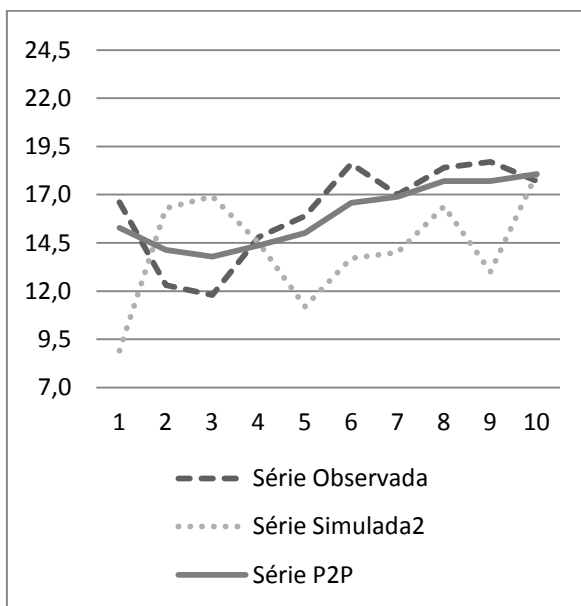
APÊNDICE D – GRÁFICOS COMPARATIVOS DOS DADOS OBSERVADOS, GERADOS PELO SIMULADOR E PELO MODELO P2P UTILIZANDO TODAS AS LOCALIDADES.

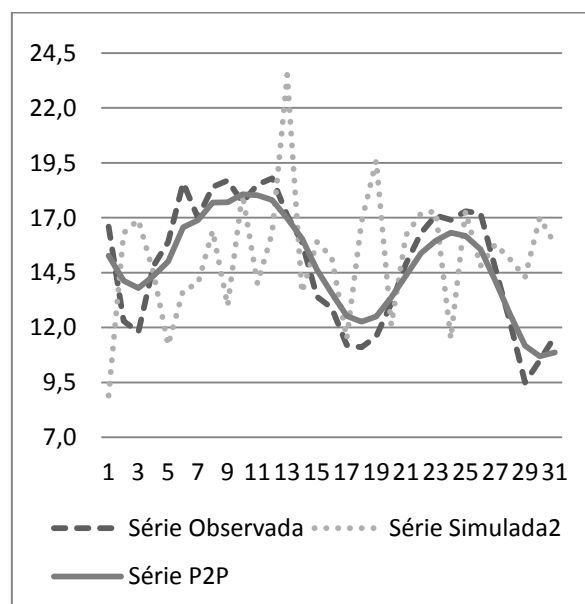
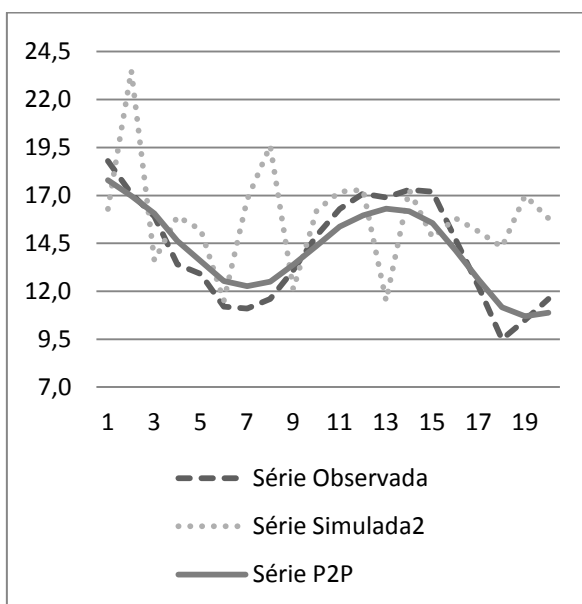
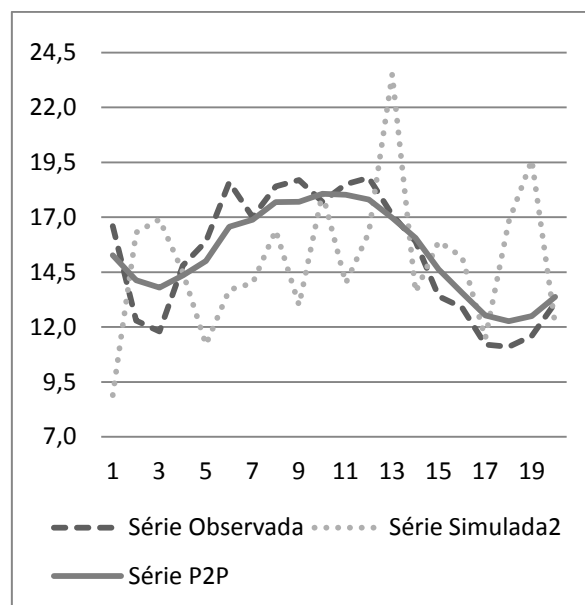
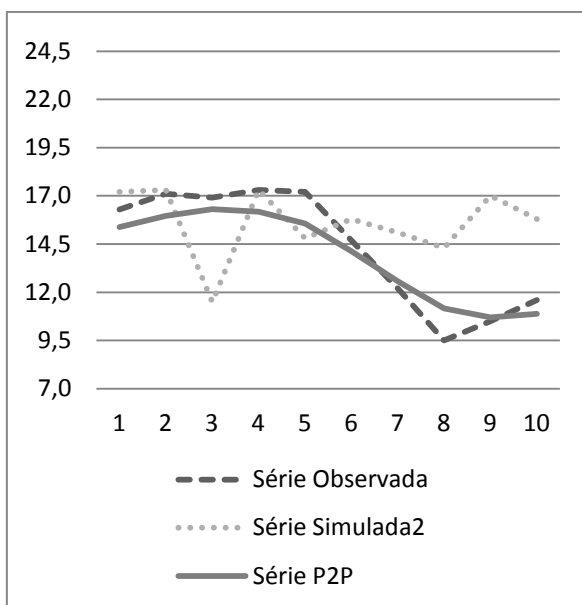
Gráficos do mês de Fevereiro para 10 dias iniciais, 10 dias do meio, 10 dias finais, 20 dias iniciais, 20 dias finais e do mês inteiros respectivamente.



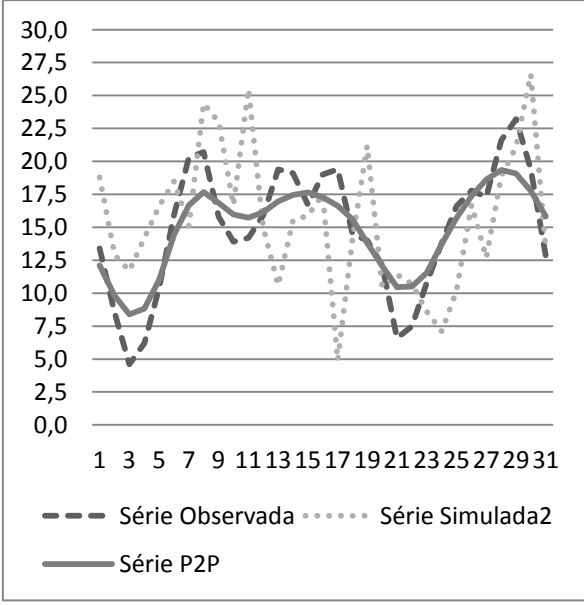
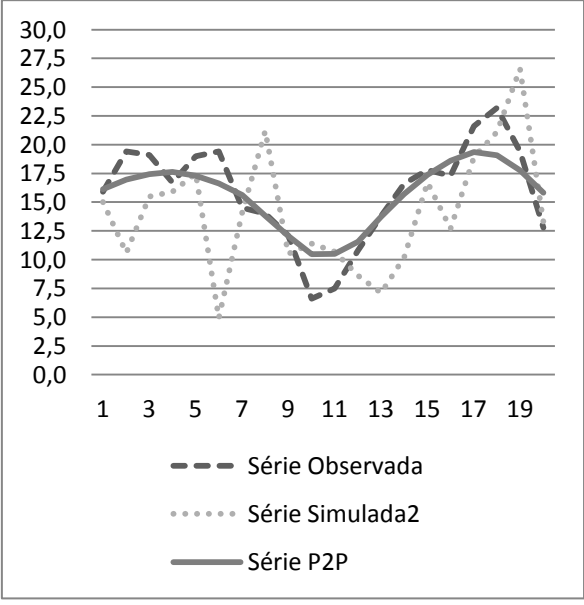
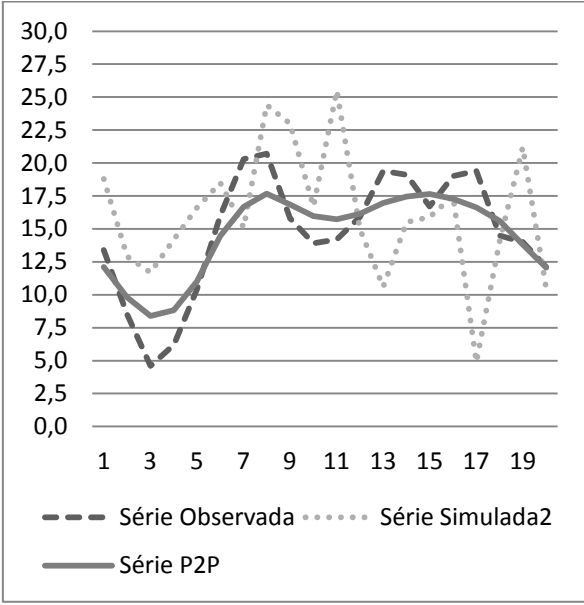
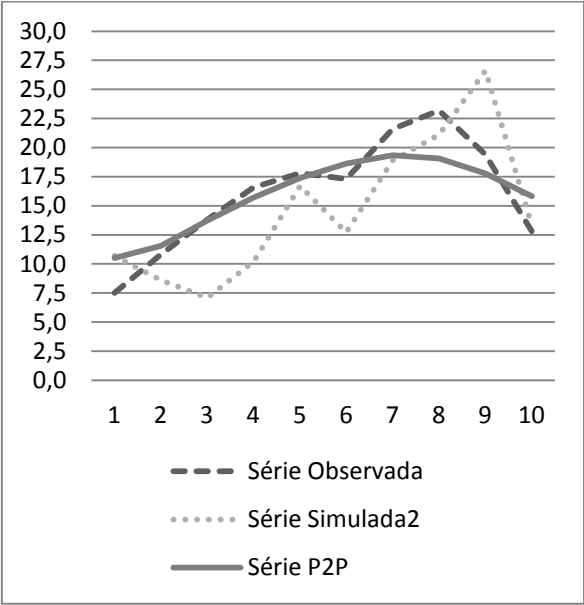
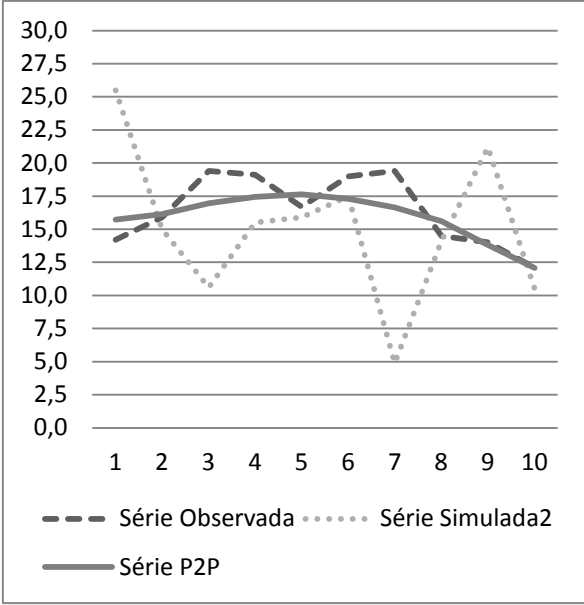
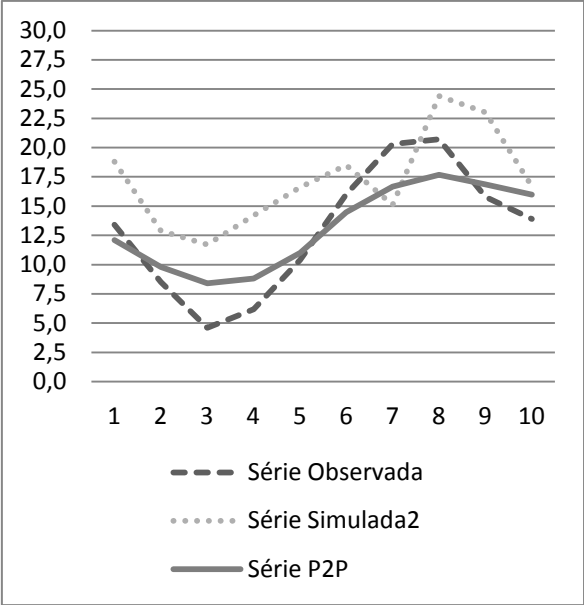


Gráficos do mês de Maio para 10 dias iniciais, 10 dias do meio, 10 dias finais, 20 dias iniciais, 20 dias finais e do mês inteiros respectivamente.





Gráficos do mês de Agosto para 10 dias iniciais, 10 dias do meio, 10 dias finais, 20 dias iniciais, 20 dias finais e do mês inteiros respectivamente.



Gráficos do mês de Novembro para 10 dias iniciais, 10 dias do meio, 10 dias finais, 20 dias iniciais, 20 dias finais e do mês inteiros respectivamente.

