

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ – UTFPR  
CURSO SUPERIOR DE TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE  
SISTEMAS

RENNAN ANDRES PALOSCHI

**APLICAÇÃO DO SOFTWARE R NA CLASSIFICAÇÃO DE ÁREAS ASSOCIADAS  
À PRODUTIVIDADE DA SOJA E VARIÁVEIS AGROMETEOROLÓGICAS POR  
MEIO DE AGRUPAMENTO FUZZY**

TRABALHO DE DIPLOMAÇÃO

MEDIANEIRA

2013

RENNAN ANDRES PALOSCHI

**APLICAÇÃO DO SOFTWARE R NA CLASSIFICAÇÃO DE ÁREAS ASSOCIADAS  
À PRODUTIVIDADE DA SOJA E VARIÁVEIS AGROMETEOROLÓGICAS POR  
MEIO DE AGRUPAMENTO FUZZY**

Trabalho de Diplomação apresentado à disciplina de Trabalho de Diplomação, do Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas – COADS – da Universidade Tecnológica Federal do Paraná – UTFPR, como requisito parcial para obtenção do título de Tecnólogo.

Orientador: Prof. Dr. Everton Coimbra de Araújo.

MEDIANEIRA

2013



---

## TERMO DE APROVAÇÃO

### APLICAÇÃO DO SOFTWARE R NA CLASSIFICAÇÃO DE ÁREAS ASSOCIADAS À PRODUTIVIDADE DA SOJA E VARIÁVEIS AGROMETEOROLÓGICAS POR MEIO DE AGRUPAMENTO FUZZY

Por

**Rennan Andres Paloschi**

Este Trabalho de Diplomação (TD) foi apresentado às 16:40 h do dia 20 de agosto de 2013 como requisito parcial para a obtenção do título de Tecnólogo no Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas, da Universidade Tecnológica Federal do Paraná, *Campus* Medianeira. O acadêmico foi argüido pela Banca Examinadora composta pelos professores abaixo assinados. Após deliberação, a Banca Examinadora considerou o trabalho aprovado com louvor e mérito.

---

Prof. Dr. Everton Coimbra de Araújo  
UTFPR – *Campus* Medianeira  
(Orientador)

---

Prof. Me. Ricardo Sobjak  
UTFPR – *Campus* Medianeira  
(Convidado)

---

Prof. Me. Nelson Miguel Betzek  
UTFPR – *Campus* Medianeira  
(Convidado)

---

Prof. Me. Juliano Rodrigo Lamb  
UTFPR – *Campus* Medianeira  
(Responsável pelas atividades de TCC)

A folha de aprovação assinada encontra-se na Coordenação do Curso.

Dedico este trabalho ao Espírito Santo  
por me iluminar durante o trabalho e ao meu  
Anjo da Guarda por me proteger no caminho.

## **AGRADECIMENTOS**

Deixo aqui meu mais sincero obrigado a todas as pessoas que de alguma forma me ajudaram a terminar mais essa etapa em minha vida, que suavizaram meu trabalho e me ampararam nos momentos em que dificilmente atravessaria sozinho.

Agradeço ao professor Dr. Everton Coimbra de Araújo por ser tão prestativo ao me orientar neste trabalho, e por todos os professores que durante todo o curso puderam parar nos corredores para escutar com atenção minhas dúvidas e queixas, em especial ao professor Márcio Matté, ao Me. Juliano Lamb e a Ma. Alessandra Hoffman.

Agradeço também ao meu pai Diomar Paloschi por tantas vezes me dar o apoio e suporte necessários para continuar o curso, mesmo no momento mais difícil de sua vida.

## RESUMO

PALOSCHI, Rennan Andres. APLICAÇÃO DO SOFTWARE R NA CLASSIFICAÇÃO DE ÁREAS ASSOCIADAS À PRODUTIVIDADE DA SOJA E VARIÁVEIS AGROMETEOROLÓGICAS POR MEIO DE AGRUPAMENTO FUZZY. Trabalho de conclusão de curso (Tecnologia em Análise e Desenvolvimento de Sistemas), da Universidade Tecnológica Federal do Paraná. Medianeira 2013.

Este trabalho teve como objetivo aplicar uma abordagem sobre a utilização da software R baseada em agrupamento difuso para classificação de áreas associadas à produtividade da soja, conjuntamente com as variáveis agrometeorológicas: precipitação pluvial média, temperatura média do ar e radiação solar global média. O estudo foi realizado envolvendo os 50 municípios da região oeste do estado do Paraná, Brasil, com dados da safra do ano-agrícola 2010/2011. Por meio do algoritmo *Fuzzy c-Means*, foi possível formar grupos de municípios similares à produtividade de soja, a identificação do número adequado de agrupamentos foi obtida utilizando a Entropia de Partição Modificada (MPE), os grupos foram validados pelo Método de Decisão pelo Maior Grau de Pertinência (MDMGP). Para mensurar o nível de similaridade de cada agrupamento, foi utilizado o Índice de Similaridade de *Clusters* (ISC<sub>i</sub>). Dentro das perspectivas deste estudo, o método empregado se mostrou adequado, permitindo identificar agrupamentos de municípios com graus de similaridades da ordem de 66 à 68%.

Palavras-chave: Autocorrelação espacial, Similaridade espacial, Regressão espacial, c-means.

## RESUMO EM LINGUA ESTRANGEIRA

PALOSCHI, Rennan Andres. APPLICATION OF R SOFTWARE IN THE CLASSIFICATION OF AREAS ASSOCIATED WITH SOY AND PRODUCTIVITY METEOROLOGICAL VARIABLES THROUGH FUZZY CLUSTERING. Completion of course work (Technology Analysis and Systems Development), Federal Technological University of Paraná. Medianeira 2013.

This study aimed to apply an approach on the use of the R software based on fuzzy clustering for classification of areas associated with soybean yield, together with meteorological variables: rainfall average, average air temperature and solar radiation average. The study was conducted involving 50 municipalities in the western region of the state of Paraná, Brazil, with data from the harvest of the agricultural year 2010/2011. Through the algorithm Fuzzy C-Means, it was possible to form groups of municipalities similar to soybean yield, the identification of the appropriate number of clusters was obtained using the Entropy Partition Modified (MPE), the groups were validated by the Method Decision by Major degree of Pertinence (MDMGP). To measure the level of similarity of each cluster, we used the Index of Cluster Similarity (ISC<sub>i</sub>). Within the perspective of this study, the method used was adequate, allowing to identify clusters of municipalities with degrees of similarities in the order of 66 to 68%.

Keywords: Spatial autocorrelation, spatial similarity, similarity index; c-means.

## LISTAS DE FIGURAS

Figura 1 - Região oeste do Paraná .....	15
Figura 2 - Estações meteorológicas .....	16
Figura 3 - Região oeste e estações meteorológicas próximas .....	23
Figura 4 - Estações próximas da região oeste do Paraná com áreas de influência demarcadas .....	24
Figura 5 - Municípios com estações influentes na região oeste do Paraná .....	24
Figura 6 - Gráfico dos resultados do MPE por número de clusters.....	26
Figura 7 - Distribuição dos municípios nos agrupamentos imposta pelo FCM .....	29
Figura 8 - Mapa temático da produtividade da soja .....	31
Figura 9 - Hiperlink CRAN do site r-project.....	38
Figura 10 - Hiperlink da versão UFPR do software R do site r-project .....	39
Figura 11 - Hiperlink da versão para Windows do software R .....	39
Figura 12 - Hiperlink da distribuição base do software R.....	40
Figura 13 - Hiperlink de download do software R .....	40
Figura 14 - R Console.....	41
Figura 15 - Interface do R-Commander .....	42
Figura 16 - Municípios plotados.....	45
Figura 17 - Municípios do PR, SC, MT e SP .....	46
Figura 18 - Botão "<No active dataset>" do R-Commander.....	47
Figura 19 - Janela para escolha do objeto Data .....	48
Figura 20 - Botão "Edit data set" do R-Commander.....	48
Figura 21 - Janela de edição de conjuntos de dados do R .....	49
Figura 22 - Estações meteorológicas .....	49
Figura 23 - Oeste do Paraná e estações .....	50
Figura 24 - Centroides .....	51
Figura 25 - Polígonos de Thiessen .....	52
Figura 26 - Mapas sobrepostos.....	53
Figura 27 - Oeste do paraná e estações.....	54
Figura 28 - Ajustando Polígonos .....	55

## LISTA DE TABELAS

Tabela 1 - Análise preliminar dos dados .....	25
Tabela 2 - Graus de inclusão entre os agrupamentos estabelecidos pelo método MDMGP ...	27
Tabela 3 - Distribuição quantitativa dos municípios nos agrupamentos .....	27
Tabela 4 - Centro das variáveis de cada agrupamento pelo método cmeans .....	28
Tabela 5 - Estatísticas para as variáveis do estudo em cada agrupamento da região de estudo.....	30

## LISTA DE ABREVIATURAS, SIGLAS E ACRÔNIMOS

CRAN	<i>Comprehensive R Archive Network</i>
DBF	<i>Data Base File</i>
FCM	<i>Fuzzy c-Means</i>
GP	Grau de Pertinência
ISC <sub>1</sub>	Índice de Similaridade de Clusters
MDMGP	Método de Decisão pelo Maior Grau de Pertinência
MPE	Entropia de Partição Modificada
PID	Identificador do polígono
PRECM	Precipitação pluvial
PROD	Produtividade da soja
RSGM	Radiação solar global média
SHP	<i>Shapefile</i>
SIG	Sistemas de Informação Geográfica
TMED	Temperatura média do ar

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>11</b>
1.1	OBJETIVO GERAL.....	12
1.2	OBJETIVOS ESPECÍFICOS.....	13
1.3	JUSTIFICATIVA.....	13
1.4	ESTRUTURA DO TRABALHO.....	14
<b>2</b>	<b>MATERIAIS E MÉTODOS.....</b>	<b>15</b>
2.1	CÁLCULO DAS ÁREAS DE INFLUÊNCIA DAS ESTAÇÕES METEOROLÓGICAS.....	18
2.2	DISTRIBUIÇÃO DOS ATRIBUTOS.....	19
2.3	PADRONIZAÇÃO DOS ATRIBUTOS.....	19
2.4	APLICAÇÃO DO <i>FUZZY C-MEANS</i> .....	20
2.5	ALOCAÇÃO DOS DADOS NOS AGRUPAMENTOS ESTABELECIDOS.....	20
2.6	MENSURAÇÃO DO NÍVEL DE SIMILARIDADE DE CADA AGRUPAMENTO.....	21
<b>3</b>	<b>RESULTADOS E DISCUSSÃO.....</b>	<b>23</b>
<b>4</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>32</b>
4.1	CONCLUSÃO.....	32
4.2	TRABALHOS FUTUROS/CONTINUAÇÃO DO TRABALHO.....	32
<b>5</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>33</b>
	<b>APÊNDICES.....</b>	<b>37</b>

## 1 INTRODUÇÃO

Nas últimas décadas houve um considerável aumento da massa de dados obtidos de diversas áreas, como: biologia, medicina, engenharia e outras, gerando uma demanda de desenvolvimento de ferramentas para analisá-los de maneira que encontrem novos padrões, que sugiram novas informações, de forma útil (MEDEIROS, 2010).

A necessidade de recursos computacionais para realizar análises estatísticas levam os pesquisadores à procura e aprendizado de softwares específicos para tais fins. Essa procura leva em conta diversos fatores, como a disponibilidade destes softwares (tipo de licença), incentivo de terceiros (indicações) e treinamento (FONSECA, 2010).

O software R é uma linguagem e um ambiente de desenvolvimento integrado para realização e obtenção de cálculos estatísticos e gráficos (THE R FOUNDATION FOR STATISTICAL COMPUTING, 2012). Os principais motivos para buscar um estudo ou especialização no R são: software gratuito; todos os códigos são abertos, reproduzíveis e adaptáveis; possibilidade de implementar e disponibilizar novos pacotes; multiplataforma (Windows, diversas plataformas UNIX e MacOS); a comunidade acadêmica vem disseminando conhecimento; empresas privadas e públicas estão buscando o R (FONSECA, 2010).

Em relação às variáveis para o estudo proposto, a soja vem sendo estudada por pesquisadores que buscam compreender as relações entre variáveis agrometeorológicas e a produtividade da cultura (CARMELLO, 2011).

No desenvolvimento da soja há alguns fatores que são de extrema importância, como:

- Disponibilidade de água: tanto o excesso como a falta de água é prejudicial para obtenção de uma boa uniformidade na população de plantas, sendo o excesso hídrico mais limitante do que o déficit (EMBRAPA, 2007);
- Temperatura: já que soja se adapta melhor às regiões onde a temperatura oscila entre 20° e 30°C, sendo que a semeadura não deve ser realizada quando a temperatura do solo estiver abaixo dos 20°C, pois a germinação e a emergência da planta ficam comprometidas (THOMAS, 1994);
- Radiação solar: além de fornecer energia luminosa para a fotossíntese, fornece sinais ambientais para uma gama de processos fisiológicos para essa cultura (THOMAS, 1994).

Com vistas a investigar a relação entre a produtividade da soja e as variáveis agrometeorológicas, a teoria de conjuntos nebulosos conhecida como teoria dos conjuntos *fuzzy* será adotada. Essa abordagem baseia-se na caracterização de classes que não possuem limites rígidos entre si (BURROUGH; MCDONNELL, 1998; VIEIRA JUNIOR *et al.*, 2011), sendo indicada quando se busca trabalhar com informações em um ambiente de incerteza, imprecisão, ambiguidades, abstrações e ambivalências em modelos matemáticos complexos, em que os limites difusos comuns em processos que ocorrem no espaço são representados (TAYLOR, MCBRATNEY; WHELAN, 2007; YAN *et al.*, 2007). “Recentemente, trabalhos com aplicações de *fuzzy set* em Análise de Agrupamento tem sido desenvolvidos em diferentes campos do conhecimento” (PIMENTEL, 2013; YANG, 1993).

Sistema híbrido é o uso combinado de características referentes a duas ou mais técnicas da Inteligência Computacional, com o objetivo de aproveitar o que há de melhor em cada técnica. Esse contexto dá origem ao algoritmo *Fuzzy c-Means*, o qual é a base para vários outros modelos híbridos aplicáveis a tarefas de agrupamento e classificação *fuzzy* (ROCHA *et al.*, 2012). Neste contexto, o objetivo deste trabalho é propor a classificação de áreas associadas à produtividade da soja na região oeste do estado do Paraná, utilizando agrupamentos *fuzzy*, mais especificamente o algoritmo *Fuzzy c-Means*, fazendo uso do software R, considerando as seguintes variáveis agro meteorológicas: precipitação pluvial (mm), temperatura média do ar (°C) e radiação solar global média ( $Wm^2$ ).

## 1.1 OBJETIVO GERAL

Aplicar o software R na preparação dos dados, confecção de mapas temáticos e identificação de agrupamentos, fazendo uso do algoritmo classificador *Fuzzy c-Means* no estudo da produtividade da soja na região oeste do Paraná.

## 1.2 OBJETIVOS ESPECÍFICOS

Os principais objetivos específicos são:

- Identificar a aplicação de pacotes que apliquem o algoritmo *Fuzzy c-Means* no R;
- Gerar Agrupamentos com variação estatística em áreas associadas ao cultivo de soja da região oeste do Paraná;
- Validar os agrupamentos obtidos por meio de estatísticas de validação da qualidade dos mesmos;
- Gerar mapas temáticos para os agrupamentos obtidos.

## 1.3 JUSTIFICATIVA

A soja é um dos principais produtos agrícolas brasileiros sendo que, somente em 2011, gerou uma receita da ordem de US\$ 24,9 bilhões (CONAB, 2008).

O agronegócio brasileiro corresponde a 24% do Produto Interno Bruto, segundo dados da Confederação Nacional da Agricultura de 2008, e a soja lidera como principal produto no setor (SILVA; LIMA; BATISTA, 2011).

Os procedimentos para geração de estimativas de safra agrícola, assim como o conhecimento da sua distribuição espacial, constituem uma importante informação para o setor da agricultura. Grande parte desses procedimentos envolvem técnicas de previsão baseadas na agrometeorologia, fundamentando-se na relação estatística entre a variável dependente, produtividade da cultura de soja, e as variáveis independentes, como: precipitação pluvial, temperatura média e radiação solar global (BERLATO; FONTANA; GONÇALVES, 1992; DALLACORT *et al.*, 2006; ASSAD *et al.*, 2007). Com relação aos fatores agro meteorológicos, como seca, excesso de chuvas, temperaturas muito altas ou muito baixas e a baixa luminosidade, ressalta-se que estes podem ocasionar reduções significativas na produtividade da soja, restringindo inclusive as áreas onde espécies comercialmente importantes podem ser cultivadas (FARIAS, 2011).

R é uma software de linguagem *open-source*, gratuita utilizada em diversas disciplinas (WAGNER; MILLER; GARIBALDI, 2011). A adoção de uma ferramenta aberta e gratuita como o R para empregar agrupamentos *Fuzzy c-Means*, pode propiciar que profissionais e pesquisadores de uma vasta gama de áreas façam uso do trabalho desenvolvido.

#### **1.4 ESTRUTURA DO TRABALHO**

O capítulo 1 deste trabalho, é composto pela introdução, objetivos a serem alcançados, a justificativa e estrutura do projeto. O capítulo 2 é composto pela apresentação dos materiais e métodos estudados e utilizados neste trabalho. No capítulo 3 foram descritas as atividades realizadas para implementar o estudo assim como os resultados obtidos por meio deste. No capítulo 4 foram descritas as conclusões do trabalho e possíveis trabalhos futuros deste projeto.

## 2 MATERIAIS E MÉTODOS

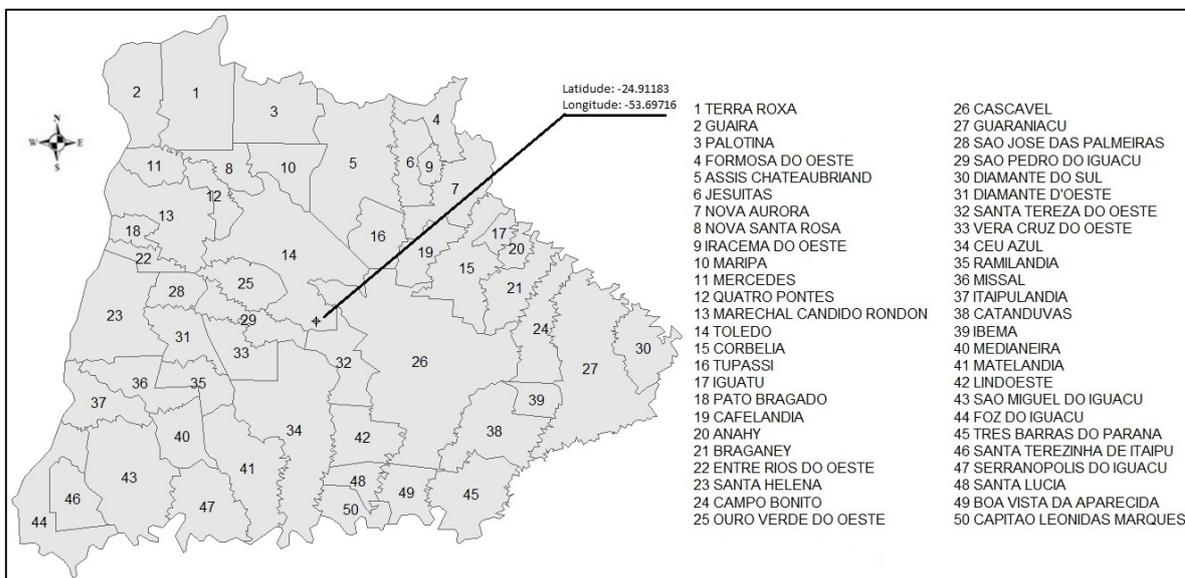
Materiais empregados neste estudo compreendem:

- Área de estudo;
- Dados espaciais e não espaciais;

Os métodos empregados neste estudo envolvem um procedimento de 6 etapas:

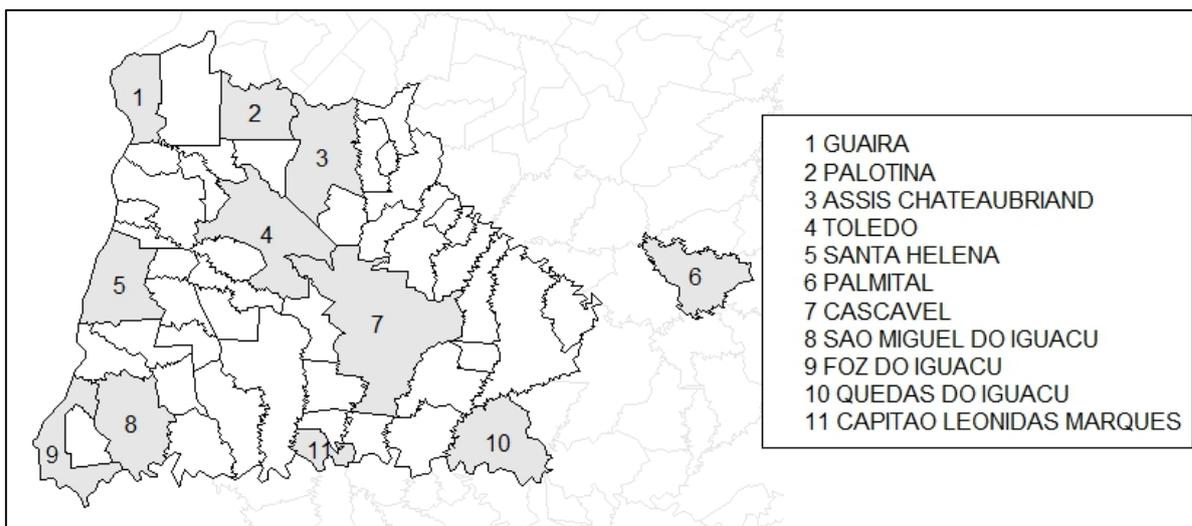
- i. Cálculo das áreas de influência das estações meteorológicas;
- ii. Distribuição dos atributos;
- iii. Padronização dos atributos;
- iv. Aplicação do *Fuzzy c-Means* (FCM);
- v. Alocação dos dados nos agrupamentos estabelecidos;
- vi. Mensuração do nível de similaridade de cada agrupamento;

Área de Estudo: compreende os 50 municípios da região oeste do estado do Paraná, Brasil, de latitude e longitude central aproximada de  $-24,9^\circ$  e  $-53,6^\circ$ , conforme ilustra a Figura 1.



**Figura 1 - Região oeste do Paraná.**

O estudo também compreende os 11 municípios da região e próximos a ela que contem estações meteorológicas, utilizados para estimar os dados meteorológicos dos municípios restantes do oeste, que não as possuem, conforme ilustra a Figura 2.



**Figura 2 - Estações meteorológicas.**

Organização dos Dados: no domínio de Sistemas de Informação Geográfica (SIG), as considerações para organização dos dados são aplicáveis de acordo com os diferentes tipos e natureza dos dados, que necessitam ser organizados e armazenados. Em geral, e também para fins deste estudo, os dados são classificados como: dado espacial e dado não-espacial.

Dado espacial: está associado a elementos geográficos ou espaciais, como o conjunto de polígonos que representam o mapa de municípios que cobrem a região de estudo (vide Figura 1). Atualmente, muitos desses mapas são fornecidos no formato digital.

Para este estudo, o mapa de municípios da região oeste do Paraná foi obtido da empresa GISMAPS que tem como fonte o IBGE, com última atualização em 2005 (GISMAPS, 2013).

Dado não-espacial: se refere a um conjunto de atributos que complementam o dado espacial, descrevendo o que está associado em um ponto, ao longo de uma linha ou em um polígono. Neste trabalho, os dados não espaciais referentes aos municípios como nome e estado, organizados em um arquivo de extensão DBF ligado ao mapa de municípios.

Arquivos \*.DBF (*Data Base File*) contém a descrição alfanumérica dos objetos geográficos (AMORIM; EBERT; HORST, 2005).

Outros atributos foram incorporados a este arquivo DBF para dar sequência ao trabalho. A definição dos atributos incorporados foi baseada no ano safra mais recente, sendo selecionado o do ano-safra 2010/2011. Assim, para cada um dos 50 municípios foram estabelecidos os seguintes atributos:

- Identificador do polígono (PID);

- Produtividade da soja (PROD), medida em ( $\text{tha}^{-1}$ ), fornecida pela Secretaria da Agricultura e do Abastecimento do Paraná (SEAB, 2013);
- Três atributos agrometeorológicos advindos do SIMEPAR (2013): precipitação pluvial (PRECM), medida em mm, temperatura média do ar (TMED), em  $^{\circ}\text{C}$ , e radiação solar global média (RSGM), em  $\text{Wm}^2$ . Inicialmente, essas variáveis estavam disponíveis somente para nove municípios da região de estudo e outros dois municípios próximos a região, conforme destacados na Figura 2, que foram incluídos para melhorar a precisão das estimativas já que os demais municípios sem medições, foram obtidas estimativas por meio do uso de Polígonos de Thiessen (ANDRADE *et al.*, 2008).

O período das safras utilizado para obtenção dos dados agrometeorológicos diários foi de 1<sup>o</sup> de outubro de 2010 até 28 de fevereiro de 2011. A precipitação pluvial média, a temperatura média e radiação solar global média utilizada foi obtida por meio da média aritmética dos dados do período.

Este trabalho foi realizado inteiramente com o auxílio do software R. Os pacotes utilizados em conjunto com o software R foram:

- *base*: pacote nativo do R, contém diversas funções básicas do mesmo;
- *stats*: pacote nativo do R, contém diversas funções estatísticas simples;
- *graphics*: pacote nativo do R, contém as funções básicas para criação gráfica (CRAN, 2013). Todos os gráficos e mapas deste trabalho foram criados através da funções *plot* (gera o gráfico), *legend* (gera a legenda no gráfico) *text* (gera texto no gráfico) e *points* (gera pontos ou símbolos no gráfico), ambas deste pacote;
- *Rcmdr*: uma plataforma independente, de estatística básica que fornece uma interface gráfica com diversas ferramentas (CRAN, 2013);
- *maptools*: ferramentas para a manipulação e leitura de dados geográficos, em particular *shapefiles*, (CRAN, 2013);
- *foreign*: funções para ler e gravar dados persistidos por pacotes estatísticos, e para ler arquivos “dBase” (DBF), (CRAN, 2013);
- *rgeos*; contém as mais famosas funções geoestatísticas, adaptadas para estudantes (RGEOS, 2013);
- *e1071*: funções para análise de classes, agrupamento difuso (*Fuzzy c-Means*), agrupamento ensacado e outros (CRAN, 2013).

## 2.1 CÁLCULO DAS ÁREAS DE INFLUÊNCIA DAS ESTAÇÕES METEOROLÓGICAS

Segundo Unwin e Unwin (1998) a região de influência de cada estação meteorológica pode ser obtida pela aplicação do método dos polígonos de Thiessen.

Os dados obtidos por cada estação meteorológica compreendem a precipitação pluviométrica, a radiação solar e a temperatura média do ar.

Para estudos hidrológicos é necessário salientar que a precipitação pluviométrica varia de forma temporal e espacial. Portanto, é necessário utilizar as informações das estações dentro dessa superfície e em suas vizinhanças, para que seja possível calcular a precipitação média de qualquer superfície. Com base nisso foram adicionadas as cidades com estação meteorológica mais próximas (Figura 2).

Existem três métodos para o cálculo da chuva média: método da média aritmética, método de Thiessen e método das Isoietas (PEDRAZZI, 1999; GOMIG, LINDNER e KOBIYAMA, 2007). É indicado o método dos Polígonos de Thiessen, para quando não há distribuição uniforme das estações pluviométricas. As análises Thiessen, conhecidas também como análises do diagrama de Voronoi, podem ser aplicadas na gestão de diversos temas como meio ambiente, *marketing*, segurança e saúde, entre outros. Dessa forma, cada polígono do tema Thiessen contém os atributos do ponto dentro dele. O procedimento pode ser manual, com o uso de softwares como o AutoCad ou ArcView GIS (GOMIG, LINDNER e KOBIYAMA, 2007).

Para este estudo os polígonos de Thiessen foram calculados utilizando somente o software R, em conjunto com a biblioteca *rgeos* e *maptools*.

Com esse processo realizado, cada polígono, com sua área calculada, representa um valor específico da variável medida. Em relação aos municípios, eles podem fazer parte de um ou mais polígonos e o valor da variável precisa ser computado para cada município

## 2.2 DISTRIBUIÇÃO DOS ATRIBUTOS

A distribuição dos atributos (PRECM, TMED, RSGM) leva em consideração sempre a área do município que faz parte do polígono, ela foi definida e ocorreu conforme Equação 1:

$$Vatr_m = \sum_{i=1}^n \frac{A_m \cap A_{ei}}{A_m} Vatr_{ei} \quad (1)$$

Em que:  $Vatr_m$  é o valor do atributo atribuído ao município;  $n$  é o número de estações meteorológicas;  $A_m$  é a área do município;  $A_{ei}$  é a área de influência da estação  $i$ ;  $Vatr_{ei}$  é o valor do atributo da estação  $i$ .

O atributo restante (PROD) não precisa ser calculado, já que o valor de cada município na região oeste é conhecido (SEAB, 2013).

## 2.3 PADRONIZAÇÃO DOS ATRIBUTOS

A Padronização dos atributos (PROD, PRECM, TMED, RSGM), ocorreu conforme segue:

$$Vatr_{pad} = \frac{(Vatr - Vatr_{min})}{(Vatr_{max} - Vatr_{min})} \quad (2)$$

Em que:  $Vatr_{pad}$  é o valor do atributo padronizado, variando de 0 a 1;  $Vatr$  é o atributo observado;  $Vatr_{min}$  e  $Vatr_{max}$  referem-se aos valores mínimos e máximos do atributo observado, respectivamente. Essa forma de padronização é necessária para o emprego do FCM, garantindo que todos os atributos tenham a mesma ordem de grandeza, isto é, variando de 0 a 1 (GOMES *et al.*, 2011).

## 2.4 APLICAÇÃO DO FUZZY C-MEANS

A aplicação do *Fuzzy c-Means* (FCM) foi utilizada para se obter agrupamentos ou grupos similares em um conjunto de dados. O FCM é um algoritmo iterativo e, a cada iteração, novos centros de agrupamentos e graus de pertinência são calculados, buscando sempre minimizar a métrica euclidiana entre cada dado e o centro do agrupamento. A responsabilidade de verificar essa convergência cabe à função objetivo.

O método c-Means segue os 4 seguintes passos (SHAHI, ATAN e SULAIMAN, 2011):

- 1- Inicializar  $u = [u_{ij}]$  matriz,  $u^{(0)}$
- 2- A cada passo: calcular o centro dos vetores  $C^{(k)} = [c_j]$  com  $u^{(k)}$

$$C_j = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m} \quad (3)$$

- 3- Atualizar  $u^{(k)}, u^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (4)$$

- 4- Se  $\|u^{(k+1)} - u^{(k)}\| < \epsilon$  então para o algoritmo; se não retorna ao passo 2

Uma descrição mais detalhada do FCM pode ser vista em Shahi, Atan e Sulaiman (2011), em Dunn (1973), Bezdek (1981) e Bezdek e Pal (1992).

## 2.5 ALOCAÇÃO DOS DADOS NOS AGRUPAMENTOS ESTABELECIDOS

A alocação dos dados nos agrupamentos estabelecidos pelo FCM foi realizada por meio do Método de Decisão pelo Maior Grau de Pertinência (MDMGP) (GUIERA *et al.*, 2005; FERREIRA *et al.*, 2008; NG *et al.*, 2008; WANG; FEI, 2009). No MDMGP, a determinação

de qual agrupamento o dado pertencerá é dada pelo maior grau de pertinência, garantindo que todos os dados sejam alocados. A partir da alocação dos dados nos agrupamentos, torna-se importante uma avaliação da qualidade do resultado obtido. Neste estudo, empregou-se a Entropia de Partição Modificada (MPE) para a indicação do melhor número de agrupamentos (BUDAYAN; DIKMEN; BIRGONUL, 2009; SUN *et al.*, 2012). Definida conforme segue:

$$MPE = \frac{-\sum_{k=1}^n \sum_{i=1}^c u_{ik} \log(u_{ik}) / n}{\log(c)} \quad (5)$$

Em que:  $c$  representa a quantidade de agrupamentos;  $n$  corresponde ao número de polígonos;  $u_{ik}$  corresponde ao grau de pertinência do polígono  $k$  do agrupamento  $i$ .

O MPE mede o grau de desorganização entre os agrupamentos de um conjunto de dados. Seu valor varia de 0 a 1. Valores próximos de zero indicam agrupamentos distintos, apresentando pequeno grau de compartilhamento entre os dados. Valores próximos de um indicam não haver agrupamentos distintos, apresentando elevado grau de compartilhamento dos dados (MCBRATNEY; MOORE, 1985; FRIDGEN *et al.*, 2004). Segundo Boydell e Mcbratney (2002), o melhor número de agrupamento de um conjunto de dados é estabelecido com base no valor mínimo de MPE (ARAÚJO, 2012).

## 2.6 MENSURAÇÃO DO NÍVEL DE SIMILARIDADE DE CADA AGRUPAMENTO

Para a Mensuração do nível de similaridade de cada agrupamento, definiu-se e empregou-se o Índice de Similaridade de Clusters ( $ISC_i$ ), definido por Araújo (2012), conforme segue:

$$ISC_i = \left[ \frac{\sum_{k=1}^{n_i} u_{ik}}{n_i} \right] 100 \quad (6)$$

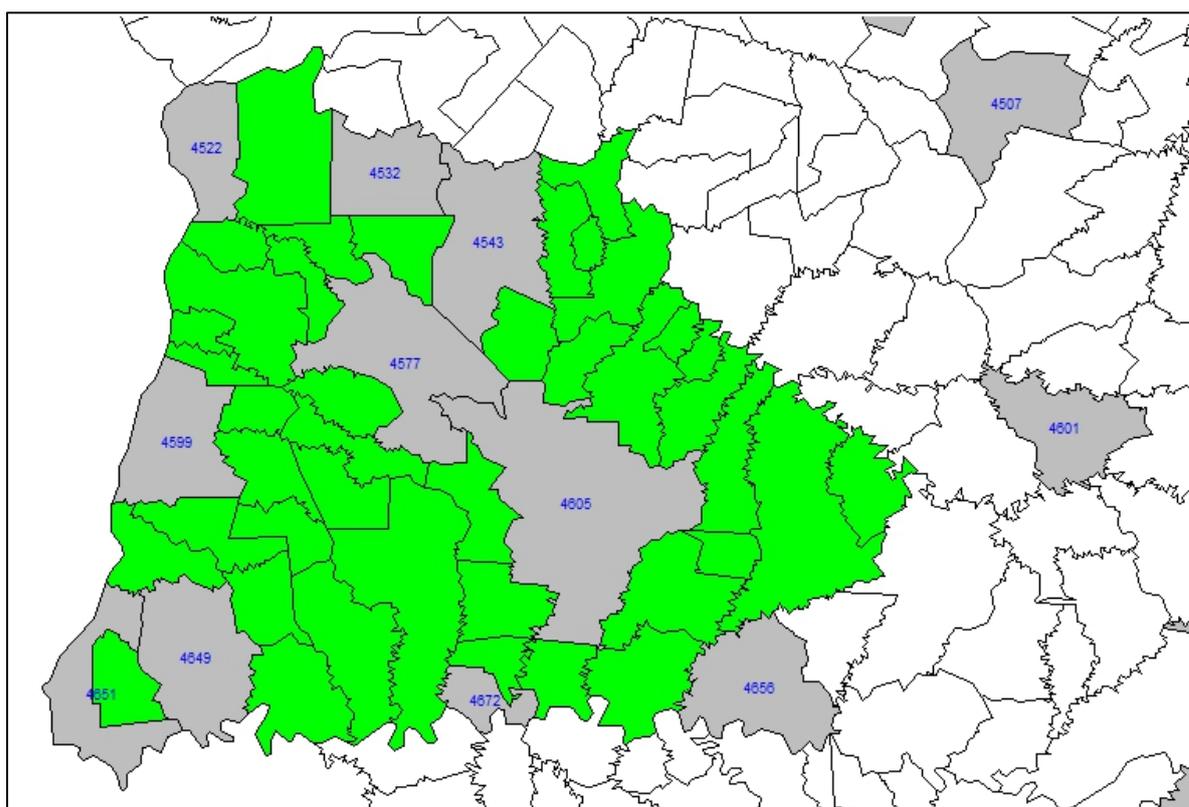
Em que:  $n_i$  corresponde ao número ( $n$ ) de polígonos do agrupamento  $n$ , e  $u_{ik}$  segue a mesma definição dada na Equação 5.

O índice mede quanto os elementos de um determinado agrupamento são similares entre si, os valores variam de 0 à 1 sendo 1 equivalente a totalmente similar e 0 a totalmente distintos (ARAÚJO, 2012).

Foi efetuado o transporte dos dados em que os resultados do FCM foram transportados para o arquivo DBF e assim conectados ao mapa de polígonos para posteriores análises.

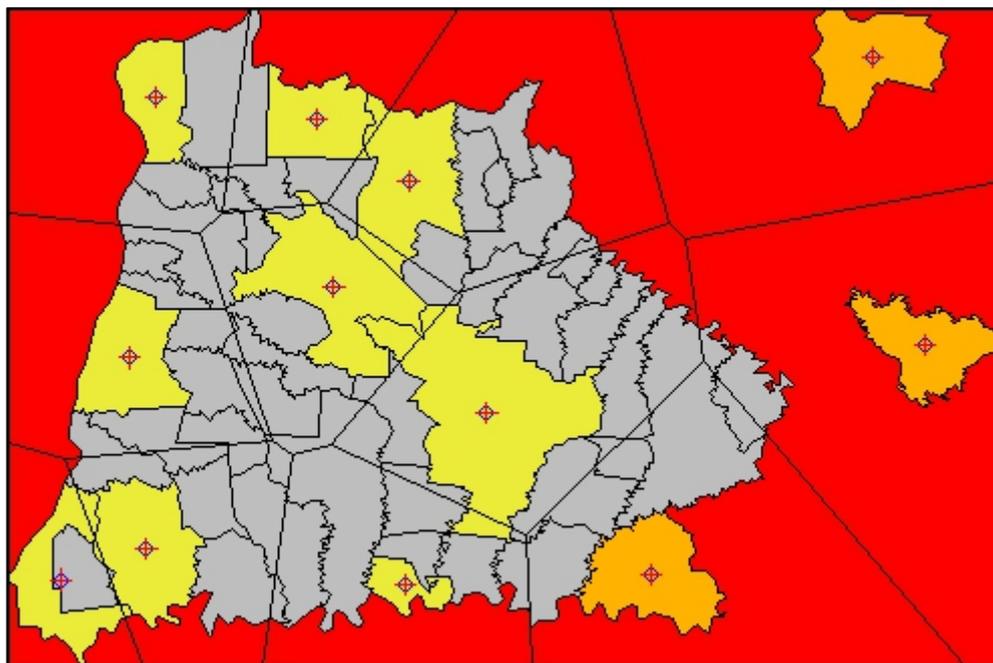
### 3 RESULTADOS E DISCUSSÃO

Fora identificados os PID's dos municípios com as estações meteorológicas mais próximos da região Oeste do Paraná, municípios estes definidos pelos dados advindos do SIMEPAR (2013). A Figura 3 ilustra a região Oeste em verde e os municípios com estações em cinza, com seus respectivos PIDs em azul.



**Figura 3 - Região oeste e estações meteorológicas próximas.**

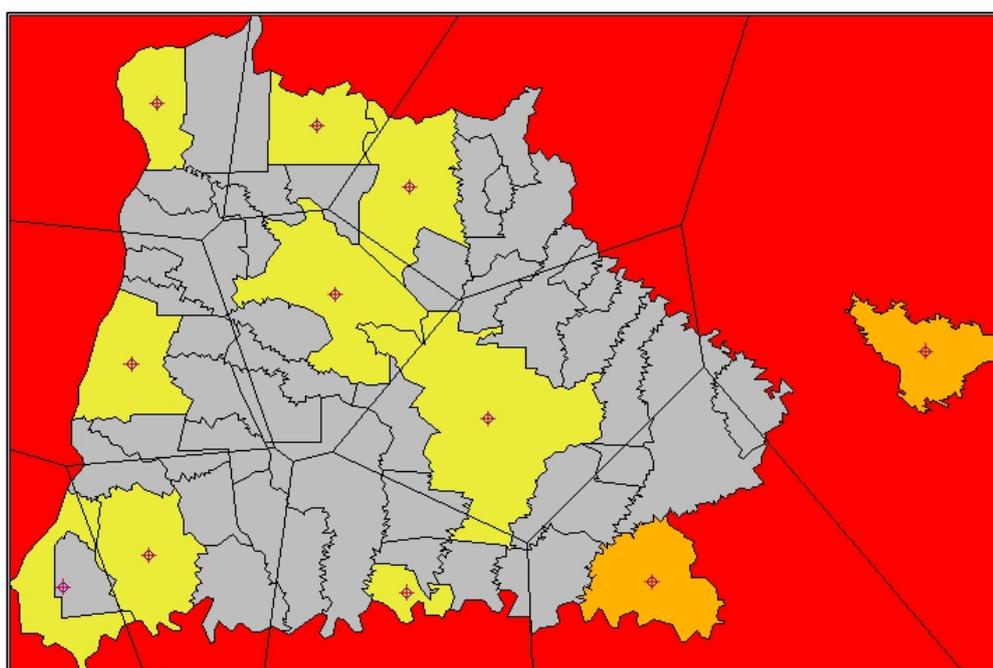
Foram obtidas as coordenadas centroides das estações meteorológicas, utilizando a função `gCcentroid()` do pacote `Rgeos`, e foi definida a função de Voronoi para gerar os polígonos de Thiessen, assim pôde-se calcular a área de influência de cada estação, processo descrito no Apêndice B. O resultado pode ser observado na Figura 4, que destaca os municípios com estações meteorológicas na região oeste em amarelo, e os fora da região em laranja, ambos com seus respectivos centroides sinalizados.



**Figura 4 - Estações próximas da região oeste do Paraná com áreas de influência demarcadas.**

A partir dos polígonos de Thiessen foi possível identificar as estações meteorológicas que tem alguma influência sobre a região oeste do Paraná. Isto possibilitou a estimativa dos dados meteorológicos, nas regiões vizinhas as das estações (PEDRAZZI, 1999; GOMIG; LINDNER; KOBİYAMA, 2007), e dessa forma nos municípios do oeste.

Os polígonos foram recalculados mantendo somente os municípios de estações influentes, utilizando a função subset() do pacote base (Figura 5).



**Figura 5 - Municípios com estações influentes na região oeste do Paraná.**

Com as áreas de influência das estações (polígonos de Thiessen) calculada pela função definida (Voronoi), foi possível fazer a distribuição (atribuição) do dados meteorológicos (PRECM, TMED, RSGM), para o restante dos municípios, levando em consideração sempre a área do município que faz parte do polígono de influência da estação. Para isto foi utilizado o código do cálculo dos dados meteorológicos (Apêndice D, Quadro 8).

Com os dados distribuídos foi feita a padronização dos dados, utilizando a função Padronizar() (Apêndice C, Quadro 5), necessária para o emprego do FCM, garantindo assim que todos os atributos tenham a mesma ordem de grandeza.

Foi realizada uma análise preliminar dos dados (feita por meio do código para o cálculo das estatísticas preliminares, Apêndice D, Quadro 9). A Tabela 1 sintetiza as principais estatísticas descritivas das variáveis observadas, bem como de seus valores padronizados.

Tabela 1 - Análise preliminar dos dados

Atributo	Média	Desvio Padrão	Coefficiente de variação	Valor Mínimo	Valor Máximo	Mediana
PROD (tha <sup>-1</sup> )	3,44	0,24	1,72	2,88	4,00	3,40
PROD padronizada	0,50	0,21	9,32	0,00	1,00	0,46
PRECM (mm)	7,13	1,51	32,04	4,14	9,73	7,36
PRECM padronizada	0,53	0,27	13,67	0,00	1,00	0,57
TMED (°C)	23,16	0,78	2,67	21,80	24,10	23,36
TMED padronizada	0,59	0,34	19,75	0,00	1,00	0,67
RSGM (Wm <sup>2</sup> )	940,83	24,93	66,07	876,30	969,40	941,55
RSGM padronizada	0,69	0,26	10,34	0,00	1,00	0,70

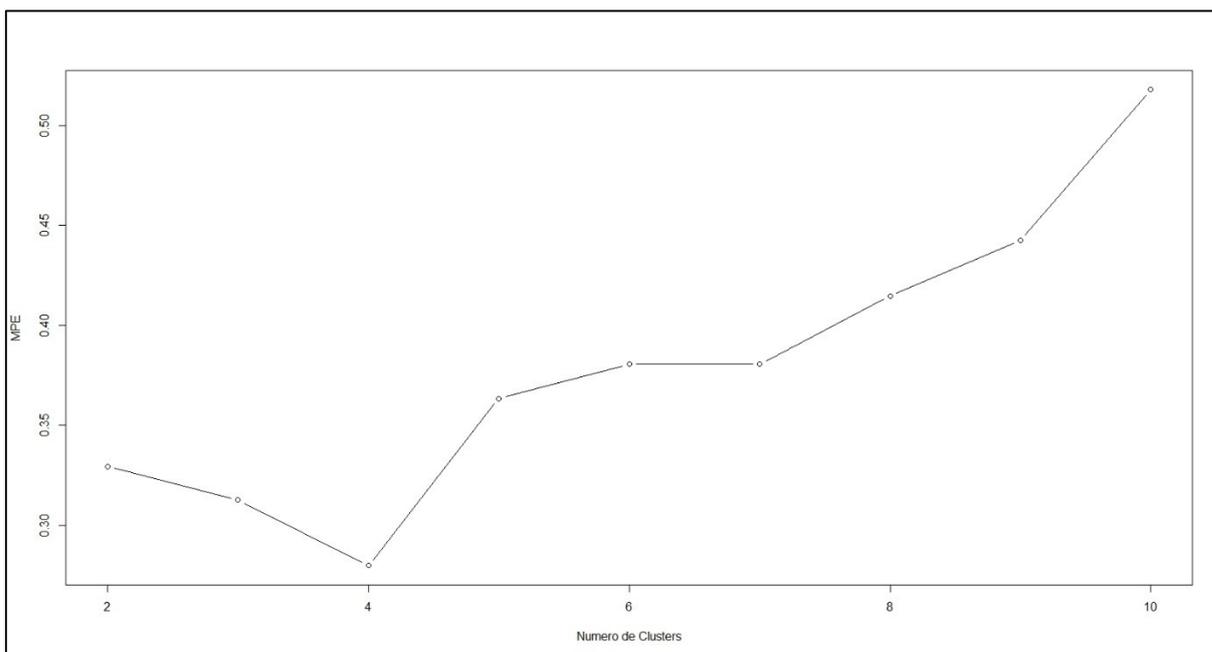
PROD.: Produtividade da soja; PRECM.: Precipitação pluviométrica média; TMED.: Temperatura média; RSGM.: Radiação solar global média.

Pela Tabela 1, verifica-se que a produtividade média obtida no ano-safra de 2010/2011 foi de 3,44 tha<sup>-1</sup>, produtividade média considerada alta em comparação com a média de produção nacional 3.106 tha<sup>-1</sup> no mesmo ano (EMBRAPA, 2013), demonstrando o potencial de produção da região oeste do Paraná. A precipitação pluviométrica média diária do período de estudo foi de 7,13 mm, favorecendo o desenvolvimento da cultura (EMBRAPA, 2007). A temperatura média de 23,16°C da região está dentro dos padrões em que a soja se adapta melhor (20 a 30°C).

Para aplicar o FCM foi utilizado a função cmeans() do pacote e1071. Das vinte tentativas realizadas para os 4 agrupamentos, o algoritmo *Fuzzy c-Means* atingiu sua melhor condição de parada em 22 iterações em que a função objetivo forneceu o valor de 0,05586909.

Esses valores corroboram a convergência dos dados, minimizando a distância de qualquer dado de um agrupamento em relação ao seu centro (CHEN; WANG, 2009; ZHU *et al.*, 2012).

Após o FMC, o resultado das alocações foi submetido ao índice MPE variando a quantidade de agrupamentos de 2 à 10, para qualificar a separação dos agrupamentos. Depois de 15 execuções do MPE, foram identificados quatro agrupamentos, como mostra o gráfico na Figura 6 (feito por meio do código para a análise dos clusters, Apêndice D, Quadro 10).



**Figura 6 - Gráfico dos resultados do MPE por número de clusters.**

De acordo com as características do MPE, buscou-se validar a indicação de um número ótimo para a quantidade de agrupamentos. Isto foi realizado por meio do Método de Decisão pelo Maior Grau de Pertinência (MDMGP), através do código para a distribuição dos municípios (Apêndice D, Quadro 11), que quantifica o grau de inclusão (sobreposição) entre agrupamentos (ARAÚJO, 2012). Os valores obtidos da sobreposição são apresentado na Tabela 2 (valores da tabela obtidos por meio do código para a análise dos clusters, Apêndice D, Quadro 10).

Tabela 2 - Graus de inclusão entre os agrupamentos estabelecidos pelo método MDMGP

(A, B)	S(A, B)	(A, B)	S(A, B)	(A, B)	S(A, B)	(A, B)	S(A, B)
1,2	0.450	2,1	0.468	3,1	0.375	4,1	0.296
1,3	0.385	2,3	0.365	3,2	0.342	4,2	0.259
1,4	0.320	2,4	0.291	3,4	0.463	4,3	0.438

Notas: S(A, B) refere-se ao grau de sobreposição entre os agrupamentos A e B.

Em relação aos valores de inclusão entre agrupamentos, quanto menor o valor identificado, mais definido é o agrupamento, pois não se aproxima de outros. Em contrapartida, valores acima de 0,5 identificam uma nebulosidade entre os agrupamentos, pois são próximos (ARAÚJO, 2012). Desta maneira, a qualidade dos quatro agrupamentos, apontada pelo MPE foi considerada satisfatória para esse estudo, já que nenhuma pertinência foi identificada acima de 0,5 como mostra a Tabela 2.

Estabelecido o número de agrupamentos e seus respectivos graus de inclusão, buscou-se quantificar a distribuição dos municípios segundo seus agrupamentos. Para tal, foram empregados os métodos de decisão (MDMGP). Os resultados são apresentados na Tabela 3, na qual a coluna (M) indica o número de municípios alocados para cada agrupamento e a coluna (%) indica o percentual de municípios alocados em relação ao total de municípios do estudo (50). Os valores da tabela foram obtidos por meio do código para a análise dos clusters, Apêndice D, Quadro 10.

Tabela 3 - Distribuição quantitativa dos municípios nos agrupamentos

Agrup. 1		Agrup. 2		Agrup. 3		Agrup. 4	
(M)	(%)	(M)	(%)	(M)	(%)	(M)	(%)
12	24	12	24	11	22	15	30

Pela característica do *Fuzzy c-Means* de reduzir a métrica euclidiana entre cada dado e o centro do agrupamento, é interessante observar este centro, como mostra a tabela 4 (valores da tabela obtidos por meio do código para o cálculo das estatísticas dos clusters parte 2, Apêndice D, Quadro 13).

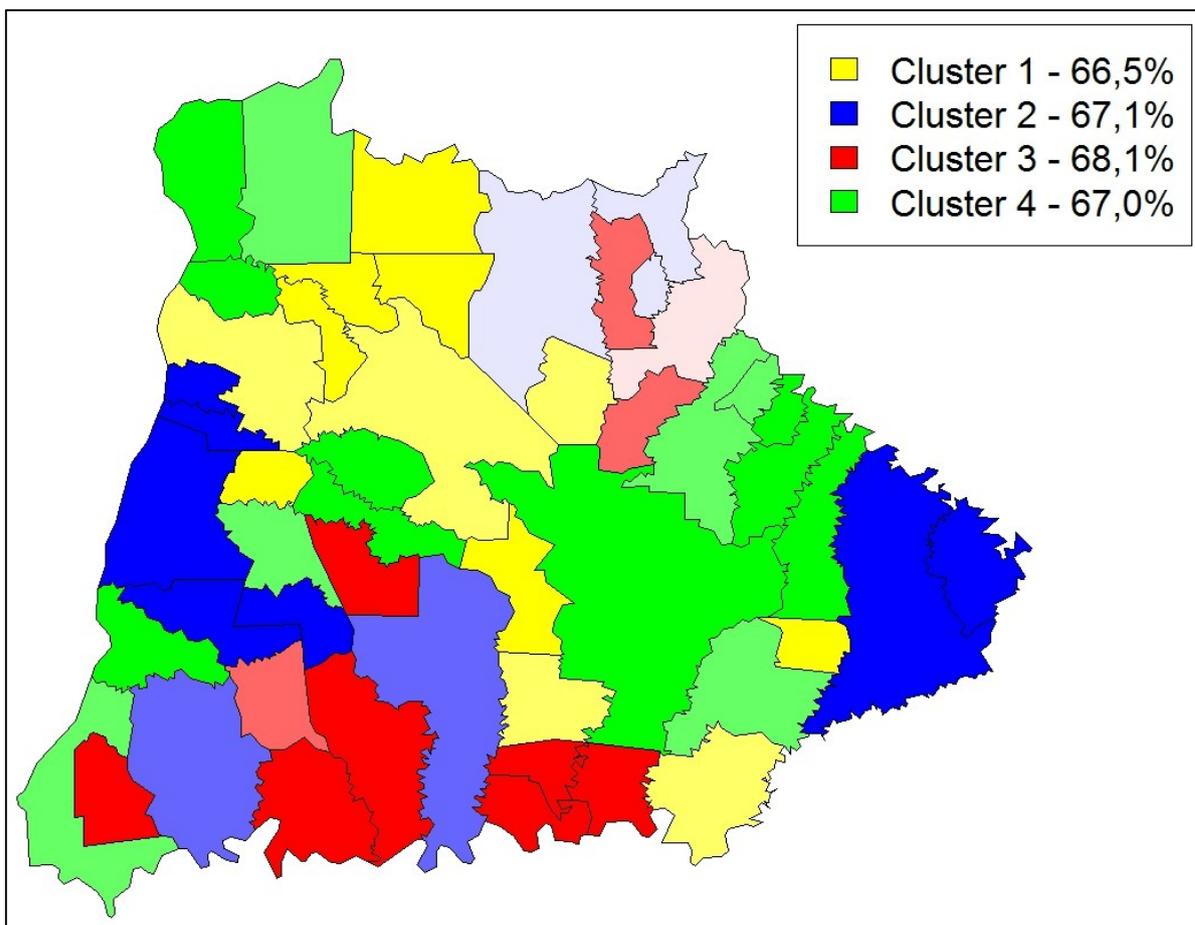
Tabela 4 - Centro das variáveis de cada agrupamento pelo método *Fuzzy c-Means*.

Agrupamento	PROD	TMED	RSGM	PRECM
1	0,507	0,789	0,338	0,565
2	0,335	0,908	0,653	0,131
3	0,636	0,507	0,882	0,875
4	0,499	0,158	0,834	0,556

PROD.: Produtividade da soja; PRECM.: Precipitação pluviométrica média; TMED.: Temperatura média; RSGM.: Radiação solar global média.

Ao se observar os centros obtidos do agrupamento difuso é possível observar algumas características de cada agrupamento, como o agrupamento 3 que é o agrupamento que possui maior precipitação média (0,875), também é o que possui maiores níveis de produção (0,636), sendo também o primeiro em radiação solar global média (0,882). Pode-se observar ainda que o segundo agrupamento possui a menor precipitação média como centro (0,131), e também a menor produção (0,335).

A distribuição dos municípios impostas pelos métodos MDMGP pode ser espacialmente visualizada na forma de mapa. Neste trabalho, os mapas foram gerados com o auxílio do software R e o pacote *maptools*. Conforme ilustra a Figura 7, em que as tonalidades das cores, da mais clara para a mais escura, denota o nível da pertinência do município para o agrupamento em que está alocado. As tonalidades foram divididas em 3 classes de pertinência: mais clara  $>0$  e  $\leq 0,35$ ; intermediária  $>0,35$  e  $\leq 0,65$ ; e a mais escura  $>0,65$  e  $\leq 1,0$ . Os valores apresentados para cada cor representam os níveis de similaridades obtidos para cada agrupamento, sendo o agrupamento 3 o mais similar, da ordem de 68%. Estes níveis de similaridade foram obtidos por meio do  $ISC_1$ . O mapa da Figura 7 foi gerado por meio do código para criação do mapa temático (Apêndice D, Quadro 16).



**Figura 7 - Distribuição dos municípios nos agrupamentos imposta pelo FCM.**

Por meio da Tabela 5, que sintetiza as principais estatísticas descritivas dentro dos agrupamentos encontrados pelo método MDMGP, buscou-se analisar a produtividade da soja e as variáveis agrometeorológicas. Identificou-se que apesar de o agrupamento 1 conter o maior produtor ( $4 \text{ tha}^{-1}$ ), a maior produtividade média está localizada no agrupamento 3, onde também foram obtidos a maior média de precipitação, em relação também à produtividade, menor desvio padrão e menor coeficiente de variação dentre todos os agrupamentos. O agrupamento número 1 também tem nele identificado a maior temperatura média, com os menores valores para desvio padrão e coeficiente de variação. Em relação à precipitação, o menor volume de chuva (total e média) ocorreu no agrupamento de número 2, que teve a menor produtividade média dentre os agrupamentos, identificando ainda, para a precipitação, o maior desvio padrão e coeficiente de variação em relação aos demais grupos. Valores da tabela obtidos por meio do código para o cálculo das estatísticas dos clusters parte 1 (Apêndice D, Quadro 12).

Tabela 5 - Estatísticas para as variáveis do estudo em cada agrupamento da região de estudo

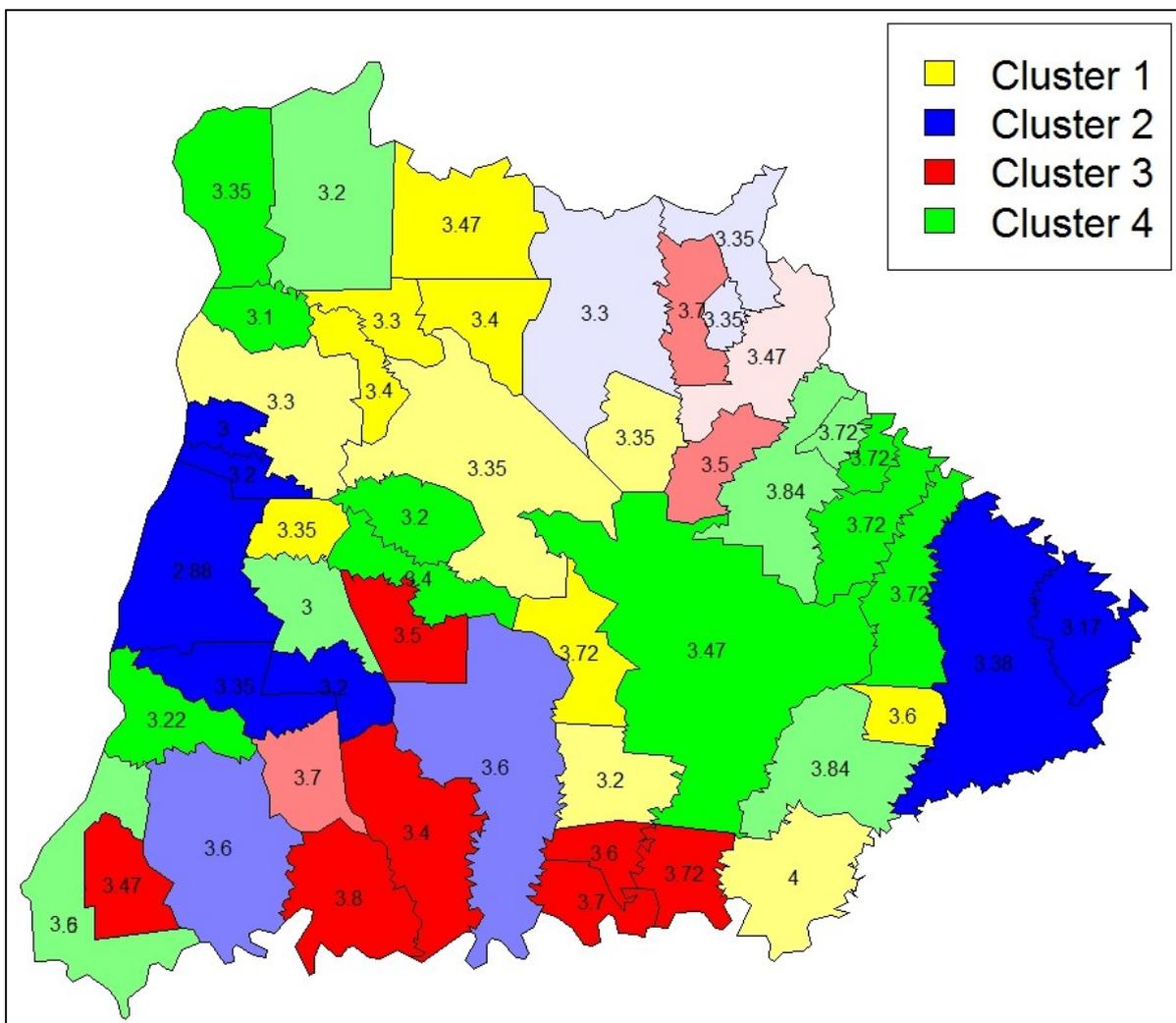
Cluster	Variável	Mínimo	Máximo	Média	Mediana	Des. Pad.	C.V. (%)
1	Prod	3.200	4.000	3.453	3.375	0.221	1.418
	Precm	6.237	8.528	7.275	7.317	0.751	7.760
	TMedia	23.112	24.037	23.653	23.602	0.288	0.350
	RSGM	876.303	939.788	905.987	906.224	18.653	38.405
2	Prod	2.880	3.595	3.281	3.324	0.210	1.348
	Precm	4.144	7.431	5.315	4.762	1.359	34.729
	TMedia	23.870	24.103	23.963	23.926	0.095	0.037
	RSGM	919.512	969.400	942.755	940.600	15.681	26.082
3	Prod	3.397	3.800	3.596	3.595	0.134	0.497
	Precm	7.367	9.734	8.842	9.115	1.038	12.188
	TMedia	22.650	24.095	23.162	22.929	0.492	0.047
	RSGM	938.777	969.400	960.405	962.300	9.329	9.062
4	Prod	3.000	3.842	3.473	3.471	0.283	2.309
	Precm	6.372	7.667	7.232	7.474	0.430	2.559
	TMedia	21.802	22.554	22.143	22.177	0.234	0.247
	RSGM	931.697	964.363	952.811	957.240	12.185	15.583

C.V.(%): Coeficiente de Variação; PROD.: Produtividade; PRECM.: Precipitação Pluviométrica Média; TMED.: Temperatura Média; RSGM.: Radiação Solar Global Média;

Para obter os valores mostrados na Tabela 5 foram utilizados as seguintes funções:

- Função *mean* pacote *base*: Obtém a média de um conjunto de dados;
- Função *median* do pacote *stats*: Obtém a mediana de um conjunto de dados;
- Função *sd* do pacote *stats*: Obtém o desvio padrão de um conjunto de dados;
- Função *var* do pacote *stats*: Obtém a variância de um conjunto de dados.

Com o objetivo de validar os agrupamentos identificados pelo FCM (Figura 7) e as estatísticas apresentadas na Tabela 5, o mapa apresentado na Figura 8 mostra a produtividade da soja em destaque para cada município.



**Figura 8 - Mapa temático da produtividade da soja.**

É possível identificar que apesar da localização geográfica não ser considerada no algoritmo de classificação difusa, o grupoamento 3 (*cluster 3*), de maior produção média encontra-se quase totalmente ao sul. Outro aspecto que foi percebido é que quanto maior a precipitação média no período, maior a produção média por hectare. Não é possível fazer alguma afirmação segura com base nos mapas apresentados, sendo necessário analisar não apenas um ano-safra mas sim um período bem maior, já que eventos isolados como uma geadas muito fortes ou uma chuvas concentradas no início da germinação podem comprometer a confiabilidade dos resultados do FCM.

## 4 CONSIDERAÇÕES FINAIS

### 4.1 CONCLUSÃO

Utilizando o software R, para aplicar do algoritmo *Fuzzy C-Means*, obteve-se uma classificação dos municípios, com graus de similaridades da ordem de 66 a 68%. Dos números de agrupamentos que poderiam ser formados o número de 4 agrupamentos foi indicado pelo método MPE e o Método de Decisão pelo Maior Grau de Pertinência (MDMGP) foi utilizado por sua característica de garantir que todos os dados pertençam a um grupo.

Com as classificações obtidas foi possível identificar diferentes similaridades, tanto nos municípios que compuseram cada agrupamento, como entre os agrupamentos obtidos. A mensuração da similaridade entre os municípios de cada agrupamento foi possível por meio do Índice de Similaridade de Clusters (ISC<sub>1</sub>). Em relação à similaridade entre agrupamentos, ela foi subsidiada pelo indicador que mede o grau de inclusão entre os agrupamentos.

O R mostrou-se, além de ser capaz de gerar diversas análises estatísticas e criar classificações e agrupamentos de dados difuso, uma ótima ferramenta gráfica, com grande potencial em estudos envolvendo dados geográficos, oferecendo diversos pacotes e ferramentas e com uma documentação satisfatória. Por ser totalmente livre e aberto as muitas funções puderam ser implementadas e ou modificadas para se adequarem melhor ao estudo.

### 4.2 TRABALHOS FUTUROS/CONTINUAÇÃO DO TRABALHO

Com os resultados obtidos neste trabalho é possível subsidiar futuros estudos com metodologias que possam, por exemplo, considerar a correlação espacial entre as unidades de áreas (municípios). As ferramentas aqui utilizadas também podem subsidiar não só estudos com base em recursos agrícolas mas como qualquer conjunto de dados georreferenciado ou não.

O estudo do software R em conjunto com as variáveis agrometeorológicas será continuado em um possível curso de pós graduação em engenharia agrícola, para aperfeiçoar os métodos e buscar melhores utilizações para as bases estatísticas, com uma janela de tempo maior, afim de relacionar com segurança possíveis padrões entre as variáveis.

## 5 REFERÊNCIAS BIBLIOGRÁFICAS

AMORIM, G. M. E.; EBERT, D.; HORST, R. Integração de informações geológicas para o geoturismo na bacia do rio corumbataí e sua divulgação na web através do Mapsserver. **Geociências**, p. 221- 238 São Paulo, 2005.

ANDRADE, N.L.R. de; XAVIER, F.V.; ALVES, E.C.R. de F.SILVEIRA, A.; OLIVEIRA, C.U.R.. Caracterização morfométrica e pluviométrica da bacia do Rio Manso – MT. **Revista Brasileira de Geociências**, São Paulo/SP, 2008, v.27, n.2, p.237-248,

ARAÚJO, E. C. D. **Análise da produtividade da soja associada a fatores agrometeorológicos, por meio de estatística espacial de área, na região oeste do estado do Paraná**. Universidade Estadual do Oeste do Paraná. Cascavel/PR. 2012.

ASSAD, E. D.; MARIN, F. R.; MEDEIROS, S. R. E.; PILAU, F. G.; FARIAS, J. R. R.; PINTO, H. S.; ZULLO JR, J.. Sistema de previsão de safra de soja para o Brasil. **Pesquisa Agropecuária Brasileira**, Brasília/DF, v. 42, p. 615-625, 2007.

BERLATO, M. A.; FONTANA, D. C.; GONÇALVES, H. M. Relação entre rendimento de grãos de soja e variáveis meteorológicas. **Pesquisa Agropecuária Brasileira**, Brasília/DF, v. 27, n. 5, p. 695-702, 1992.

BEZDEK, J. C. Pattern recognition with fuzzy objective function algorithms. New York: **Plenum**, 1981.

BEZDEK, J. C.; PAL, S. K. **Fuzzy Models for Pattern Recognition**. IEEE Press. New York: [s.n.]. 1992.

BUDAYAN, C.; DIKMEN, I.; BIRGONUL, M. T. Comparing the performance of traditional cluster analysis, self-organizing maps and fuzzy C-means method for strategic grouping. **Expert Systems With Applications - ESWA**, v. 36, n. 9, p. 11772-11781, 2009.

BURROUGH, P. A.; MCDONNELL, R. A. Principles of Geographical Information Systems. **OXFORD UNIVERSITY PRESS**, 1998.

CARMELLO, V. VULNERABILIDADE AGRÍCOLA DA PRODUÇÃO DE SOJA NA REGIÃO METROPOLITANA DE LONDRINA – PR: ANALISE DA SAFRA DE 2005/06. **Revista Geográfica de América Central**, Costa Rica, v. 2, n. 47E, p. 1-16, 2011.

CHEN, W.-C.; WANG, M.-S. A fuzzy c-means clustering-based fragile watermarking scheme for image authentication. **Department of Engineering Science, National Cheng Kung University**, Taiwan, v. 36, n. 2, p. 1300–1307, Março 2009.

CONAB. Acompanhamento da Safra Brasileira. **Conab**, 2008. Disponível em: <[http://www.conab.gov.br/conabweb/download/safra/12\\_levantamento\\_set2008.pdf](http://www.conab.gov.br/conabweb/download/safra/12_levantamento_set2008.pdf)>. Acesso em: 15 Junho 2013.

CRAN. Related Directories 2013. Disponível em: <<http://cran.r-project.org/web/packages/>>. Acesso em: 27 Julho 2013.

DALLACORT, R. FREITAS, P. S. L. de; FARIA, R. T. de; GONÇALVES, A. C. A.; REZENDE, R.; BERTONHA, A.; Utilização do modelo Cropgro-soybean na determinação de melhores épocas de semeadura da cultura da soja, na região de Palotina, estado do Paraná. **Acta Scientiarum.Agronomy**, Maringá - PR, v. 28, p. 583-589, 2006.

DUNN, J. C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. **Journal Cybernetics and Systems**, França, v. 3, p 32-57, 1973.

EMBRAPA. Tecnologias de produção de soja, **Embrapa**, 2007. Disponível em: <[http://www.cnpso.embrapa.br/download/tpsoja\\_2007\\_pr.pdf](http://www.cnpso.embrapa.br/download/tpsoja_2007_pr.pdf)>. Acesso em: 28 Julho 2013.

EMBRAPA. Soja em números (safra 2010/2011), **Embrapa**, 2013. Disponível em: <[http://www.cnpso.embrapa.br/index.php?cod\\_pai=2&op\\_page=294](http://www.cnpso.embrapa.br/index.php?cod_pai=2&op_page=294)>. Acesso em: 28 Julho 2013.

FARIAS, J. R. B. **Limitações climáticas à obtenção de rendimentos máximos de soja**. Quinto congreso de la Soja en el Mercosur. 2011.

FERREIRA, G. C. N.; GAMA, R. A. T. S. da; CAVALCANTI, M. C.; MOURA, A. M. de C. **Organização automática de páginas Web para exibição em portais semânticos**. WebMedia '08 Companion Proceedings of the XIV Brazilian Symposium on Multimedia and the Web. New York/USA: [s.n.]. 2008. p. 161-163.

FONSECA, B. H. F. A escolha do R como Software Estatístico. **Leg**, 2010. Disponível em: <<http://www.leg.ufpr.br/lib/exe/fetch.php/pessoais:bruno:dissertacao:palestra.pdf>>. Acesso em: 15 Junho 2013.

FRIDGEN, J. J.; KITCHEN, N. R.; SUDDUTH, K. A.; DRUMMOND, S. T.; WIEBOLD, W. J.; FRAISSE, C. W. Management Zone Analyst (MZA): software for sub-field management zone delineation. **Agron**, Madison/USA, v. 96, p. 100-108, 2004.

GISMAPS, 2013. Disponível em: <<http://www.gismaps.com.br/english/shape.htm>>. Acesso em: 15 Junho 2013.

GOMES, A. da S.; PIRES, M. de M.; ALMEIDA, V. M.de; ROSADO, P. L.; SANTOS, P. R. P. SÃO JOSÉ, A. R. Análise dos territórios da região sudoeste da Bahia na perspectiva do desenvolvimento rural. **Revista Desenharia**, Salvador/BA, 2011, p. 59-82.

GOMIG, K.; LINDNER, E.; KOBAYAMA, M. Áreas de influência das estações pluviométricas na bacia rio do Peixe/SC pelo método de polígonos de Thiessen utilizando imagem de satélite e SIG. **Anais XIII Simpósio Brasileiro de Sensoriamento Remoto**, Florianópolis, Brasil, 21 Abril 2007, p. 3373-3380.

GUIERA, A. J. A.; Centeno, T. M.; DELGADO, M. R.; MULLER, M. Segmentação por Agrupamentos Fuzzy C-means em Imagens LiDAR Aplicados na Identificação de Linhas de Transmissão de Energia Elétrica. **Espaço Energia**, Paraná, Outubro 2005. 24-31.

JUNIOR, F. A. R. et al. Geração de zonas de manejo para cafeicultura empregando-se sensor SPAD e análise foliar. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 15, n. 8, p. 778 - 787, Agosto 2011.

MCBRATNEY, A. B.; MOORE, A. W. Application of fuzzy sets to climatic classification. **Agricultural and Forest Meteorology - AGR FOREST METEOROL**, v. 35, p. 165-185, 1985.

MEDEIROS, D. M. R. D. **Análise de dados por meio de agrupamento fuzzy semi-supervisionado e mineração de textos**. Instituto de Ciências Matemáticas e de Computação. São Carlos. 2010.

NG, H. P. et al. Fuzzy c-means algorithm with local thresholding for gray-scale images. **International Journal on Artificial Intelligence Tools**, Reino Unido, v. 17, p. 765–775, 2008.

PEDRAZZI, J. A. FACENS – Hidrologia Aplicada, 1999. Disponível em: <<http://www.facens.br/site/alunos/download/hidrologia>>. Acesso em: 25 Junho 2013.

PIMENTEL, B.. **Agrupamento de Dados Simbólicos usando Abordagem Possibilistic**. UFPE. [S.l.]. 2013.

RGEOS. Geostatistical R Package – Home. 2013. Disponível em: <<http://rgeos.free.fr/>>. Acesso em: 25 de junho 2013.

SEAB. Secretaria da Agricultura e do Abastecimento do Paraná. [S.l.]. 2013.

SHAHI, A.; ATAN, R. B.; SULAIMAN, N. B. **Efficient Decision Support Systems - Practice and Challenges From Current to Future**. Malaysia: InTech, 2011.

SILVA, A. C. D.; LIMA, É. P. C. D.; BATISTA, H. R. A IMPORTÂNCIA DA SOJA PARA O AGRONEGÓCIO BRASILEIRO, 2011.

Sun, Xiao-Lin; Zhao, Yu-Guo; Wang, Hui-Li; Yang, Lin; Qin, Cheng-Zhi ; Zhu, A-Xing; Zhang, Gan-Lin; Pei, Tao; Li, Bao-Lin. Sensitivity of digital soil maps based on FCM to the fuzzy exponent and the number of clusters. **Geoderma**, Pequim/China, v. 171-172, p. 24-34, 2012.

TAYLOR, J. A.; MCBRATNEY, A. B.; WHELAN, B. M. Establishing management classes for broadacre agricultural production. **Agronomy Journal**, Madison/USA, v. 99, p. 1366-1376, 2007.

THE R FOUNDATION FOR STATISTICAL COMPUTING. R: Software Development Life Cycle A Description of R's Development, Testing, Release and Maintenance Processes, 2012. Disponível em: <<http://www.r-project.org/doc/R-SDLC.pdf>>. Acesso em: 16 Junho 2013.

THOMAS, J. F. Ontogenetic and morphological plasticity in crop plants. **In: BOOTE, K.J. et al.(Comp.). Physiology and determinations of crop yield.**, Madison: ASA/CSSA/SSSA, p. 181-185, 1994.

UNWIN, A.; UNWIN, D. Spatial Data Analysis with Local Statistics. **Journal of the Royal Statistical Society: Series D (The Statistician)**, Londres/Inglaterra, v. 47, p. 415–421, 1998.

WAGNER, C.; MILLER, S.; GARIBALDI, J. M. **A fuzzy toolbox for the R programming language**. IEEE International Conference on. Taipei: [s.n.]. 2011. p. 1185 - 1192.

WANG, H.; FEI, B. A modified fuzzy C-means classification method using a multiscale diffusion filtering scheme. **Medical image analysis**, Cleveland/USA, 2009, p. 193-202.

YAN, L. et al. Delineation of site specific management zones using fuzzy clustering analysis in a coastal saline land. **Computers and Electronics in Agriculture**, p. 174-186, 2007.

YANG, M. S. **A Survey of Fuzzy Clustering**. Chung Yuan Christian University. Taiwan, v. 18, p. 1-16. 1993.

ZHU, W. et al. Clustering Algorithm Based on Fuzzy C-means and Artificial Fish Swarm. **Procedia Engineering**, Tongliao/China, v. 29, p. 3307–3311, 2012.

## APÊNDICES

## APÊNDICE A – ADQUIRINDO E UTILIZANDO O SOFTWARE R

O software R, estudado e utilizado neste trabalho, é um software estatístico grátis e aberto. Sua linguagem é totalmente orientada a objetos e fracamente tipada. Serão mostrados agora os passos necessários para se obter a distribuição adequada do R, como utilizá-la em conjunto com a interface R-Commander do pacote “Rcmdr”, como obter a documentação dos pacotes e funções, e como criar uma função.

### Instalação do R

Para fazer *download* do instalador R deve-se acessar o hiperlink CRAN (Comprehensive R Archive Network) do site, “<http://www.r-project.org>” como mostra a Figura 9.

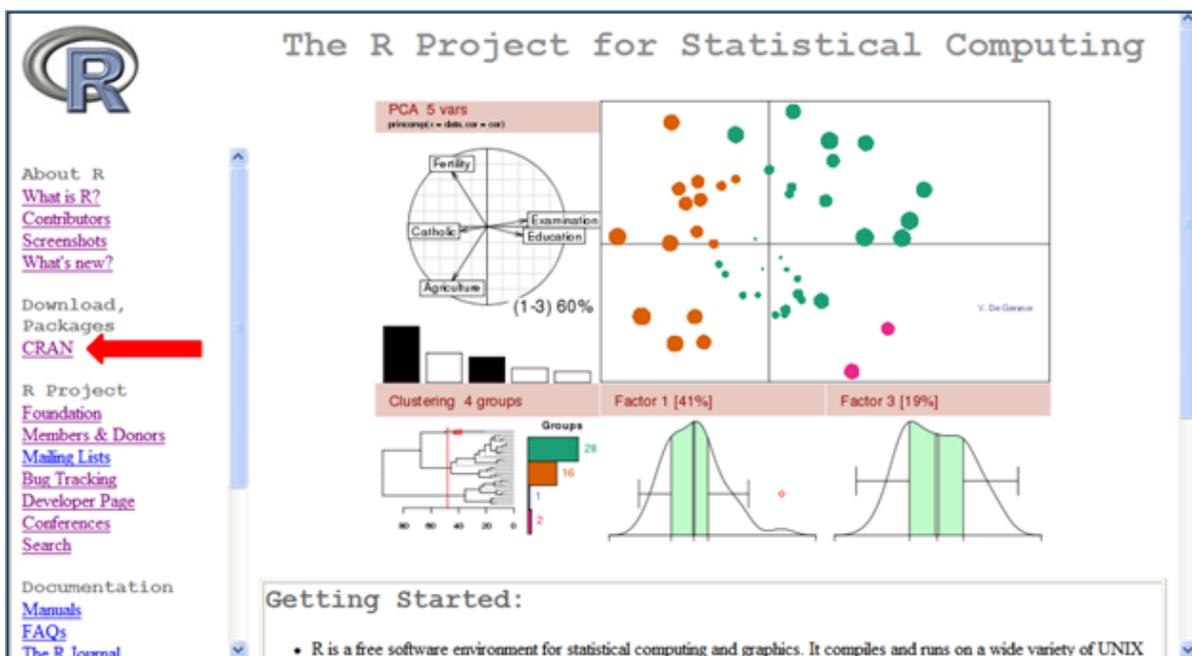


Figura 9 - Hiperlink CRAN do site r-project.

Deve se então procurar a seção “Brazil” e é recomendável que seja baixada a versão da UFPR, acessando o hiperlink correspondente, como mostra a Figura 10.

The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some statistics on the status of the mirrors can be found [here](#).

Country	URL	Institution
Australia	<a href="http://cran.ms.unimelb.edu.au/">http://cran.ms.unimelb.edu.au/</a>	University of Melbourne
Austria	<a href="http://cran.at-r-project.org/">http://cran.at-r-project.org/</a>	Wirtschaftsuniversitaet Wien
Belgium	<a href="http://www.freeststatistics.org/cran/">http://www.freeststatistics.org/cran/</a>	K.U.Leuven Association
Brazil	<a href="http://cran-r.c3sl.ufpr.br/">http://cran-r.c3sl.ufpr.br/</a>	Universidade Federal do Parana
	<a href="http://cran.fiocruz.br/">http://cran.fiocruz.br/</a>	Oswaldo Cruz Foundation, Rio de Janeiro
	<a href="http://www.vps.fmvz.usp.br/CRAN/">http://www.vps.fmvz.usp.br/CRAN/</a>	University of Sao Paulo, Sao Paulo
	<a href="http://brieger.esalq.usp.br/CRAN/">http://brieger.esalq.usp.br/CRAN/</a>	University of Sao Paulo, Piracicaba
Canada	<a href="http://cran.stat.sfu.ca/">http://cran.stat.sfu.ca/</a>	Simon Fraser University, Burnaby
	<a href="http://mirror.its.dal.ca/cran/">http://mirror.its.dal.ca/cran/</a>	Dalhousie University, Halifax
	<a href="http://probability.ca/cran/">http://probability.ca/cran/</a>	University of Toronto
	<a href="http://cran.skazkaforyou.com/">http://cran.skazkaforyou.com/</a>	iWeb, Montreal
	<a href="http://cran.parentingamerica.com/">http://cran.parentingamerica.com/</a>	iWeb, Montreal
Chile	<a href="http://dirichlet.mat.puc.cl/">http://dirichlet.mat.puc.cl/</a>	Pontificia Universidad Catolica de Chile, Santiago
China		

Figura 10 - Hiperlink da versão UFPR do software R do site r-project.

Após escolhida a distribuição é necessário escolher a versão correspondente ao seu sistema operacional. Neste caso foi utilizada a versão para Windows, como mostra a Figura 11.

The Comprehensive R Archive Network  
Frequently used pages

CRAN  
[Mirrors](#)  
[What's new?](#)  
[Task Views](#)  
[Search](#)

About R  
[R Homepage](#)  
[The R Journal](#)

Software  
[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Other](#)

Documentation  
[Manuals](#)  
[FAQs](#)  
[Contributed](#)

**Download and Install R**

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Linux](#)
- [MacOS X](#)
- [Windows](#)

**Source Code for all Platforms**

Windows and Mac users most likely want the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- **The latest release** (2010-10-15): [R-2.12.0.tar.gz](#) (read [what's new](#) in the latest version).
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features](#) and [bug fixes](#) before filing corresponding feature requests or bug reports.

Figura 11 - Hiperlink da versão para Windows do software R.

Logo após deve ser escolhida a distribuição base do R para download, essa distribuição consiste em um arquivo de instalação, as outras versões contém informações e

ferramentas mais avançadas do R, as quais não terão foco nesse estudo, como mostra a Figura 12.

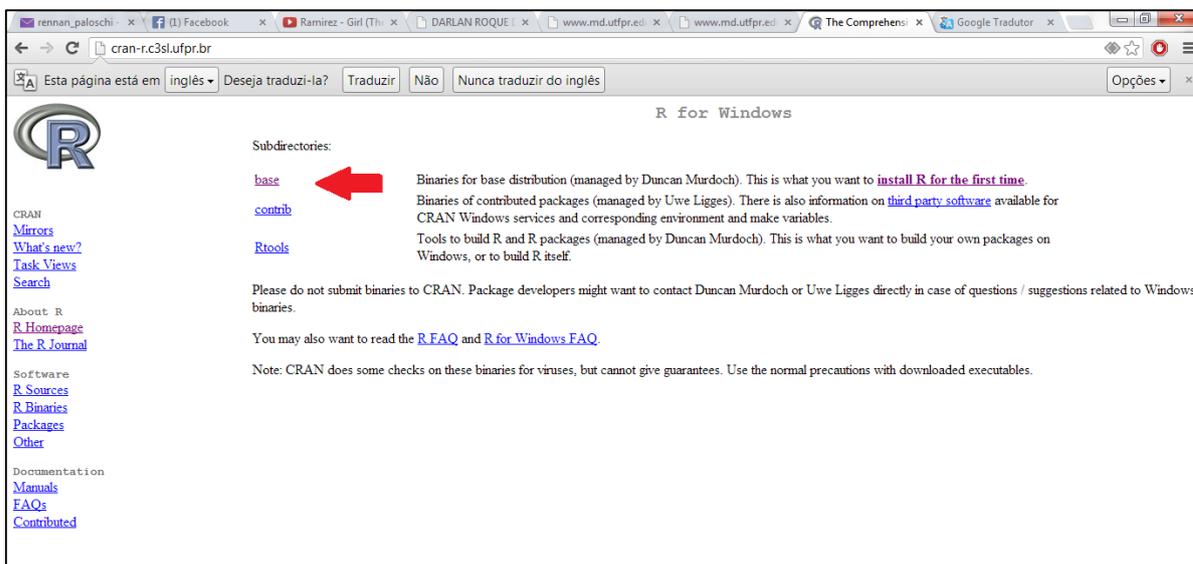


Figura 12 - Hiperlink da distribuição base do software R.

Somente então é disponibilizado o link de download da versão mais recente da configuração escolhida. Neste estudo a versão utilizada foi a 3.0.1. Deve ser feito download da distribuição acessando o link de download como mostra a Figura 13.

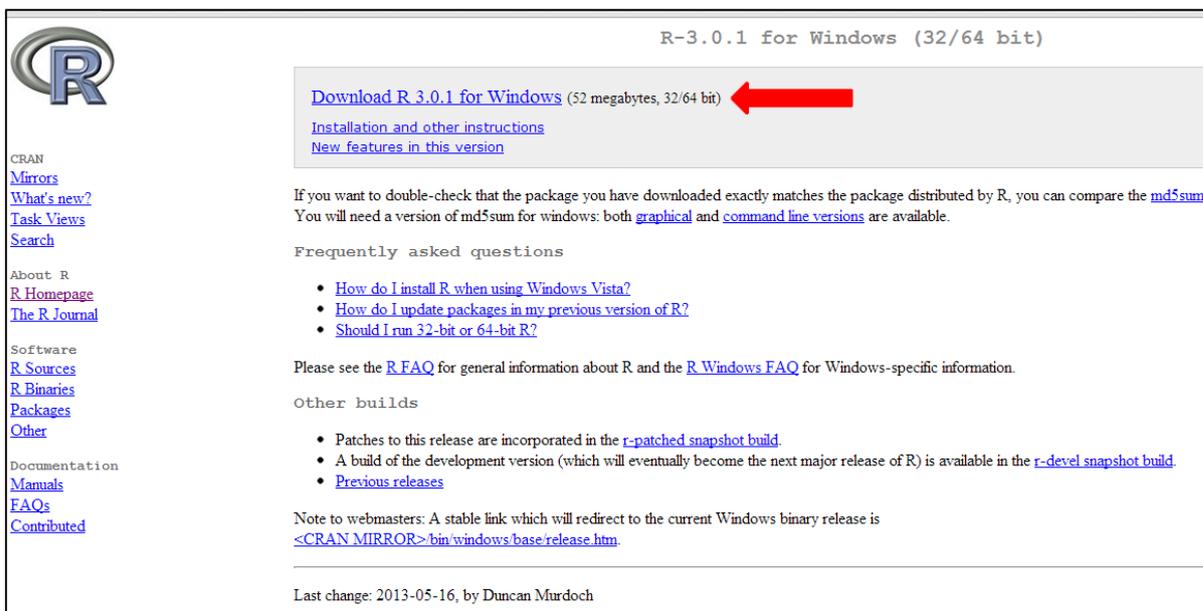
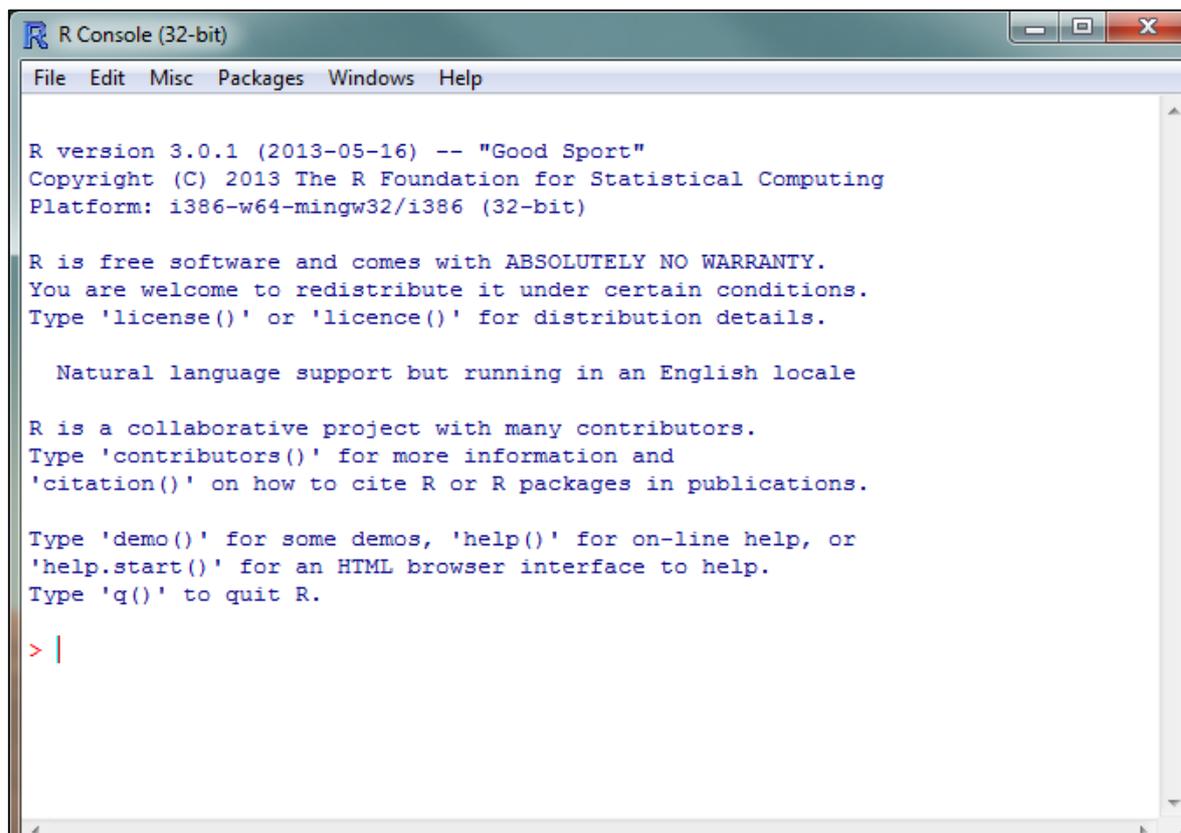


Figura 13 - Hiperlink de download do software R.

O executável pode ser copiado e executado a partir de qualquer diretório, este é um instalador. Sua instalação é rápida e não exige configurações adicionais, porém é interessante

mudar o idioma para português quando for ofertado a janela de idiomas, do contrário apenas é necessário avançar a instalação até seu término. A instalação criará um atalho para o software.

Ao executar o atalho, um console é aberto e a partir dele todo o trabalho já pode ser feito, como mostra a Figura 14.

The image shows a screenshot of the R Console (32-bit) window. The window title is "R Console (32-bit)" and it has a menu bar with "File", "Edit", "Misc", "Packages", "Windows", and "Help". The main area contains the following text:

```
R version 3.0.1 (2013-05-16) -- "Good Sport"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

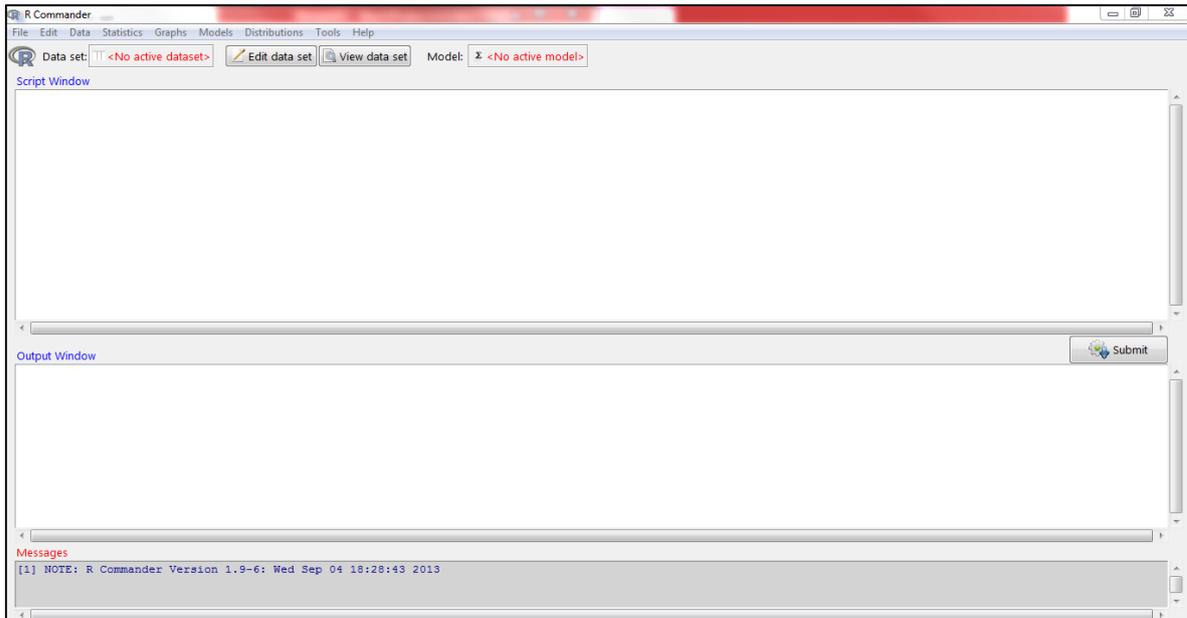
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

Figura 14 - R Console.

### Instalação do R-Commander e estrutura de pacotes

Para facilitar a edição dos códigos, neste estudo foi utilizado a interface estatística R-Commander. O R-Commander é um pacote disponível para o R que facilita a utilização do mesmo, para isto, ele fornece uma interface gráfica (janela). A interface fornecida contém uma janela *script* (editor de texto), também uma janela de resultados e uma janela para o retorno de mensagens, e um menu superior também em português que fornece diversos recursos estatísticos e gráficos, dentre outros. Essa interface é exibida na Figura 15.



**Figura 15 - Interface do R-Commander.**

Antes de obter a interface é necessário conhecer como funciona a estrutura de pacotes do software R.

Pacotes são estruturas que contêm diversas funções e códigos, podem ser chamadas de bibliotecas.

Ao instalar o R, diversos pacotes básicos são automaticamente baixados e instalados e ao iniciar o R, alguns pacotes ficam disponíveis para atualização.

Muitos pacotes, apesar de serem instalados automaticamente não são carregados para a memória do programa e assim suas funções e recursos não são disponíveis no console. Para carregar tais pacotes é necessário utilizar uma das funções a seguir:

- `library(nomeDoPacote)`
- `require(nomeDoPacote)`

Porém muitos pacotes não são instalados automaticamente, como o “Rcmdr”, que contém a interface R-Commander. Para baixar e instalar um pacote novo deve ser utilizado a função `install.packages()`.

O nome do pacote do R-Commander que já foi apresentado, é “Rcmdr”. Como dito anteriormente ele não é instalado automaticamente no R, portanto precisa ser instalado através do comando `install.packages("Rcmdr")` (o nome do pacote deve estar entre aspas já que representa um objeto *string* - objeto que contém texto).

Para se utilizar os seus recursos e ativar sua interface, o Rcmdr precisa ser carregado para a memória do R, através das funções `require(Rcmdr)` ou `library(Rcmdr)`.

Todos os códigos e funções foram desenvolvidos utilizando-se dessa interface (R-Commander).

### **Obter ajuda sobre o R**

Para obter a documentação de algum pacote ou uma função basta utilizar a função `help()` das seguintes formas:

- `help(Rcmdr)` – No caso do pacote ou função específica
- `help.search("clusters")` – No caso de procurar por uma função característica. Este recurso disponibiliza uma lista contendo o nome do pacote, nome da função e uma breve descrição da função.

Nem todos os pacotes vêm com uma descrição, mas todas as funções de todos os pacotes são devidamente descritas e documentadas. Mesmo os pacotes que não vêm automaticamente com sua descrição, possuem descrição, que pode ser facilmente encontrada em sites relacionados como o próprio site [cran.r-project.org](http://cran.r-project.org).

### **Criar função no R**

No software R uma função é um bloco executável que pode receber parâmetros e/ou devolver algum objeto como resposta. Para criar esse bloco executável basta criar uma referência e atribuir os comandos para a mesma, utilizando a palavra reservada “function” seguida dos parâmetros a serem recebidos entre parênteses, com o bloco logo a pós entre colchetes, como pode ser visto no Apêndice D.

Detalhes sobre a sintaxe do R, construção de vetores, objetos, listas, etc., podem ser obtidos em vários de sites a respeito, como no documento contido neste endereço <http://www.sph.umich.edu/csg/abecasis/class/815.04.pdf>.

## APÊNDICE B – ANOTAÇÕES

### **Leitura e plotagem do *shape*:**

O mapa virtual chamado de *shape*, contendo todos os municípios do Brasil, é composto por arquivo de extensão SHP, chamado de *shapefile*, que contém os dados espaciais do mapa, o *shape* pode vir acompanhado de mais arquivos, arquivos estes que contém os dados não espaciais do mapa, como os nomes dos municípios por exemplo, estes arquivos devem ter o mesmo nome apenas com as extensões diferentes, este *shape* ser encontrado no endereço <http://www.gismaps.com.br/english/shape.htm>.

Para a leitura do *shapefile* foi utilizado a função `readShapePoly()` do pacote `maptools` passando por parâmetro apenas o domínio do arquivo. O pacote `maptools`, pacote que contém ferramentas para a manipulação e leitura de dados geográficos, em particular *shapefiles*, por padrão não vem com o R, portanto é preciso instalá-lo e carregá-lo.

A função deve ser invocada (executada), passando por parâmetro o endereço completo (com o nome) do arquivo entre aspas, mas não é necessário informar a extensão, já que ao carregar o arquivo *shapefile* os outros arquivos contidos junto, de mesmo nome, como o arquivo DBF que contém informações sobre o *shapefile* (dados não espaciais), também são carregados.

Exemplo:

```
municipios <- readShapePoly("municipios_br")
```

Ao executar essa função, “municipios” passa a ser um objeto *shape*, caso a função tenha sucesso nenhuma resposta é retornada no console, do contrário aparecerá mensagens de erro.

Para plotar o mapa (imprimi-lo na tela), já que objeto `municipios` pode ser interpretado graficamente, basta utilizar a função `plot()`, passando por parâmetro o conjunto de dados que será plotado.

Exemplo:

```
plot(municipios)
```

Uma janela interna surgirá no ambiente do software R, e não no editor do R-Commander, contendo a imagem gráfica gerada.

A Figura 16 mostra o *shapefile* dos municípios do Brasil plotado.

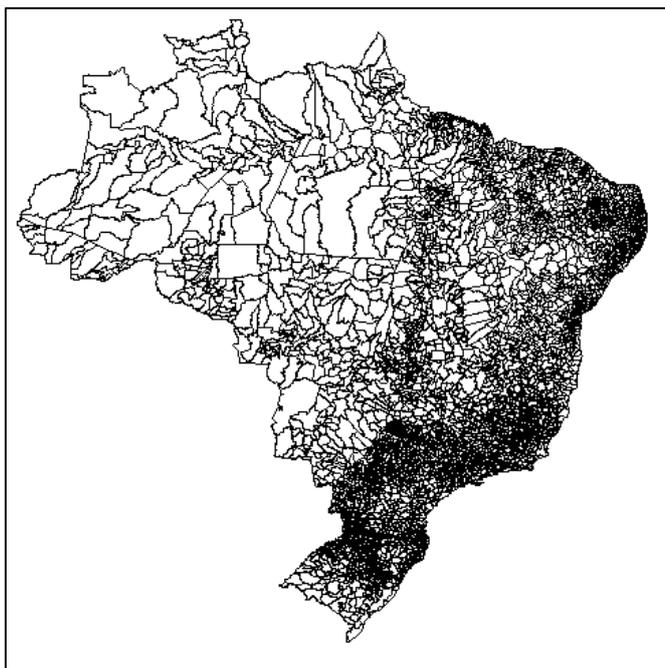


Figura 16 - Municípios plotados.

## Redução do espaço de amostragem

Para este estudo não é necessário manter todos os municípios do Brasil, sendo assim deve ser mantido, até para questões de desempenho, apenas os municípios de interesse, que são os municípios do estado do Paraná, Santa Catarina, São Paulo e Mato Grosso.

Para reduzir o número de polígonos, foi utilizado a função `subset(x, subset)` criando um subconjunto com apenas os municípios dos estados desejados, à qual foram passados alguns parâmetros:

- Parâmetro `x`: objeto que será subdividido.
- Parâmetro `subset`: expressão lógica que indica os elementos ou linhas que serão selecionados, valores resultantes em falso não serão selecionados.

Exemplo:

```
sub <- subset(municipios_br, NOMEUF=="PARANA" |
NOMEUF=="SANTA CATARINA" | NOMEUF=="SAO PAULO" | NOMEUF=="MATO
GROSSO DO SUL")
```

Informações sobre os operadores lógicos do software R pode ser vista em <http://www.statmethods.net/management/operators.html>.

A Figura 17 mostra o subconjunto dos estados PR, SC, MT e SP plotados.

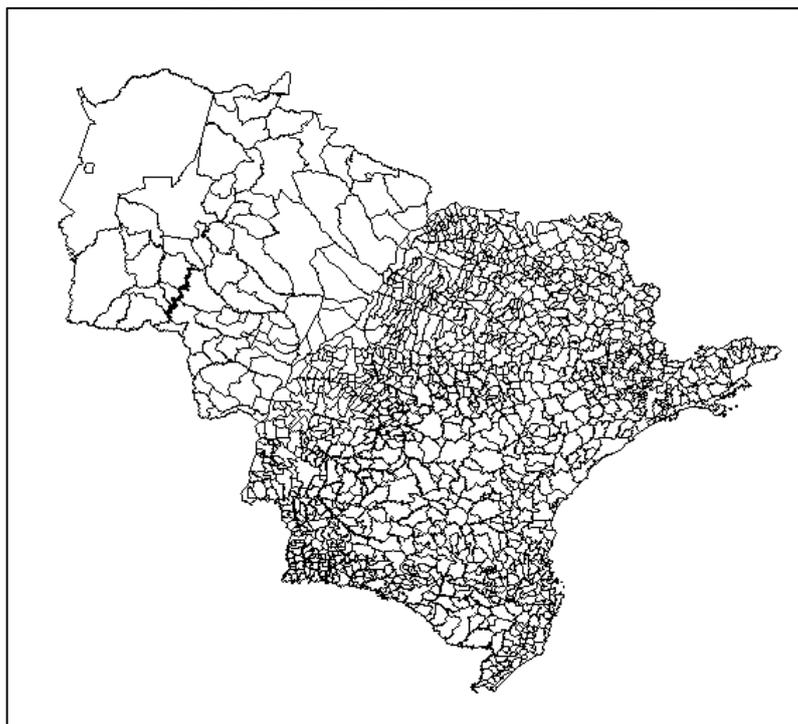


Figura 17 - Municípios do PR, SC, MT e SP.

### **Salvando o novo *shape***

Para salvar o novo *shape*, obtido anteriormente, no formato *shapefile* (SHP) foi utilizada a função `writePolyShape()`, a qual deve ser passado por parâmetro o objeto e em sequência o domínio, sem extensão, ao qual será gravado.

### **Selecionando áreas de interesse**

A área de interesse deste estudo é a região oeste do Paraná, mas também é necessário que se mantenha as áreas dos municípios que contém estações meteorológicas.

O *shape* criado e salvo anteriormente gerando um novo *shapefile*, contendo os municípios dos 4 estados (PR, SC, MT e SP) contém um arquivo DBF. O arquivo DBF do *shapefile*, que deve ter o mesmo nome do mesmo, contém todas as informações relativas aos polígonos, no caso os municípios.

Esse arquivo DBF foi editado adicionando-se uma nova coluna lógica (booleana), chamada “ESTACAO”, a qual foi realizado um *set* manual para T (*true*), à todas as linhas referentes aos municípios com existência de estação meteorológica. Desta forma foi possível

criar um novo subconjunto de polígonos, utilizando novamente a função `subset()`, que contém a região oeste do Paraná e também as estações meteorológicas dos quatro estados. Porém é possível utilizar o função `subset()`, passando na expressão lógica o nome de todas as cidades desejadas.

Para realizar esse *set* é necessário recuperar o arquivo DBF do *shape* salvo. Para isso o pacote “foreign” faz-se necessário, é dele a função `read.dbf()` utilizada para leitura do arquivo, portando é necessário instalá-lo e carrega-lo.

Diferentemente da função `readShapepoly()` a função `read.dbf()` necessita que seja informada a extensão do arquivo (DBF).

Exemplo:

```
dados <- read.dbf("C:/R/shapes/municipios/shape.dbf")
```

Após a execução da função a interface do R-Commander deve ser configurada para identificar o conjunto de dados que será editado clicando em “<No active dataset>”, como mostra a Figura 18. Caso tenha escolhido a configuração em língua portuguesa na instalação do R, os menus aparecerão também em português.

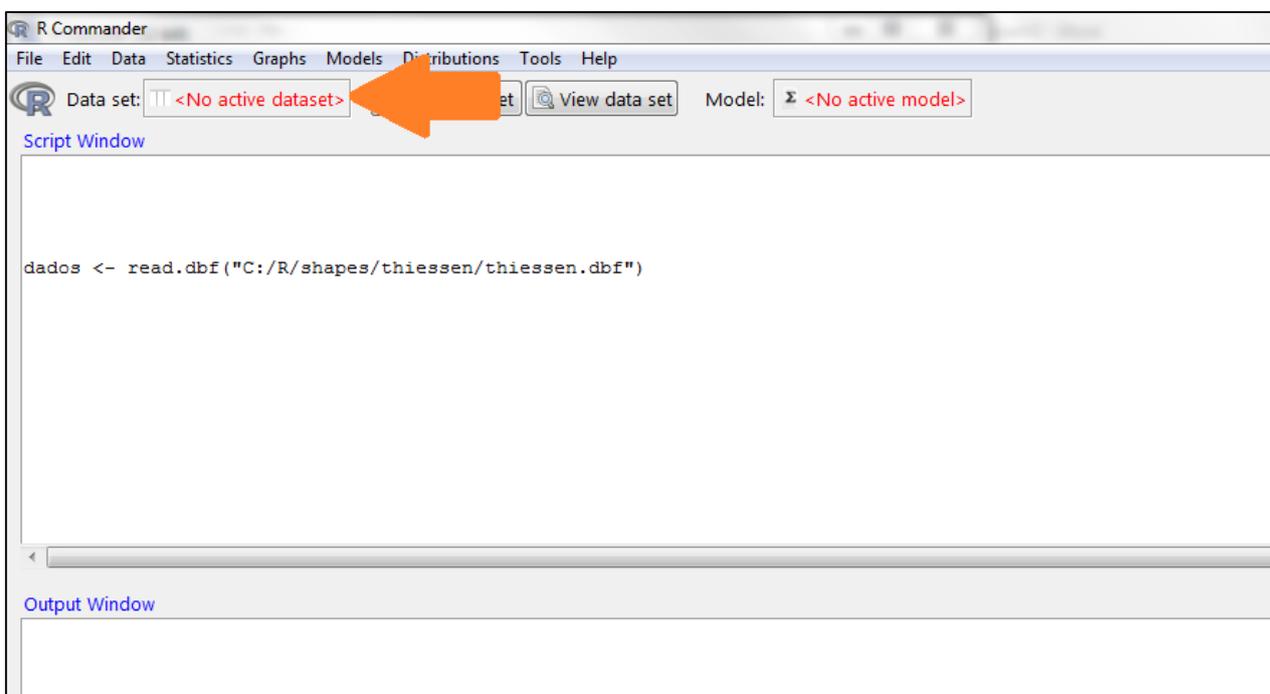
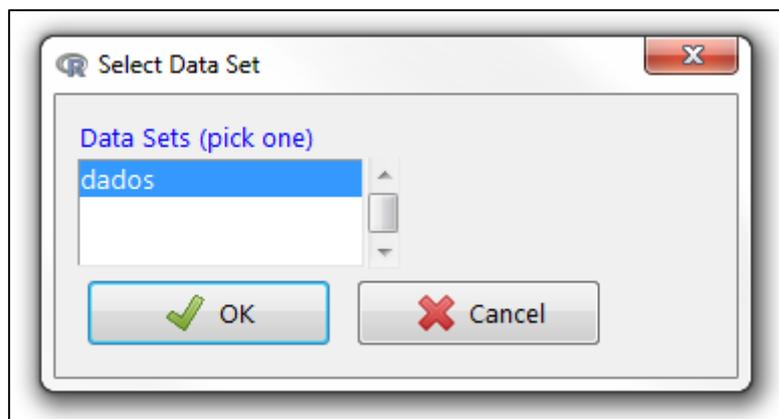


Figura 18 - Botão "<No active dataset>" do R-Commander.

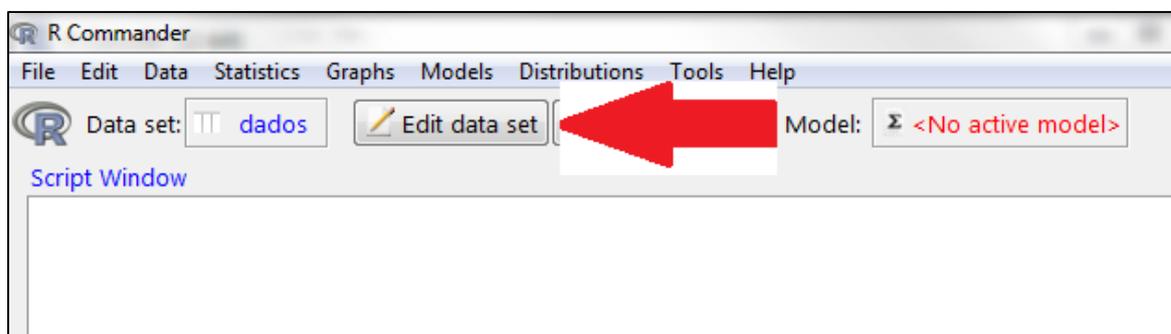
Será exibida uma janela contendo todos os objetos Data gerados, estes objetos podem ser gerados por diversas funções a partir de diversos tipos de arquivos, até mesmo serem construídos no próprio console, neste caso ele foi obtido pela função `read.dbf()`. Nesta

janela estarão todos estes objetos Data que estão carregados em memória para que um seja escolhido, como mostra a Figura 19.



**Figura 19 - Janela para escolha do objeto Data.**

Logo escolhido o objeto já é possível fazer tanto a visualização como edição dos dados. Para editar o conjunto é necessário clicar no botão “Edit data Set” como mostra a Figura 20.



**Figura 20 - Botão “Edit data set” do R-Commander.**

Uma janela para a edição do conjunto de dados será aberta, através dela todos os dados do *shape* já podem ser editados como mostra a Figura 21. Para salvar, basta fechar a janela. Vale lembrar que o R-Commander, do pacote “Rcmdr”, atua apenas como uma interface, sendo assim as janelas de edição e gráficos gerados, podem ser invocados por ele, mas não são gerados por ele e sim pelos outros pacotes do R, sua maioria nativos.

SP_ID	ID	NOMEUF	REGIAO	CODMESO	NOMEMESO	CODMICRO	NOMEMICRO	NOME_1	
1	2690	2761	MATO GROSSO DO SUL	CENTRO-OESTE	01	FANTANAL SUL MATO-GROSSENSE	01	BAIXO FANTANAL	CORUMBA
2	2699	2770	MATO GROSSO DO SUL	CENTRO-OESTE	02	CENTRO NORTE DE MATO GROSSO DO SUL	03	ALTO TAQUARI	SONORA
3	2740	2811	MATO GROSSO DO SUL	CENTRO-OESTE	02	CENTRO NORTE DE MATO GROSSO DO SUL	03	ALTO TAQUARI	PEDRO GOMES
4	2742	2813	MATO GROSSO DO SUL	CENTRO-OESTE	02	CENTRO NORTE DE MATO GROSSO DO SUL	03	ALTO TAQUARI	COXIM
5	2765	2836	MATO GROSSO DO SUL	CENTRO-OESTE	02	CENTRO NORTE DE MATO GROSSO DO SUL	03	ALTO TAQUARI	ALCINOPOLIS
6	2784	2855	MATO GROSSO DO SUL	CENTRO-OESTE	03	LESTE DE MATO GROSSO DO SUL	05	CASSILANDIA	COSTA RICA
7	2806	2877	MATO GROSSO DO SUL	CENTRO-OESTE	02	CENTRO NORTE DE MATO GROSSO DO SUL	03	ALTO TAQUARI	RIO VERDE DE MATO GROSSO
8	2851	2923	MATO GROSSO DO SUL	CENTRO-OESTE	02	CENTRO NORTE DE MATO GROSSO DO SUL	03	ALTO TAQUARI	CAMAFUA
9	2874	2946	MATO GROSSO DO SUL	CENTRO-OESTE	01	FANTANAL SUL MATO-GROSSENSE	02	AQUIDAUANA	AQUIDAUANA
10	2881	2953	MATO GROSSO DO SUL	CENTRO-OESTE	03	LESTE DE MATO GROSSO DO SUL	05	CASSILANDIA	CHAPADA DO SUL
11	2896	2968	MATO GROSSO DO SUL	CENTRO-OESTE	02	CENTRO NORTE DE MATO GROSSO DO SUL	03	ALTO TAQUARI	SAO GABRIEL DO OESTE
12	2900	2972	MATO GROSSO DO SUL	CENTRO-OESTE	03	LESTE DE MATO GROSSO DO SUL	05	CASSILANDIA	CASSILANDIA
13	2944	3016	MATO GROSSO DO SUL	CENTRO-OESTE	01	FANTANAL SUL MATO-GROSSENSE	01	BAIXO FANTANAL	LADARIO
14	2966	3038	MATO GROSSO DO SUL	CENTRO-OESTE	03	LESTE DE MATO GROSSO DO SUL	07	TRES LAGOAS	AGUA CLARA
15	2967	3039	MATO GROSSO DO SUL	CENTRO-OESTE	03	LESTE DE MATO GROSSO DO SUL	06	PARANAIBA	PARANAIBA
16	2975	3047	MATO GROSSO DO SUL	CENTRO-OESTE	02	CENTRO NORTE DE MATO GROSSO DO SUL	04	CAMPO GRANDE	RIO NEGRO
17	2976	3048	MATO GROSSO DO SUL	CENTRO-OESTE	03	LESTE DE MATO GROSSO DO SUL	06	PARANAIBA	INOCENCIA
18	3024	3096	MATO GROSSO DO SUL	CENTRO-OESTE	02	CENTRO NORTE DE MATO GROSSO DO SUL	04	CAMPO GRANDE	CORSUINHO
19	3038	3110	MATO GROSSO DO SUL	CENTRO-OESTE	03	LESTE DE MATO GROSSO DO SUL	07	TRES LAGOAS	RIBAS DO RIO PARDO
20	3040	3112	MATO GROSSO DO SUL	CENTRO-OESTE	02	CENTRO NORTE DE MATO GROSSO DO SUL	04	CAMPO GRANDE	BANDEIRANTES
21	3062	3134	MATO GROSSO DO SUL	CENTRO-OESTE	03	LESTE DE MATO GROSSO DO SUL	07	TRES LAGOAS	TRES LAGOAS
22	3063	3135	MATO GROSSO DO SUL	CENTRO-OESTE	01	FANTANAL SUL MATO-GROSSENSE	02	AQUIDAUANA	MIRANDA
23	3087	3159	MATO GROSSO DO SUL	CENTRO-OESTE	02	CENTRO NORTE DE MATO GROSSO DO SUL	04	CAMPO GRANDE	ROCHEDO
24	3098	3170	SAO PAULO	SUDESTE	01	SAO JOSE DO RIO PRETO	01	JALES	POPULINA
25	3101	3173	SAO PAULO	SUDESTE	01	SAO JOSE DO RIO PRETO	02	FERNANDOPOLIS	OUROESTE
26	3103	3175	MATO GROSSO DO SUL	CENTRO-OESTE	03	LESTE DE MATO GROSSO DO SUL	06	PARANAIBA	APARECIDA DO TABOADO
27	3121	3193	SAO PAULO	SUDESTE	01	SAO JOSE DO RIO PRETO	01	JALES	MESOPOLIS
28	3123	3195	SAO PAULO	SUDESTE	01	SAO JOSE DO RIO PRETO	02	FERNANDOPOLIS	INDIAPORA
29	3125	3197	SAO PAULO	SUDESTE	01	SAO JOSE DO RIO PRETO	02	FERNANDOPOLIS	MIRA ESTRELA
30	3132	3204	SAO PAULO	SUDESTE	01	SAO JOSE DO RIO PRETO	04	SAO JOSE DO RIO PRETO	PAULO DE FARIA
31	3133	3205	SAO PAULO	SUDESTE	01	SAO JOSE DO RIO PRETO	01	JALES	SANTA ALBERTINA
32	3134	3206	MATO GROSSO DO SUL	CENTRO-OESTE	03	LESTE DE MATO GROSSO DO SUL	06	PARANAIBA	SELVIRIA

Figura 21 - Janela de edição de conjuntos de dados do R.

A partir dessa janela toda a edição pode ser feita. Existem programas que oferecem melhores recursos para esse tipo de edição como o DBF Viewer & Editor.

Na Figura 22 pode ser visto o mapa formado apenas pelas cidades que contém estações meteorológicas.

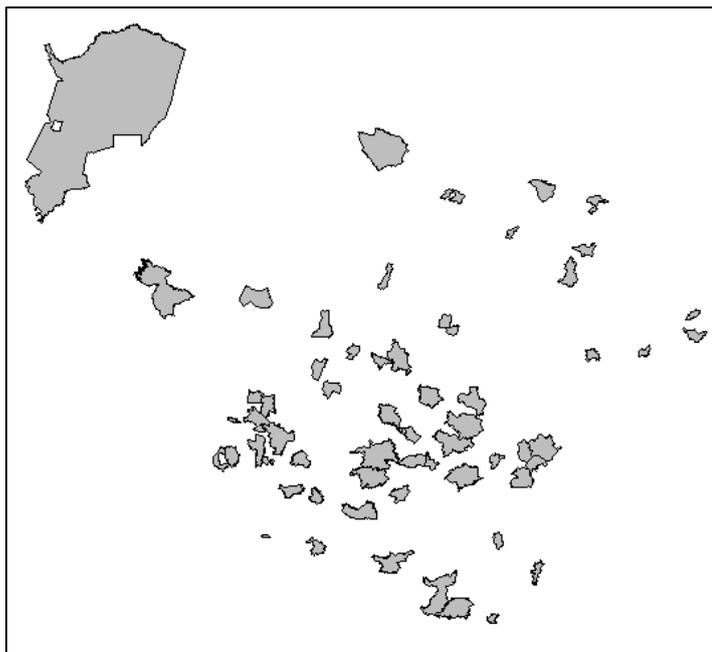


Figura 22 - Estações meteorológicas.

Adicionando o Oeste do Paraná é obtido mapa que pode ser observado na Figura 9, com as cidades do Paraná que não contém estações destacadas em amarelo. Os parâmetros adicionais para as opções de cor, transparência do *shape* ou cor da linha e muitas outras podem ser vistas na documentação da função `plot()`. Para fazer uma plotagem sobreposta como a da Figura 23, basta adicionar o parâmetro “`add=TRUE`” (sem aspas). Pode-se observar que no R ao utilizar-se desse tipo de atribuição, não é necessário que os parâmetros estejam em ordem, já que ocorre um apontamento direto para a referência desejada, no caso a “`add`” que passa a apontar para o objeto lógico `TRUE` de valor verdadeiro. A última plotagem fica sempre sobreposta.

Exemplo:

```
plot(oeste, col="Yellow")  
plot(estacoes, col="Gray", add=TRUE)
```

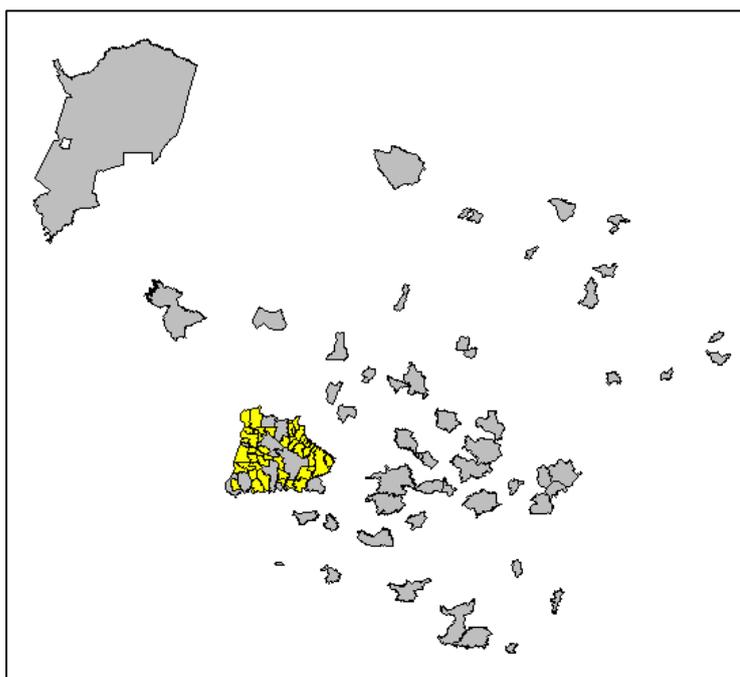


Figura 23 - Oeste do Paraná e estações.

### Calculando Área de Influência das Estações Meteorológicas

Para se calcular a área de influência das estações meteorológicas sobre os outros municípios que não as possuem, foram obtidos os polígonos de Thiessen. Para isto é necessário obter o centro exato de cada polígono (centroide).

## Calculando centroides

Para calcular o centroide dos polígonos foi utilizado a função `gCentroid()` do pacote “`rgeos`”, o qual deve ser instalado e carregado em memória para ser utilizado.

A função `points()` apresenta os pontos, o parâmetro “`spgeom`” indica o objeto que representa o *shape* com os polígonos, o “`pch`” indica o tipo de figura a ser mostrada e o “`col`” a cor que ela terá.

Exemplo:

```
centroides = gCentroid(spgeom=estacoes, byid=TRUE)
plot(estacoes, col="grey")
points(centroides, pch=1, col = "blue" )
points(centroides, pch=3, col = "Red")
```

Na Figura 24 pode ser observado o resultado gerado, mostrando os municípios que possuem estações em cinza e seus respectivos centroides em vermelho.

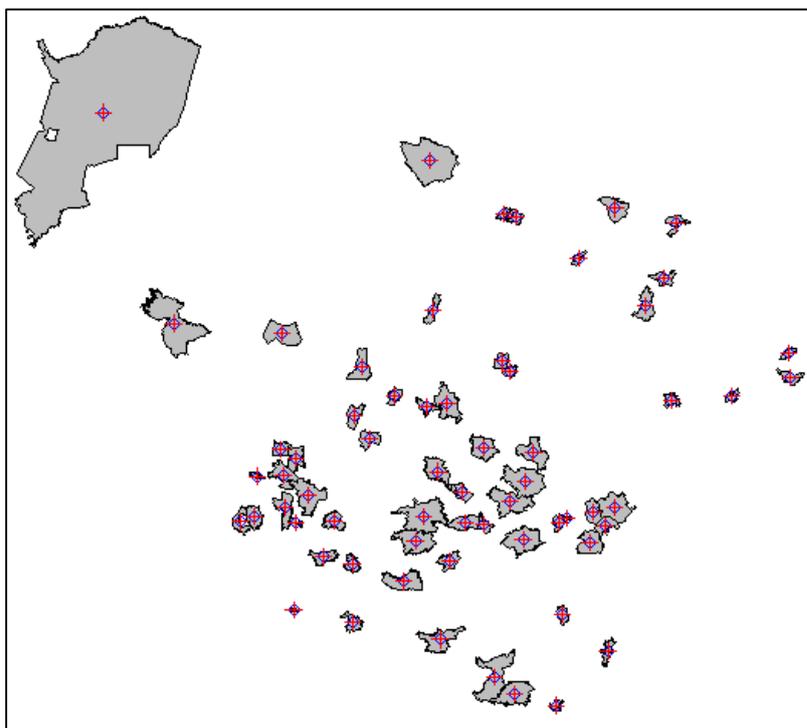


Figura 24 - Centroides.

## Thiessen – Voronoi

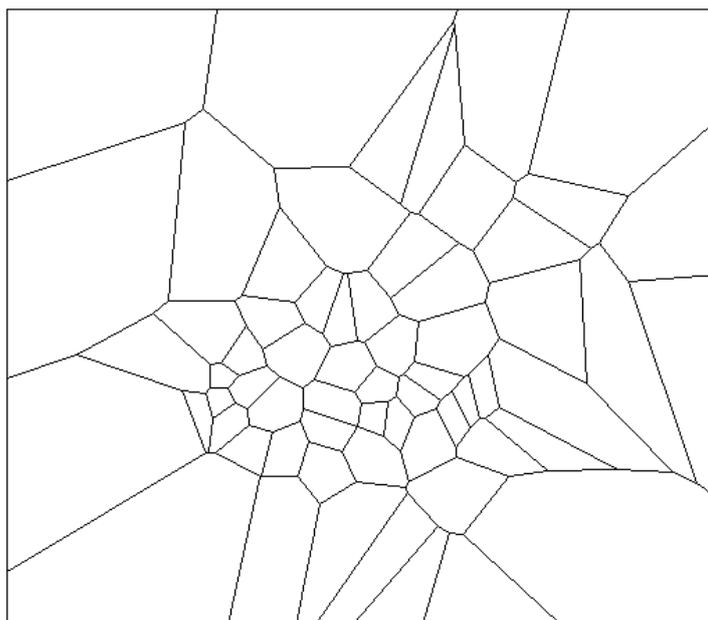
Para calcular os polígonos de Thiessen a partir das coordenadas foi definida e utilizada a função de Voronoi (`VoronoiPolygons()`, vide Apêndice C). Para que a mesma funcione é necessário que o pacote “deldir” seja instalado.

Aplicando a função de Voronoi sobre as coordenadas obtidas dos polígonos das cidades que contêm estações meteorológicas é obtido os polígonos de Thiessen.

Exemplo:

```
cents <- coordinates(municipios)
thiessen <- voronoiPolygons(cents, municipios)
```

Os polígonos gerados podem ser observados na Figura 25.



**Figura 25- Polígonos de Thiessen.**

Sobrepondo as imagens como na Figura 26, pode ser visualizado os municípios que contêm estações em cinza, com seus respectivos centroides em vermelho, e o restante dos municípios do Oeste do Paraná em amarelo, tudo isso sob os polígonos de Thiessen.

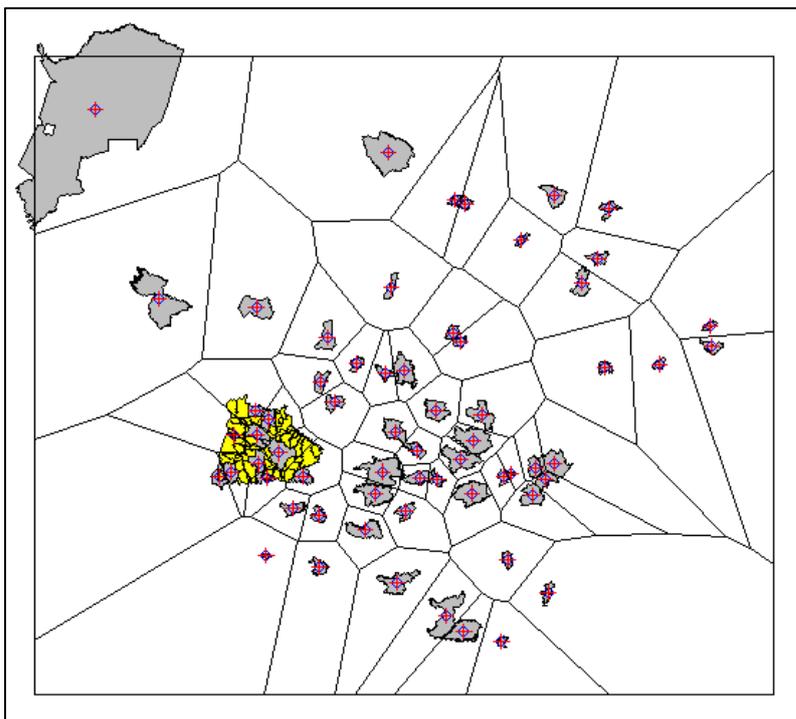


Figura 26 - Mapas sobrepostos.

### Recuperando atributos

A função de Voronoi utilizada devolve um objeto do tipo `SpatialPolygonsDataFrame`, porém todos os atributos do *shape* antigo são perdidos já que o novo *shape* Thiessen é construído apenas baseado nos vetores de coordenadas x e y, sendo assim nem o atributo identificador permanece, o algoritmo possui um `for` que cria uma nova numeração identificadora, portanto é necessário salvar o novo *shape* e novamente alterar o arquivo DBF, colocando os atributos identificadores corretos.

Com os atributos identificadores iguais, ou seja igualando os objetos, já há portanto, uma maneira de diferenciar qual cidade corresponde a qual polígono no outro *shape* e vice-versa, dessa forma podemos construir um novo objeto do tipo `SpatialPolygonsDataFrame`, passando o atual objeto criado pela função de Voronoi, mas informando, por parâmetro, os dados (*data*) do outro *shape*.

Dessa forma:

```
estacoes <- readShapePoly("C:/shapes/Estacoes/estacoes")
thiessen <- readShapePoly("C:/shapes/thiessen/thiessen")
thiessen <- SpatialPolygonsDataFrame ( thiessen , data =
estacoes@data )
```

O construtor do objeto `SpatialPolygonsDataFrame` automaticamente identificará a informação correspondente de cada polígono utilizando o atributo identificador.

### Diminuindo espaço de amostragem

Pode-se observar na Figura 27, as estações que ficaram longe da área de influência, e os polígonos calculados por essas estações também. Como o objetivo é calcular a influência apenas no oeste do Paraná, todas as estações que não atingiram o oeste devem ser retiradas.

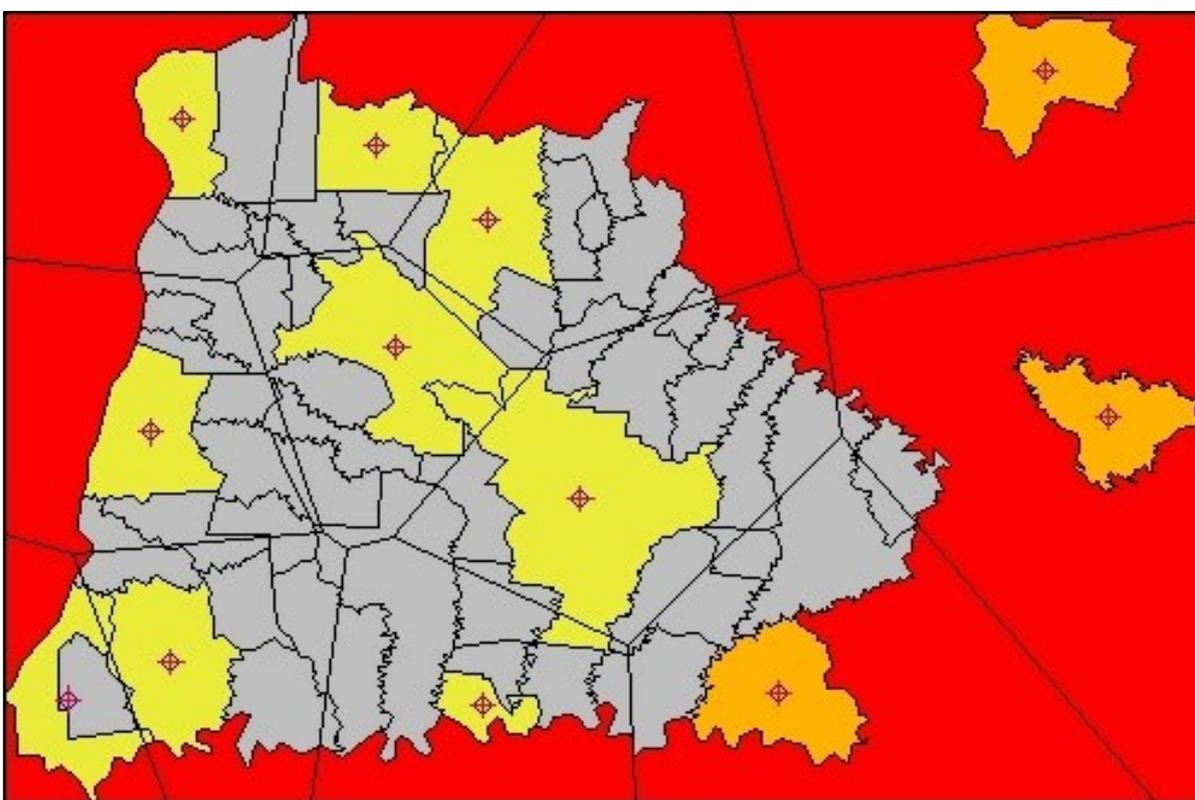
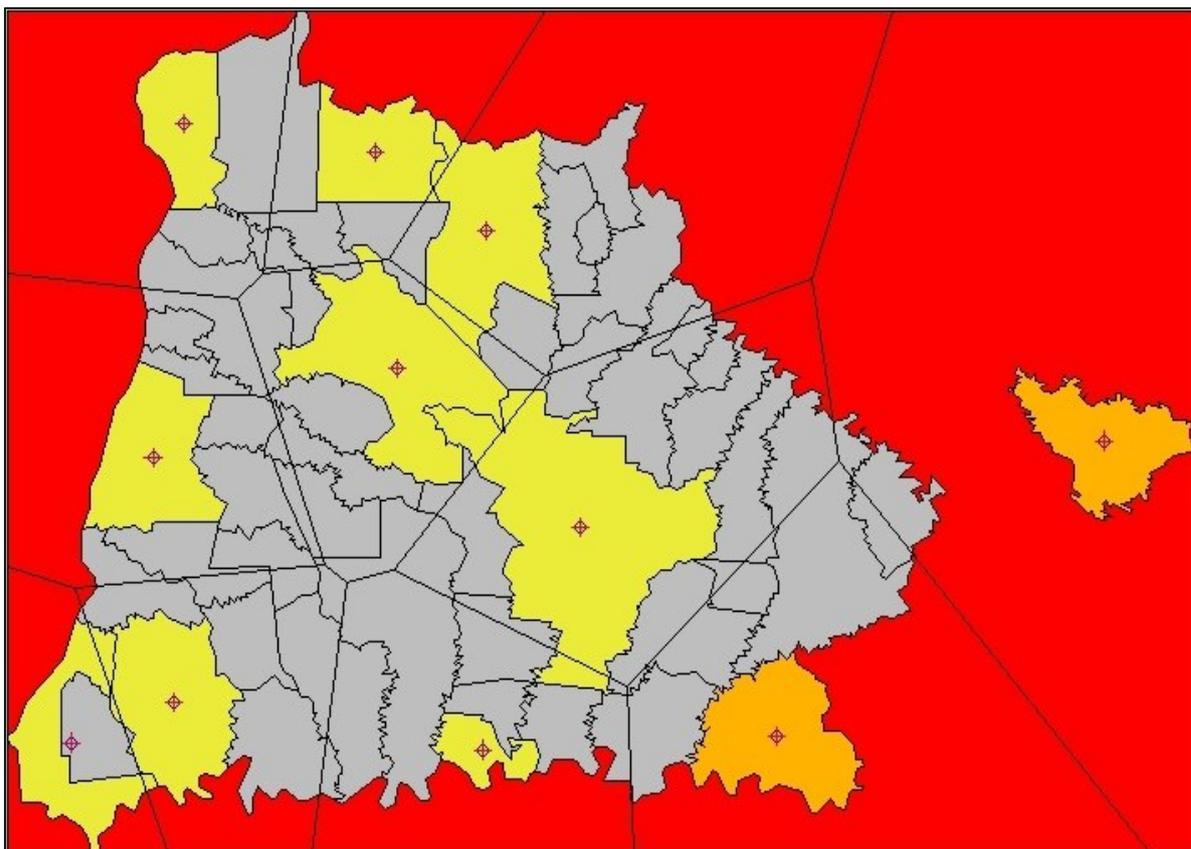


Figura 27 - Oeste do paran e estaoes.

### Filtrando polgonos

Para uma melhor visualizaao todo o processo foi repetido, mas desta vez, considerando apenas as estaoes cuja suas reas de atuaao incidem sobre a regiao oeste. O codigo utilizado foi codigo para a filtragem dos mapas de municpios (Apndice D, Quadro 14). A Figura 13 mostra o resultado gerado, o oeste em cinza, as cidades com estaoes meteorolgicas fora do oeste em laranja, e as dentro do oeste em amarelo.



**Figura 28 - Ajustando Polígonos.**

O novo *shape* contido no objeto obtido pelo código de Voronoi ainda não possui as informações dos municípios. Para isso é necessário modificar novamente os IDs e atribuí-los, o código utilizado foi o código para a transferência dos dados meteorológicos (Apêndice D, Quadro 17).

A partir deste estágio todo o processo está devidamente detalhado no capítulo 2 desta monografia e os códigos utilizados em cada processo encontram-se no Apêndice D.

## APÊNDICE C – FUNÇÕES UTILIZADAS

- Função CalcularGP: calcula Grau de Pertinência (GP) do cluster A no cluster B;

Quadro 1 – Função CalcularGP.

```

CalcularGP <- function(UMatrix, ANumber, BNumber) {
  diferencaClusters <- matrix(0, 1, length(UMatrix[,1]));
  gp <- 0;
  for(i in 1:length(UMatrix[,ANumber])){
    diferencaClusters[1,i]=UMatrix[i, ANumber] - UMatrix[i, BNumber];
    diferencaClusters[1,i] = max(0, diferencaClusters[1,i]);
    gp=gp+diferencaClusters[1,i];
  }
  gp=1/Cardinalidade(UMatrix,ANumber)*(Cardinalidade(UMatrix,ANumber)-gp);
  return(gp)
}
Cardinalidade <- function(UMatrix, vector ) {
  somaGrausDePertinencia <- 0;
  for(i in 1:length(UMatrix[,vector])){
    somaGrausDePertinencia <- somaGrausDePertinencia + UMatrix[i, vector];
  }
  c <- somaGrausDePertinencia;
  return(c)
}

```

- Função `CalcularIndices`: retorna um objeto com vetores, um para cada agrupamento, contendo os índices representando qual elemento pertence a determinado agrupamento;

#### Quadro 2 – Função `CalcularIndices`.

```

CalcularIndices <- function(pertinencia_dados) {
indices <- list(c(),c(),c(),c(),c(),c(),c(),c())
names(indices) <- c("c1", "c2", "c3", "c4")
for(i in 1:length(pertinencia_dados$membership[,1])){
  maior_grau <- 0
  cluster <- 0
  indice <- 0
  for(ii in 1:length(pertinencia_dados$membership[1,])){
    pertinencia <- pertinencia_dados$membership[i , ii]
    if(maior_grau < pertinencia){
      maior_grau <- pertinencia
      #dados$CLUSTER[i] <- ii
      cluster <- ii
      indice <- i
    }
  }
  indices[[cluster]] <- c(indices[[cluster]], indice)
}
return(indices)
}

```

- Função `CalcularISC`: calcula o índice ISC;

#### Quadro 3 – Função `CalcularISC`.

```

CalcularISC <- function(matrizU, clusterNumero, clusterIndices) {
somaticaK <- 0;
for (k in 1:length(clusterIndices)){
  somaticaK <- somaticaK + matrizU[clusterIndices[k], clusterNumero]
}
ISC <- somaticaK/length(clusterIndices)
return(ISC)
}

```

- Função CalcularMPE: calcula o índice MPE;

#### Quadro 4 – Função CalcularMPE.

```

CalcularMPE <- function(cNumeroClusters, nDataPoints, uMatriz) {
  uikSum = 0;
  for (i in 1:cNumeroClusters){
    for (k in 1:nDataPoints){
      uikSum <- uikSum + (uMatriz[k,i]*log(uMatriz[k,i])/nDataPoints)
    }
  }
  mpe <- (uikSum*-1)/log(cNumeroClusters)
  return(mpe)
}

```

- Função Padronizar: padroniza padronizar os dados entre 0 e 1;

#### Quadro 5 – Função Padronizar.

```

Padronizar <- function(dados) {
  dados_padronizados <- dados
  for(i in 1:length(dados)){
    vetor <- dados[,i]
    max <- Maximo(vetor)
    min <- Minimo(vetor)
    for(ii in 1:length(vetor)){
      dados_padronizados[ii,i] <- (vetor[ii]-min) / (max-min)
    }
  }
  return(dados_padronizados)
}

```

- Funções Máximo e Mínimo: retornam o maior e menor valor, respectivamente;

Quadro 6 – Funções Máximo e Mínimo.

```
Maximo <- function(x) {  
    maximo <- x[1]  
    for(i in 2:length(x)){  
        if(x[i]>maximo)  
            maximo <-x[i]  
    }  
    return( maximo )  
}  
Minimo <- function(x) {  
    minimo <- x[1]  
    for(i in 2:length(x)){  
        if(x[i]<minimo)  
            minimo <-x[i]  
    }  
    return( minimo )  
}
```

- Função Voronoipolygons: calcula os polígonos de Thiessen a partir de coordenadas geográficas.

#### Quadro 7 – Função voronoipolygons.

```

voronoipolygons <- function(x,poly) {
  require(deldir)
  if (.hasSlot(x, 'coords')) {
    crds <- x@coords
  } else crds <- x
  bb = bbox(poly)
  rw = as.numeric(t(bbox(poly)))
  z <- deldir(crds[,1], crds[,2],rw=rw)
  w <- tile.list(z)
  polys <- vector(mode='list', length=length(w))
  require(sp)
  for (i in seq(along=polys)) {
    pcrds <- cbind(w[[i]]$x, w[[i]]$y)
    pcrds <- rbind(pcrds, pcrds[1,])
    polys[[i]] <- Polygons(list(Polygon(pcrds)), ID=as.character(i))
  }
  SP <- SpatialPolygons(polys)
  voronoi <- SpatialPolygonsDataFrame(SP, data=data.frame(x=crds[,1],
                                                         y=crds[,2], row.names=sapply(slot(SP, 'polygons'),
                                                         function(x) slot(x, 'ID'))))
  return(voronoi)
}

```

## APÊNDICE D – CÓDIGOS UTILIZADOS

- Cálculo dos dados meteorológicos dos municípios vizinhos as estações meteorológicas distribuindo fazendo a distribuição dos atributos em relação a área de influência das estações;

Quadro 8 – Código do cálculo dos dados meteorológicos.

```

require(foreign)
require(rgeos)
require(maptools)

data <- read.dbf("C:/R/shapes2/m_estacoes_sobrepostas_oeste cdM/m_estacoes_sobrepostas_oeste.dbf")
oeste <- readShapePoly("C:/R/shapes/regiao_oeste_pr/regiao_oeste_pr")
areas_estacoes <- readShapePoly("C:/R/shapes2/t_estacoes_sobrepostas_cdM/t_estacoes_sobrepostas_cdM2")

for (ii in seq(along=oeste)){
  m <- oeste[ii,]
  m_area <- gArea(m)
  temp <- 0
  radia <- 0
  precip <- 0

  for (i in seq(along=areas_estacoes)){
    a <- areas_estacoes[i,]
    if(gIntersects(m, a)){
      intersection <- gIntersection(m,a)
      i_area <- gArea(intersection)
      porcentagem <- i_area/m_area

      temp <- temp + porcentagem * as.numeric(a@data[,"TEMP"])
      radia <- radia + porcentagem * as.numeric(a@data[,"RADIA"])
      precip <- precip + porcentagem * as.numeric(a@data[,"PRECIP"])

      data[ii,"TEMP"] <- temp
      data[ii,"RADIA"]<- radia
      data[ii,"PRECIP"]<- precip
    }
  }
}
write.dbf(data, "C:/R/shapes2/m_estacoes_sobrepostas_oeste cdM/m_estacoes_sobrepostas_oeste2.dbf")

```

- Cálculo das estatísticas preliminares dos atributos PROD, PRECM, TMED e RSGM, as quais são a média, mediana, desvio padrão, coeficiente de variação, valor mínimo e valor máximo, mostradas na Tabela 1;

Quadro 9 – Código para o cálculo das estatísticas preliminares.

```

require(foreign) # pacote usado para leitura, edição e gravação do arquivo DBF
require(e1071) # pacote para cmeans
dados <- read.dbf("C:/R/Dados/dados.dbf")
dados_padronizados <- Padronizar(dados)
d<-dados_padronizados
d<-dados
#----- Média -----
mean(d$PRODUCAO)
mean(d$PRECIP)
mean(d$TEMP)
mean(d$RADIA)
#----- Desvio Padrão -----
sd(d$PRODUCAO)
sd(d$PRECIP)
sd(d$TEMP)
sd(d$RADIA)
#----- Coeficiente de Variância -----
var(d$PRODUCAO) * 100 /mean(d$PRODUCAO)
var(d$PRECIP)* 100 /mean(d$PRECIP)
var(d$TEMP)* 100 /mean(d$TEMP)
var(d$RADIA)* 100 /mean(d$RADIA)
#----- Minimo -----
min(d$PRODUCAO)
min(d$PRECIP)
min(d$TEMP)
min(d$RADIA)
#----- Maximo -----
max(d$PRODUCAO)
max(d$PRECIP)
max(d$TEMP)
max(d$RADIA)
#----- Mediana -----
median(d$PRODUCAO)
median(d$PRECIP)
median(d$TEMP)
median(d$RADIA)

```

- Análise dos Clusters. Essa análise envolve o cálculo de MPE, geração do gráfico mostrado na Figura 6, cálculo do índice ISC e cálculo do índice MDMGP;

Quadro 10 – Código para a análise dos clusters.

```
#----- Inicialização -----
require(foreign) # pacote usado para leitura, edição e gravação do arquivo DBF
require(e1071) # pacote para cmeans
dados <- read.dbf("C:/R/Dados/dados.dbf")
n_cluster <- 4
#----- Processamento / MPE -----
dados_padronizados <- Padronizar(dados)
wss <- (nrow(dados_padronizados)-1)*sum(apply(dados_padronizados,2,var))
for (i in 2:n_cluster){
  pertinencia_dados <- cmeans(dados_padronizados, # matriz dos dados
    n_cluster, # numero de clusters
    verbose = FALSE, # não imprime histórico
    dist = 'euclidean', # distância empregada
    method = 'cmeans') # método
  wss[(i-1)] <- CalcularMPE(i, 50, pertinencia_dados$membership)
}
plot(2:n_cluster, wss, type="b", xlab="Numero de Clusters", ylab="MPE")
indices <- CalcularIndices(pertinencia_dados)
#----- ISC -----
CalcularISC(pertinencia_dados$membership, 1, indices$c1)
CalcularISC(pertinencia_dados$membership, 2, indices$c2)
CalcularISC(pertinencia_dados$membership, 3, indices$c3)
CalcularISC(pertinencia_dados$membership, 4, indices$c4)
#----- MDMGP -----
CalcularGP(pertinencia_dados$membership, 1, 2)
CalcularGP(pertinencia_dados$membership, 1, 3)
CalcularGP(pertinencia_dados$membership, 1, 4)
CalcularGP(pertinencia_dados$membership, 2, 1)
CalcularGP(pertinencia_dados$membership, 2, 3)
CalcularGP(pertinencia_dados$membership, 2, 4)
CalcularGP(pertinencia_dados$membership, 3, 1)
CalcularGP(pertinencia_dados$membership, 3, 2)
CalcularGP(pertinencia_dados$membership, 3, 4)
CalcularGP(pertinencia_dados$membership, 4, 1)
CalcularGP(pertinencia_dados$membership, 4, 2)
CalcularGP(pertinencia_dados$membership, 4, 3)
#----- Quantia de elementos de cada cluster -----
pertinencia_dados$size
```

- Distribuição dos municípios nos clusters. Com base no FCM os municípios são classificados e distribuídos em seus agrupamentos de maior pertinência, e então salvados em um arquivo para a montagem dos mapas temáticos;

Quadro 11 – Código para a distribuição dos municípios.

```
#----- Inicialização -----
require(foreign) # pacote usado para leitura, edição e gravação do arquivo DBF
require(e1071) # pacote para cmeans
dados <- read.dbf("C:/R/Dados/dados.dbf")
n_cluster <- 4
#----- Processamento / MPE -----
dados_padronizados <- Padronizar(dados)
pertinencia_dados <- cmeans(dados_padronizados, # matriz dos dados
                           n_cluster, # numero de clusters
                           verbose = FALSE, # não imprime histórico
                           dist = 'euclidean', # distância empregada
                           method = 'cmeans') # método
#----- Distribuição dos agrupamentos -----
dados_cluster <- read.dbf("C:/R/shapes3/cluster/dados.dbf")
for(i in 1:length(pertinencia_dados$membership[,1])){
  maior_grau <- 0
  for(ii in 1:length(pertinencia_dados$membership[1,])){
    pertinencia <- pertinencia_dados$membership[i, ii]
    if(maior_grau < pertinencia){
      maior_grau <- pertinencia
      dados_cluster$CLUSTER[i] <- ii
      dados_cluster$GRAU[i] <- maior_grau
    }
  }
}
write.dbf(dados_cluster, "C:/R/shapes3/cluster/dados.dbf")
```

- Cálculo das estatísticas dos clusters mostradas na Tabela 5, tais são a média, mediana, desvio padrão, coeficiente de variação e valores mínimo e máximo, obtenção dos valores de centro dos agrupamentos (mostrados na Tabela 4) e a distribuição quantitativa dos municípios (mostrados na Tabela 3).

#### Quadro 12 – Código para o cálculo das estatísticas dos clusters parte 1.

```

dados <- read.dbf("C:/R/shapes3/cluster/dados.dbf")
#----- pegando os dados -----
#Estrutura -----
var <- list(c(),c(),c(),c())
names(var) <- c("Prod", "Precm", "TMedia", "RSGM")
dados_clusters <- list(var, var, var, var)
names(dados_clusters) <- c("c1", "c2", "c3", "c4")
#Processamento -----
for(i in 1:length(dados[,1])){
  dados_clusters[[dados$CLUSTER[i]]]$Prod <- c(dados_clusters[[dados$CLUSTER[i]]]$Prod,
dados$PRODUCAO[i])
  dados_clusters[[dados$CLUSTER[i]]]$Precm <- c(dados_clusters[[dados$CLUSTER[i]]]$Precm,
dados$PRECIP[i])
  dados_clusters[[dados$CLUSTER[i]]]$TMedia <- c(dados_clusters[[dados$CLUSTER[i]]]$TMedia,
dados$TEMP[i])
  dados_clusters[[dados$CLUSTER[i]]]$RSGM <- c(dados_clusters[[dados$CLUSTER[i]]]$RSGM,
dados$RADIA[i])
}
#----- Estatísticas ds dados -----
#Estrutura -----
estatisticas <- list(c(),c(),c(),c(),c(),c())
names(estatisticas) <- c("Min", "Max", "Media", "Mediana", "DesPad", "CV" )
variaveis <- list(estatisticas,estatisticas,estatisticas,estatisticas)
names(variaveis) <- c("Prod", "Precm", "TMedia", "RSGM")
clusters <- list(variaveis, variaveis, variaveis, variaveis)
names(clusters) <- c("c1", "c2", "c3", "c4")
#Processamento -----
for(i in 1:4){
  for(ii in 1:4){
    clusters[[i]][[ii]]$Min <- min(dados_clusters[[i]][[ii]])
    clusters[[i]][[ii]]$Max <- max(dados_clusters[[i]][[ii]])
    clusters[[i]][[ii]]$Media <- mean(dados_clusters[[i]][[ii]])
    clusters[[i]][[ii]]$Mediana <- median(dados_clusters[[i]][[ii]])
    clusters[[i]][[ii]]$DesPad <- sd(dados_clusters[[i]][[ii]])
    clusters[[i]][[ii]]$SCV <- var(dados_clusters[[i]][[ii]]) * 100 / mean(dados_clusters[[i]][[ii]])
  }
}
}

```

Quadro 13 – Código para o cálculo das estatísticas dos clusters parte 2.

```
#----- Centragem -----
pertinencia_dados$center
#----- Quantia de elementos de cada cluster -----
pertinencia_dados$size
```

- Filtrando mapas de municípios, reduzindo o espaço de amostragem e selecionando apenas os municípios cujos polígonos de Thiessen formados pelas estações meteorológicas incidem sobre municípios do Oeste do Paraná;

Quadro 14 – Código para a filtragem dos mapas de municípios.

```
municipios <- readShapePoly("C:/R/shapes/municipios + dados estacao simepar +
inmet/municipios_pr_sp_sc_mt_estacao")
estacoes <- subset(municipios, ESTACAO==TRUE)
estacoes_pr <- subset(municipios, ESTACAO==TRUE & NOMEUF=="PARANA")
oeste <- readShapePoly("C:/R/shapes/regiao oeste_pr/regiao_oeste_pr")
plot(oeste, col="green")
plot(estacoes, add=TRUE, col="GRAY")
plot(municipios, add=TRUE)
text(coordinates(estacoes), labels=estacoes@data[,1], cex=0.6, col="blue")
estacoes_sobrepostas <- subset(estacoes,
((SP_ID=="4522") | (SP_ID=="4532") | (SP_ID=="4543") | (SP_ID=="4601")
| (SP_ID=="4577") | (SP_ID=="4599") | (SP_ID=="4605") | (SP_ID=="4651")
| (SP_ID=="4672") | (SP_ID=="4656") | (SP_ID=="4649")))
estacoes_sobrepostas_oeste <- subset(municipios,
((SP_ID=="4522") | (SP_ID=="4532") | (SP_ID=="4543") | (SP_ID=="4601")
| (SP_ID=="4577") | (SP_ID=="4599") | (SP_ID=="4605") | (SP_ID=="4651")
| (SP_ID=="4672") | (SP_ID=="4656") | (SP_ID=="4649") | (NOMEMESO=="OESTE PARANAENSE")))
cents <- coordinates(estacoes_sobrepostas)
thiessen <- voronoiPolygons(cents, estacoes_sobrepostas_oeste)
plot(thiessen, col="RED")
plot(oeste, add=TRUE, col="GREY", )
plot(estacoes_sobrepostas, add=TRUE, col=rgb(1, 1, 0, 0.7))
plot(thiessen, add=TRUE)
points(cents, pch=1, col="blue")
points(cents, pch=3, col="Red")
writePolyShape(estacoes, "C:/R/shapes2/m_estacoes/m_estacoes")
writePolyShape(estacoes_pr, "C:/R/shapes2/m_estacoes_pr/m_estacoes_pr")
writePolyShape(estacoes_sobrepostas, "C:/R/shapes2/m_estacoes_sobrepostas/m_estacoes_sobrepostas")
writePolyShape(estacoes_sobrepostas_oeste,
"C:/R/shapes2/m_estacoes_sobrepostas_oeste/m_estacoes_sobrepostas_oeste")
writePolyShape(oeste, "C:/R/shapes2/m_oeste_pr/m_oeste_pr")
writePolyShape(thiessen, "C:/R/shapes2/t_estacoes_sobrepostas_sd/t_estacoes_sobrepostas_sd")
```

- Criando mapas com legenda e numeração, e extraindo as coordenadas do mapa;

Quadro 15 – Código para criação dos mapas com legenda.

```

require(maptools)
require(rgeos)
#----- OESTE -----
#----- Numeração -----
oeste <- readShapePoly("C:/R/shapes/regiao oeste_pr/regiao_oeste_pr")
numeracao <- seq(along=oeste)
plot(oeste, border=gray(0.4), col=gray(0.9))
points(gCentroid(oeste), pch=3)
points(gCentroid(oeste), pch=1)
text(coordinates(oeste), labels=seq(along=oeste))
#----- Legenda -----
plot(oeste, border="white")
legenda <- paste(numeracao, oeste@data[, "NOME_1"])
legend("right", cex=1, legend=legenda, ncol=2, col="black")
#----- Coordenada -----
gCentroid(oeste)
#----- ESTAÇÕES -----
#----- Numeração -----
estacoes <- readShapePoly("C:/R/shapes2/m_estacoes_sobrepostas/m_estacoes_sobrepostas")
parana <- readShapePoly("C:/R/shapes/municipios pr_sp_sc_mt/municipios_pr_sp_sc_mt")
oeste <- readShapePoly("C:/R/shapes/regiao oeste_pr/regiao_oeste_pr")
parana$NOMEUF <- subset(parana, NOMEUF=="PARANA")
numeracao <- seq(along=estacoes)

plot(estacoes, add=TRUE, col=gray(0.9))
plot(parana, add=TRUE, border=gray(0.9))
plot(oeste, add=TRUE, border="black")
points(gCentroid(estacoes), pch=3)
points(gCentroid(estacoes), pch=1)
text(coordinates(estacoes), labels=seq(along=estacoes))
#----- Legenda -----
plot(estacoes, border="white")
legenda <- paste(numeracao, estacoes@data[, "NOME_1"])
legend("right", cex=1, legend=legenda, ncol=1, col="black")
#----- Coordenada -----
gCentroid(oeste)

```

- Criando mapa temático dos clusters a partir de um shapefile contendo as classificações e graus de pertinência geradas pelo FCM;

Quadro 16 – Código para criação do mapa temático.

```
#----- Inicialização -----
write.dbf(dados_cluster, "C:/R/shapes3/cluster/dados.dbf")
oeste <- readShapePoly("C:/R/shapes3/cluster/dados")
#----- Montagem do mapa -----
plot(oeste)
cores <- c("yellow", "blue", "red", "green", "purple", "black")
for(i in 1:length(oeste)){
  plot(oeste[i,], add=TRUE, col=cores[oeste$CLUSTER[i]])
  if(oeste$GRAU[i] > 0 & oeste$GRAU[i] <= 0.35){
    plot(oeste[i,], add=TRUE, col=rgb(1, 1, 1, 0.9))
  }
  if(oeste$GRAU[i] > 0.35 & oeste$GRAU[i] <= 0.65){
    plot(oeste[i,], add=TRUE, col=rgb(1, 1, 1, 0.4))
  }
}
legenda <- c("Cluster 1 - 66,5%", "Cluster 2 - 67,1%", "Cluster 3 - 68,1%", "Cluster 4 - 67,0%")
legend("topright", cex=2, legend=legenda, ncol=1, col="black", fill=cores)
gCentroid(oeste)
```

- Transferindo dados não espaciais do shapefile contendo os municípios para os novos polígonos gerados a partir da função de Voronoi (Quadro 6) e salvando os dados.

Quadro 17 – Código para a transferência dos dados meteorológicos.

```
estacoes <- readShapePoly("C:/R/shapes2/m_estacoes_sobrepostas/m_estacoes_sobrepostas")
thiessen_sd <- readShapePoly("C:/R/shapes2/t_estacoes_sobrepostas_sd/t_estacoes_sobrepostas_sd")
thiessen_cd <- SpatialPolygonsDataFrame(thiessen_sd, data=estacoes@data)
writePolyShape(thiessen_cd, "C:/R/shapes2/t_estacoes_sobrepostas_cd/t_estacoes_sobrepostas_cd")
```