



**ANA LÚCIA SOUZA SILVA MATEUS**

**PROPOSIÇÃO DE NOVAS METODOLOGIAS  
PARA ANÁLISE DE ALEATORIEDADE EM  
PROCESSOS PONTUAIS NO ESPAÇO-TEMPO**

**LAVRAS – MG  
2013**

**ANA LÚCIA SOUZA SILVA MATEUS**

**PROPOSIÇÃO DE NOVAS METODOLOGIAS PARA ANÁLISE DE  
ALEATORIEDADE EM PROCESSOS PONTUAIS NO ESPAÇO-TEMPO**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Orientador

Dr. João Domingos Scalon

**LAVRAS – MG**

**2012**

**Ficha Catalográfica Elaborada pela Divisão de Processos Técnicos da  
Biblioteca da UFLA**

Mateus, Ana Lúcia Souza Silva.

Proposição de novas metodologias para análise de aleatoriedade  
em processos pontuais no espaço-tempo / Ana Lúcia Souza Silva  
Mateus. – Lavras : UFLA, 2013.

138 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2012.

Orientador: João Domingos Scalon.

Bibliografia.

1. Função *K*. 2. Estatística *scan*. 3. Métodos estatísticos. 4.  
Simulação Monte Carlo. 5. Testes estatísticos. 6. Interação espaço-  
tempo. I. Universidade Federal de Lavras. II. Título.

CDD – 519.282

**ANA LÚCIA SOUZA SILVA MATEUS**

**PROPOSIÇÃO DE NOVAS METODOLOGIAS PARA ANÁLISE DE  
ALEATORIEDADE EM PROCESSOS PONTUAIS NO ESPAÇO-TEMPO**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Aprovada em 27 de dezembro de 2012.

Dr. Antônio Policarpo Souza Carneiro	UFV
Dr. Mauricio Sergio Zacarias	Embrapa Café
Dr. Marcelo Silva de Oliveira	UFLA
Dr. Renato Ribeiro Lima	UFLA

Dr. João Domingos Scalon  
(Orientador)

**LAVRAS – MG**

**2012**

DEDICO

*A Deus,  
força maior de todo ser humano, por ter  
me dado tudo o que eu sempre precisei;  
À minha família;  
Ao meu esposo Wilson.*

## AGRADECIMENTOS

A Deus, que me preparou dando toda força e sabedoria necessárias para os estudos e situações adversas deste desafio;

Ao meu esposo Wilson, companheiro que nunca me abandonou, estando comigo, ora nos momentos difíceis, ora nas vitórias;

À minha mãe, que embora não esteja mais no plano físico, sinto sempre a sua presença e felicidade na realização deste sonho;

Ao meu pai, pelo apoio e incentivo;

Às minhas sobrinhas, Gleice e Franciene, pela força e carinho.

Aos meus queridos irmãos, Shaine e Luíz Cláudio, pela amizade, carinho e incentivo;

À minha tia, Maria DeLourdes, pelo exemplo de vida, força, incentivo e dedicação;

Ao professor, orientador, amigo e inspirador, Dr. João Domingos Scalon, pela grande relevância neste trabalho e em minha vida. Obrigado pelo rico tesouro do conhecimento;

Aos colegas, Ana Paula, Tânia, Ed Carlos, Paulo Emiliano, Moysés, Fernanda, Adriana, Luzia, Leila, Isabel, Crysttian, Vanessa, Liliane e Rosi, pelo companheirismo, amizade e momentos de alegria. Em especial, Augustão, Leandro, Gisele, pela troca de conhecimentos, amizade e apoio nas horas difíceis;

À Universidade Federal de Lavras (UFLA), em especial ao Departamento de Ciências Exatas (DEX), berço de grandes profissionais que contribuíram imensamente em minha formação;

A todos os funcionários do DEX, pelos serviços prestados, em especial à Josi (P e G), Selminha, Edila e Maria;

À banca composta pelos professores Antônio Policarpo Souza Carneiro, Mauricio Sergio Zacarias, Marcelo Silva de Oliveira e Renato Ribeiro de Lima por aceitaram o convite que lhes foi feito e, dessa forma, colaboraram para conclusão deste projeto;

Aos pesquisadores do Projeto PROCAD/CAPES “Seguro Agrícola: Modelagem Estatística e Precificação”, em especial aos professores Marcelo A. Costa (UFMG) e Elias T. Krainski (UFPR), pelo auxílio neste trabalho;

A CAPES, pelo apoio financeiro no Projeto PROCAD;

À Universidade Federal do Amazonas (UFAM), pela minha liberação integral para cursar o doutorado;

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pelo apoio financeiro;

À Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM), pela bolsa de doutorado;

Enfim, a todos que contribuíram, direta ou indiretamente, meus eternos agradecimentos.

## RESUMO

A informação sobre a dinâmica espaço-temporal de doenças de plantas é de fundamental importância para desenvolver tecnologias apropriadas ao manejo das doenças em sistemas de produção. Assim, nos últimos anos, vários estudos envolvendo padrões espaciais e temporais estão sendo desenvolvidos para buscar uma melhor compreensão dos mecanismos e da dinâmica de doenças presentes na agricultura. Entretanto, a análise da interação espaço-tempo, utilizando métodos de processos pontuais, tem sido pouco abordada nessa área. O objetivo deste trabalho é apresentar métodos baseados na estatística de segunda ordem e na estatística de varredura para analisar processos pontuais a fim de detectar a presença de agrupamentos no tempo, no espaço, bem como agrupamentos oriundos das interações espaço-tempo. Com o trabalho propõe-se também um novo teste de hipótese para detectar agrupamento no tempo, utilizando a função distribuição das contagens dos tempos de ocorrência de eventos e um teste de hipótese para detectar interação espaço-tempo. Os dois testes de hipóteses são baseados em técnicas de Monte Carlo. Os métodos apresentados foram testados em quatro realizações geradas por meio de simulação computacional e aplicados em dados mensais de incidência de casos de Morte Súbita dos Citros (MSC) em laranjeiras. Os resultados obtidos a partir das análises dos dados simulados e dos dados de MSC mostraram que tanto as metodologias disponíveis na literatura quanto as propostas nesta tese foram eficientes para detectar agrupamentos no tempo, no espaço e no espaço-tempo.

Palavras-chave: Função  $K$ . Estatística *scan*. Testes estatísticos. Simulação Monte Carlo. Interação espaço-tempo.



## ABSTRACT

The information on the spatiotemporal dynamics of plant diseases is of paramount importance to develop appropriate technologies to the management of diseases in production systems. Thus, in recent years, several studies involving spatial and temporal patterns are being conducted to get a better understanding of the mechanisms and dynamics of diseases in agriculture. However, the analysis of the space-time interaction, using point processes methods, has been scarcely discussed in this area. The aim of this paper is to present methods based on both the second-order statistic and the scan statistic to analyze point processes to detect the presence of clusters in time, space as well as clusters from the spatial and temporal interactions. The paper also proposes a new hypothesis test to detect clustering in time, using the distribution function of time counts of the occurrence of events, and a hypothesis test to detect space-time interaction. The two hypotheses tests are based on Monte Carlo techniques. The methods presented were tested in the four realizations generated through computer simulation and applied to monthly data on incidence of Citrus Sudden Death (SCD) in orange. The results from analyzes of both simulated and actual data from SCD showed that the methodologies available in the literature and suggested in this thesis were effective to detect clusters in time, in space and space-time.

Key words:  $K$ -function. Scan statistic. Statistical tests. Monte Carlo simulation. Space-time interaction.

## LISTA DE ILUSTRAÇÕES

Figura 1	Ideia gráfica da função K espacial.....	24
Figura 2	Mecanismo de estimação da função K temporal.....	30
Figura 3	Exemplo hipotético do método de varredura espacial.....	39
Figura 4	Procedimento da varredura espaço-temporal.....	44
Figura 5	Coloração verde pálida a amarela das folhas (a) e planta com sinais da MSC ao lado de uma planta sadia (b).....	48
Figura 6	Raízes e radículas de uma planta sadia (esquerda), podridão das raízes e ausência de radículas de uma planta doente (direita).....	49
Figura 7	Detalhe da coloração amarela da parte interna da casca da árvore na parte do porta enxerto.....	50
Gráfico 1	Cenário 1: processo com agrupamento no espaço, sem agrupamento no tempo e sem interação. (a) 1º trimestre, (b) 2º trimestre, (c) 3º trimestre e (d) 4º trimestre.....	54
Gráfico 2	Cenário 2: processo sem agrupamento no espaço, com agrupamento no tempo e sem interação espaço-tempo. (a) 1º trimestre, (b) 2º trimestre, (c) 3º trimestre e (d) 4º trimestre.....	56
Gráfico 3	Cenário 3: processo sem agrupamento no espaço, sem agrupamento no tempo e sem interação espaço-tempo. (a) 1º trimestre, (b) 2º trimestre, (c) 3º trimestre e (d) 4º trimestre.....	57
Gráfico 4	Cenário 4: processo com agrupamento no espaço, agrupamento no tempo e interação espaço-tempo. (a) 1º trimestre, (b) 2º trimestre, (c) 3º trimestre e (d) 4º trimestre.....	58

Gráfico 5	Croqui do talhão experimental.....	59
Gráfico 6	Função de distribuição observada (linha contínua em vermelho), função de distribuição teórica sob a suposição de completa aleatoriedade temporal (linha tracejada em preto) para as quatro realizações típicas.....	73
Gráfico 7	Função de distribuição observada (linha contínua em vermelho), função de distribuição teórica sob a suposição de completa aleatoriedade temporal (linha tracejada em preto), para o primeiro (a), segundo (b), terceiro (c) e quarto (d) cenários simulados.....	75
Gráfico 8	Histograma da distribuição de frequência empírica de $D(s,t)$ , sob a suposição de ausência de interação espaço-tempo, com a estatística $D(s,t)$ para a realização observada (linha vertical preta), para o primeiro (a), segundo (b), terceiro (c) e quarto (d) cenários simulados.	80
Gráfico 9	Função K observada (linha contínua em azul) e envelope de simulação com 99% de confiança (linhas tracejadas em vermelho), representado os limites da completa aleatoriedade espacial, para o primeiro (a), segundo (b), terceiro (c) e quarto (d) cenários simulados.....	82
Gráfico 10	Função K temporal observada (linha contínua azul), função K temporal teórica (linha contínua preta) e envelope de simulação com 99% de confiança (linhas tracejadas em vermelho), representado os limites da completa aleatoriedade temporal, para o primeiro (a), segundo (b), terceiro (c) e quarto (d) cenários simulados.	84
Gráfico 11	Gráfico de dispersão dos resíduos generalizados $R(s,t)$ versus o produto das funções K temporal e espacial	

	separadamente (a); Gráfico de superfície do grau de agrupamento espaço-tempo adicional $D(s,t)$ (b), para o cenário 1.....	87
Gráfico 12	Gráfico de dispersão dos resíduos generalizados $R(s,t)$ versus o produto das funções $K$ temporal e espacial separadamente (a); Gráfico de superfície do grau de agrupamento espaço-tempo adicional $D(s,t)$ , (b), para o cenário 2.....	88
Gráfico 13	Gráfico de dispersão dos resíduos generalizados $R(s,t)$ versus o produto das funções $K$ temporal e espacial separadamente (a); Gráfico de superfície do grau de agrupamento espaço-tempo adicional $D(s,t)$ , (b); para o cenário 3.....	89
Gráfico 14	Gráfico de dispersão dos resíduos generalizados $R(s,t)$ versus o produto das funções $K$ temporal e espacial separadamente (a); Gráfico de superfície do grau de agrupamento espaço-tempo adicional $D(s,t)$ (b); para o cenário 4.....	91
Gráfico 15	Gráfico representando os dois agrupamentos mais prováveis e seus respectivos valores-p de acordo com o cenário 1.....	93
Gráfico 16	Gráfico representando os quatro agrupamentos e seus respectivos valores-p de acordo com o cenário 4.....	94
Gráfico 17	Gráfico representando os quatro agrupamentos espaciais mais prováveis simulados de acordo com o cenário 4, no período de um ano e seus respectivos valores-p.....	97
Gráfico 18	Configurações espaciais da incidência da MSC no talhão, no período de novembro de 2001 a julho de	

	2003, em que cada círculo representa uma planta com MSC.....	100
Gráfico 19	Incidência mensal de plantas doentes com a MSC no período de novembro de 2001 até julho de 2003.....	102
Gráfico 20	Função K observada (linha contínua em azul) e envelope de simulação com 99% de confiança (linhas tracejadas em vermelho), representando os limites da completa aleatoriedade espacial.....	103
Gráfico 21	Gráfico representando os agrupamentos, estatisticamente significativos identificados pela varredura espacial nos dados MSC: agrupamento mais provável (vermelho), segundo agrupamento mais provável (amarelo), terceiro agrupamento mais provável (azul) e seus respectivos valores-p.....	105
Gráfico 22	Função K observada (linha azul contínua), função K teórica (reta preta contínua) e envelopes de simulação com 99% de confiança (linhas tracejadas em vermelho) representando os limites da CAT.....	106
Gráfico 23	Função distribuição observada (linha contínua em vermelho) e função teórica sob a hipótese de CAT (linha tracejada em preto).....	107
Gráfico 24	Gráfico de dispersão dos resíduos padronizados $R(s,t)$ versus o produto das funções K temporal e espacial (a), histograma da distribuição de frequência empírica de $D(s,t)$ , sob a suposição de ausência de interação espaço-tempo, com a estatística $D(s,t)$ para a realização observada (linha vertical preta) (b), Gráfico de	

	superfície do grau de agrupamento espaço-tempo adicional $D(s,t)$ (c).....	110
Gráfico 25	Gráfico representando o agrupamento espacial mais provável (vermelho) identificado na varredura espaço-temporal, no período de janeiro a julho de 2003, com o respectivo valor-p.....	111

## LISTA DE TABELAS

Tabela 1	Cenários propostos de processos pontuais espaço-temporal.....	52
Tabela 2	Contagens e função de distribuição observada (F(t)) para o vetor de tempos de ocorrência dos eventos $t = [1, 2, 2, 2, 2, 2, 3, 3, 4, 5, 5]$ .....	69
Tabela 3	Resultados da varredura espacial para o cenário 1.....	93
Tabela 4	Resultados da varredura espacial para o cenário 4.....	94
Tabela 5	Resultados da varredura espaço-temporal para o cenário 4.	97
Tabela 6	Resultados da varredura espacial para os dados de morte súbita dos citros.....	104

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>17</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO.....</b>	<b>20</b>
<b>2.1</b>	<b>Processos estocásticos pontuais em uma e duas dimensões.....</b>	<b>20</b>
<b>2.2</b>	<b>Função K espacial.....</b>	<b>23</b>
<b>2.3</b>	<b>Função K temporal.....</b>	<b>29</b>
<b>2.4</b>	<b>Análise espaço-temporal baseada na função K.....</b>	<b>33</b>
<b>2.5</b>	<b>Método de varredura de Kulldorff para análise espacial.....</b>	<b>38</b>
<b>2.6</b>	<b>Método de varredura de Kulldorff para análise temporal.....</b>	<b>43</b>
<b>2.7</b>	<b>Método de varredura de Kulldorff para análise espaço-temporal.....</b>	<b>43</b>
<b>2.8</b>	<b>A citricultura.....</b>	<b>45</b>
<b>2.9</b>	<b>Morte súbita do citros.....</b>	<b>47</b>
<b>3</b>	<b>MATERIAL E MÉTODOS.....</b>	<b>52</b>
<b>3.1</b>	<b>Cenários típicos simulados de configurações pontuais no espaço e tempo.....</b>	<b>52</b>
<b>3.2</b>	<b>Dados de morte súbita dos citros.....</b>	<b>59</b>
<b>3.3</b>	<b>Análise exploratória da morte súbita dos citros.....</b>	<b>60</b>
<b>3.4</b>	<b>Análise de completa aleatoriedade espacial usando a função K.....</b>	<b>60</b>
<b>3.5</b>	<b>Análise de completa aleatoriedade temporal usando a função K.....</b>	<b>61</b>
<b>3.6</b>	<b>Análise gráfica espaço-temporal usando a função K.....</b>	<b>61</b>
<b>3.7</b>	<b>Análise espacial usando método de varredura de Kulldorff.....</b>	<b>62</b>
<b>3.8</b>	<b>Análise temporal usando método de varredura de Kulldorff.....</b>	<b>62</b>
<b>3.9</b>	<b>Análise espaço-temporal usando método de varredura de Kulldorff.....</b>	<b>63</b>
<b>3.10</b>	<b>Softwares.....</b>	<b>63</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO.....</b>	<b>65</b>
<b>4.1</b>	<b>Novo teste para a hipótese de completa aleatoriedade temporal usando a função distribuição de ocorrência de eventos.....</b>	<b>66</b>
<b>4.2</b>	<b>Análise para completa aleatoriedade temporal nos cenários típicos simulados usando o novo teste.....</b>	<b>74</b>
<b>4.3</b>	<b>Novo teste para interação espaço-tempo.....</b>	<b>77</b>
<b>4.4</b>	<b>Análise da interação espaço-tempo nos cenários típicos</b>	



	simulados usando o novo teste.....	78
4.5	<b>Análise dos cenários típicos simulados usando a função K</b>	80
4.5.1	Análise espacial.....	81
4.5.2	Análise temporal.....	83
4.5.3	Análise da interação espaço-temporal.....	86
4.6	<b>Análise dos cenários típicos simulados usando a estatística de varredura de Kulldorff.....</b>	92
4.6.1	Análise de varredura espacial.....	92
4.6.2	Análise de varredura temporal.....	95
4.6.3	Análise de varredura espaço-temporal.....	96
4.7	<b>Análise da morte súbita dos citros.....</b>	98
4.7.1	Análise exploratória.....	99
4.7.2	Análise da distribuição espacial.....	102
4.7.3	Análise da distribuição temporal.....	105
4.7.4	Análise da interação espaço-temporal.....	108
4.8	<b>Considerações finais.....</b>	112
5	<b>CONCLUSÃO.....</b>	115
	<b>REFERÊNCIAS.....</b>	116
	<b>ANEXOS.....</b>	121

## 1 INTRODUÇÃO

Atualmente é comum ter acesso a dados coletados em diferentes localizações de uma região geográfica que foram observados em diferentes instantes de tempo. Porém, nem sempre, essas informações são utilizadas na análise de forma conjunta. Em geral, as análises são feitas utilizando apenas os dados na dimensão do tempo, caracterizando uma análise por técnicas de séries temporais, ou na dimensão espacial, caracterizando uma análise puramente espacial. Entretanto, para uma completa descrição da interação espacial e temporal dos dados é necessário utilizar técnicas que não somente consideram a localização espacial dos eventos, mas também, a dependência entre os tempos de coleta.

Um dos primeiros estudos sobre análise de agrupamento espaço-temporal em processos pontuais foi proposto por Knox (1963). O autor desenvolveu um teste de significância que foi aplicado em dados de casos de leucemia infantil no nordeste da Inglaterra, encontrando evidências significativas de agrupamento espaço-temporal. Besag e Newell (1991) e Whittemore et al. (1987) também sugeriram testes para analisar agrupamento proveniente da interação espaço-tempo.

Alguns anos depois, Diggle et al. (1995) propuseram utilizar uma estatística de segunda ordem, dada pela função  $K$  de Ripley (1977), para analisar a interação espaço-tempo e temporal em padrões pontuais em dados epidemiológicos. Outra proposta encontrada na literatura é a estatística de varredura (*scan statistics*) proposta por Kulldorff e Nagarwalla (1995). Este método prioriza uma região formada pelas áreas dos quadrantes cujos centroides caem dentro de um círculo. Variando o raio e o centro do círculo, os possíveis agrupamentos podem ser identificados a partir da razão entre as verossimilhanças dos riscos de ocorrência do evento nas áreas internas e

externas do círculo. A diferença entre as metodologias propostas por Diggle et al. (1995) e Kulldorff e Nagarwalla (1995) é que no primeiro trabalho os autores tem o objetivo de detectar o agrupamento espaço-temporal para diferentes escalas de tempo e espaço enquanto no segundo trabalho os autores apontam quem são os agrupamentos mais prováveis e onde eles estão localizados.

A informação sobre a interação espaço-temporal de doenças de plantas apresenta fundamental importância na agricultura, pois permite entender a dinâmica e o padrão de distribuição das doenças e, conseqüentemente, pode auxiliar no desenvolvimento de tecnologias apropriadas ao manejo dessas doenças em sistemas de produção. Mais especificamente, existe o interesse em saber se os casos da doença estão próximos tanto no espaço quanto no tempo, ou seja, o pesquisador teria interesse em saber se os casos de doenças exibem agrupamentos espaço-temporal.

Tendo em vista a importância de detectar agrupamentos espaço-temporal de casos de doenças na agricultura, objetiva-se com esta tese apresentar métodos baseados na estatística de segunda ordem, propostos por Diggle et al. (1995) e na estatística de varredura proposta por Kulldorff e Nagarwalla (1995) para analisar processos pontuais a fim de detectar a presença de agrupamentos no tempo, no espaço, bem como agrupamentos obtidos a partir das interações espaciais e temporais. Com o trabalho propõem-se também um novo teste de hipótese para detectar agrupamento no tempo, utilizando a função distribuição das contagens dos tempos de ocorrência da doença e um teste de hipótese para detectar interação espaço-tempo. Os dois novos testes de hipóteses propostos são baseados em simulação de Monte Carlo.

Todos os métodos apresentados, tanto os existentes na literatura quanto os propostos nesta tese, são avaliados em quatro realizações espaço-temporais típicas, geradas por meio de simulação computacional. Os métodos também são

aplicados em dados mensais de incidência de casos de Morte Súbita dos Citros em laranjeiras.

## 2 REFERENCIAL TEÓRICO

Neste capítulo serão apresentados alguns conceitos necessários para alcançar os objetivos propostos.

### 2.1 Processos estocásticos pontuais em uma e duas dimensões

Segundo Cressie (1993), um processo estocástico pontual é uma família de variáveis aleatórias denotadas por:  $\{Z(s) : s \in D\}$ , em que  $Z(s)$  é um atributo de interesse na localização espacial e  $s$  é uma quantidade aleatória que varia sobre o conjunto indexado  $D \subset \mathfrak{R}^d$  em que  $D$  é assumido por um subconjunto fixo de  $\mathfrak{R}^d$  ( $d = 1, 2, \dots$ ). Quando  $d=1$  tem-se processo estocástico com domínio no tempo, chamado de processo temporal,  $d=2$ , tem-se o processo espacial e  $d=3$  um processo espaço-temporal.

Um processo estocástico pontual em uma dimensão (no tempo) é um modelo usual para representar uma sequência de tempos aleatórios de ocorrências de um particular evento, como por exemplo, os tempos de ocorrência de casos de uma doença em plantas. Observe que cada caso de planta doente ocorre em um instante, ou ponto, de tempo. Assim, existirá um determinado número aleatório de plantas doentes em qualquer período de tempo, onde esses casos ocorrerão em instantes aleatórios de tempos. Um tratamento teórico rigoroso sobre processos estocásticos pontuais em uma dimensão pode ser encontrado em Baddeley (2006) e Daley e Vere-Jones (1988).

Um processo estocástico pontual espacial é um modelo usual para representar uma configuração aleatória de pontos no espaço  $d$ -dimensional, onde  $d \geq 2$ . Por exemplo, tem-se o caso de um mapa das localizações de todos os casos de plantas doentes em um particular dia ou mês. Este mapa constitui uma configuração aleatória de pontos de duas dimensões. Observe que existirá um

número aleatório desses pontos e suas localizações também serão aleatórias. Baddeley (2006) e Cressie (1993) apresentam um tratamento teórico rigoroso sobre processos estocásticos pontuais em duas dimensões.

Em algumas situações podem-se ter, simultaneamente, as localizações e os tempos de ocorrência de casos de plantas doentes. Essas situações podem ser modeladas por um processo estocástico pontual em três dimensões (espaço  $\times$  tempo) denominado de processo espaço-temporal. Alternativamente, essa situação também pode ser trabalhada como um processo estocástico pontual em duas dimensões, onde cada ponto (localização da planta doente) é marcado por um número (tempo de ocorrência de doença na planta). Baddeley (2006) apresenta uma boa apresentação teórica sobre processos estocásticos pontuais em três dimensões.

Segundo Baddeley (2006) e Daley e Vere-Jones (1988), um processo estocástico pontual pode ser descrito em termos dos efeitos de primeira e segunda ordem. Os efeitos de primeira ordem, considerados globais ou de larga escala, correspondem a variações no valor médio do processo estocástico pontual. Neste caso, estamos interessados na intensidade do processo, isto é, no número de eventos por unidade de tempo/área. Os efeitos de segunda ordem, também denominados locais ou de pequena escala, representam a dependência do processo estocástico, proveniente da estrutura de correlação. Para medir a dependência espacial, procura-se estimar o relacionamento entre pares de pontos (por unidade de área/tempo) no espaço/tempo, o que corresponde a uma aproximação do cálculo da covariância entre as variáveis aleatórias.

A maioria das técnicas para análise de configurações pontuais supõe, explícita, ou implicitamente, um comportamento estacionário e isotrópico do processo estocástico pontual subjacente aos pontos analisados. Esses conceitos estão relacionados aos efeitos de primeira e segunda ordem. Um processo estocástico pontual é estacionário se os efeitos de primeira e segunda ordem são

constantes em toda área/tempo do processo e, portanto, a média e a variância são constantes em qualquer subárea/segmento de tempo e a covariância entre dois pontos quaisquer depende somente de sua localização relativa. Um processo estocástico pontual é isotrópico se a covariância depende apenas da distância entre dois pontos e não da direção entre eles. Um processo estocástico pontual que é estacionário e isotrópico é denominado de movimento invariante (BADDELEY, 2006; DALEY; VERE-JONES, 1988).

A análise estatística dos padrões de distribuições de pontos requer um modelo teórico do processo estocástico pontual de referência, base para o desenvolvimento de métodos formais que checam a significância dos resultados exploratórios. O modelo teórico mais simples é conhecido como completa aleatoriedade pontual. Assim, têm-se os modelos de completa aleatoriedade pontual espacial, completa aleatoriedade pontual temporal e completa aleatoriedade pontual espaço-temporal.

Cada um desses modelos de completa aleatoriedade pontual divide a região/tempo de estudo em subáreas/segmentos de tempos e modela a distribuição de eventos pontuais como um processo aleatório. Neste caso, consideramos o número de eventos que ocorrem na subárea/segmento de tempo. No modelo de completa aleatoriedade espacial/temporal, considera-se que as ocorrências de pontos em cada subárea/segmento de tempo são não correlacionadas e homogêneas, e estão associadas à mesma distribuição de probabilidade, ou seja, a um processo de Poisson homogêneo. Assim, numa visão intuitiva, pode-se considerar que as posições dos pontos são independentes e que apresentam a mesma probabilidade de ocorrência em toda a região/tempo (CRESSIE, 1993; DALEY; VERE-JONES, 1988). Esta formulação permite estabelecer uma base de comparação entre uma distribuição pontual completamente aleatória, que seria gerada por um processo de Poisson homogêneo, e os dados coletados em campo.

O conceito de completa aleatoriedade pontual é utilizado para caracterizar os efeitos de segunda ordem em distribuições de pontos utilizando vários métodos, como as funções baseadas em vizinhos mais próximos ( $F$ ,  $G$  e  $J$ ) e a função  $K$ . Estas funções são utilizadas em vários métodos para a detecção de agrupamentos de pontos, principalmente, no espaço. O presente trabalho concentra-se na função  $K$  que será apresentada em detalhes a seguir. Maiores detalhes teóricos sobre as funções  $F$ ,  $G$  e  $J$  são apresentados em Baddeley (2006), Cressie (1993) e Diggle (2003).

## 2.2 Função $K$ espacial

A origem da função  $K$  pode ser encontrada em Bartlett (1964). Contudo, a sua importância como resumo de dependência espacial sobre uma ampla gama de escalas foi primeiramente realizada e desenvolvida por Ripley (1976, 1977). Devido ao fato de permitir uma quantificação da dependência espacial entre diferentes regiões do processo, muitas vezes ela é chamada de análise de segunda ordem reduzida. A função  $K$  tem sido geralmente utilizada para espaços pontuais bidimensionais, mas também pode ser usada em espaços pontuais tridimensionais e unidimensionais (domínio da reta).

Sob a suposição de isotropia, o que garante que os efeitos direcionais entre dois pontos, numa particular distância, sejam invariantes, a função  $K$  espacial é definida como:  $\lambda K(s) = E(\text{número de pontos contidos a uma distância } s \text{ de um evento arbitrário})$ , em que  $E(\cdot)$  é a esperança, e  $\lambda$  é a intensidade, ou número médio de pontos por unidade de área.

A estimação da função  $K$  baseia-se na variação do número médio de outros pontos em torno de um ponto qualquer, à medida que se varia o raio de consideração. O mais simples e natural estimador de  $K(s)$ , proposto por Ripley (1977), que corrige o efeito de bordas, é dado por:

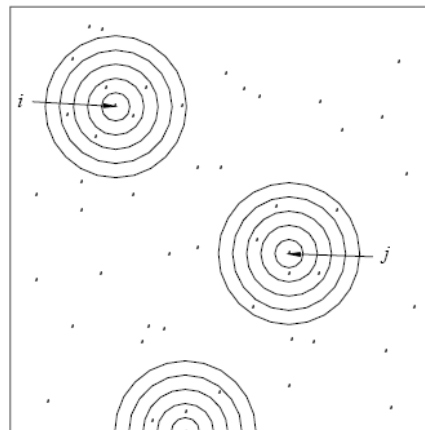


$$\hat{K}(s) = \frac{A}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n \frac{I_s(d_{ij})}{w_{ij}}. \quad (1)$$

em que  $d_{ij}$  é a distância entre os pontos  $i$  e  $j$  ( $i \neq j$ ),  $I_s(d_{ij})$  é uma função indicadora igual a 1 quando  $d_{ij}$  é menor que a distância  $s$  e 0, caso contrário,  $n$  é o número de pontos no mapa analisado com área  $A$  e  $w_{ij}$  é um fator de correção que representa a proporção da circunferência ao redor de um ponto  $i$ , passando sobre o ponto  $j$  que esta dentro da área  $A$ .

A distância  $s$  não possui regra fixa de determinação, depende do objeto de interesse e geralmente é estabelecida pelo pesquisador. O mecanismo da estimação da função  $K$  é mostrado pela Figura 1.

7 pontos dentro da distância  $s$  de 5 unidades, a partir do ponto  $i$ .



6 pontos dentro da distância  $s$  de 5 unidades, a partir do ponto  $j$ .

Figura 1 Ideia gráfica da função  $K$  espacial  
Fonte: Bailey e Gatrell (1995)

Quando se estuda o padrão pontual espacial com base numa determinada região  $A$ , a aleatoriedade do padrão pontual é definida pela completa aleatoriedade pontual espacial (CAE), isto é, o processo que gerou o padrão é aleatório em qualquer escala espacial que, como foi mostrado anteriormente, é dado pelo Processo Poisson homogêneo. Neste caso, as funções de intensidade são constantes, isto é,

$$\lambda_1(s) \equiv \lambda$$

$$\lambda_2(x, y) = \lambda_2(\|x - y\|) = \lambda_2(s) \equiv \lambda^2$$

em que  $\|x - y\|$  é a distância euclidiana entre os pontos  $x$  e  $y$  (DIGGLE, 2003).

Como consequência, a função  $K$ , sob a hipótese de CAE, que reflete uma disposição casual dos pontos, sem interações entre eles, é expressa pela equação:

$$K(s) = \frac{2\pi}{\lambda^2} \int_0^s \lambda_2(s) \varpi d\varpi$$

$$K(s) = \frac{2\pi}{\lambda^2} \int_0^s \lambda^2 \varpi d\varpi$$

$$K(s) = \pi s^2. \quad (2)$$

Para um processo pontual agrupado, os pontos têm em média mais vizinhos do que o esperado sob a hipótese nula de CAE e, conseqüentemente, as

estimativas de  $K(s)$  serão maiores que  $\pi s^2$ . Assim, para configurações que exibem *padrões agrupados*, cada ponto estará provavelmente mais próximo do entorno de indivíduos do mesmo grupo e, conseqüentemente, para valores pequenos de  $s$ ,  $K(s)$  será relativamente grande.

Contrariamente, para um processo pontual que exhibe regularidade, os pontos têm, em média, menos vizinhos do que o esperado sob a hipótese nula. As estimativas de  $K(s)$  serão menores que  $\pi s^2$ , o que significa que se os eventos são regularmente espaçados, formando espaços vazios entre os pontos e, conseqüentemente, para valores pequenos de  $s$ ,  $K(s)$  será relativamente pequeno.

Entre as principais vantagens da função  $K$  estão: (a) permite a detecção do padrão espacial pontual em diferentes escalas de distâncias simultaneamente; (b) o padrão espacial pontual observado pode ser comparado a modelos de processos estocásticos conhecidos para diferentes configurações pontuais (DIGGLE, 2003).

Conforme foi apresentado na Seção 2.1, ao se estudar o padrão pontual espacial é necessário distinguir entre o padrão espacial observado e o processo espacial que o gerou. Nesta tese, tem-se interesse em verificar se a configuração pontual espacial observada foi gerada por um processo estocástico espacial completamente aleatório (processo de Poisson homogêneo) e, portanto, o interesse está em testar a hipótese nula de completa aleatoriedade pontual espacial da configuração pontual.

A literatura apresenta diversas formas de testar a hipótese nula de completa aleatoriedade pontual espacial de uma configuração pontual usando a função  $K$  (DIGGLE, 2003). Todos esses métodos utilizam alguma forma para comparar as funções das equações (1) e (2). Existem basicamente duas formas (teste formal e gráfica) para fazer esta comparação e, conseqüentemente, testar a hipótese nula de que os pontos seguem um processo de Poisson homogêneo, ou seja, que os pontos estão aleatoriamente distribuídos na área.

Para a construção formal do teste de completa aleatoriedade espacial, Diggle (2003) sugere um procedimento de Monte Carlo, que é descrito em detalhes em Scalco et al. (2003), como sendo:

i) Define-se  $r_i$  como a medida da discrepância entre  $\hat{K}(s)$  e  $K(s)$  sobre toda distância  $s_0$ . Essa medida pode ser obtida tanto pela distância de Kolmogorov-Smirnov como pela distância de Cramer-vonMises dadas, respectivamente, por:

$$rm_i = \max_{0 \leq s \leq s_0} |\hat{K}(s) - K(s)| \quad (3)$$

$$rc_i = \int_0^{s_0} [\hat{K}(s) - K(s)]^2 ds, \quad (4)$$

em que  $\hat{K}(s)$  é a função  $K$  espacial estimada pela equação (1) e  $K(s)$  é a função  $K$  teórica dada pela equação (2).

ii) Calcula-se  $r_0$ , em que  $r_0$  é a medida para o processo pontual observado.

iii) Calcula-se  $r_i : i = 2, \dots, m$ , em que  $m$  é o número total de simulações e  $r_i$  é a medida para os processos simulados sob a hipótese de CAE para uma das medidas de teste ( $rm_i$  ou  $rc_i$ ).

iv) Ordena-se  $r_i$ . Se obtivermos  $z_0$  em um dos extremos da distribuição dos  $r_i : i = 1, 2, \dots, m$ , então existem evidências estatísticas para rejeitar a hipótese de completa aleatoriedade espacial. Suponha que,  $r_{(1)} = r_{(j)}$  para algum  $j \in \{1, \dots, m\}$ , então se rejeita a hipótese de independência se,  $\text{valor} - p = ((m + 1 - j) / m) \leq \alpha$ , em que  $\alpha$  é o nível de significância.

O teste de hipótese descrito não indica se a direção da rejeição de hipótese nula é a favor da regularidade ou agrupamento. A fim de determinar

essa direção, sugere-se a construção de envelopes de simulação. Para tal, Diggle (2003) descreve o seguinte procedimento:

i) Calcula-se a função  $\hat{K}_1(s)$  a partir dos dados de uma configuração pontual espacial observada.

ii) Calcula-se a função de distribuição empírica  $\hat{K}_i(s): i = 2, 3, \dots, m$ , para cada  $(m-1)$  simulações independentes de  $n$  pontos independentes e identicamente distribuídos na região observada sob a suposição de completa aleatoriedade espacial (processo de Poisson homogêneo).

iii) Calcula-se os envelopes de simulação superiores e inferiores como sendo:

$$\begin{aligned} K_U(s) &= \max\{\hat{K}_i(s)\}, i = 1, \dots, m. \\ K_L(s) &= \min\{\hat{K}_i(s)\}, i = 1, \dots, m. \end{aligned} \quad (5)$$

iv) Plota-se um gráfico com os valores de  $\hat{K}_1(s)$ ,  $K_U(s)$  e  $K_L(s)$  contra a distância  $s$ . Esse gráfico apresenta a propriedade:

$$P\{\hat{K}_1(s) > K_U(s)\} = P\{\hat{K}_1(s) < K_L(s)\} = m^{-1}.$$

O gráfico resultante pode ser interpretado da seguinte maneira: se a linha de  $\hat{K}_1(s)$  estiver dentro dos envelopes inferior e superior, existe evidência de completa aleatoriedade espacial na escala considerada. se a linha de  $\hat{K}_1(s)$  estiver acima do envelope superior, existe evidência da ocorrência de agrupamento pontual na escala considerada. Caso  $\hat{K}_1(s)$  fique abaixo do envelope inferior, há evidências estatísticas de regularidade dos pontos nessa escala.

A função  $K$  espacial tem sido utilizada exhaustivamente, para análise de configurações pontuais no espaço em diferentes áreas do conhecimento como

epidemiologia, ciências florestais e biologia conforme pode ser verificado em Bailey e Gatrell (1995), Cressie (1993) e Diggle (2003).

A função  $K$  espacial também tem sido usada com frequência para análise da distribuição de casos de doenças na citricultura. Por exemplo, Spósito (2003) avaliou a incidência de plantas com mancha preta dos citros em quatro áreas com diferentes intensidades da doença, variedades e porta-enxertos no período da pré-colheita. A observação dos resultados revelou que existe um padrão de agregação de plantas com frutos sintomáticos independentemente da incidência de plantas doentes. Spósito et al. (2007) mediram a incidência de plantas com mancha preta dos citros em quatro diferentes pomares do Estado de São Paulo, anualmente, de 1999 a 2001 no período da pré-colheita. Com os estudos os autores concluíram que existe um padrão de agregação da doença para todos os pomares analisados e a dependência entre as árvores sintomáticas variou de 2 a 3 árvores vizinhas. Gottwald et al. (2002) analisaram cinco áreas urbanas em Miami afetadas com o cancro cítrico. Para cada área estudada, o padrão espacial foi analisado para as medições de distâncias focais entre as árvores infectadas para vários períodos temporais de 30, 60, 90 e 120 dias de duração. Foi observado o padrão agregado para todas as áreas estudadas.

### **2.3 Função $K$ temporal**

A função  $K$  temporal é definida como:  $\lambda K(t) = E(\text{número de eventos contidos a um tempo } t \text{ de um evento arbitrário})$ , em que  $E(.)$  é a esperança, e  $\lambda$  é a intensidade, ou número médio de eventos por unidade de tempo (DIGGLE et al., 1995). Sob a hipótese de completa aleatoriedade pontual temporal (CAT), ou seja, o processo subjacente é um processo de Poisson homogêneo no tempo, as funções de intensidade são constantes, isto é,

$$\lambda_1(u) \equiv \lambda$$

$$\lambda_2(u_i, u_j) = \lambda_2(|u_i - u_j|) \equiv \lambda$$

em que,  $u_i$  e  $u_j$  são dois pontos localizados no tempo  $t$ ,  $|u_i - u_j|$  é a distância absoluta entre os pontos  $u_i$  e  $u_j$ .

Considerando que a função  $K$  trabalha com vizinhanças, é necessário tratar o tempo em duas direções (Figura 2), simetricamente, conseqüentemente a função  $K$  sobre a hipótese de completa aleatoriedade temporal, é expressa pela equação:

$$K(t) = 2\lambda^{-1} \int_0^t \lambda du$$

$$K(t) = 2t. \quad (6)$$

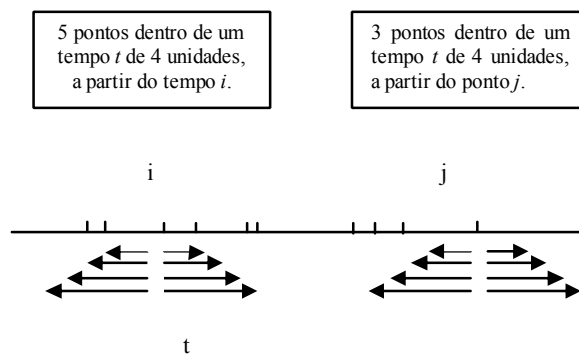


Figura 2 Mecanismo de estimação da função  $K$  temporal

O estimador da função  $K$  temporal é baseado na contagem do número de pares de eventos situados num intervalo de tempo inferior ou igual a  $t$  e, em seguida, é realizada a contagem média em relação ao número de pontos no intervalo de tempo observado. Um estimador que corrige o efeito de borda é dado pela equação:

$$\hat{K}(t) = \frac{T}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n \frac{I_t(u_{ij})}{v_{ij}}, \quad (7)$$

em que  $u_{ij}$  é a distância no tempo entre os pontos  $i$  e  $j$  ( $i \neq j$ ),  $I_t(u_{ij})$  é uma função indicadora igual a 1 quando  $u_{ij}$  é menor que o tempo  $t$  e 0, caso contrário,  $n$  é o número de pontos observados no intervalo de tempo  $T$  e  $v_{ij}$  é um fator de correção que representa um seguimento de reta, centrado em  $i$ , de comprimento  $u_{ij}$ , que se encontra dentro do tempo total  $T$  (DIGGLE et al., 1995).

Conhecido o estimador da função  $K$  temporal, pode-se verificar se a configuração pontual temporal observada foi gerada por um processo estocástico temporal completamente aleatório (processo de Poisson homogêneo no tempo). Portanto, o interesse está em testar a hipótese nula de completa aleatoriedade pontual temporal da configuração pontual.

Por analogia com o caso espacial, no presente trabalho defende-se que o estimador da função  $K$  temporal também pode ser utilizado de duas formas (gráfico e teste formal) para testar a hipótese nula de que os eventos seguem um processo de Poisson homogêneo no tempo, ou seja, os pontos estão aleatoriamente distribuídos na reta.

Assim, os procedimentos para a obtenção do teste formal e dos gráficos apresentados para a função  $K$  espacial devem ser semelhantes aos procedimentos utilizando a função  $K$  temporal. Neste caso, define-se  $h_i$  como a medida da



discrepância entre  $\hat{K}(t)$  e  $K(t)$  sobre todo o tempo  $t$ . Essa medida pode ser obtida por várias formas de distâncias. Utilizaram-se as distâncias de Kolmogorov-Smirnov e Cramer-vonMises. Usando estas duas distâncias, obtém as estatísticas de teste:

$$hm_i = \max_{0 \leq t \leq t_0} |\hat{K}(t) - K(t)| \quad (8)$$

$$hc_i = \int_0^{t_0} [\hat{K}(t) - K(t)]^2 dt, \quad (9)$$

em que  $\hat{K}(t)$  é a função  $K$  temporal estimada, dada pela equação (7) e  $K(t)$  é a função  $K$  temporal teórica, dada pela equação (6).

A partir das equações (8) e (9) pode-se construir um teste formal contra a hipótese de completa aleatoriedade pontual temporal usando o mesmo procedimento de Monte Carlo que foi descrito para o caso espacial.

Assim como no caso espacial, o teste de hipótese proposto anteriormente não indica se a direção da rejeição de hipótese nula é a favor da regularidade ou agrupamento dos pontos no tempo. A fim de determinar essa direção, pode-se construir envelopes de simulação de maneira análoga ao caso espacial, onde substitui-se  $\hat{K}_1(s)$  por  $\hat{K}_1(t)$ , a função calculada a partir dos dados de uma configuração pontual temporal observada e  $K_U(t)$  e  $K_L(t)$  são os envelopes superiores e inferiores respectivamente, obtidos de forma análoga a equação (5). Esses valores são plotados contra o tempo  $t$ .

O gráfico resultante pode ser interpretado da seguinte maneira: se a linha de  $\hat{K}_1(t)$  estiver dentro dos envelopes inferior e superior, existe evidência de completa aleatoriedade temporal na escala de tempo considerada. se a linha

de  $\hat{K}_1(t)$  estiver acima do envelope superior, existe evidência da ocorrência de agrupamento pontual na escala de tempo considerada. Caso  $\hat{K}_1(t)$  fique abaixo do envelope inferior, há evidências estatísticas de regularidade dos pontos nessa escala de tempo.

A função  $K$  temporal vem sendo utilizada com frequência apenas no contexto de análises espaço-temporal e tem sido pouco explorada como ferramenta estatística de análise de processos pontuais no tempo (CANNON; CRESSIE, 1995). Não foi encontrado na literatura, salvo melhor juízo, trabalhos que utilizem a função  $K$  temporal na análise de configurações pontuais no tempo em agricultura. Assim, esta tese também inova ao explorar o potencial dessa função em diversas situações, inclusive na análise de dados agrícolas.

#### 2.4 Análise espaço-temporal baseada na função $K$

Diggle et al. (1995) propuseram um método baseado na clássica função  $K$  de Ripley (1977) para investigar a interação espaço-tempo para dados contínuos no tempo e no espaço. O método tem como base a função  $K(s, t)$  que representa o número esperado de pontos dentro de uma distância espacial  $s$  e um intervalo de tempo  $t$  a partir de um ponto arbitrário, ponderado pelo número esperado de pontos por unidade de espaço e tempo, ou seja,  $\lambda K(s, t) = E(\text{número de pontos contidos a uma distância } s \text{ e um tempo } t \text{ de um ponto arbitrário})$ .

As propriedades de segunda ordem de um processo pontual espaço-temporal são dadas em termos de funções de intensidade de segunda ordem  $\lambda_2(s, t)$ . Para uma região espacial  $S \subseteq \mathfrak{R}^2$  em um intervalo de tempo  $(a, b)$ , sendo  $N(S, a, b)$  o número de eventos em  $S \times (a, b)$  e a intensidade do processo é  $\lambda = E[N(S, a, b)] / [|S|(b - a)]$ , para qualquer  $S$ ,  $a$  e  $b$ . Considerando elementos de área infinitesimais  $dx$  e  $dy$  separados por uma distância  $s$  e elementos de tempo

infinitesimais  $dt$  e  $du$  separados por uma distância  $t$  tem-se, segundo Diggle et al. (1995), que

$$\lambda_2(s, t) = \lim_{\substack{|dx|, |dy| \rightarrow 0 \\ dt, du \rightarrow 0}} \left\{ \frac{E[N(dx, 0, dt)N(dy, t, t + du)]}{|dx||dy|dtdu} \right\}.$$

Para um processo de Poisson homogêneo no espaço-tempo, em que as componentes espaciais e temporais são independentes em  $\mathfrak{R}^2 \times \mathfrak{R}$ , respectivamente, a função  $K(s, t)$  é dada por:

$$K(s, t) = 4\pi\lambda^{-1} \int_0^t \int_0^s \lambda_2(u, v)ududv$$

$$K(s, t) = 2\pi s^2 t \quad (10)$$

A equação (10) está relacionada ao número cumulativo de pontos dentro de um cilindro espaço-tempo com raio  $s$  e altura  $2t$  centrado em um ponto do processo.

Diggle et al. (1995) mostram que um estimador da função  $K(s, t)$ , com correção para o efeito de borda, é dada pela equação:

$$\hat{K}(s, t) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n \frac{I_s(d_{ij})I_t(u_{ij})}{w_{ij}v_{ij}}_{AT} \quad (11)$$

em que  $n$  é o número total de eventos dentro da região espacial-temporal  $A \times (0, T)$ ,  $d_{ij}$  é a distância espacial entre os eventos  $i$  e  $j$  ( $i \neq j$ ),  $u_{ij}$  é a distância temporal entre os eventos  $i$  e  $j$  ( $i \neq j$ ),  $I_s$  é uma função indicadora igual a 1

quando  $d_{ij}$  é menor que a distancia  $s$  e 0, caso contrário,  $I_t$  é uma função indicadora igual a 1 quando  $u_{ij}$  é menor que um tempo  $t$  e 0, caso contrário.

Devido a presença do efeito de borda espacial e temporal que pode acarretar distorções nas estimativas nas proximidades das bordas da área  $A$  e tempo  $T$ , é necessário introduzir correções para esses efeitos. A correção  $w_{ij}$  representa a proporção da circunferência de um círculo com centro em  $i$ , passando pelo ponto  $j$ , que se encontra dentro da área  $A$ . Por analogia,  $v_{ij}$  representa um seguimento de reta, centrado em  $i$ , de comprimento  $u_{ij}$ , que se encontra dentro do tempo total  $T$ . Se ambas as extremidades do intervalo, o comprimento  $2a_{ij}$  e centro  $u_i$ , sendo  $a_{ij} = |u_i - u_j|$  estiverem dentro do intervalo  $(0, T)$ ,  $v_{ij} = 1$ , caso contrário  $v_{ij} = 2$ .

Segundo Diggle et al. (1995), se os processos temporal e espacial operam independentemente então a função  $K$  apresentada na equação (11) pode ser fatorada da seguinte forma,  $\hat{K}(s, t) = \hat{K}(s)\hat{K}(t)$ , em que  $\hat{K}(s)$  e  $\hat{K}(t)$  são os estimadores da função  $K$  no espaço e no tempo dadas pelas equações (1) e (7), respectivamente. Dessa forma, os autores sugerem que havendo independência entre os dois processos (espacial e temporal) existe uma ausência de interação espaço-temporal, ou seja, se  $\hat{K}(s, t) \neq \hat{K}(s)\hat{K}(t)$  então existe uma interação espaço-tempo. A partir dessa propriedade, os autores propõem duas novas medidas para estimar o grau de interação espaço-tempo. A primeira delas é o grau de agrupamento espaço-tempo adicional dado por:

$$\hat{D}(s, t) = \hat{K}(s, t) - \hat{K}(s)\hat{K}(t). \quad (12)$$

Para a hipótese nula de independência espaço-temporal, esperam-se valores de zero para (12). Quanto mais afastado de zero estiverem os valores de  $\hat{D}(s,t)$ , maior é a interação, portanto maior é o aumento do agrupamento no espaço-tempo. Diggle et al. (1995) argumentam que a presença de interação espaço-temporal pode ser observada por meios de picos no gráfico da superfície tridimensional de  $\hat{D}(s,t)$  contra a distância espacial  $s$  e a sequência de tempo  $t$ . Outra forma de analisar o gráfico é observar se na superfície os valores de  $s$  e  $t$  aumentam e os valores de  $\hat{D}(s,t)$  também aumentam. Esse comportamento também pode ser um indicio de interação.

A segunda medida é o aumento proporcional do agrupamento espaço-tempo que é devido à interação espaço-temporal. Esta medida é calculada por

$$\hat{D}_0(s,t) = \hat{D}(s,t) / [\hat{K}(s)\hat{K}(t)]. \quad (13)$$

Similarmente à função  $\hat{D}(s,t)$ , a superfície tridimensional da função  $\hat{D}_0(s,t)$  pode ser colocada em um gráfico contra a distância espacial  $s$  e a sequência de tempo  $t$  para verificar a presença de interação espaço-temporal.

Outra forma de diagnosticar a interação espaço-tempo é realizando um gráfico dos resíduos padronizados  $\hat{R}(s,t)$  versus o produto das funções  $K$  temporal e espacial (DIGGLE et al., 1995). Os resíduos padronizados  $\hat{R}(s,t)$  são estimados por:

$$\hat{R}(s,t) = \hat{D}(s,t) / \sqrt{\hat{V}(s,t)}, \quad (14)$$

em que  $\hat{V}(s,t)$  é a variância de  $\hat{D}(s,t)$ , que é dada por:

$$\begin{aligned}
\text{Cov}[\hat{D}(s,t), \hat{D}(s',t')] = & |A|^2 T^2 (n_2^{-2}) [(n_4)^{-1} (W_1 W_1' - 4W_2 + 2W_3) (V_1 V_1' - \\
& - 4V_2 + 2V_3) + 4(n_3)^{-1} (W_2 - W_3) (V_2 - V_3) + \\
& + 2(n_2)^{-1} W_3 V_3 - (n_2)^{-2} W_1 V_1 W_1' V_1'] \quad (15)
\end{aligned}$$

em que para cada  $s$  e  $t$ , os pesos são definidos simetricamente por:

$$W_{ij} = 1/2(w_{ij} + w_{ji})I(d_{ij} \leq s); \quad V_{ij} = 1/2(v_{ij} + v_{ji})I(u_{ij} \leq t); \quad W_{ii} = 0; \quad V_{ii} = 0.$$

Similarmente, define-se  $W'_{ij}$  e  $V'_{ij}$  substituindo  $s$  e  $t$  por  $s'$  e  $t'$ ,

$$\text{respectivamente; } W_1 = \sum_{i=1}^n \sum_{j=1}^n W_{ij}; \quad W_1' = \sum_{i=1}^n \sum_{j=1}^n W'_{ij}; \quad W_2 = \sum_{i=1}^n \left( \sum_{j=1}^n W_{ij} \right) \left( \sum_{k=1}^n W'_{ik} \right)$$

$$\text{e } W_3 = \sum_{i=1}^n \sum_{j=1}^n W_{ij} W'_{ij}. \text{ De forma análoga define-se } V_1, \quad V_1', \quad V_2 \text{ e } V_3 \text{ e}$$

$n_r = n(n-1)\dots(n-r+1)$ . Maiores detalhes da demonstração podem ser encontrados em Diggle et al. (1995).

Diggle et al. (1995) afirmam que se a maioria dos valores de  $\hat{R}(s,t)$  estiver fora dos limites de  $\pm 2$  desvios-padrão, então existem indícios de interação espaço-tempo.

A função  $K$  espacial-temporal vem sendo utilizada com frequência em áreas tão diversas do conhecimento como epidemiologia (DIGGLE et al., 1995), ecologia (CARSLAKE et al., 2005) e planejamento urbano (ARBIA et al., 2010). Na área agrícola, Gottwald et al. (2002) analisaram áreas urbanas de Miami afetadas com o cancro cítrico em vários períodos temporais de 30, 60, 90 e 120 dias de duração. Foi observado o padrão agregado para todas as áreas estudadas. Entretanto, esses autores utilizaram a função  $K$  em cada tempo, considerando independência entre os tempos, o que não é recomendado. Assim, salvo melhor juízo, não foram encontrados trabalhos na literatura que utilizaram a função  $K$  espacial-temporal na análise de configurações pontuais de doenças

agrícolas. Assim, esta tese procura avançar um pouco mais sobre o potencial dessa função na análise de dados agrícolas.

## **2.5 Método de varredura de Kulldorff para análise espacial**

O método de varredura espacial idealizado por Kulldorff e Nagarwalla (1995) prioriza uma região formada pelas áreas cujos centroides pertencem a um círculo. Os possíveis agrupamentos dos pontos são constituídos alterando o raio, que variam de zero até o máximo de 50% da população total da área e o centro do círculo. Para cada círculo é calculado a razão entre a verossimilhança sob a hipótese alternativa de que o risco de ocorrência do evento é maior dentro do círculo do que fora do mesmo. Sob a hipótese nula, espera-se que a diferença entre as duas verossimilhanças sejam devido ao acaso.

O método de varredura espacial permite detectar agrupamentos supondo que uma variável de interesse siga um dos modelos de probabilidade: Bernoulli, Poisson, Ordinal, Exponencial e Normal. A escolha do modelo depende da distribuição da variável de interesse em estudo. Para dados oriundos de contagens os modelos indicados são os de Poisson e Permutação espaço-temporal. O modelo ordinal é indicado quando o número de casos apresenta categorias hierárquicas, ou seja, é possível associar o número de casos observados a categorias, como por exemplo, nível baixo, médio e alto. No caso de números contínuos o modelo normal é o recomendado. Quando os casos são de análise de sobrevivência o modelo exponencial é utilizado. Para dados oriundos de processos pontuais, o modelo de Bernoulli deve ser utilizado (KULLDORFF, 2011). As pesquisas desta tese concentram-se neste último modelo.

Para a realização da varredura espacial é importante considerar a notação:  $N$  representa o número total de indivíduos na população,  $C$  é o número total de casos observados na região de estudo,  $Z$  é o conjunto das  $z$  áreas

candidatas a formarem um agrupamento. Estes  $z$  candidatos são círculos de raio  $r$  arbitrários centrados em cada uma das  $n$  coordenadas (ponto central dos círculos) (Figura 3). Estes círculos, em tese, são infinitos, no entanto limitados ao raio que determina que o agrupamento  $z$  contenha no máximo 50% da população total da área (KULLDORFF, 1997).

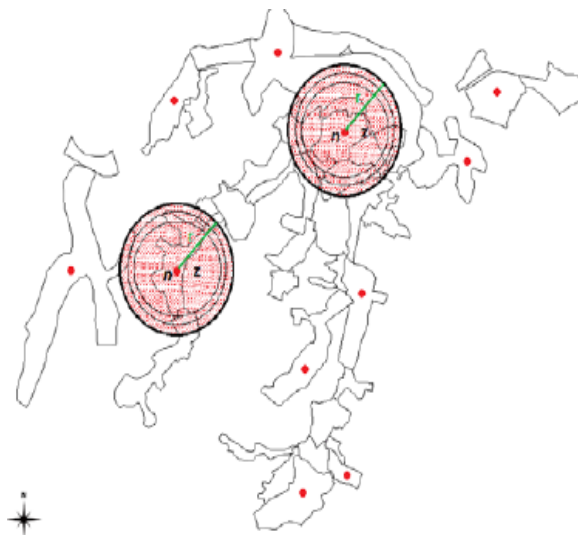


Figura 3 Exemplo hipotético do método de varredura espacial  
Fonte: Balieiro (2008)

Toda descrição para a obtenção da estatística de teste para a varredura espacial, apresentada a seguir, é dada em Kulldorff (1997). Considere uma região  $A$  particionada em pequenas áreas e seja  $c_i$  e  $n_i$  o número de casos observados e o número de indivíduos da área  $i$ ,  $i = 1, 2, \dots, A$ , respectivamente. Sob completa aleatoriedade espacial, o número de casos na  $i$ -ésima unidade pode ser modelado como variáveis aleatórias independentes,  $H_0: c_i \sim \text{Binomial}(n_i, p)$ .



Sob a hipótese nula, o estimador da máxima verossimilhança de  $\tau$  é dada por  $\hat{\tau} = C / N$ , em que  $C$  é o número total de casos na região,  $C = \sum_{i=1}^n c_i$ , e  $N$  é a população total na região  $A$ , incluindo casos e não casos,  $N = \sum_{i=1}^n n_i$ .

A hipótese alternativa assume que existe um agrupamento espacial para uma determinada localização não especificada. Seja  $Z$  um conjunto composto por todos os possíveis agrupamentos espaciais  $z$ . O parâmetro da função de verossimilhança sob a hipótese alternativa inclui o candidato a agrupamento  $z \in Z$ , logo  $p$  é a probabilidade de que um indivíduo dentro de uma zona  $z$  venha ser um caso enquanto que  $q$  é a mesma probabilidade, mas considerando o indivíduo fora da zona  $z$ . Resumidamente, as hipóteses a serem consideradas no teste de razão de verossimilhança são dadas por:  $H_a: p > q$ , sendo  $z \in Z$  e  $H_0: p = q$ .

Assumindo dois vetores de variáveis de Bernoulli independentes, cada um com casos e não casos para as populações dentro e fora do agrupamento  $z$ , a função de máxima verossimilhança usando o modelo de Bernoulli pode ser escrito como:

$$L(\hat{z}, \hat{p}, \hat{q}) = \max_{p > q} p^{n_z} (1-p)^{n_z - c_z} q^{C - c_z} (1-q)^{(N - n_z) - (C - c_z)}, \quad (16)$$

em que  $n_z$  é o número de indivíduos na zona  $z$  e  $c_z$  o valor observado da variável aleatória  $C_z$ , que representa o número de casos na zona  $z$ . Os estimadores da função de máxima verossimilhança para  $p$  e  $q$ , são dados por:  $\hat{p} = c_z / n_z$  e  $\hat{q} = (C - c_z) / (N - n_z)$ .

Para identificar a zona mais provável de ser um agrupamento, dentre todas as possíveis, usa-se a razão de verossimilhança:

$$\tau = \frac{\sup_{z \in Z, p > q} L(\hat{z}, \hat{p}, \hat{q})}{\sup_{p=q} L(\hat{z}, \hat{p}, \hat{q})}, \quad \{p, q \in (0,1)\}. \quad (17)$$

Sob algumas condições a equação (17) pode ser reescrita como

$$\tau = \frac{\sup_z L(\hat{z}, \hat{p}, \hat{q})}{L_0} = \begin{cases} \frac{L(\hat{z}, \hat{p}, \hat{q})}{L_0}, & \text{se } \frac{c_z}{n_z} > \frac{C - c_z}{N - n_z} \\ 1, & \text{se } \frac{c_z}{n_z} \leq \frac{C - c_z}{N - n_z} \end{cases}, \quad (18)$$

em que  $L_0$  e  $L(\hat{z}, \hat{p}, \hat{q})$  são dadas respectivamente por:

$$L_0 = \sup_{p \in (0,1)} p^C (1-p)^{N-C} = \frac{C^C (N-C)^{N-C}}{N^N} \quad (19)$$

e

$$L(\hat{z}, \hat{p}, \hat{q}) = \sup_{p > q} p^{n_z} (1-p)^{n_z - c_z} q^{C - c_z} (1-q)^{(N - n_z) - (C - c_z)}$$

e

$$L(\hat{z}, \hat{p}, \hat{q}) = \begin{cases} \left( \frac{c_z}{n_z} \right)^{n_z} \left( \frac{n_z - c_z}{n_z} \right)^{n_z - c_z} \left( \frac{C - c_z}{N - n_z} \right)^{C - c_z} \left( \frac{N - n_z - (C - c_z)}{N - n_z} \right)^{(N - n_z) - (C - c_z)}, & \text{se } p > q \\ \frac{C^C (N - C)^{N - C}}{N^N}, & \text{se } p \leq q. \end{cases} \quad (20)$$

Para detectar a zona como sendo o agrupamento mais provável é escolhida a zona  $\hat{z}$  para a qual a  $L(\hat{z}, \hat{p}, \hat{q})$  é maximizada.

A distribuição de  $\tau$  depende da distribuição da população e é, virtualmente, impossível de ser obtida analiticamente. A solução deste problema

é dada por Kulldorff e Nagarwalla (1995) através de simulação de Monte Carlo que reproduz a análise para um grande número de repetições aleatórias do conjunto de dados original sob a hipótese nula, de completa aleatoriedade espacial, e condicionada ao número total de casos.

Para cada uma dessas repetições a estatística de teste  $\tau_i$  ( $i = 1, 2, \dots, m$ ) é computada de igual forma para os dados originais. Se o valor de  $\tau$  dos dados originais excede 95% dos valores da simulação Monte Carlo, então, o agrupamento é considerado significativo ao nível de 5% de significância. Os agrupamentos secundários também são obtidos por comparações de  $\tau$  com a distribuição simulada. Em termos gerais, o *valor-p* empírico é dado por  $\text{valor-p} = R / (m + 1)$ , em que  $R$  é a posto da estatística  $\tau$  associado ao conjunto de dados original.

A análise puramente espacial é executável por meio da análise retrospectiva (KULLDORFF et al., 1998). Esta análise visa detectar em um espaço e período fixados tanto os agrupamentos ativos, aqueles que permanecem até a data final do período em estudo, quanto os agrupamentos históricos, aqueles que deixaram de existir antes do final do período em estudo.

A análise de varredura espacial vem sendo muito aplicada em diversas áreas do conhecimento. Entretanto, a maioria dessas aplicações é realizada para dados de contagem por área, utilizando o modelo de Poisson como pode ser visto nos trabalhos de Bhattarai et al. (2010) e Oviedo et al. (2009). A análise de varredura espacial não tem sido utilizada com frequência na análise de dados originários de processos pontuais. Não foi encontrado na literatura, salvo melhor juízo, trabalhos que utilizem a estatística de varredura espacial para a análise de configurações pontuais de doenças agrícolas.

## 2.6 Método de varredura de Kulldorff para análise temporal

Todo o procedimento descrito na Seção 2.5 para a realização da varredura espacial é tomado para a varredura temporal. Neste tipo de análise a varredura é realizada para o intervalo de tempo desejado, sendo que este tempo pode ser agregado, por exemplo, por dia, mês ou ano.

A análise puramente temporal é executável por meio da análise retrospectiva e pela análise prospectiva. Conforme Kulldorff (2001), a análise prospectiva tem como objetivo detectar somente os agrupamentos ativos, ou seja, sua varredura visa somente detectar agrupamentos que não cessaram de ocorrer até a data final do período estudado (são os agrupamentos que nunca deixaram de existir naquele determinado período). O cálculo do *valor-p* é dado pela mesma expressão referenciada na Seção 2.5.

A análise de varredura puramente temporal não vem sendo utilizada, salvo melhor juízo, em processos pontuais e, conseqüentemente, ainda não foi aplicada para a análise de configurações pontuais de doenças agrícolas.

## 2.7 Método de varredura de Kulldorff para análise espaço-temporal

A estatística de varredura espaço-temporal tem como princípio a utilização de uma janela cilíndrica com uma base geográfica circular e altura indicado pelo período de tempo. A base é centrada nos vários centroides da área estudada, com raios variando de tamanho de forma constante. A altura é determinada por um intervalo de tempo menor ou igual à metade do período total estudado. Dessa forma, a janela é deslocada no espaço e no tempo para cada localização geográfica e período de tempos possíveis (Figura 4), tendo como resultados um número infinito de cilindros sobrepostos de tamanhos e formas distintas, cobrindo conjuntamente a região de estudo como um todo,

onde cada cilindro reflete um possível agrupamento (KULLDORFF et al., 1998).

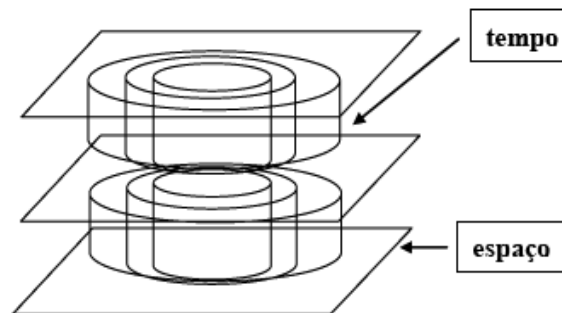


Figura 4 Procedimento da varredura espaço-temporal  
Fonte: Balieiro (2008)

Para cada cilindro o número de pontos dentro e fora dele é verificado juntamente com o número de eventos esperados, com base nestes números, a verossimilhança é calculada para cada cilindro. De modo análogo a estatística  $\tau$ , o cilindro com a máxima verossimilhança e com mais que o número de casos esperados é designado agrupamento mais verossímil (KULLDORFF, 1997; KULLDORFF et al., 1998).

Para testar a significância são geradas várias permutações aleatórias dos casos no espaço e no tempo, misturando a data/tempo e associando ao conjunto inicial de localizações de casos, garantindo que as duas marginais; tanto a espacial quanto a temporal, mantenham-se inalteradas. Logo em seguida, calcula-se a razão de verossimilhança, para cada conjunto simulado, da mesma maneira como foi feito para os dados reais. A significância estatística é avaliada utilizando teste de significância de Monte Carlo, em que a hipótese nula da

inexistência de aglomerados é rejeitada ao nível de 0,05 exatamente quando o *valor-p* simulado for menor ou igual a 0,05 para o agrupamento mais verossímil.

A detecção de agrupamentos no espaço-tempo pode ser feita considerando os modelos probabilísticos citados na Seção 2.5 e pelo modelo de permutação espaço-temporal. A análise espaço-temporal pode ser executada tanto pela análise retrospectiva como pela análise prospectiva.

As vantagens principais da aplicação do teste de varredura espaço-temporal são: procura agrupamentos sem especificar previamente a localização e tamanho; se a hipótese nula (ausência de agrupamentos ou aleatoriedade completa) é rejeitada, o teste fornece a localização do agrupamento mais verossímil que levou a rejeição e, finalmente, evita o problema de testes múltiplos fornecendo um *valor-p* real. Tendo em vista essas vantagens, a análise de varredura espaço-temporal vem sendo muito aplicada em diversas áreas do conhecimento. Mais uma vez, a maioria dessas aplicações é realizada para dados de contagem por área, utilizando o modelo de Poisson, como pode ser visto nos trabalhos de Bhattarai (2010) e Oviedo et al. (2009).

A análise de varredura espaço-temporal não tem sido utilizada com frequência na análise dados originários de processos pontuais. Não foi encontrado na literatura, salvo melhor juízo, trabalhos que utilizem a estatística de varredura espacial-temporal para a análise de configurações pontuais de doenças agrícolas.

## **2.8 A citricultura**

Os métodos estatísticos apresentados anteriormente para análise de processos pontuais no tempo, espaço e espaço-tempo estão sendo pouco ou não utilizados para análises de dados agrícolas. Assim, esta tese pretende explorar o potencial desses métodos nesse tipo de análise. Para tal exploração, escolheu-se

atacar o problema da morte súbita do citros que vem intrigando os pesquisadores da área agrícola. Nesse sentido, neste e na próxima Seção apresenta-se uma breve revisão bibliográfica sobre a citricultura e a morte súbita do citros.

O Brasil detém mais de 50% da produção mundial de suco de laranja e é responsável por 85% das exportações mundiais do produto. Enquanto que os EUA, também grandes produtores de suco de laranja, concentram-se no mercado interno, o Brasil exporta 98% da sua produção. Na safra 2009/10, São Paulo e Triângulo Mineiro (Cinturão Citrícola) produziram 317 milhões de caixas de laranja e a Flórida 134 milhões (CITRUS BR, 2012).

A citricultura brasileira apresenta números expressivos que traduzem a grande importância econômica e social que a atividade tem para a economia do país: a área plantada está ao redor de um milhão de hectares e a produção de frutas supera 19 milhões de toneladas, a maior no mundo há alguns anos. O país é o maior exportador de suco concentrado congelado de laranja cujo valor das exportações, juntamente com as de outros derivados, tem gerado cerca de 1,5 bilhões de dólares anuais (EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA, 2012).

Os principais produtores mundiais de citros têm destinos diferentes para a produção, compondo um *mix* entre processamento industrial (produção de suco), consumo doméstico de fruta *in natura* e exportação de fruta *in natura*. O Brasil destina 70% de sua produção para processamento industrial, sendo que São Paulo e o Triângulo Mineiro destinam 86% de suas produções para esse mercado. No Brasil, nenhuma outra fruta é produzida para fins industriais em volume semelhante (NEVES et al., 2012).

De acordo com Neves et al. (2012) são muitos os desafios internos e externos do setor. Porém, durante a última década, quatro doenças foram responsáveis pela erradicação de 39 milhões de árvores do parque citrícola de São Paulo e Triângulo Mineiro, são elas: o cancro cítrico que é uma doença

bacteriana que causa queda prematura das folhas e frutas; a clorose variegada dos citros (CVC), uma doença que afeta o sistema vascular das árvores reduzindo o tamanho das frutas ao tamanho de uma bola de golfe; a morte súbita dos citros (MSC), uma doença vascular capaz de matar a árvore em 12 meses e finalmente tem-se o *greening*, a mais recente doença bacteriana, considerada a doença mais importante em muitos países.

## 2.9 Morte súbita do citros

Relatos datados de 1997 indicam a ocorrência da morte de muitas laranjeiras doces enxertadas sobre limoeiro Cravo, por motivos desconhecidos, no município do Prata, Minas Gerais. De acordo com Bassanezi et al. (2003), as primeiras observações da doença foram feitas no município de Comendador Gomes, no Triângulo Mineiro em 1999. Entretanto, o primeiro relato oficial dessa doença foi feito em fevereiro de 2001, quando os seus sinais característicos foram descritos (GIMENES-FERNANDES; BASSANEZI, 2001). A partir de então, novos relatos de ocorrência foram surgindo e, em setembro de 2001, devido à rapidez com que as plantas afetadas entravam em declínio, colapso e morte, a doença recebeu o nome de morte súbita dos citros (MSC).

A MSC é uma doença de combinação copa/porta-enxerto (BASSANEZI; YAMAMOTO, 2007) e até o momento, o diagnóstico da doença é realizado com base nos sinais apresentados pelas plantas doentes. Ainda não existe um teste de diagnóstico que permita saber se a planta está com MSC antes que ela manifeste os sinais típicos da doença.

O primeiro sinal da MSC é a perda generalizada do brilho das folhas, que ficam sem viço, de aspecto pálido (Figura 5-a). Pode ocorrer perda de turgidez das folhas, de intensidade variável. Conforme os sinais vão evoluindo,



ocorre desfolha parcial da planta e, em estágio mais avançado, a desfolha quase total e morte da planta (Figura 5-b). Pode ocorrer falta de brotações novas e vigorosas, seca de ponteiros, colapso das plantas e retenção de frutos, principalmente em variedades tardias. A morte do sistema radicular é consequência do bloqueio dos vasos do floema, que transportam a seiva elaborada das folhas para as raízes (BASSANEZI; YAMAMOTO, 2007).



(a)



(b)

Figura 5 Coloração verde pálida a amarela das folhas (a) e planta com sinais da MSC ao lado de uma planta sadia (b)

Fonte: Bassanezi e Yamamoto (2007)

O sistema radicular apresenta grande quantidade de raízes mortas e podres, quase não apresenta radículas que, quando presentes, estão localizadas próximas à base do tronco (Figura 6). Isso acontece porque os vasos do floema do porta-enxerto, que transportam os produtos da fotossíntese para as raízes, ficam bloqueados. No período das chuvas pode ocorrer ligeira emissão de

radicelas, refletindo uma leve recuperação das plantas que, posteriormente, voltam a definhir (FUNDECITRUS, 2012).



Figura 6 Raízes e radicelas de uma planta sadia (esquerda), podridão das raízes e ausência de radicelas de uma planta doente (direita)  
Fonte: Fundecitrus (2012)

A presença de coloração amarelo-alaranjada nos tecidos internos da casca do porta-enxerto, logo abaixo da zona de enxertia, com limite nítido que a separa dos tecidos de coloração branca a creme da casca da copa é o sinal que permite o diagnóstico da MSC (Figura 7) (GIMENES-FERNANDES; BASSANEZI, 2001).

A velocidade de progresso da doença é variável entre variedades, talhões e propriedades. Apesar de variável, a evolução da doença, a partir do aparecimento dos primeiros sintomas, pode ser considerada rápida. Em muitos talhões, após 1 a 2 anos do aparecimento dos primeiros sintomas, de 15 a 30% das plantas já estavam sintomáticas. Na maioria dos casos, a maior parte das

plantas apresenta sintomas iniciais da doença nos primeiros dois anos após o aparecimento das primeiras plantas doentes (FUNDECITRUS, 2012 ).

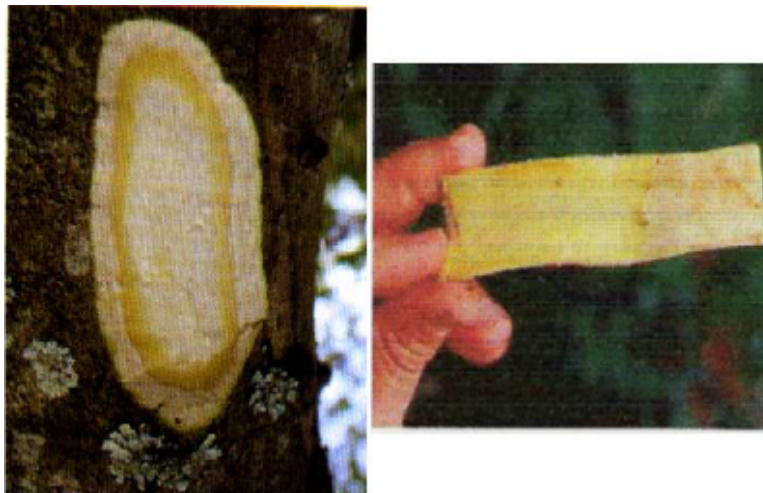


Figura 7 Detalhe da coloração amarela da parte interna da casca da árvore na parte do porta enxerto  
Fonte: Bassanezi e Yamamoto (2007)

Sendo que o padrão de distribuição espacial de plantas com MSC é semelhante ao padrão observado para plantas com CTV (*Citrus tristeza vírus*, família, *Closteroviridae*) (BASSANEZI et al., 2003). Os autores asseguram que as duas doenças afetam plantas com porta-enxertos específicos, o limão ‘Cravo’ no caso da MSC e a laranja ‘Azeda’ no caso da tristeza, sendo que, com a troca dos porta-enxertos, as plantas se recuperam e as duas doenças apresentam semelhanças na sintomatologia, progresso das doenças e distribuição espacial das plantas doentes.

Entretanto, o agente causal da doença ainda não foi confirmado. Os pesquisadores da *Allelyx Applied Genomics* identificaram o vetor responsável

pela transmissão do vírus da morte súbita dos citros, ou vírus da morte súbita dos citros, um novo tipo de vírus da família *Tymoviridae* descoberto em 2003 por esta empresa e que pode estar associado à MSC (TERENCIO; MILORI; MARTIN-NETO, 2005). O vírus foi encontrado em três espécies de pulgões (*Toxoptera citricida*, *Aphis spiraecola* e *Aphis gossypii*) coletados nas regiões norte de São Paulo e sul do Triângulo Mineiro. Porém, a pesquisa ainda não foi encerrada e não está totalmente comprovado que este vírus seja realmente o agente causal da MSC.

Segundo Bassanezi et al. (2003), estudos sobre o padrão espacial e o comportamento, ou o progresso, da MSC ao longo do tempo fornecem informações importantes que podem auxiliar no esclarecimento da etiologia da doença. Portanto, o estudo da dinâmica espacial e temporal da MSC é muito importante para que se possa auxiliar na descrição da dinâmica da doença, levantar ou avaliar hipóteses biológicas sobre mecanismos de propagação, formular ou dar suporte a hipóteses levantadas em relação à sua etiologia.

### 3 MATERIAL E MÉTODOS

Nesta seção são descritos a metodologia e os dados utilizados neste trabalho.

#### 3.1 Cenários típicos simulados de configurações pontuais no espaço e tempo

Neste tópico serão apresentados quatro cenários típicos de processos pontuais no espaço e tempo com o intuito de permitir um entendimento mais didático dos métodos apresentados anteriormente e dos dois novos métodos que serão propostos no capítulo 4. O cenário 1 é caracterizado por apresentar agrupamento no espaço, sem agrupamento no tempo e sem interação espaço-tempo. O cenário 2 não apresenta agrupamento no espaço, apresenta agrupamento no tempo e não apresenta interação espaço-tempo. O cenário 3 é caracterizado por não apresentar agrupamento no espaço e nem no tempo e nenhuma interação espaço-tempo. Finalmente, o cenário 4 constitui-se de um processo pontual que exhibe agrupamento no espaço e no tempo e interação espaço-tempo. Um resumo desses cenários é apresentado na Tabela 1.

Tabela 1 Cenários propostos de processos pontuais espaço-temporal.

Cenários	Agrupamento no espaço	Agrupamento no tempo	Interação espaço-tempo
1	Sim	Não	Não
2	Não	Sim	Não
3	Não	Não	Não
4	Sim	Sim	Sim

Deve-se observar que esses cenários típicos apresentados na Tabela 1 não representam todas as possibilidades de processos pontuais no espaço e tempo, porém, são suficientes para mostrar, de forma didática, como os métodos apresentados nesta tese podem ser utilizados e interpretados.

Para a simulação dos quatro cenários de processos pontuais no espaço e tempo foi considerado que em um período de 365 dias ocorreram 400 casos de uma particular doença, distribuídos em um plano quadrado de  $100 \times 100$  m, com área total de  $10000 \text{ m}^2$ . Esses cenários estão representados nos gráficos de 1 a 4.

Deve-se observar que, apenas para uma melhor visualização do comportamento espaço-temporal dos cenários, os dados de ocorrência em todos os cenários foram agrupados em quatro trimestres. Assim, no quarto trimestre sempre terá 400 pontos.

O gráfico 1 apresenta uma realização de um processo pontual com agrupamento no espaço, sem agrupamento no tempo e sem interação espaço-tempo. Para simular esse processo, primeiramente, geram-se dois pontos principais localizados, aproximadamente, nos centros do primeiro e quarto quadrantes. Esses dois pontos serão os centros dos agrupamentos espaciais. Em seguida, cada ponto principal é substituído por um número aleatório de pontos que segue uma distribuição de Poisson homogêneo com intensidade suficiente para obter agrupamentos com um número de pontos aproximadamente igual ao número de ocorrências no referido trimestre. Nesta realização foram gerados 50 pontos, para cada ponto principal, por trimestre.

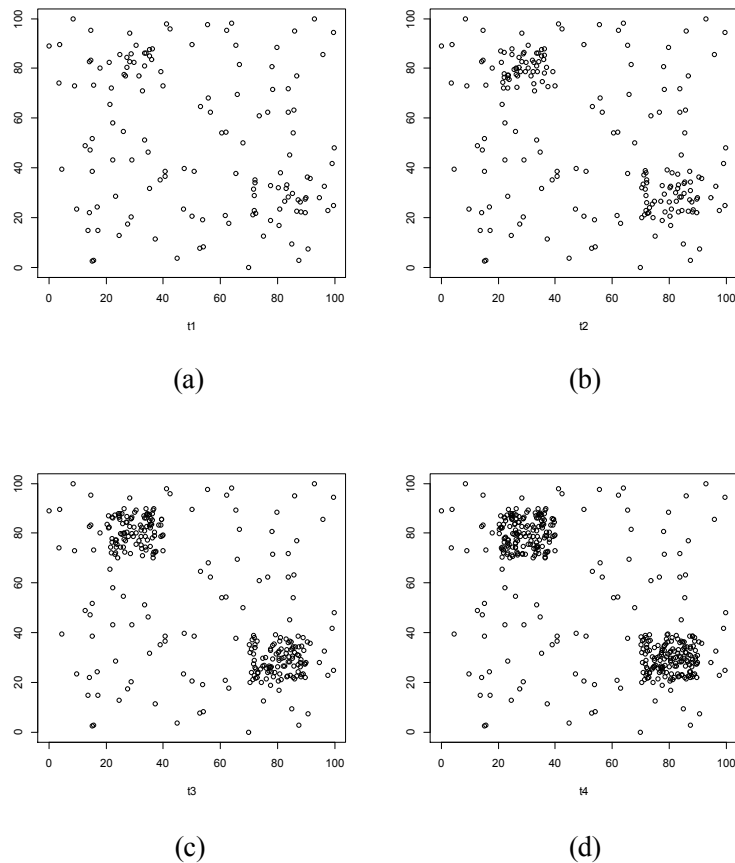


Gráfico 1 Cenário 1: processo com agrupamento no espaço, sem agrupamento no tempo e sem interação. (a) 1<sup>o</sup> trimestre, (b) 2<sup>o</sup> trimestre, (c) 3<sup>o</sup> trimestre e (d) 4<sup>o</sup> trimestre

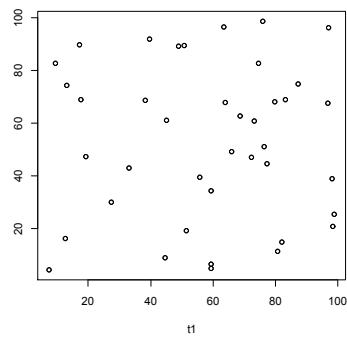
O gráfico 2 apresenta um cenário de um processo pontual que não apresenta agrupamento no espaço, apresenta agrupamento no tempo e não apresenta interação espaço-tempo. Para obter uma realização desse processo, basta gerar um número aleatório de pontos que segue uma distribuição de Poisson homogêneo com intensidade suficiente para obter um número de pontos

aproximadamente igual ao número de ocorrências no trimestre desejado. Nesta realização foram gerados um número crescente de pontos, por trimestre, de tal forma que no último trimestre tenham sido adicionados, aproximadamente, 400 pontos. Isso garante um agrupamento no final do período de tempo sem agrupamento no espaço e sem interação espaço-tempo.

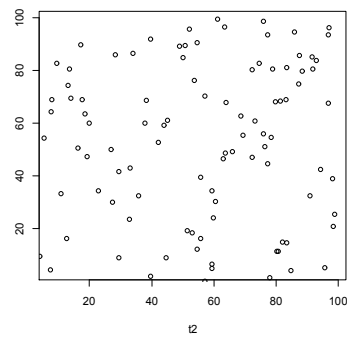
O gráfico 3 apresenta um cenário de processo pontual que não apresenta agrupamento no espaço, não apresenta agrupamento no tempo e não apresenta interação espaço-tempo. Para obter uma realização desse processo, basta gerar um número aleatório de pontos que segue uma distribuição de Poisson homogêneo com intensidade suficiente para obter um número de pontos aproximadamente igual ao número de ocorrências no trimestre desejado. Nesta realização foram gerados 100 pontos, por trimestre. Isso garante uma realização que não apresenta agrupamentos no tempo, sem agrupamento no espaço e sem interação espaço-tempo. Observe que apesar de a realização apresentada no gráfico 3 ser semelhante, em aparência, a realização mostrada no gráfico 2, os processos pontuais que as originaram, conforme mostrado, são completamente diferentes.

Finalmente, o gráfico 4 representa o cenário de uma realização de um processo pontual com agrupamento no espaço e no tempo e ainda com interação espaço-tempo. Para simular esse processo, no primeiro trimestre gera-se um ponto principal localizado, aproximadamente, no centro do primeiro quadrante. Em seguida, esse ponto principal é substituído por 100 pontos que segue uma distribuição de Poisson homogêneo com intensidade suficiente para obter um agrupamento nesse quadrante. Na configuração obtida no primeiro trimestre é adicionada uma nova configuração gerada de modo análoga a primeira configuração, com a diferença que ponto principal é localizado, aproximadamente, no centro do segundo quadrante. O processo continua até que se obtenha os 400 pontos distribuídos em quatro agrupamentos.

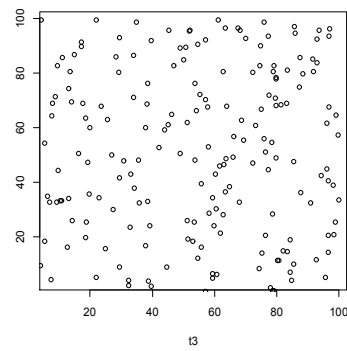




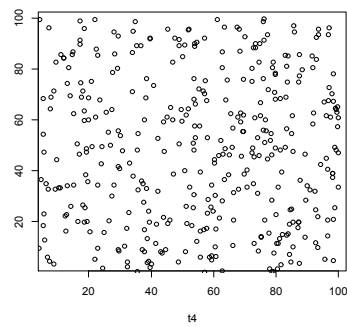
(a)



(b)

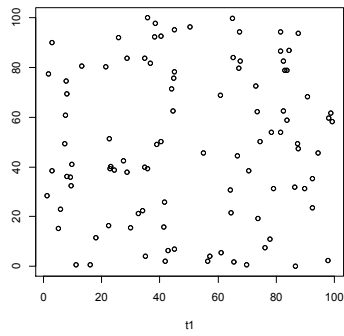


(c)

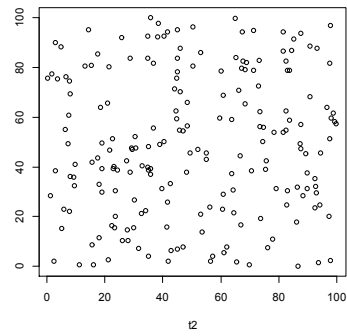


(d)

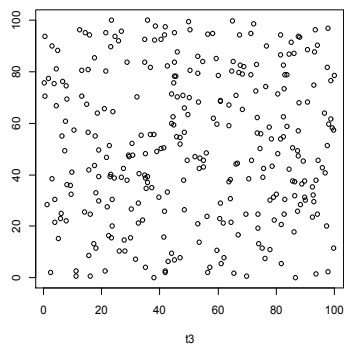
Gráfico 2 Cenário 2: processo sem agrupamento no espaço, com agrupamento no tempo e sem interação espaço-tempo. (a) 1º trimestre, (b) 2º trimestre, (c) 3º trimestre e (d) 4º trimestre



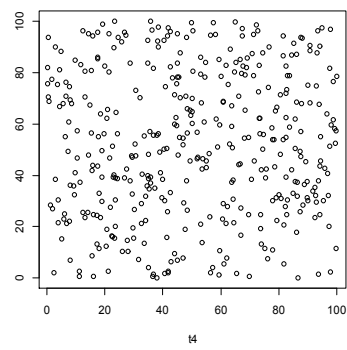
(a)



(b)



(c)



(d)

Gráfico 3 Cenário 3: processo sem agrupamento no espaço, sem agrupamento no tempo e sem interação espaço-tempo. (a) 1º trimestre, (b) 2º trimestre, (c) 3º trimestre e (d) 4º trimestre

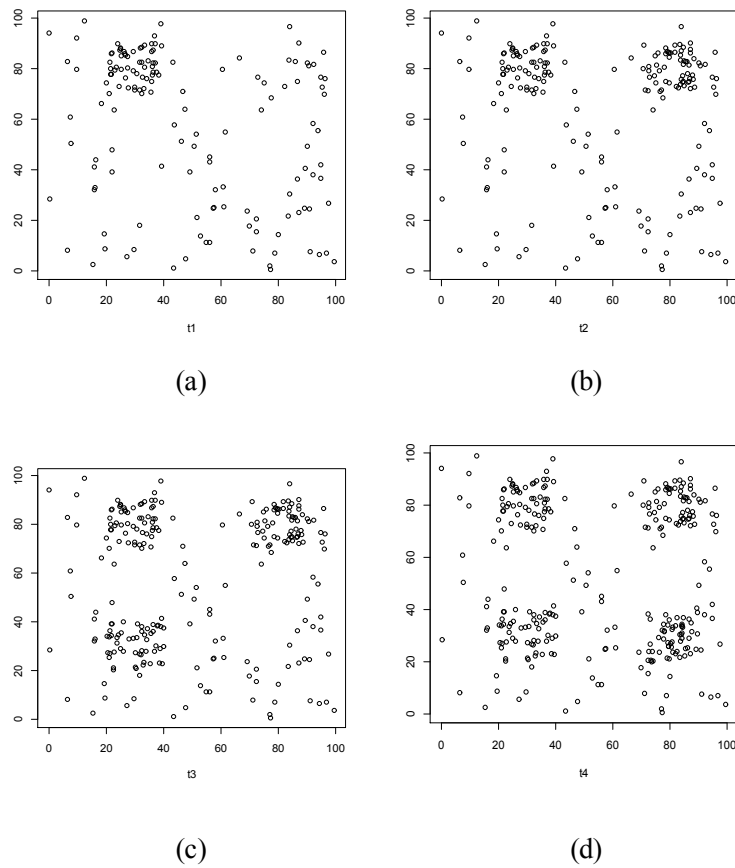


Gráfico 4 Cenário 4: processo com agrupamento no espaço, agrupamento no tempo e interação espaço-tempo. (a) 1º trimestre, (b) 2º trimestre, (c) 3º trimestre e (d) 4º trimestre

### 3.2 Dados de morte súbita dos citros

Os dados de morte súbita do citros (MSC) são provenientes de um talhão de laranjeira doce (*Citrus sinensis* L. Osbeck) da variedade “Natal”. O experimento foi conduzido na fazenda Vale Verde situada no município de Comendador Gomes/MG.

O talhão era constituído de 20 fileiras com 48 plantas por fileira as coordenadas  $x$  e  $y$  são dadas em metros. O espaçamento entre linhas de cultivo é 7,5 metros e nas linhas é de 4 metros entre plantas. Foram realizadas 18 avaliações correspondentes ao período de novembro de 2001 a julho de 2003, sendo que o ano de plantio foi em 1990. O Gráfico 5 mostra um croqui da representação, aproximada, do posicionamento das plantas no talhão e não tem a intenção de apresentar as coordenadas exatas de cada planta.

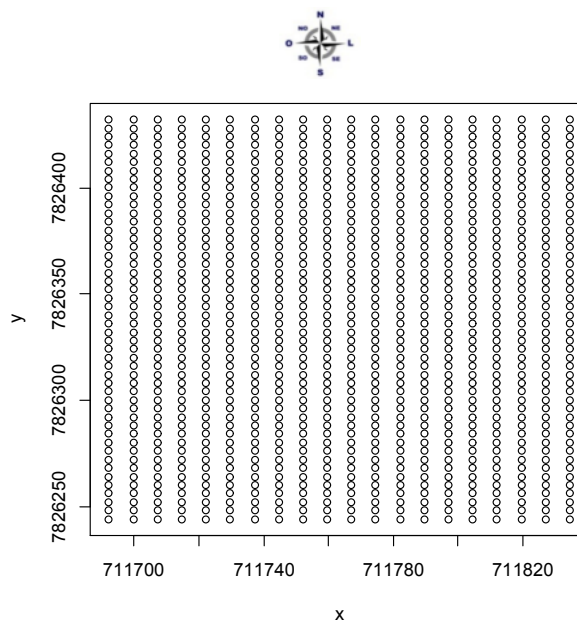


Gráfico 5 Croqui do talhão experimental

Os dados da incidência de MSC fazem parte de um conjunto de dados coletados pelo FUNDECITRUS e estão dispostos no pacote Rcitrus. Esse pacote foi desenvolvido em uma estrutura de pacote adicional da linguagem R (R DEVELOPMENT CORE TEAM, 2012), implementado e adaptado para metodologias de análise estatística para dados de incidência de doenças em plantas. Seu desenvolvimento foi motivado pela necessidade de automatizar a análise estatística de dados de MSC (KRAINSKI et al., 2005).

Os autores usaram os seguintes códigos para identificar o *status* da doença dentro do talhão: 0, 1, 2, e 3, indicando, respectivamente: planta sadia, estado inicial da doença, estado avançado da doença e morte da planta. Neste trabalho considerou-se 0 para identificar que a planta estava sadia e 1 planta doente (englobando os *status* 1, 2 e 3).

### **3.3 Análise exploratória da morte súbita dos citros**

Com o objetivo de observar o comportamento geral dos dados de MSC, construiu-se um diagrama de pontos representando a localização mensal dos casos de MSC no talhão e um gráfico de barras para o total de casos mensais de incidência da doença.

### **3.4 Análise de completa aleatoriedade espacial usando a função $K$**

Para testar a hipótese de completa aleatoriedade espacial para cada cenário simulado foram realizadas 1000 simulações Monte Carlo tanto para obter os *valores-p* das estatísticas de teste quanto para obter os envelopes de confiança. As escalas espaciais ( $s$ ) atribuídas para a geração dos valores de  $\hat{K}(s)$  foram de 45 valores, igualmente espaçados, entre 0 e 60 metros.

Para a análise dos dados de MSC foram realizadas 1000 simulações de Monte Carlo. Foram adotados 30 valores, igualmente espaçados, entre 0 e 45 metros.

A escolha desses valores que foram utilizados na análise foi feita tendo como base os estudos empíricos de Diggle (2003) e evidências empíricas obtidas neste estudo que possibilitaram indicar que tais valores permitem uma descrição adequada dos processos espaciais pontuais analisados.

### **3.5 Análise de completa aleatoriedade temporal usando a função $K$**

Para testar a hipótese de completa aleatoriedade temporal para cada cenário simulado foram realizadas 1000 simulações de Monte Carlo tanto para obter as estatísticas de teste e os respectivos *valores-p* quanto para obter os envelopes de confiança. As escalas temporais ( $t$ ) atribuídas para a geração dos valores de  $\hat{K}(t)$  foram de 45 valores, igualmente espaçados, entre 0 e 90 dias.

Para a análise dos dados de MSC, foram realizadas 1000 simulações de Monte Carlo. Para as escalas temporais ( $t$ ), utilizadas para o cálculo de  $\hat{K}(t)$ , foram adotados 30 valores, igualmente espaçados, entre 0 e 10 meses.

A escolha desses valores utilizados na análise foi feita tendo como base evidências empíricas obtidas a partir deste estudo que possibilitaram indicar que tais valores permitem uma descrição adequada dos processos temporais pontuais analisados.

### **3.6 Análise gráfica espaço-temporal usando a função $K$**

Para a construção dos gráficos descritos na Seção 2.4 para os cenários simulados utilizou-se como escalas temporais e espaciais, respectivamente, 45

valores, igualmente espaçados, compreendidos entre 0 e 60 metros e 45 valores, igualmente espaçados, compreendidos entre 0 e 90 dias.

Para a análise dos dados da MSC utilizou-se como escalas temporais e espaciais, respectivamente, 30 valores, igualmente espaçados, entre a distância de 0 e 45 metros e 30 valores, igualmente espaçados, entre 0 e 10 meses.

A escolha desses valores utilizados nesta análise foi feita tendo como base os estudos empíricos de Diggle et al. (1995) e evidências empíricas obtidas neste estudo que possibilitaram indicar que esses valores permitem uma descrição adequada dos processos espaciais e temporais pontuais analisados.

### **3.7 Análise espacial usando método de varredura de Kulldorff**

Para a análise de varredura espacial, foi considerada a análise retrospectiva com um raio de busca de zero até o máximo de 50% da população total da área. De acordo com a configuração dos dados adotou-se o modelo Bernoulli. Para a obtenção dos *valores-p* e da estatística de teste realizaram-se 999 simulações Monte Carlo. Esta forma de executar as análises foi adotada tanto para analisar as simulações quanto para os dados de MSC. Foi adotado um nível de significância  $\alpha = 0,05$ . Esses valores foram escolhidos tendo como base os trabalhos de Kulldorff (1997, 2011).

### **3.8 Análise temporal usando método de varredura de Kulldorff**

Na referida análise utilizou-se o modelo probabilístico Bernoulli, a análise retrospectiva com um raio de busca de zero até o máximo de 50% da população total da área. Como a varredura temporal exige que as informações temporais estejam agregadas para um dado intervalo de tempo, adotou-se a agregação por dia para os dados simulados e mês para os dados de MSC.

Para a obtenção dos *valores-p* e da estatística de teste realizaram-se 999 simulações Monte Carlo. Foi adotado um nível de significância  $\alpha = 0,05$ . Esses valores foram escolhidos tendo como base os trabalhos de Kulldorff (1997, 2011).

### **3.9 Análise espaço-temporal usando método de varredura de Kulldorff**

A análise de varredura espaço-temporal foi realizada utilizando a análise retrospectiva com o modelo Bernoulli e raio de busca de zero até o máximo de 50% da população total da área.

Da mesma forma que as análises temporais, as análises espaço-temporal envolve a informação do tempo em que os dados foram obtidos. Assim, para os dados simulados adotou-se o dia e para os dados de MSC o mês.

Para a obtenção dos *valores-p* e da estatística de teste realizaram-se 999 simulações Monte Carlo. Foi adotado um nível de significância  $\alpha = 0,05$ . Esses valores foram escolhidos tendo como base sugestões apresentadas nos trabalhos de Kulldorff (1997, 2011).

### **3.10 Softwares**

Nas análises utilizando a função  $K$ , nos cálculos dos envelopes e nos testes da hipótese de completa aleatoriedade espacial e temporal, foram utilizadas funções desenvolvidas e/ou disponíveis no *software* R (R DEVELOPMENT CORE TEAM, 2012) conjuntamente com as bibliotecas *splancs* (ROWLINGSON; DIGGLE, 1993), *Rcitrus* (KRAINSKI; RIBEIRO JÚNIOR, 2007) e *spatstat* (BADDELEY; TURNER, 2005). Para realizar as varreduras espaciais, temporais e espaço-temporais utilizou-se *software* SaTScan



(KULLDORFF, 2011). As funções desenvolvidas estão disponíveis nos anexos de A a G.

## 4 RESULTADOS E DISCUSSÃO

Este capítulo tem início com a apresentação de um novo teste de hipótese para detectar agrupamentos no tempo, utilizando a função distribuição dos tempos de ocorrência dos casos, conforme descrito na Seção 4.1.

Na Seção 4.2 são apresentados e discutidos os resultados obtidos a partir da aplicação deste novo teste nos quatro cenários típicos simulados (veja Seção 3.1).

Na Seção 4.3 é apresentado um novo teste de hipótese para detectar interação espaço-tempo em processos pontuais, enquanto na Seção 4.4 este novo teste é aplicado e discutido nos quatro cenários típicos simulados.

Na Seção 4.5 apresentam-se os resultados obtidos a partir da aplicação dos métodos baseados na função  $K$  descritos nas seções 2.2-2.4 nos quatro cenários típicos simulados.

Na Seção 4.6 apresentam-se os resultados obtidos a partir da aplicação dos métodos baseados na estatística de varredura descritos nas seções 2.5-2.7 nos quatro cenários típicos simulados.

As seções 4.5 e 4.6 têm por objetivo mostrar como esses métodos se comportam em situações bem controladas (simuladas) para permitir um entendimento mais didático dos mesmos. Evidentemente, esses exemplos simulados não representam uma exaustiva situação de todas as possibilidades possíveis de acontecer na prática, mas apenas algumas situações típicas que, espera-se, possam fornecer uma ideia do funcionamento dos métodos apresentados.

Na Seção 4.7 apresentam-se os resultados e as discussões obtidas a partir da aplicação de todos os métodos apresentados nesta tese (disponíveis na literatura e novos propostos nesta tese) nos dados de morte súbita do citros.

Finalmente, na Seção 4.8 são apresentadas algumas observações sobre os métodos apresentados nesta tese e algumas possibilidades de trabalhos futuros.

#### **4.1 Novo teste para a hipótese de completa aleatoriedade temporal usando a função distribuição de ocorrência de eventos**

Nesta Seção é proposto um método alternativo para testar a hipótese de completa aleatoriedade temporal (CAT) utilizando a função distribuição das contagens dos tempos de ocorrência de casos de um determinado evento. Nesta tese, o evento é a ocorrência de um caso de doença na planta. Assim, definem-se as seguintes hipóteses:

$H_0$ : Existe CAT na ocorrência dos eventos.

$H_1$ : Não existe CAT na ocorrência dos eventos.

Para testar  $H_0$  tem-se que, primeiramente, definir duas funções: uma função que representa a distribuição acumulada observada dos tempos de ocorrência de eventos e uma função que representa a distribuição acumulada teórica dos tempos de ocorrência de eventos sob a hipótese nula.

Para obter a função que represente a distribuição acumulada teórica, considere  $t$  como sendo uma variável que representa o tempo de ocorrência dos eventos. Suponha que o experimento começa em  $t = 0$ . Os eventos ocorrem aleatoriamente, em que o primeiro caso ocorre em  $T_1$ , o segundo em  $T_2$  e assim por diante. A variável aleatória  $T_i$  denota o tempo em que o  $i$ -ésimo evento ocorre. Os valores  $t_i$ , assumidos pelas realizações de  $T_i$ , são chamados de tempos de ocorrência dos eventos.

Um processo estocástico  $\{N(t); t \geq 0\}$  é chamado de processo de contagem, se  $N(t)$  representa o número de eventos total ocorridos no intervalo  $(0, t)$ . Esse processo, segundo Daley e Vere-Jones (2003), deve satisfazer as seguintes propriedades:

1.  $N(t) \geq 0$  e  $N(0) = 0$ .
2.  $N(t)$  é um número inteiro.
3.  $N(s) \leq N(t)$  se  $s < t$ .
4.  $N(t) - N(s)$  é igual ao número de eventos que ocorreram no intervalo  $(s; t)$ .

Ainda segundo Daley e Vere-Jones (2003), um processo de contagem  $N(t)$  é dito ter incrementos independentes se o número de casos ocorridos, em intervalos de tempo disjuntos, é independente. Um processo de contagem  $N(t)$  é dito ter incrementos estacionários se o número de eventos no intervalo  $(s + h; t + h)$  tem a mesma distribuição do número de eventos  $(s; t)$ , para todo  $s < t$  e  $h > 0$ .

Um processo de contagem  $N(t)$ , segundo Daley e Vere-Jones (2003), é dito ser um processo de Poisson homogêneo com intensidade  $\lambda > 0$  se:

1.  $N(0) = 0$ .
2.  $N(t)$  tem incrementos estacionários independentes.
3.  $\lim_{h \rightarrow 0} \frac{P[N(t+h) - N(t) = 1]}{h} = \lambda$ .
4.  $\lim_{h \rightarrow 0} \frac{P[N(t+h) - N(t) \geq 2]}{h} = 0$ .

Daley e Vere-Jones (2003) mostram que em um processo de Poisson homogêneo, tem-se que  $E[N(t)] = \lambda t$  e  $V[N(t)] = \lambda t$ . Portanto, o valor esperado do número de eventos no intervalo unitário  $(0; 1)$ , ou em qualquer outro intervalo de tamanho unitário, é igual a  $\lambda$ .

O parâmetro intensidade  $\lambda$  é o valor esperado de ocorrências de casos em certo intervalo de duração  $t$ , dividido por  $t$ , dado por  $\lambda = \frac{E[N(t)]}{t}$ . Assim,  $\lambda$  pode ser interpretado como taxa média do número de ocorrências de casos por unidade de tempo.

Outra propriedade muito importante do processo de Poisson homogêneo é que os tamanhos dos intervalos de tempo  $T_i - T_{i-1}$  desse processo  $N(t)$ , com intensidade  $\lambda$ , são variáveis aleatórias exponenciais com taxa  $\lambda$  independentes entre si. Também, o número de casos que ocorrem em um intervalo de tamanho  $t$  em um processo de Poisson homogêneo é uma variável aleatória discreta de Poisson com taxa  $\lambda t$  (DALEY; VERE-JONES, 2003).

Para definir a função que representa a função distribuição observada dos tempos de ocorrência de eventos basta notar que esta função pode ser obtida diretamente da distribuição acumulada das contagens dos tempos de ocorrência dos eventos. Para exemplificar como esta função distribuição, denotada por  $F(t)$ , pode ser obtida, considere um vetor de tempos de ocorrência de eventos ( $t$ ), dado por  $t = [1, 2, 2, 2, 2, 2, 3, 3, 4, 5, 5]$ . Observe que neste vetor, no tempo 1 ocorreu apenas um evento, enquanto no tempo 2 ocorreram 5 eventos e, assim, sucessivamente, conforme pode ser visto na coluna “Contagem” da Tabela 2.

Tabela 2 Contagens e função de distribuição observada ( $F(t)$ ) para o vetor de tempos de ocorrência dos eventos  $t = [1, 2, 2, 2, 2, 2, 3, 3, 4, 5, 5]$ .

Tempo	Contagem	$F(t)$
0	0	0
1	1	1
2	5	6
3	2	8
4	1	9
5	2	11

A Tabela 2 também apresenta a função de distribuição dos tempos de ocorrência de eventos,  $F(t)$ , para  $t = 1, 2, \dots, 5$ . Observe que no tempo inferior a 1 (ou seja  $t=0$ ) não tem ocorrências de eventos, no tempo inferior a dois tem-se uma ocorrência, no tempo inferior a 3 tem-se seis ocorrências e, assim sucessivamente, até que no tempo inferior a 6 temos todas as ocorrências, ou seja, 11 ocorrências de eventos.

Os resultados anteriores mostram que a função distribuição observada  $F(t)$  é uma realização de um processo de contagem. Assim, para testar  $H_0$  tem-se que medir a discrepância entre  $F(t)$  e  $N(t)$  sob a suposição da hipótese nula de CAT. Para tal, deve-se recordar que se  $N(t)$  segue um processo de Poisson homogêneo, então sua média é dada pela reta  $\lambda t$  que passa pela origem e apresenta inclinação  $\lambda$ . Assim, qualquer inclinação de  $F(t)$  que se afaste da inclinação  $\lambda$  fornece indícios para rejeitar a hipótese nula de CAT na ocorrência dos eventos.

Tendo como base o que foi apresentado até o momento, podem-se sugerir os seguintes passos para testar a hipótese nula  $H_0$  de que existe CAT na ocorrência dos eventos.

- a) Define-se  $q_j$  como a medida da discrepância entre  $F(t)$  e  $\hat{\lambda}t$  para  $t = 1, 2, \dots, p$ , em que  $p$  é o tempo que representa a última ocorrência de um

evento e  $\hat{\lambda}$  é o estimador de  $\lambda$ . O estimador  $\hat{\lambda}$ , pode ser obtido facilmente a partir da razão entre o “número de ocorrências do evento no intervalo de tempo” e o “intervalo de tempo”. Essa medida pode ser obtida pela distância de Kolmogorov-Sminorv dada por:

$$q_k = \max_{1 \leq t \leq p} |F(t) - \hat{\lambda}t|; \quad (21)$$

b) Calcula-se  $q_1$ , em que  $q_1$  é a medida de discrepância entre os valores observados de  $F(t)$  e  $\hat{\lambda}t$ . Neste ponto deve-se observar que  $q_k$  não tem distribuição amostral conhecida. Assim, para obter o teste de hipótese recorre-se ao método de Monte Carlo que é descrito nos passos seguintes;

c) Calcula-se  $q_k$  para  $k = 2, 3, \dots, m$ , em que  $m$  é o número total de simulações e  $q_k$  é a medida de discrepância entre os valores simulados de  $F(t)$ , sob a hipótese de CAT, e  $\hat{\lambda}t$ ;

d) Ordena-se  $q_k$ . Se  $q_1$  ocorrer em um dos extremos da distribuição dos  $q_k$  para  $k = 2, 3, \dots, m$ , então existem evidências estatísticas para rejeitar a hipótese de CAT. Suponha que,  $q_{(1)} = q_{(j)}$  para algum  $j \in \{1, \dots, m\}$ , então se rejeita a hipótese de independência se,  $valor - p = ((m + 1 - j) / m) \leq \alpha$ , em que  $\alpha$  é o nível de significância.

Uma maneira de visualizar a existência (ou não) de CAT na ocorrência de eventos e também verificar o período de tempo que ocorre (ou não) a existência de CAT é utilizar um procedimento gráfico. Para tal, nesta tese sugere-se construir o gráfico de  $F(t)$  e  $\hat{\lambda}t$  contra os valores de tempo  $t = 1, \dots, p$ .

Para realizar a interpretação do gráfico proposto deve-se lembrar de que a função distribuição teórica é representada por uma reta que passa pela origem e apresenta inclinação  $\hat{\lambda}$ . Assim, qualquer inclinação de  $F(t)$  que se afaste da

inclinação da reta  $\hat{\lambda}_t$  fornece indícios para rejeitar a hipótese nula de CAT na ocorrência dos eventos.

Para uma melhor interpretação do novo método proposto para testar a hipótese nula de CAT, consideram-se quatro casos típicos de ocorrências de eventos em um intervalo de tempo: (a) presença de CAT em todo o período de tempo; (b) presença de agrupamento temporal (não existência de CAT) no início do período de tempo; (c) presença de agrupamento temporal (não existência de CAT) na metade do período de tempo e (d) presença de agrupamento temporal (não existência de CAT) no final do período de tempo.

Para obter os quatro casos típicos, considerou-se um vetor composto de 400 tempos de ocorrências de eventos distribuídos dentro do período de um ano, ou seja, (0, 365). As estatísticas de teste e os respectivos valores- $p$  foram obtidos a partir de 999 simulações ( $m=1000$ ) Monte Carlo sob a suposição de CAT, ou seja, sob a suposição de um processo pontual de Poisson homogêneo com parâmetro  $\lambda=1,9$ . Os resultados obtidos são apresentados no Gráfico 6.

O primeiro caso típico refere-se a uma realização que apresenta CAT em todo intervalo de tempo. O teste de hipótese mostrou que não existem evidências estatísticas para rejeitar a hipótese nula de CAT ( $q_l = 15,3014$ ;  $p = 0,586$ ). Pelo Gráfico 6-a percebe-se que a função de distribuição observada encontra-se próxima da reta que representa a hipótese de CAT e ambas apresentam, aproximadamente a mesma inclinação. Assim, com a análise do Gráfico 6-a observa-se visualmente que a hipótese de CAT não deve ser rejeitada.

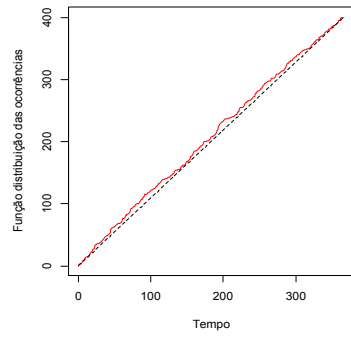
O segundo caso típico representa uma realização pontual cujo agrupamento temporal ocorre no início do período de tempo. O teste de hipótese mostrou que existem fortes evidências estatísticas para rejeitar a hipótese nula de CAT ( $q_l = 202,411$ ;  $p = 0,001$ ). Observe que no Gráfico 6-b, a função de distribuição observada encontra-se acima da reta que representa a hipótese nula de CAT. Além disso, observe que taxa média de ocorrências sobe até o tempo



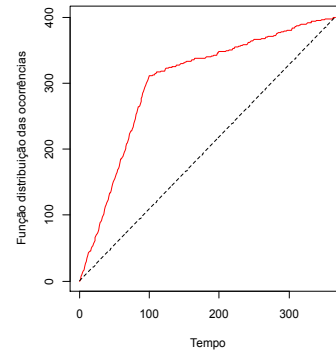
100 (dada pela inclinação positiva da reta) e depois essa taxa decresce (dada pela inclinação negativa da reta).

O terceiro caso típico representa uma realização pontual cujo agrupamento temporal ocorre no meio do período de tempo. O teste de hipótese mostrou que existem fortes evidências estatísticas para rejeitar a hipótese nula de CAT ( $q_1 = 80,8219$ ;  $p = 0,001$ ). O Gráfico 6-c exibe uma forma típica de um processo pontual cujo agrupamento ocorre no meio do intervalo de tempo. Note que nesse gráfico a função de distribuição observada inicia-se abaixo da reta e termina acima da reta que representa a hipótese nula de CAT. Observe que nos períodos de tempo (0, 150) e (200, 365) a taxa média de ocorrências decresce (dada pela inclinação negativa da reta), enquanto no período de tempo entre 150 e 200 a taxa média de ocorrências cresce (dada pela inclinação positiva da reta).

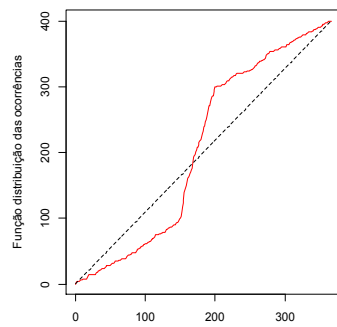
O quarto caso típico representa uma realização pontual cujo agrupamento temporal ocorre no final do período de tempo. O teste de hipótese mostrou que existem fortes evidências estatísticas para rejeitar a hipótese nula de CAT ( $q_1 = 79,8630$ ;  $p = 0,001$ ). O Gráfico 6-d representa uma forma típica de um processo pontual cujo agrupamento temporal ocorre no final do período de tempo observado. Note que nesse gráfico a função de distribuição observada encontra-se abaixo da reta que representa a hipótese nula de CAT. Além disso, observe que taxa média de ocorrências decresce até o tempo 300 (dada pela inclinação negativa da reta) e a partir desse tempo a taxa começa a crescer (dada pela inclinação positiva da reta).



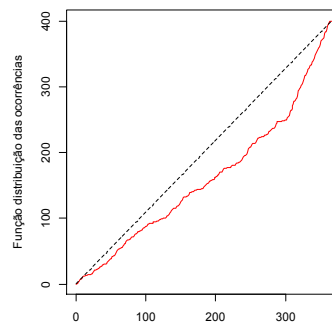
(a)



(b)



(c)



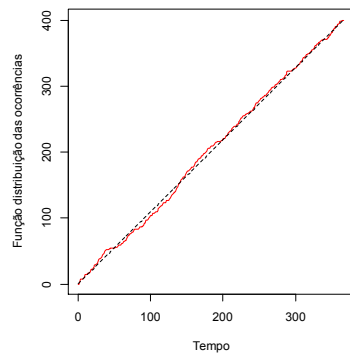
(d)

Gráfico 6 Função de distribuição observada (linha contínua em vermelho), função de distribuição teórica sob a suposição de completa aleatoriedade temporal (linha tracejada em preto) para as quatro realizações típicas

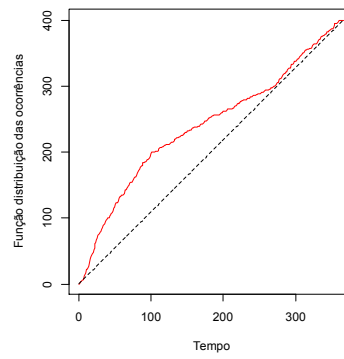
#### **4.2 Análise para completa aleatoriedade temporal nos cenários típicos simulados usando o novo teste**

Nesta Seção serão apresentados os resultados obtidos a partir da aplicação do teste proposto na Seção 4.1 nos quatro cenários típicos simulados apresentados na Seção 3.1. Para testar a hipótese nula de não existência de completa aleatoriedade temporal para cada configuração simulada, foram realizadas 999 simulações Monte Carlo sob a suposição de CAT (processo pontual de Poisson homogêneo com parâmetro  $\lambda=1,9$ ).

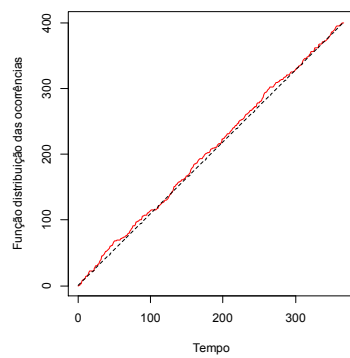
O primeiro e o terceiro cenário foram simulados para caracterizar ausência de agrupamentos temporais e o teste não rejeitou a hipótese nula de CAT nessas realizações ( $q_1 = 10,7260$ , *valor-p* = 0,901;  $q_1 = 12,68493$ , *valor-p* = 0,7765). Os Gráficos 7-a e 7-c comprovam visualmente estes resultados onde se pode observar que a função distribuição dos casos observados encontra-se muito próxima da reta que representa a hipótese nula de CAT.



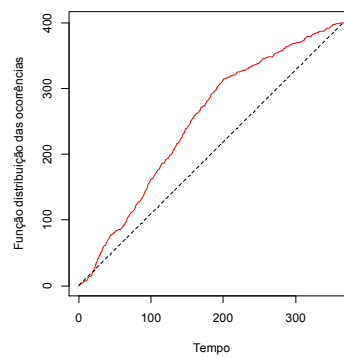
(a)



(b)



(c)



(d)

Gráfico 7 Função de distribuição observada (linha contínua em vermelho), função de distribuição teórica sob a suposição de completa aleatoriedade temporal (linha tracejada em preto), para o primeiro (a), segundo (b), terceiro (c) e quarto (d) cenários simulados

O segundo e o quarto cenário foram simulados para caracterizar existência de agrupamentos no tempo e o teste proposto rejeitou a hipótese nula de CAT nessas realizações ( $q_1 = 89,3151$ ,  $valor-p = 0,001$ ) e ( $q_1 = 94,7260$ ,  $valor-p = 0,001$ ). Os Gráficos 7-b e 7-d comprovam visualmente estes resultados onde se pode observar que a função distribuição dos casos observados encontra-se acima da reta da função de distribuição teórica, indicando que o número de casos observados no período inicial do tempo analisado é maior do que o esperado sob a hipótese de CAT. Os gráficos ainda mostram que o no segundo cenário o agrupamento temporal ocorre até 100 dias (Gráfico 7-b), enquanto no quarto cenário o agrupamento ocorre até 200 dias (Gráfico 7-d).

Os resultados obtidos nos cenários simulados parecem indicar que o método proposto nessa tese pode ser uma alternativa eficiente para detectar agrupamento no tempo. Evidentemente, para uma melhor avaliação do método seria necessário analisar o poder do teste (probabilidade de aceitar a hipótese alternativa quando ela é verdadeira) e o erro do tipo I (probabilidade de rejeitar a hipótese nula verdadeira) e comparar esses resultados com os obtidos a partir da função  $K$  temporal e da estatística de varredura temporal de Kulldorff.

O teste proposto foi baseado na medida de discrepância entre  $F(t)$  e  $\hat{\lambda}t$ . Uma medida alternativa seria substituir na equação (21) a função  $F(t)$  pela função  $\hat{\lambda}_o t$ , onde  $\hat{\lambda}_o$  poderia ser estimado pelo método de mínimos quadrados a partir dos dados observados. Os demais passos seriam os mesmos apresentados anteriormente.

Outra possibilidade seria comparar os coeficientes angulares das duas retas, ou seja, comparar  $\hat{\lambda}$  com  $\hat{\lambda}_o$ . Sob certas condições, a distribuição amostral desta estatística de teste é bastante conhecida (CHARNET et al., 2008) e, conseqüentemente, o teste poderia ser realizado facilmente.

Mais uma vez, para uma melhor avaliação dessas propostas sugeridas, seria necessário analisar o poder do teste e o erro do tipo I e comparar esses

resultados com a proposta desenvolvida em detalhes nesta tese e os outros métodos disponíveis na literatura.

### 4.3 Novo teste para interação espaço-tempo

Na Seção 2.4 foram definidos os resíduos padronizados como sendo  $\hat{R}(s,t) = \hat{D}(s,t) / \sqrt{\hat{V}(s,t)}$ . Diggle et al. (1995) sugerem que a soma dos resíduos padronizados, dada por  $U = \sum_s^n \sum_t^n R(s,t)$ , pode ser utilizada para testar a existência de interação espaço-temporal de duas formas: (a) comparando a estatística de teste  $U / \sqrt{\text{Var}(U)}$  com os valores críticos da  $N(0,1)$  e (b) usando  $U$  em um teste Monte Carlo.

Nesta tese, sugere-se utilizar diretamente a estatística  $\hat{D}(s,t) = \hat{K}(s,t) - \hat{K}(s)\hat{K}(t)$ , definida na equação (12), para testar a hipótese  $H_0$ : Não existe interação espaço-temporal contra a hipótese alternativa  $H_1$ : Existe interação espaço-temporal.

A ideia por trás desta proposta foi apresentada na Seção 2.4, ou seja, se os processos temporais e espaciais operam independentemente, e existe uma ausência de interação espaço-temporal, conseqüentemente, esperam-se valores zero para  $\hat{D}(s,t)$ . Quanto mais afastado de zero estiverem os valores de  $\hat{D}(s,t)$ , maior é a interação espaço-temporal. O problema está em determinar o quanto  $\hat{D}(s,t)$  deve estar afastado de zero para considerar a interação espaço-temporal estatisticamente significativa.

A estatística apresentada na equação (12) não apresenta distribuição amostral conhecida e, assim, mais uma vez, utiliza-se Monte Carlo seguindo os seguintes passos:

a) Define-se a estatística  $P = \sum_s^n \sum_t^n \hat{D}(s, t)$ ;

b) Calcula-se  $P_0$ , em que  $P_0$  é a medida para a configuração espaço-temporal observada;

c) Calcula-se,  $P_k$  para  $k = 1, 2, \dots, m$ , em que  $m$  é o número total de simulações e  $P_k$  é a medida para as configurações no espaço-tempo observadas sob a hipótese de não existência de interação espaço-temporal. Para obter essas configurações, as coordenadas dos tempos dos casos são reordenadas aleatoriamente em relação às coordenadas de localização. Este procedimento elimina qualquer interação espaço-tempo presente nos dados;

d) Ordena-se  $P_k$ . Se obtivermos  $P_0$  em um dos extremos da distribuição dos se  $P_k$  para  $k = 1, 2, \dots, m$ , então existem evidências estatísticas para rejeitar a hipótese de completa aleatoriedade espacial. Suponha que,  $P_{(j)} = P_{(0)}$  para algum  $j \in \{1, \dots, m\}$ , então se rejeita a hipótese nula de não existência de interação espaço-temporal se,  $\text{valor} - p = ((m + 1 - j) / m) \leq \alpha$ .

#### 4.4 Análise da interação espaço-tempo nos cenários típicos simulados usando o novo teste

Nesta Seção serão apresentados os resultados obtidos a partir da aplicação do teste proposto na Seção 4.3 nos quatro cenários simulados. Para testar a hipótese nula de não existência de interação espaço-temporal para cada configuração simulada, foram realizadas 1000 simulações Monte Carlo. As escalas temporais ( $t$ ) e espaciais ( $s$ ) atribuídas para a obtenção dos valores de  $\hat{D}(s, t)$  foram, respectivamente, de 45 valores, igualmente espaçados, compreendidos entre 0 e 60 metros e 45 valores, igualmente espaçados, compreendidos entre 0 e 90 dias.

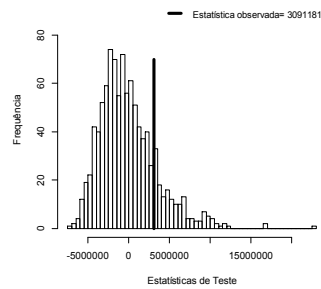
Os três primeiros cenários foram simulados para caracterizar ausência de agrupamentos resultantes da interação espaço-temporal e o teste proposto mostrou que a interação espaço-tempo é estatisticamente não significativa para essas realizações com os seguintes resultados: primeiro cenário ( $P_0 = 3091181$ ,  $p = 0,169$ ), segundo cenário ( $P_0 = 3539925$ ,  $p = 0,217$ ) e terceiro cenário ( $P_0 = 343092,8$ ,  $p = 0,439$ ). Esses resultados confirmam que essas três realizações não apresentam agrupamentos oriundos da interação espaço-temporal.

Esses resultados podem ser visualizados nos Gráficos 8-a, 8-b e 8-c, onde se observa claramente que as estatísticas observadas  $P_0$  estão dentro das distribuições amostrais simuladas sob a hipótese nula de ausência de interação espaço-temporal.

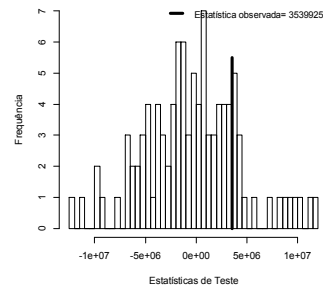
O cenário quatro foi simulado para caracterizar a presença de agrupamentos marcantes resultantes da interação espaço-temporal. O teste proposto mostrou que a interação espaço-tempo é estatisticamente significativa para essa realização ( $P_0 = 118545707$ ,  $p = 0,001$ ). Esse resultado confirma a presença de agrupamentos nessa realização oriundos da interação espaço-temporal.

Esse resultado pode ser analisado no Gráfico 8-d, onde se observa claramente que a estatística observada  $P_0$  está muito afastada da distribuição amostral simulada sob a hipótese nula de ausência de interação espaço-temporal.

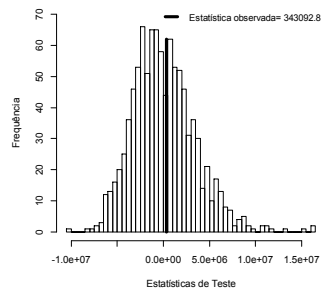




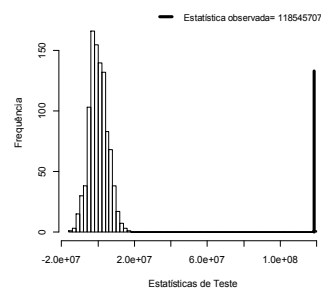
(a)



(b)



(c)



(d)

Gráfico 8 Histograma da distribuição de frequência empírica de  $D(s,t)$ , sob a suposição de ausência de interação espaço-tempo, com a estatística  $D(s,t)$  para a realização observada (linha vertical preta), para o primeiro (a), segundo (b), terceiro (c) e quarto (d) cenários simulados

#### 4.5 Análise dos cenários típicos simulados usando a função $K$

Nesta Seção são apresentados os resultados obtidos a partir da aplicação dos métodos baseados na função  $K$ , descritos nas seções 2.2-2.4, nos quatro cenários típicos simulados. O objetivo desta Seção é mostrar como esses

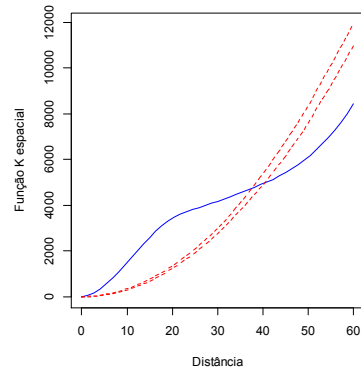
métodos se comportam em situações simuladas para permitir um entendimento didático dos mesmos.

#### 4.5.1 Análise espacial

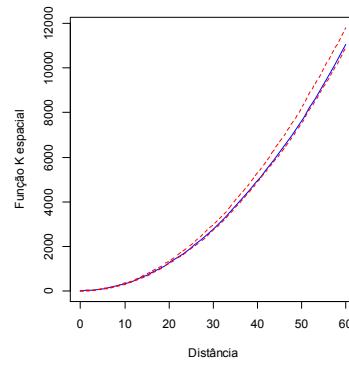
Inicia-se esta Seção apresentando os resultados obtidos a partir do uso dos métodos baseados na função  $K$  espacial, descritos na Seção 2.2, para testar a hipótese de completa aleatoriedade espacial (CAE) nos dados simulados.

Os resultados obtidos para o cenário 1 (caracterizado por apresentar agrupamento no espaço, sem agrupamento no tempo e nenhuma interação espaço-tempo) foram: para distância de Kolmogorov-Smirnov ( $rm_i = 2893,795$ ,  $p = 0,001$ ) e para a distância de Cramer-vonMises ( $rc_i = 385506620$ ,  $p = 0,001$ ), ou seja, existe evidências estatisticamente significativas para rejeitar a hipótese nula de CAE. O Gráfico 9-a mostra que a função  $K$  observada se encontra acima do envelope superior até a uma distância de 38 metros, mostrando que a hipótese de CAE é rejeitada em favor de uma configuração com agrupamentos até aquela distância.

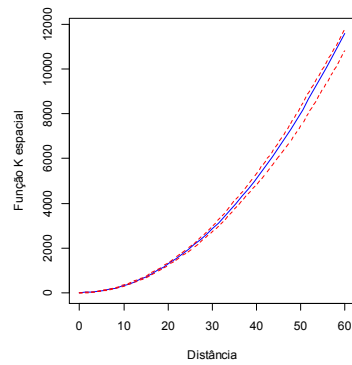
O segundo e terceiro cenário foram geradas para representar realizações de configurações que não apresentam agrupamento no espaço. Os resultados obtidos foram os seguintes: cenário 2 (Kolmogorov-Smirnov:  $rm_i = 261,3957$ ,  $p = 0,126$ ; Cramer-vonMises:  $rc_i = 2314977$ ,  $p = 0,103$ ); cenário 3 (Kolmogorov-Smirnov:  $rm_i = 291,7556$ ,  $p = 0,061$ ; Cramer-vonMises:  $rc_i = 2023200$ ,  $p = 0,105$ ). Assim, a hipótese de CAE é aceita nas duas configurações. O Gráfico 9-b e o Gráfico 9-c corroboram os resultados dos testes de hipóteses, mostrando que a função  $K$  observada está contida dentro dos envelopes de simulação e, portanto, caracterizando que as duas realizações são originárias de um processo pontual espacial (Poisson homogêneo).



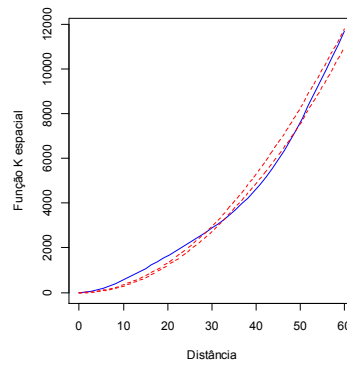
(a)



(b)



(c)



(d)

Gráfico 9 Função  $K$  observada (linha contínua em azul) e envelope de simulação com 99% de confiança (linhas tracejadas em vermelho), representado os limites da completa aleatoriedade espacial, para o primeiro (a), segundo (b), terceiro (c) e quarto (d) cenários simulados

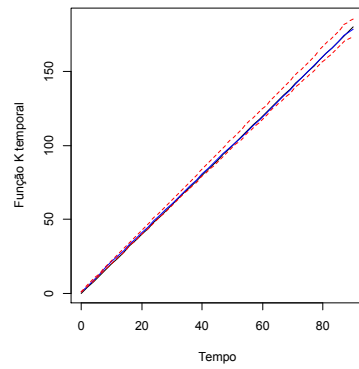
O cenário quatro caracteriza uma realização de um processo espacial com agrupamentos. Os resultados obtidos pela distância de Kolmogorov-Smirnov ( $rm_i = 448,8911$ ,  $p = 0,005$ ) e pela distância de Cramer-vonMises ( $rc_i = 13160275$ ,  $p = 0,001$ ) mostram evidências estatísticas significativas para rejeitar a hipótese nula de CAE. No Gráfico 9-d nota-se um agrupamento até a distância de aproximadamente 30 metros.

A função  $K$  espacial demonstrou-se eficiente para detectar os agrupamentos esperados para a primeira e quarta simulação e em não rejeitar a hipótese de CAE para a segunda e terceira simulação. Diggle (2003) afirma que a função  $K$  é igualmente poderosa para detectar tanto agrupamentos como regularidade de pontos em diferentes escalas no espaço. A função  $K$  é considerada poderosa, inclusive, para configurações com pequeno número de pontos (SCALON; SILVA, 2006).

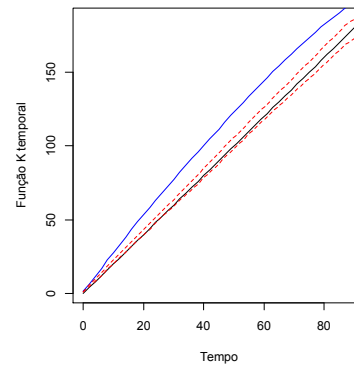
#### 4.5.2 Análise temporal

Os métodos para testar a hipótese nula de completa aleatoriedade temporal (CAT), utilizando a função  $K$ , estão descritos na Seção 2.3. Neste capítulo apresentam-se os resultados da aplicação desses métodos nos cenários simulados.

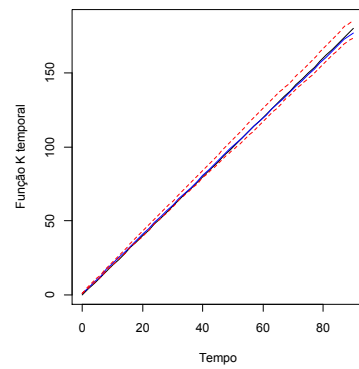
O cenário 1, caracterizada por não apresentar agrupamento no tempo, apresentou os seguintes resultados: para distância de Kolmogorov-Smirnov ( $hm_i = 1,5990$ ,  $p = 0,868$ ) e para a distância de Cramer-vonMises ( $hc_i = 92,1298$ ,  $p = 0,898$ ), ou seja, não existem evidências estatísticas para rejeitar a hipótese nula de completa aleatoriedade temporal (CAE). Pelo Gráfico 10-a percebe-se que a função  $K$  temporal observada se encontra dentro dos envelopes, corroborando o aceite da hipótese nula de CAT.



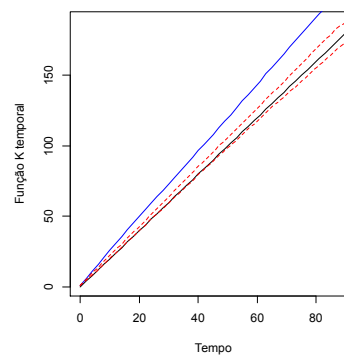
(a)



(b)



(c)



(d)

Gráfico 10 Função  $K$  temporal observada (linha contínua azul), função  $K$  temporal teórica (linha contínua preta) e envelope de simulação com 99% de confiança (linhas tracejadas em vermelho), representado os limites da completa aleatoriedade temporal, para o primeiro (a), segundo (b), terceiro (c) e quarto (d) cenários simulados

Os resultados obtidos para cenário 2, caracterizado por apresentar agrupamento no tempo, foram: para distância de Kolmogorov-Smirnov ( $hm_i = 24,4677$ ,  $p = 0,001$ ) e para a distância de Cramer-vonMises ( $hc_i = 63277,73$ ,  $p = 0,001$ ), ou seja, existe evidência estatisticamente significativa para rejeitar a hipótese nula de CAT. No Gráfico 10-b nota-se que a função  $K$  temporal observada se encontra acima do envelope superior para todas as escalas analisadas, mostrando que a hipótese de CAT é rejeitada em favor de uma configuração com agrupamentos dentro dessas escalas.

Para o cenário 3, também caracterizada por não apresentar agrupamento no tempo, obteve-se os seguintes resultados: para distância de Kolmogorov-Smirnov ( $hm_i = 3,0215$ ,  $p = 0,332$ ) e utilizando a distância de Cramer-vonMises ( $hc_i = 125,4609$ ,  $p = 0,829$ ), ou seja, não existem evidências estatísticas para rejeitar a hipótese nula de CAT. Pelo Gráfico 10-c observa-se que a função  $K$  temporal observada se encontra dentro dos envelopes, corroborando o aceite da hipótese nula de CAT.

Os resultados obtidos para o cenário 4, caracterizado por apresentar agrupamento no tempo, foram: para distância de Kolmogorov-Smirnov ( $hm_i = 32,7264$ ,  $p = 0,001$ ) e para a distância de Cramer-vonMises ( $hc_i = 67264,77$ ,  $p = 0,001$ ), ou seja, existem evidências estatisticamente significativas para rejeitar a hipótese nula de CAT. No Gráfico 10-d a função  $K$  temporal observada se encontra bem acima do envelope superior em todas as escalas de tempo analisadas, mostrando que a hipótese de CAT é rejeitada em favor de uma configuração temporal com agrupamento.

Os resultados obtidos nesta Seção mostram que a função  $K$  temporal pode ser uma metodologia importante para a análise de completa aleatoriedade temporal em dados de processos pontuais. Assim, causa estranheza o fato desta função ser praticamente ignorada como ferramenta estatística de análise de processos pontuais no tempo. Poucos trabalhos a exploraram teoricamente

(BADDELEY, 2006; CANNON; CRESSIE, 1995) e não foi encontrado na literatura, salvo melhor juízo, trabalhos que a utilizem em dados agrícolas. Talvez, o pouco uso da função  $K$  temporal seja devido à maior familiaridade dos pesquisadores com os métodos de séries temporais e outros métodos para processos pontuais como descritos em Daley e Vere-Jones (2003). Espera-se que os resultados obtidos nesta tese possam aumentar o interesse dos pesquisadores de diversas áreas, inclusive agrícola, sobre o uso da função  $K$  temporal.

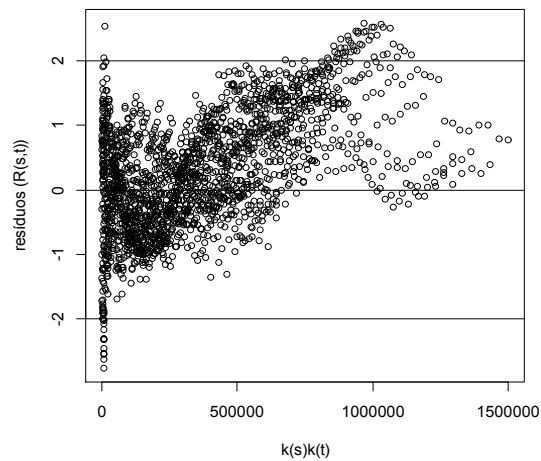
#### 4.5.3 Análise da interação espaço-temporal

Os métodos para testar a hipótese nula de ausência de interação espacial-temporal, utilizando a função  $K$ , estão descritos na Seção 2.4. Neste capítulo apresentam-se os resultados da aplicação desses métodos nos cenários simulados.

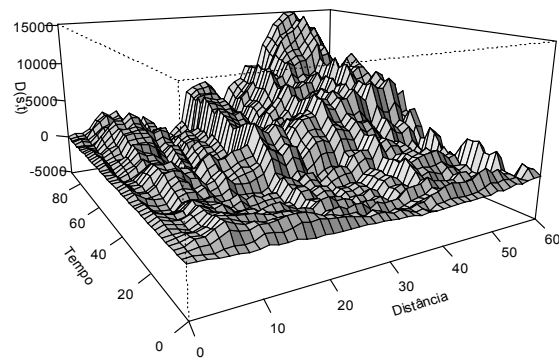
Os Gráficos de 11 a 14 estão apresentados todos os resultados da análise espaço-temporal para todos os cenários, em que os cenários um, dois e três, foram simuladas para não apresentarem interação espaço-tempo, enquanto o cenário 4 foi simulado para apresentar interação espaço-tempo.

Uma forma de diagnosticar a interação espaço-tempo é por meio do gráfico dos resíduos padronizados  $R(s,t)$  (equação 14) contra o produto das funções  $K$  temporal e espacial. Pelos Gráficos 11-a, 12-a e 13-a percebe-se que os resíduos padronizados estão distribuídos entre  $\pm 2$  desvios-padrão e, conforme recomendam Diggle et al. (1995), não se deve rejeitar a hipótese nula de interação espaço-temporal para os três primeiros cenários. Esses resultados são corroborados pelos gráficos de superfície tridimensional de  $D(s,t)$  (Gráficos 11-b, 12-b, 13-b) que apesar de apresentarem picos esses não são acentuados o suficiente para levar a rejeição da hipótese de interação. Portanto, esses

resultados mostram que não existem evidências estatísticas para rejeitar a hipótese de interação espaço-tempo nos cenários simulados 1, 2 e 3.



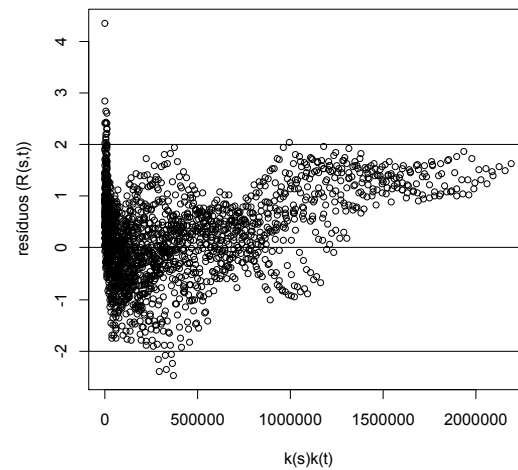
(a)



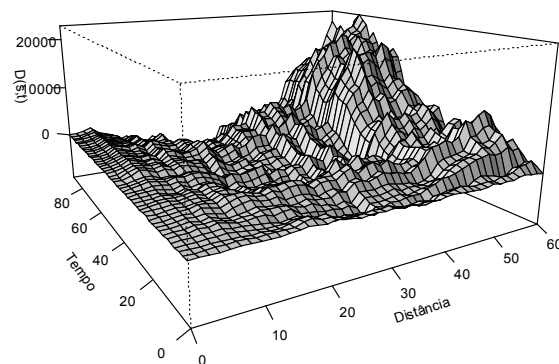
(b)

Gráfico 11 Gráfico de dispersão dos resíduos generalizados  $R(s,t)$  versus o produto das funções  $K$  temporal e espacial separadamente (a); Gráfico de superfície do grau de agrupamento espaço-tempo adicional  $D(s,t)$  (b); para o cenário 1



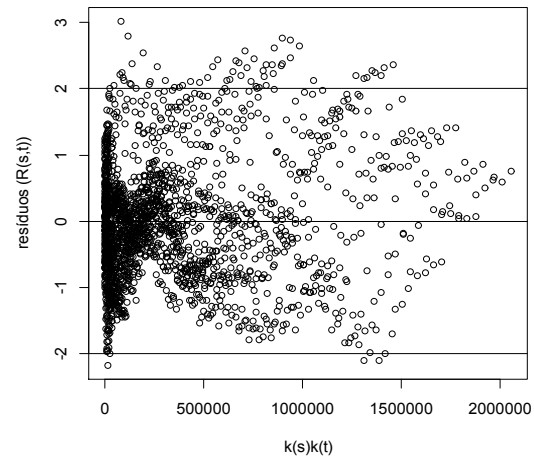


(a)

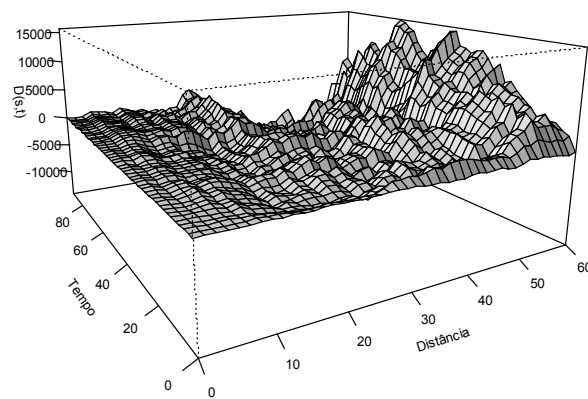


(b)

Gráfico 12 Gráfico de dispersão dos resíduos generalizados  $R(s,t)$  versus o produto das funções  $K$  temporal e espacial separadamente (a); Gráfico de superfície do grau de agrupamento espaço-tempo adicional  $D(s,t)$ , (b); para o cenário 2



(a)

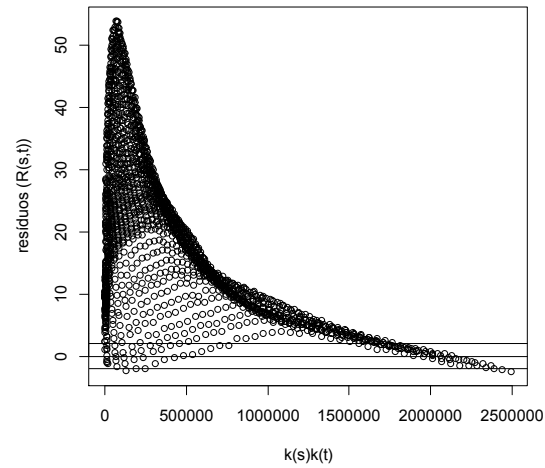


(b)

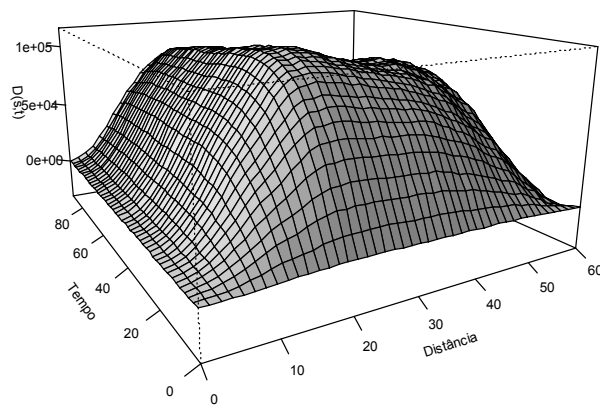
Gráfico 13 Gráfico de dispersão dos resíduos generalizados  $R(s,t)$  versus o produto das funções  $K$  temporal e espacial separadamente (a); Gráfico de superfície do grau de agrupamento espaço-tempo adicional  $D(s,t)$ , (b); para o cenário 3

No Gráfico 14 estão apresentados os resultados da análise espaço-temporal para o cenário 4 que considera a presença de agrupamentos no espaço, no tempo e na interação espaço-tempo. O gráfico dos resíduos padronizados  $R(s,t)$  (Gráfico 14-a) observa-se que a interação espaço-tempo foi significativa, visto que a grande maioria dos resíduos padronizados estão acima de 2 desvios-padrão, ou seja, rejeita a hipótese nula de independência espaço-temporal. Esses resultados são corroborados pelo gráfico de superfície (Gráfico 14-b), onde a superfície de  $D(s,t)$  apresenta uma elevação significativa por volta da distância de 20 metros para um período de tempo, aproximado, entre 30 e 70 dias.

Os resultados obtidos nesta Seção mostram o que outros trabalhos já mostraram anteriormente, ou seja, a função  $K$  espacial-temporal proposta por Diggle et al. (1995) é uma metodologia extremamente importante para análise de agrupamentos oriundos da interação espacial-temporal em dados de processos pontuais. Assim, não é surpresa que esta metodologia esteja sendo utilizada com frequência em áreas tão diversas do conhecimento como epidemiologia (DIGGLE et al., 1995), ecologia (CARSLAKE et al., 2005) e planejamento urbano (ARBIA et al., 2010). A surpresa está no fato de que, salvo melhor juízo, esta metodologia ainda não foi utilizada na análise de configurações pontuais de doenças agrícolas. Assim, espera-se que esta tese mostre um pouco mais sobre o potencial dessa função na análise de dados agrícolas e mude esse cenário.



(a)



(b)

Gráfico 14 Gráfico de dispersão dos resíduos generalizados  $R(s,t)$  versus o produto das funções  $K$  temporal e espacial separadamente (a); Gráfico de superfície do grau de agrupamento espaço-tempo adicional  $D(s,t)$  (b); para o cenário 4

#### **4.6 Análise dos cenários típicos simulados usando a estatística de varredura de Kulldorff**

Nesta Seção são apresentados os resultados obtidos a partir da aplicação dos métodos baseados na estatística de varredura de Kulldorff (descritos nas seções 2.5-2.6) nos quatro cenários típicos simulados. O objetivo com esta Seção é mostrar como esses métodos se comportam em situações conhecidas para permitir um entendimento dos mesmos.

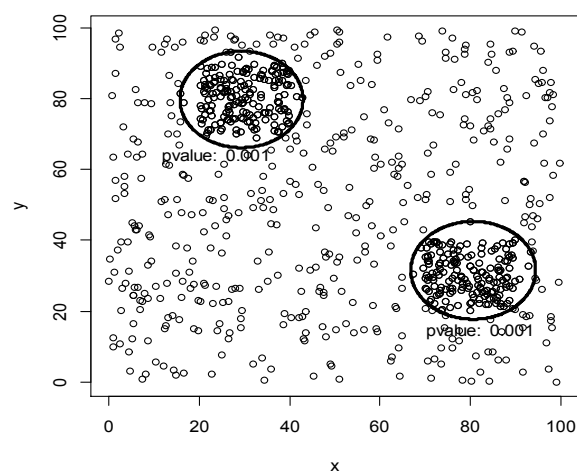
##### **4.6.1 Análise de varredura espacial**

Na Tabela 3 estão apresentados os resultados para o cenário 1, que apresenta agrupamentos no espaço, tendo como base a teoria apresentada na Seção 2.5. Dentre todos os agrupamentos percorridos pela estatística de varredura, os dois agrupamentos identificados na Tabela 3 foram os únicos que apresentaram um número excessivo e estatisticamente significativo de casos observados em relação ao número de casos esperados. O teste de razão de verossimilhança e os *valores-p* para os dois agrupamentos indicam uma rejeição da hipótese nula, ou seja, existe evidência significativa de agrupamentos no espaço. Os agrupamentos podem ser visualizados no Gráfico 15.

A segunda e a terceira simulações são cenários caracterizados por não apresentarem agrupamentos no espaço. Os resultados obtidos a partir do teste da razão de verossimilhança foram os seguintes: segunda ( $\tau = 4,8829$ , *valor-p* = 0,987) e terceira ( $\tau = 7,7013$ , *valor-p* = 0,3) e, portanto, não mostram evidências estatísticas de agrupamentos no espaço para o período analisado.

Tabela 3 Resultados da varredura espacial para o cenário 1.

Casos observados nos agrupamentos	Razão de verossimilhança	Valor- <i>p</i>
158	81,5604	< 0,001
159	69,1104	< 0,001

Gráfico 15 Gráfico representando os dois agrupamentos mais prováveis e seus respectivos valores-*p* de acordo com o cenário 1

Na Tabela 4 estão apresentados os resultados para o cenário 4 (agrupamento no espaço) tendo como base a teoria apresentada na Seção 2.5. Dentre todos os possíveis agrupamentos varridos pelo método de Kulldorff, os quatro agrupamentos identificados na Tabela 4 foram os únicos que apresentaram um número excessivo de casos observados em relação ao esperado.

Tabela 4 Resultados da varredura espacial para o cenário 4.

Casos observados nos agrupamentos	Razão de verossimilhança	Valor- <i>p</i>
66	25,2544	< 0,001
54	17,9399	< 0,001
60	12,3436	0,003
49	9,5211	0,044

O teste de razão de verossimilhança e os *valores-p*, apresentados na Tabela 4, para os quatro agrupamentos indicam uma rejeição da hipótese nula de completa aleatoriedade espacial, ou seja, existe evidência significativa de agrupamento espacial. Os grupamentos podem ser visualizados no Gráfico 16.

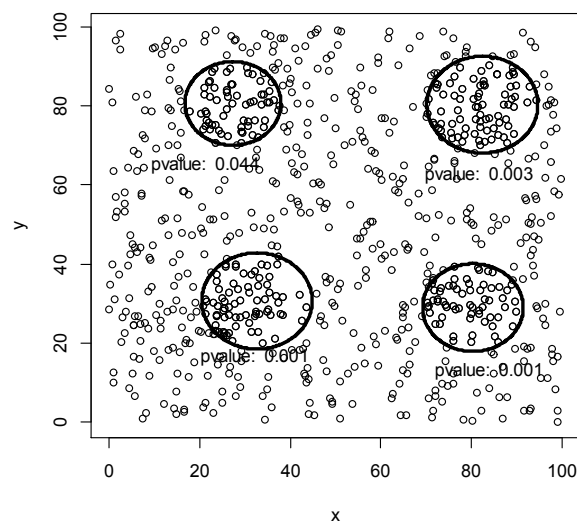


Gráfico 16 Gráfico representando os quatro agrupamentos e seus respectivos *valores-p* de acordo com o cenário 4

A análise de varredura espacial vem sendo muito aplicada em diversas áreas do conhecimento, principalmente, para dados de contagem por área, utilizando o modelo de Poisson (BALIEIRO, 2008; BHATTARAI, 2010; KULLDORFF et al., 1998; OVIEDO et al., 2009). O uso mais frequente em dados de área deve-se ao fato que o método foi criado originariamente para esse tipo de dado (KULLDORFF; NAGARWALLA, 1995). Apesar do método de varredura apresentar uma teoria bem desenvolvida para processos pontuais, o mesmo não tem sido utilizada com frequência na análise desse tipo de dados (KULLDORFF, 1997). O próprio manual do SatScan não mostra como essa análise pode ser conduzida em dados de processos pontuais (KULLDORFF, 2011). Como não foi encontrado na literatura, salvo melhor juízo, trabalhos que utilizem a estatística de varredura espacial para a análise de configurações pontuais de doenças agrícolas, espera-se que os resultados obtidos nesta tese possam incrementar o uso dessa estatística nesta área.

#### **4.6.2 Análise de varredura temporal**

Nesta Seção vamos apresentar os resultados obtidos a partir da análise de varredura temporal tendo como base a teoria apresentada na Seção 2.6.

Os cenários 1 e 3 foram simulados para caracterizar ausência de agrupamentos temporais. A análise de varredura temporal, pelo teste da razão de verossimilhança, mostrou resultados estatisticamente não significativos para a realização do cenário 1 ( $\tau = 6,3229$ ,  $p = 0,241$ ) e resultados estatisticamente significativos para a realização do cenário 3 ( $\tau = 11,2536$ ,  $p = 0,007$ ).

Os cenários 2 e 4 foram simulados para caracterizar a presença de agrupamentos temporais. Para o cenário 2, a análise de varredura temporal apresentou um resultado estatisticamente significativo ( $\tau = 16,4755$ ,  $p = 0,001$ ), mostrando a presença de um agrupamento temporal entre o 2<sup>o</sup> e 85<sup>o</sup> dia. Para o



cenário 4, a análise de varredura temporal também apresentou um resultado estatisticamente significativo ( $\tau = 17,4793$ ,  $p = 0,001$ ), mostrando a presença de um agrupamento temporal entre o 18<sup>o</sup> e 197<sup>o</sup> dia.

Os resultados apresentados anteriormente mostram que a estatística de varredura para processos pontuais pode detectar agrupamentos temporais quando esses não estão presentes nos dados. Esta possibilidade de falha pode explicar o porquê da análise de varredura puramente temporal não ter sido encontrada na literatura, salvo melhor juízo, como método para análise de dados oriundos de processos pontuais temporais.

#### **4.6.3 Análise de varredura espaço-temporal**

Na Seção 2.7 foi apresentada a teoria da análise de varredura espaço-temporal que foi utilizada nesta Seção para analisar os quatro cenários simulados.

Os três primeiros cenários foram simulados para caracterizar ausência de interação espaço-temporal. A análise de varredura espaço-temporal mostrou-se estatisticamente não significativa para esses cenários com os seguintes resultados: primeiro cenário ( $\tau = 15,5498$ ,  $p = 0,276$ ), segundo cenário ( $\tau = 13,9909$ ,  $p = 0,106$ ) e terceiro cenário ( $\tau = 13,6237$ ,  $p = 0,119$ ) e, portanto, confirmando que esses cenários simulados não apresentam interação espaço-tempo.

O cenário 4 foi simulado para caracterizar a presença de interação espaço-tempo. A Tabela 5 mostra os resultados da varredura espaço-tempo de onde se pode concluir pelo teste da razão de verossimilhança que existem agrupamentos no tempo estatisticamente significativos. Os prováveis agrupamentos espaciais podem ser observados no Gráfico 17. Congregando

essas duas informações pode-se concluir que existem agrupamentos resultantes da interação entre o tempo e o espaço.

Tabela 5 Resultados da varredura espaço-temporal para o cenário 4.

Período (dias) da ocorrência de agrupamentos	Ordem de importância dos agrupamentos prováveis	Razão de verossimilhança	Valor-p
127 <sup>o</sup> ao 190 <sup>o</sup>	Primeiro	38,3286	< 0,001
274 <sup>o</sup> ao 360 <sup>o</sup>	Segundo	31,7796	< 0,001
121 <sup>o</sup> ao 268 <sup>o</sup>	Terceiro	30,5543	< 0,001
118 <sup>o</sup> ao 245 <sup>o</sup>	Quarto	28,4139	0,002

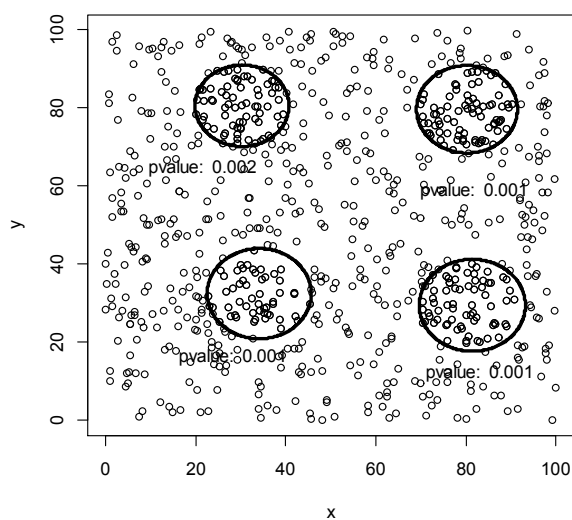


Gráfico 17 Gráfico representando os quatro agrupamentos espaciais mais prováveis simulados de acordo com o cenário 4, no período de um ano e seus respectivos valores-p

A análise de varredura espacial-temporal vem sendo aplicada, principalmente, para dados de contagem por área, utilizando o modelo de Poisson (BALIEIRO, 2008; BHATTARAI, 2010; KULLDORFF et al., 1998; OVIEDO et al., 2009). Mais uma vez, o uso mais frequente em dados de área deve-se ao fato que o método foi criado originariamente para esse tipo de dado (KULLDORFF; NAGARWALLA, 1995).

O método de varredura espacial-temporal apresenta uma teoria bem desenvolvida para processos pontuais tendo como base o modelo de Bernoulli (KULLDORFF, 1997). Entretanto, não foi encontrado na literatura, salvo melhor juízo, trabalhos que utilizem a estatística de varredura espacial-temporal para a análise de dados espaços-temporais para configurações pontuais. Essa ausência de trabalhos pode ser devido ao fato de que o *software* disponível para conduzir a análise de varredura (SatScan) não apresenta um módulo satisfatório para a visualização da interação espacial-temporal (KULLDORFF, 2011).

#### **4.7 Análise da morte súbita dos citros**

A morte súbita dos citros (MSC) é uma doença vascular que ataca laranjeiras enxertadas com limão cravo e que está trazendo muita preocupação para os citricultores brasileiros, pois é capaz de matar a árvore em 12 meses. Além disso, a doença pode se espalhar rapidamente dentro do pomar e tomar pomares e/ou talhões vizinhos com extrema facilidade. Assim, diversas pesquisas estão em curso ou necessitam ser realizadas para tentar entender como esta doença funciona e, assim, tomar medidas preventivas de controle.

Dentre essas diversas pesquisas, Bassanezi et al. (2003) defendem que estudos sobre o padrão espacial e o comportamento, ou o progresso, da MSC ao longo do tempo fornecem informações importantes que podem auxiliar no esclarecimento da etiologia da doença. Nesse sentido, esta tese utiliza diversos

métodos estatísticos, apresentados anteriormente, para detectar agrupamentos de casos de MSC que ocorrem no espaço, no tempo e a partir da interação espaço-tempo. O objetivo da análise é auxiliar na descrição da dinâmica da doença, levantar ou avaliar hipóteses biológicas sobre mecanismos de propagação, formular ou dar suporte a hipóteses levantadas em relação à sua etiologia.

#### **4.7.1 Análise exploratória**

Inicia-se a análise da dinâmica espacial e temporal da MSC com uma análise exploratória. O Gráfico 18 exibe as configurações espaciais, da incidência da MSC no talhão avaliado durante diferentes datas no período de novembro de 2001 a junho de 2003. Cada data representa uma avaliação realizada e cada círculo representa uma planta com MSC. Nos meses de agosto de 2002, novembro e dezembro de 2003 não foram coletados dados.

Pode-se observar no gráfico 18 um aumento elevado do número de plantas doentes a partir da 12<sup>a</sup> avaliação (realizada em janeiro de 2003). Esse comportamento já era esperado, pois de acordo com Bassanezi et al. (2003), após a constatação das primeiras plantas com sintomas de MSC, a doença pode se manifestar em 60 a 100% das plantas do talhão em até dois anos.

No Gráfico 19 é exposto um gráfico de barras da incidência de plantas doentes com a MSC para cada mês em que ocorreu avaliação. Observa-se que da primeira (novembro de 2001) ao décimo primeiro (outubro de 2002) mês tem-se um aumento gradual do número de plantas doentes, mas a partir da décima segunda avaliação (dezembro de 2002) praticamente todas as plantas estavam doentes.

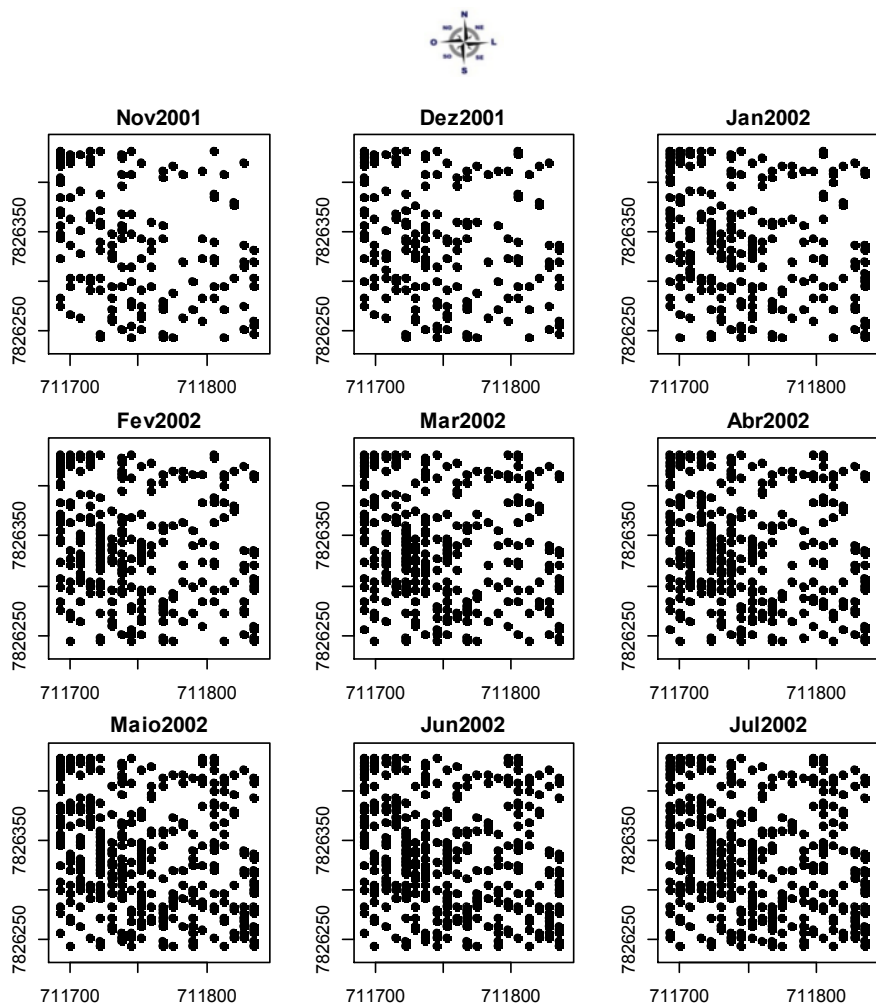


Gráfico 18 Configurações espaciais da incidência da MSC no talhão, no período de novembro de 2001 a julho de 2003, em que cada círculo representa uma planta com MSC

Continuação ...

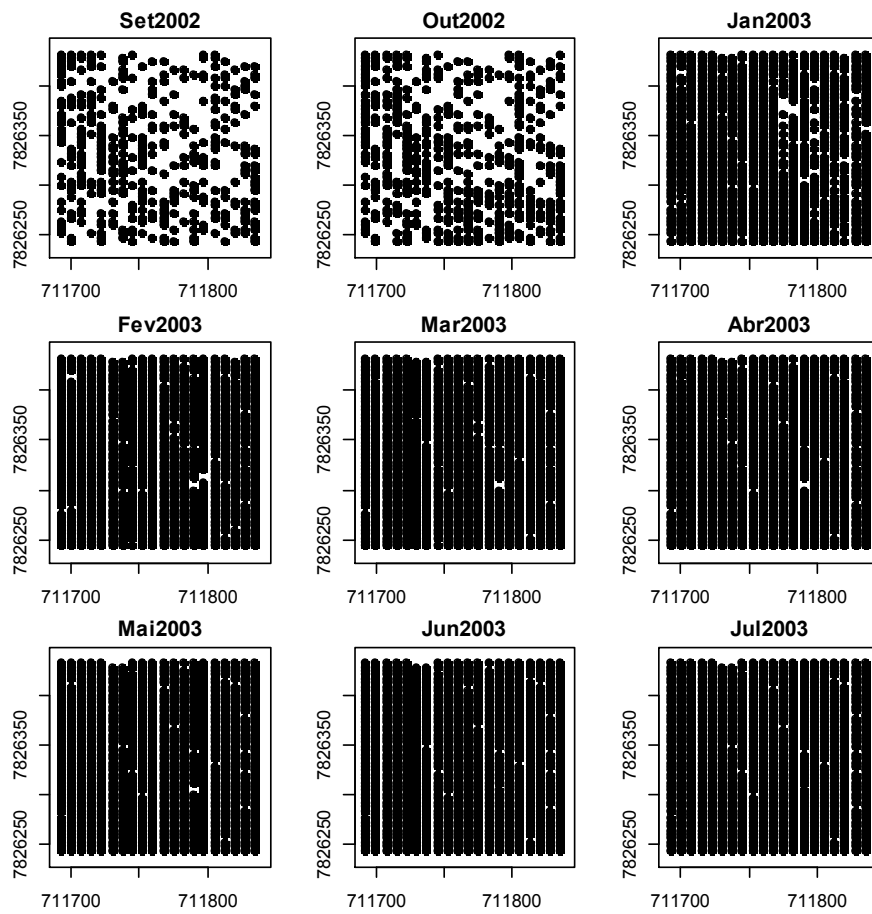


Gráfico 18 Configurações espaciais da incidência da MSC no talhão, no período de novembro de 2001 a julho de 2003, em que cada círculo representa uma planta com MSC

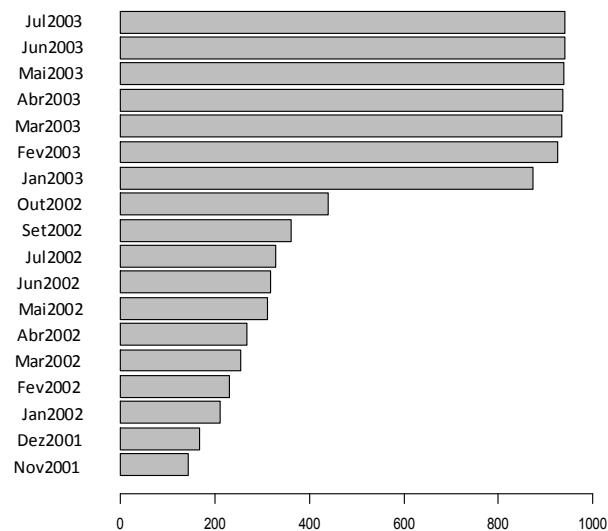


Gráfico 19 Incidência mensal de plantas doentes com a MSC no período de novembro de 2001 até julho de 2003

#### 4.7.2 Análise da distribuição espacial

Testando a hipótese de aleatoriedade espacial para os dados de MSC através da função  $K$  espacial obtiveram-se resultados (Kolmogorov-Smirnov = 1,8469,  $p = 1,0$ ; Cramer-vonMises = 68314689,  $p = 1,0$ ) que mostram não existir evidências estatística para rejeitar a hipótese nula de completa aleatoriedade espacial (CAE).

Pelo Gráfico 20 observa-se que a função  $K$  observada está ou abaixo do envelope inferior ou dentro dos envelopes simulados sob a hipótese de CAE. Este comportamento indica que a MSC não apresenta agrupamentos no espaço, mas indica uma ligeira tendência para a existência de uma distribuição regular nos casos de MSC no talhão o que, de certa forma, contradiz os resultados dos testes apresentados anteriormente.

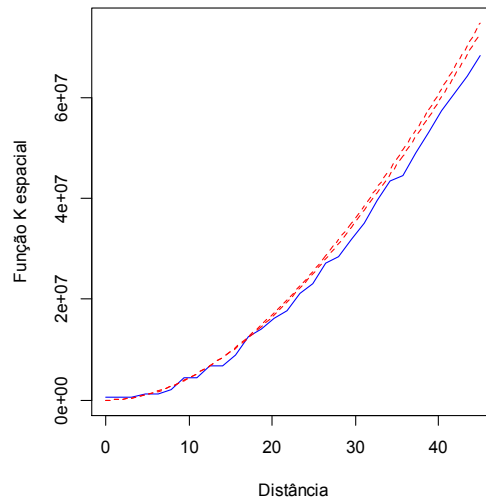


Gráfico 20 Função  $K$  observada (linha contínua em azul) e envelope de simulação com 99% de confiança (linhas tracejadas em vermelho), representado os limites da completa aleatoriedade espacial

Deve-se fazer uma observação sobre essa aparente contradição dos resultados obtidos entre as duas formas de análise (teste e gráfico) usando a função  $K$ . A função  $K$  espacial é recomendada para analisar configurações de pontos que são continuamente distribuídos no espaço (RIPLEY, 1977). Como se sabe, as laranjeiras são plantadas mantendo-se uma distância mínima entre elas o que, naturalmente, impõe uma distância mínima de ocorrência de casos de MSC no talhão, gerando assim, pelo menos, uma regularidade mínima (distância mínima entre duas árvores). Assim, pode até ser que os casos de MSC estejam distribuídos de forma aleatória (ou agrupada) em algumas escalas, mas a regularidade natural da plantação pode comprometer a análise espacial usando a função  $K$ .

Na Tabela 6 estão apresentados os resultados da análise espacial dos casos de MCS utilizando o método de varredura espacial de Kulldorff tendo



como base a teoria apresentada na Seção 2.5. Dentre todos os agrupamentos possíveis, os três agrupamentos, identificados na Tabela 6, foram os únicos que apresentaram um número excessivo de casos observados em relação ao número esperado pelo teste da razão de verossimilhança ( $p < 0,001$ ). Esses agrupamentos estatisticamente significativos são mostrados no Gráfico 21.

Tabela 6 Resultados da varredura espacial para os dados de morte súbita dos citros.

Casos de MSC observados dentro dos agrupamentos	Razão de verossimilhança	Valor- $p$
194	100,3023	< 0,001
1945	92,1224	< 0,001
125	33,8926	< 0,001

Tendo como base os resultados obtidos a partir da aplicação da estatística de varredura espacial de Kulldorff pode-se afirmar que a distribuição dos casos de MSC no talhão analisado apresenta agrupamentos espaciais.

Krainski e Ribeiro Júnior (2004) utilizam o índice de dispersão Poisson, por meio da análise de contagem por *quadrats*, para avaliar o padrão espacial da incidência de MSC em outro talhão da mesma plantação. Os autores obtiveram resultados indicando padrões aleatórios da 1ª até a 6ª avaliação e regular da 7ª até a 18ª avaliação para o tamanho de *quadrats* 2x2. Utilizando *quadrats* 7x3, os autores obtiveram resultados indicando configurações aleatórias da 1ª até a 11ª avaliação e regulares da 12ª até a 18ª. Apesar de esses autores utilizarem uma metodologia que é questionada por Cressie (1993) e Diggle (2003), por ser

influenciada pelo tamanho e número de *quadrats*, de maneira geral, os resultados obtidos na presente tese, usando a função *K* espacial e a estatística de varredura espacial de Kuldorff, corroboram os resultados obtidos por Krainski e Ribeiro Júnior (2004).

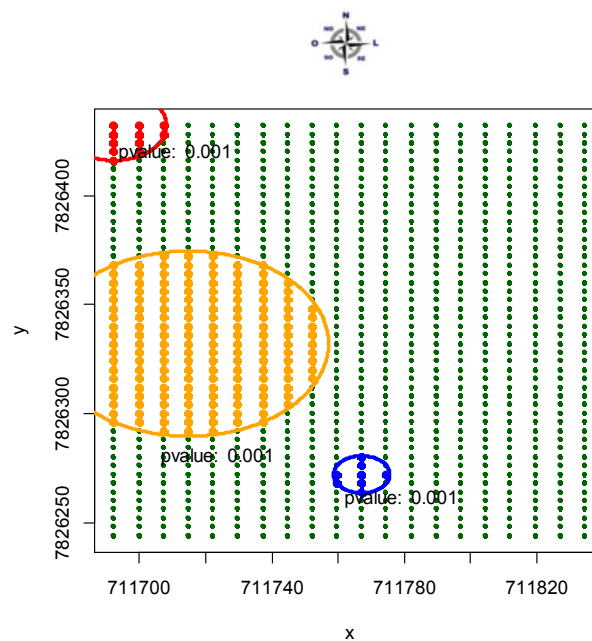


Gráfico 21 Gráfico representando os agrupamentos, estatisticamente significativos identificados pela varredura espacial nos dados MSC: agrupamento mais provável (vermelho), segundo agrupamento mais provável (amarelo), terceiro agrupamento mais provável (azul) e seus respectivos *valores-p*

#### 4.7.3 Análise da distribuição temporal

Testando a hipótese nula de aleatoriedade temporal (CAT) para os tempos de ocorrência dos casos de MSC, utilizando a função *K* temporal,

obtiveram-se os seguintes resultados: para a distância de Kolmogorov-Smirnov ( $hm_i = 216,6373$ ,  $p = 0,002$ ) e para a distância de Cramer-vonMises ( $hc_i = 5,1647$ ,  $p = 0,002$ ). Esses resultados levam a rejeição da hipótese nula de CAT.

Pelo Gráfico 22 percebe-se que a rejeição da hipótese nula de CAT foi em direção de agrupamento, uma vez que a função  $K$  temporal observada encontra-se acima do envelope superior para toda a escala de tempo analisada.

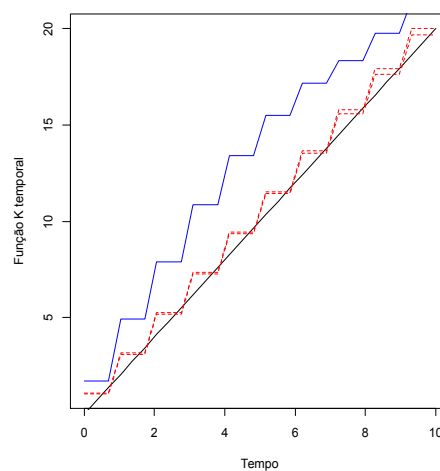


Gráfico 22 Função  $K$  observada (linha azul contínua), função  $K$  teórica (reta preta contínua) e envelopes de simulação com 99% de confiança (linhas tracejadas em vermelho) representando os limites da CAT

Utilizando o método da varredura temporal, o agrupamento mais provável de ocorrências de MSC no talhão apresentou significância estatística para o teste de razão de verossimilhança para o período de janeiro a julho de 2003 ( $\tau = 4560,7884$ ,  $p = 0,001$ ). Assim, este método também levou a uma rejeição enfática da hipótese de CAT dos casos de MSC no talhão analisado.

O teste contra a hipótese de CAT, baseado no novo método da função distribuição dos tempos de ocorrência dos casos produziu resultados ( $q_1 = 3771,286, p = 0,001$ ) que também levam a uma rejeição da hipótese de CAT dos casos de MSC.

Pelo Gráfico 23 observa-se claramente que a função de distribuição observada está muito abaixo da reta sob a hipótese nula de CAT e, portanto, existem evidências estatísticas de existência de agrupamento de ocorrências de casos de MSC no final do período de tempo analisado.

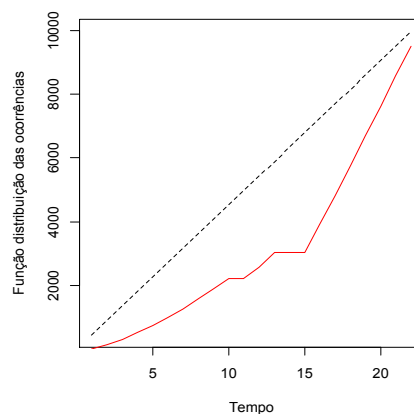


Gráfico 23 Função distribuição observada (linha contínua em vermelho) e função teórica sob a hipótese de CAT (linha tracejada em preto)

Observa-se que os três métodos (função  $K$  temporal, varredura temporal e função distribuição) levaram a mesma conclusão, ou seja, a uma rejeição enfática da hipótese nula de CAT dos casos de MSC em direção a uma agregação da ocorrência dos casos no tempo. Além disso, tanto o método da varredura temporal como o novo método da função distribuição, detectou que a

presença de agrupamento de ocorrências de MSC ocorreu no período final da coleta de dados, ou seja, entre janeiro e julho de 2003.

Não foi encontrado na literatura trabalhos com esse enfoque de diagnosticar a presença de agrupamento de casos de doenças agrícolas. Em geral, trabalhos relacionados à análise temporal estão interessados em analisar o comportamento da evolução da doença ao longo do tempo por meio de modelos não lineares, como é o caso de dos trabalhos de Bassanezi, Gimenes-Fernandes e Yamamoto (2003) que utilizaram o modelo de Gompertz e Peixoto, Barros e Bassanezi (2008) que adotaram o modelo de autômatos celulares para estudar a evolução temporal dos casos de MSC em laranjeiras.

#### **4.7.4 Análise da interação espaço-temporal**

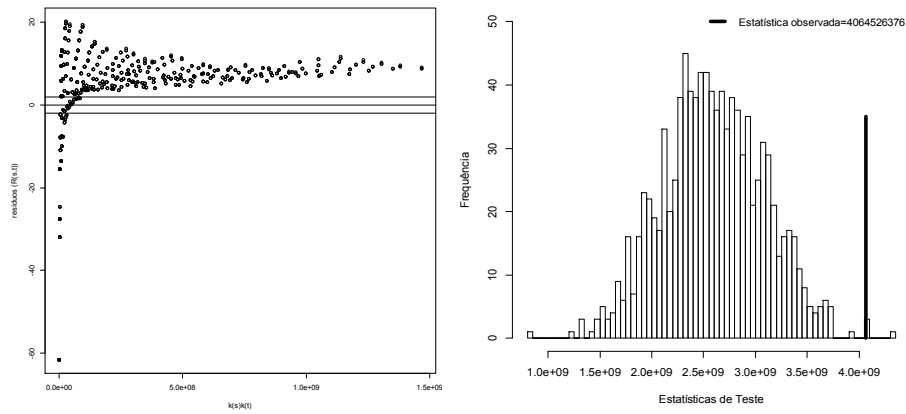
Inicia-se essa análise da interação espaço-temporal para os dados de MSC utilizando os métodos baseados na função  $K$ . Observa-se no Gráfico 24-a dos resíduos padronizados  $R(s,t)$  que a maioria dos resíduos padronizados se encontram acima do limite de dois desvios-padrão e, portanto, deve-se rejeitar a hipótese nula de independência espaço-temporal. Assim, existe a presença de agrupamentos oriundos da interação espaço-temporal. Esse resultado é corroborado pelo Gráfico 24-c de contorno que mostra o aumento de todos os valores de  $D(s,t)$  à medida que as escalas espaciais e temporais também aumentam, o que indica um claro indício de rejeição da hipótese nula de ausência de interação espaço-tempo dos casos de MSC.

Para a análise da interação espaço-tempo dos casos de MSC utilizando o novo teste apresentado na Seção 4.3 foram realizadas 999 simulações Monte Carlo sob a hipótese nula, utilizando como escalas temporais ( $t$ ) e espaciais ( $s$ ) para o cálculo de  $\hat{D}(s,t)$ , respectivamente, 30 valores de tempo, igualmente

espaçados, entre 0 e 10 meses e 30 valores de distância, igualmente espaçadas, entre 0 e 45 metros.

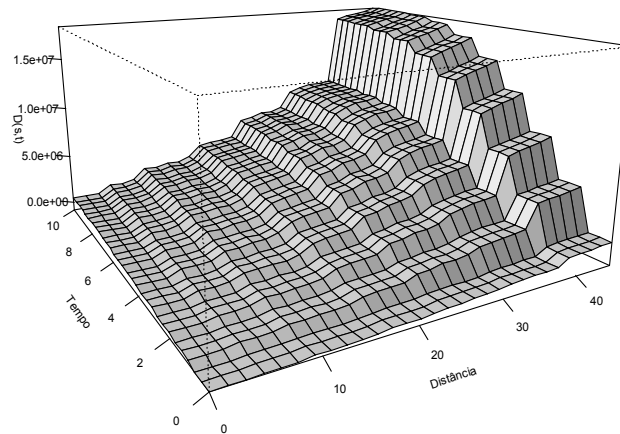
A aplicação do novo método resultou em um  $P_0 = 4064526376$ , com *valor-p* = 0,002, o que leva a rejeição da hipótese nula de ausência de interação espacial-temporal nos casos de MSC no talhão analisado.

Esse resultado pode ser visualizado no Gráfico 24-b, onde se observa claramente que a estatística observada  $P_0$  esta muito afastada da distribuição amostral simulada sob a hipótese nula de ausência de interação espaço-temporal.



(a)

(b)



(c)

Gráfico 24 Gráfico de dispersão dos resíduos padronizados  $R(s,t)$  versus o produto das funções  $K$  temporal e espacial (a), histograma da distribuição de frequência empírica de  $D(s,t)$ , sob a suposição de ausência de interação espaço-tempo, com a estatística  $D(s,t)$  para a realização observada (linha vertical preta) (b), Gráfico de superfície do grau de agrupamento espaço-tempo adicional  $D(s,t)$  (c)

A análise da interação espaço-temporal dos casos de MSC via varredura espaço-temporal de Kulldorff produziu como resultado uma estatística de teste de  $\tau = 2024,6107$  e *valor-p* = 0,001, ou seja, a interação espaço-temporal mostrou-se estatisticamente significativa durante o período de janeiro a julho de 2003. Somente nesse período foram registrados 9512 casos de MSC entre as 17280 plantas.

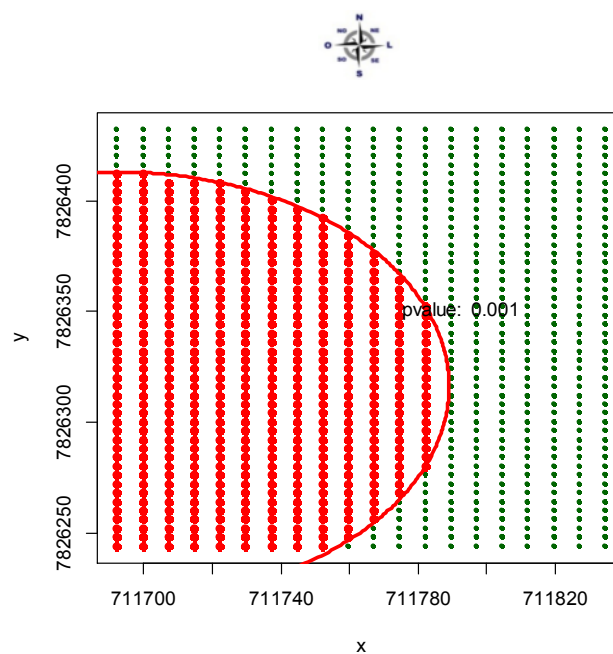


Gráfico 25 Gráfico representando o agrupamento espacial mais provável (vermelho) identificado na varredura espaço-temporal, no período de janeiro a julho de 2003, com o respectivo *valor-p*

Os resultados obtidos a partir da aplicação dos métodos baseados na função  $K$  espaço-temporal, no novo teste proposto nesta tese e na estatística de varredura espaço-temporal levaram à mesma conclusão, ou seja, existe agrupamento de casos de MSC resultantes da interação espaço-temporal. Esse



agrupamento espaço-temporal de casos de MSC ocorre, espacialmente, no lado oeste do talhão no período de janeiro a julho de 2003. Pode-se dizer que a partir do momento que a doença atinge a plantação, em algum ponto, ela começa a se espalhar entre as plantas mais próximas, formando um agrupamento espacial. Esse agrupamento se intensifica e torna cada vez maior com o passar do tempo.

Os resultados encontrados nesta tese corroboram os resultados obtidos por Krainski et al. (2008) e Lima (2005). Lima (2005) analisou o comportamento espacial e temporal da MSC utilizando técnicas diferentes das utilizadas nesta tese e também verificou que com o tempo, o principal mecanismo de disseminação da doença é a infecção de plantas saudáveis por plantas vizinhas infectadas mais próximas. Krainski et al. (2008) apresentam uma aplicação do modelo autológico a dados coletados em 11 pontos no tempo em um campo de citros afetados pela MSC e verificaram que a partir de um determinado tempo ocorre um aumento perceptível da probabilidade de uma planta se tornar doente na presença de vizinhos infectados.

#### **4.8 Considerações finais**

A análise para detectar agrupamentos de casos de doença em plantas no espaço, no tempo e na interação espaço-tempo é de extrema importância na agricultura, pois permite entender a dinâmica da doença e auxiliar o manejo em sistemas de produção. Neste sentido, objetivou-se com as pesquisas registradas nesta tese apresentar e aplicar métodos disponíveis na literatura que são baseados na função  $K$  (DIGGLE et al., 1995) e na estatística de varredura (KULLDORFF; NAGARWALLA, 1995) para analisar processos espaço-temporais pontuais. Esta tese também propôs um novo teste de hipótese para detectar agrupamento no tempo, utilizando a função distribuição dos tempos de ocorrência da doença e um novo teste de hipótese para detectar interação espaço-

tempo. Os dois novos testes de hipóteses propostos foram baseados em simulação Monte Carlo. Todos os métodos apresentados foram testados em quatro realizações geradas por meio de simulação computacional e em dados mensais de incidência de casos de Morte Súbita dos Citros (MSC) em laranjeiras. Conforme os resultados apresentados anteriormente, de uma maneira geral, todos os métodos convergiram para as mesmas conclusões. Entretanto, esses métodos merecem alguns comentários mais específicos.

A função  $K$  temporal foi originariamente proposta para analisar processos pontuais contínuos no tempo (DIGGLE et al., 1995), enquanto a estatística de varredura foi proposta para analisar processos pontuais discretos (ex. dia, mês, etc.) no tempo (KULLDORFF; NAGARWALLA, 1995). As novas técnicas baseadas na função distribuição dos tempos de ocorrência de pontos podem ser aplicadas tanto em tempos contínuos como discretos. Como os dados (simulados e reais) utilizados nesta tese são originários de processos temporais discretos, poderia se esperar um melhor rendimento dos métodos baseados na estatística de varredura e na função distribuição nesses dados. Esse comentário citado para a função  $K$  temporal pode ser estendido para as análises realizadas com a função  $K$  no espaço e no espaço tempo.

Vale ressaltar, que os métodos descritos nesta tese não podem, em princípio, serem considerados concorrentes e sim complementares. Cada método apresenta sua funcionalidade particular. A função  $K$  detecta a presença do agrupamento em uma determinada escala enquanto a varredura de Kulldorff informa quem são os agrupamentos e onde eles se encontram. Por exemplo, a metodologia proposta por Diggle et al. (1995) mostra que existe a presença de interação espaço-temporal nos casos de MSC enquanto o método de varredura de Kulldorff vem somar esta informação proferindo que o agrupamento proveniente dessa interação está localizado no lado oeste do talhão no período de janeiro a julho de 2003. O mesmo vale para os demais métodos.

Evidentemente, em uma tese muitos trabalhos foram pensados, mas não foram executados, em geral, por falta de tempo e/ou porque a tese tinha que ter um fim. Assim, a partir dos estudos conduzidos nesta tese, podem-se ter várias perspectivas de estudos futuros que devem ser atacados. São eles:

- a) Aprimorar as novas formas de teste para agrupamento temporal usando a função distribuição dos tempos de ocorrências de eventos que foram sugeridas na Seção 4.2;
- b) Testar o poder do teste e o erro tipo I nos testes de agrupamento temporal usando todas as formas da função distribuição dos tempos de ocorrências de eventos e comparar os resultados com os resultados obtidos a partir dos testes disponíveis na literatura;
- c) Testar o poder do teste e o erro tipo I no teste de hipótese para detectar interação espaço-tempo e comparar esses resultados com os resultados obtidos a partir dos testes disponíveis na literatura;
- d) Desenvolver novos métodos como os apresentados para a função  $K$  (espacial, temporal e espaço-tempo) utilizando outras funções com  $F$ ,  $G$  e  $J$ ;
- e) Desenvolver os testes apresentados para a função  $K$  (espacial, temporal e espaço-tempo) utilizando as funções  $F$ ,  $G$  e  $J$ . Considerando o espaço e o tempo discreto da mesma maneira que fez Carslake et al. (1995) para a função  $K$ ;
- f) Aplicar as metodologias apresentadas neste trabalho (novas propostas nesta tese e as encontradas na literatura) para analisar a distribuição espacial e temporal de outras doenças e/ou pragas que afetam plantações.

## 5 CONCLUSÃO

Os novos testes de hipóteses propostos nesta tese tanto para detectar agrupamento no tempo, utilizando a função distribuição dos tempos de ocorrência da doença quanto para detectar agrupamentos oriundos da interação espaço-tempo mostraram-se, no mínimo, competitivos com os atuais métodos disponíveis na literatura.

Os resultados obtidos a partir da aplicação dos métodos disponíveis na literatura e dos métodos propostos nesta tese nos dados simulados levaram as mesmas conclusões, ou seja, que todos os métodos conseguiram identificar corretamente os processos estocásticos pontuais que geraram os dados no espaço e tempo.

Esses mesmos métodos também levaram, em geral, às mesmas conclusões com relação à dinâmica espacial e temporal dos casos de morte súbita dos citros no talhão, ou seja, existe a presença de agrupamento no tempo, de agrupamento no espaço e agrupamento resultante da interação espaço-temporal. Assim, pode-se concluir que os casos de MSC ocorrem, espacialmente, no lado oeste do talhão no período de janeiro a julho de 2003. Essa conclusão é compatível com o comportamento de uma doença que atinge a plantação em algum ponto e começa a se espalhar entre as plantas mais próximas, formando um agrupamento espacial. Esse agrupamento se intensifica e se torna cada vez maior com o passar do tempo até que toda a plantação esteja tomada pela doença.

## REFERÊNCIAS

ARBIA, G. et al. Detecting the existence of space-time clustering of firms. **Regional Science and Urban Economics**, Bogotá, v. 40, n. 5, p. 311-323, Nov. 2010.

BADDELEY, A. Spatial point processes and their applications. In: BADDELEY, A. et al. (Ed.). **Stochastic geometry**: lectures given at the C.I.M.E. summer school held in Martina Franca, Italy, september 13-18, 2004, lecture notes in mathematics 1892. Firenze: Fondazione C.I.M.E., 2006. p. 1-75.

BADDELEY, A. J.; TURNER, R. Spatstat: an R package for analyzing spatial point patterns. **Journal of Statistical Software**, Los Angeles, v. 12, n. 6, p. 1-42, June 2005.

BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. Essex: Longman Scientific, 1995. 409 p.

BALIEIRO, A. A. S. **Detecção de conglomerados dos alertas de desmatamentos no Estado do Amazonas usando estatística de varredura espaço-temporal**. 2008. 87 f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Viçosa, Viçosa, MG, 2008.

BARTLETT, M. S. The spectral analysis of two-dimensional point processes. **Biometrika**, London, v. 51, p. 299-311, 1964.

BASSANEZI, R. B. et al. Spatial and temporal analyses of citrus sudden death as a tool to generate hypotheses concerning its etiology. **Phytopathology**, Saint Paul, v. 93, n. 4, p. 502-512, 2003.

BASSANEZI, R. B.; GIMENES-FERNANDES, N.; YAMAMOTO, P. T. **Morte súbita dos citros**. Araraquara: Fundecitrus, 2003. 54 p. (Boletim Citricola, 24).

BASSANEZI, R. B.; YAMAMOTO, P. T. **Manual da morte súbita dos citros**. Araraquara: Fundecitrus, 2007. 12 p.

BESAG, J.; NEWEL, J. The detection of clusters in rare diseases. **Journal of the Royal Statistical Society**, London, v. 154, n. 1, p. 143-155, June 1991.

BHATTARAI, N. R. et al. Domestic animals and epidemiology of visceral leishmaniasis, Nepal. **Emerging Infectious Diseases**, Atlanta, v. 16, n. 2, p. 231-237, Oct. 2010.

CANNON, A.; CRESSIE, N. Temporal analogues to spatial  $K$  functions. **Biometrical Journal**, Berlin, v. 37, n. 3, p. 351-373, June 1995.

CARSLAKE, D. et al. Space time clustering of cowpox virus infection in wild rodent populations. **Journal of Animal Ecology**, Oxford, v. 74, n. 4, p. 647-655, Dec. 2005.

CHARNET, R. et al. **Análise de modelos de regressão linear: com aplicações**. Campinas: UNICAMP, 2008. 368 p.

CITRUS BR. **O panorama da citricultura no mundo**. Disponível em: <[http://www.citros.com.br/download/Apresentacao\\_CitrusBR\\_Parana](http://www.citros.com.br/download/Apresentacao_CitrusBR_Parana)>. Acesso em: 7 nov. 2012.

CRESSIE, N. A. C. **Statistics for spatial data**. New York: J. Wiley, 1993. 900 p.

DALEY, D. J.; VERE-JONES, D. **An introduction to the theory of point processes**. Berlin: Springer, 1988. 552 p.

DIGGLE, P. J. **Statistical analysis of spatial point patterns**. London: Arnold, 2003. 153 p.

DIGGLE, P. J. et al. Second order analysis of space-time clustering. **Statistical Methods in Medical Research**, Urbana, v. 4, n. 2, p. 124-136, 1995.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Importância econômica**. Disponível em: <<http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Citros/CitrosNordeste/importancia.htm>>. Acesso em: 6 nov. 2012.

FUNDECITRUS. **Identificação da doença**. Disponível em: <<http://www.fundecitrus.com.br/pagina/morte-subita-dos-citros.html>>. Acesso em: 6 nov. 2012.

GIMENES-FERNANDES, N.; BASSANEZI, R. B. Doença de causa desconhecida afeta pomares cítricos no norte de São Paulo e sul do Triângulo Mineiro. **Summa Phytopathologica**, Jaguariúna, v. 27, p. 93, 2001. Resumo.

GOTTWALD, T. R. et al. Geo-referenced spatiotemporal analysis of the urban citrus canker epidemic in Florida. **Epidemiology**, Oxford, v. 92, n. 4, p. 361-377, 2002.

KNOX, G. Detection of low density epidemicity. **British Journal of Preventative and Social Medicine**, London, v. 17, p. 121-27, 1963.

KRAINSKI, E. T. et al. Rcitrus: autologistic model with na application to the citrus sudden death disease. **Scientia Agricola**, Piracicaba, v. 65, n. 5, p. 541-547, set./out. 2008.

\_\_\_\_\_. Rcitrus: pacote para análise de padrões espaciais de doenças de citrus. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 50., 2005, Londrina. **Anais...** Londrina: RBRAS, 2005. 1 CD-ROM.

KRAINSKI, E. T.; RIBEIRO JÚNIOR, P. J. **Padrões espaciais na epidemiologia da morte súbita dos citros: 2<sup>o</sup> relatório das análises estatísticas**. Curitiba: [s.n.], 2004. 84 p.

\_\_\_\_\_. **Rcitrus**: functions for the analysis of citrus disease data. Version 0.3-0. Curitiba: UFPR/LEG, 2007. Software.

KULLDORFF, M. Prospective time-periodic geographical disease surveillance using a scan statistic. **Journal of the Royal Statistical Society**, London, v. 164, n. 1, p. 61-72, Sept. 2001.

\_\_\_\_\_. **SaTScanTM user guide**. Disponível em: <<http://www.satscan.org/>>. Acesso em: 15 out. 2011.

\_\_\_\_\_. Spatial scan statistic. **Communications in Statistics. Theory and Methods**, New York, v. 26, n. 6, p. 1481-1496, 1997.

KULLDORFF, M. et al. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. **American Journal of Public Health**, Boston, v. 88, n. 9, p. 1377-1380, Sept. 1998.

KULLDORFF, M.; NARGAWALLA, N. Spatial disease clusters: detection and inference. **Statistics in Medicine**, New York, v. 14, p. 799-810, Apr. 1995.

LIMA, R. R. **Modelagem espaço-temporal para dados de incidência de doenças em plantas**. 2005. 149 f. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 2005.

NEVES, F. M. et al. **O retrato da citricultura brasileira**. Disponível em: <[http://www.citrusbr.com.br/download/biblioteca/o\\_retrato\\_da\\_citricultura.pdf](http://www.citrusbr.com.br/download/biblioteca/o_retrato_da_citricultura.pdf)>. Acesso em: 3 nov. 2012.

OVIEDO, M. et al. Evaluation of mass vaccination programmes: the experience of hepatitis A in Catalonia, in Spanish. **Revista Española de Salud Pública**, Madrid, v. 83, n. 5, p. 697-709, mayo 2009.

PEIXOTO, M. S.; BARROS, L. C. de; BASSANEZI, R. C. A model of cellular automata for the spatial and temporal analysis of citrus sudden death with the fuzzy parameter. **Ecological Modelling**, Amsterdam, v. 214, n. 1, p. 45-52, Mar. 2008.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2012. Disponível em: <<http://www.r-project.org>>. Acesso em: 15 jan. 2012.

RIPLEY, B. D. Modelling spatial patterns: with discussion. **Journal of the Royal Statistical Society, Series B**, London, v. 39, n. 2, p. 172-212, 1977.

\_\_\_\_\_. The second-order analysis of stationary point patterns. **Journal of Applied Probability**, Sheffield, v. 13, p. 255-266, 1976.

ROWLINGSON, B.; DIGGLE, P. J. Splancs: spatial point patterns analysis code in S-PLUS. **Computers and Geosciences**, Ottawa, v. 19, n. 5, p. 627-655, 1993.

SCALON, J. D. et al. Spatial pattern analysis of second-phase particles in composite materials. **Material Science and Engineering**, Lausanne, v. 356, n. 1/2, p. 245-257, Sept. 2003.

SCALON, J. D.; SILVA, F. M. Power of tests for spatial randomness in patterns with small number of events. **Revista de Ciência e Tecnologia**, Rio de Janeiro, v. 12, n. 2, p. 7-14, fev. 2006.



SPÓSITO, M. B. **Dinâmica temporal e espacial da mancha preta (*Guignardiacitricarpa*) e quantificação dos danos causados à cultura dos citros**. 2003. 112 f. Tese (Doutorado em Agronomia) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 2003.

SPÓSITO, M. B. et al. Spatial patterns of trees affected by black spots in citrus groves in Brazil. **Plant Disease**, Quebec, v. 91, n. 1, p. 36-40, Jan. 2007.

TERENCIO, J. C.; MILORI, D. M. B. P.; MARTIN-NETO, L. Identificação de morte súbita dos citros usando espectroscopia de fluorescência. **Boletim de Pesquisa e Desenvolvimento**, Brasília, v. 11, n. 1, p. 1-12, 2005.

WHITTEMORE, A. S. et al. A test to detect clusters of disease. **Biometrika**, London, v. 74, p. 631-635, 1987.

## ANEXOS

<b>ANEXO A</b>	Programa Ksenvelop: essa função implementa simulações Monte Carlo para criar e plotar envelopes de simulação para testar a hipótese de completa aleatoriedade espacial (CAE) usando a função $K$ espacial .....	123
<b>ANEXO B</b>	Programa Kstest: essa função implementa simulações Monte Carlo para testar a hipótese de completa aleatoriedade espacial (CAE) usando a função $K$ espacial .....	125
<b>ANEXO C</b>	Programa Ktenvelop: essa função implementa simulações Monte Carlo para criar e plotar envelopes de simulação para testar a hipótese de completa aleatoriedade temporal (CAT) usando a função $K$ temporal .....	128
<b>ANEXO D</b>	Programa Kttest: essa função implementa simulações Monte Carlo para testar a hipótese de completa aleatoriedade temporal (CAT) usando a função $K$ temporal .....	130
<b>ANEXO E</b>	Programa Tempotest: essa função implementa simulações Monte Carlo para testar a hipótese de completa aleatoriedade temporal (CAT) usando a função distribuição do tempo .....	133
<b>ANEXO F</b>	Programa Mctest: essa função implementa simulações Monte Carlo para testar a hipótese de completa aleatoriedade espaço-temporal e retorna o <i>valor-p</i>	

.....	135
<b>ANEXO G</b> Programa Clustersatscan: essa função plota os círculos (agrupamentos identificados) no mapa, no R por meio de saídas geradas pelo <i>software</i> SaTScan	
.....	137

**ANEXO A** - Programa Ksenvelop: essa função implementa simulações Monte Carlo para criar e plotar envelopes

```
function(event, times, poly, s, tm, nsim = 10, quiet=FALSE){
  #
  # This function implements Monte Carlo simulations
  # to create and plot simulation envelopes to test
  # against the hypothesis of complete temporal randomness (CTR)
  # of an observed spatial-temporal point pattern.
  # It uses the spatial k-function and the cramer-vonMises measure of distance.
  #
  # OBS: You need splancs loaded
  #
  # Where:
  # - event: A 2xn matrix that contains the coordinates of events
  # - times: A 1xn vector of times
  # - poly: A polygon enclosing the points
  # - s: A vector of spatial distances for the analysis.
  # - tm: A vector of times for the analysis
  # - nsim: Number of Monte Carlo simulations
  # - quiet: If quiet=TRUE then no output is produced,
  # otherwise the function prints the number of simulations
  #
  # This is crude code by J.D.Scalon. No warranty! (14/11/12)
  #
  library(splancs)
  #
  # Starting values
  #
  maxx<-max(event[,1])
  maxy<-max(event[,2])
  minx<-min(event[,1])
  miny<-min(event[,2])
  lent<-length(event[,1])
  tl<-cbind(min(times),max(times))
  #
  # Get the K-function for space and time of the observed point pattern
  #
  ksto <- stkhat(pts=event, times=times, poly=poly, tlimits=tl, s=s, tm=tm)
  ko <- ksto$ks
}
```

```

#
# Doing Monte Carlo simulation under the CSR
#
hold<-matrix(0,nsim,length(s))
for(i in (1:nsim)) {
  if (!quiet)
    cat("Doing simulation ", i, "\n")
  tt<-floor(runif(lent,min(times),max(times)))
  xx<-runif(lent,minx,maxx)
  yy<-runif(lent,miny,maxy)
  pont<-cbind(xx,yy)
  tll<-cbind(min(tt),max(tt))
  kstm<-stkhat(pts=pont, times=tt, poly=poly, tlimits=tll, s=s, tm=tm)
  kko<-kstm$ks
  hold[i,]<-kko
}
#
up<- apply(hold,2,max)
dow<- apply(hold,2,min)
#
# Creating plot
#
par(pty = "s")
plot(s, up, type = "n", xlab = "Distância ", ylab = "Função K espacial")
lines(s, ko, col="blue")
lines(s, up, lty=2, col="red")
lines(s, dow,lty=2,col="red")
}

```

**ANEXO B** - Programa Kstest: essa função implementa simulações Monte Carlo para testar a hipótese de CAE

```
function(event, times, poly, s, tm, nsim = 10, quiet=FALSE){
#
# This function implements Monte Carlo simulations
# for getting a test against the hypothesis of complete spatial
# randomness (CSR)of an observed spatial-time point pattern.
#
# It uses the spatial K-function, Cramer-vonMises and Kolmogorov-Smirnov
# measure of distances for getting statistic of test.
#
# Where:
#
# - event: A 2xn matrix that contains the coordinates of the events
# - times: A 1xn vector with times
# - poly: A polygon enclosing the points
# - s: A vector of spatial distances for the analysis.
#     WARNING: s MUST BE CHOSEN WITH CAUTION.
#     Use the function "anaksenv" for helping your choice
# - tm: A vector of times for the analysis
# - nsim: Number of Monte Carlo simulations
# - quiet: If quiet=TRUE then no output is produced,
#     otherwise the function prints the number of simulations
#
# This is crude code by J.D.Scalon. No warranty! (16/11/12)
#
library(splancs)
#
# Starting values
#
maxx<-max(event[,1])
maxy<-max(event[,2])
minx<-min(event[,1])
miny<-min(event[,2])
lent<-length(event[,1])
tl<-cbind(min(times),max(times))
#
# Get the K-function for space and time of the observed point pattern
#
```

```

ksto <- stkhath(pts=event, times=times, poly=poly, tlimits=tl, s=s, tm=tm)
ko <- ksto$ks
#
#
#   Get the statistics for the observed spatial point pattern
#
lente<- length(tm)
DC<-(ko-pi*s^2)^2
DK<-max(abs(ko-pi*s^2))
step<-max(tm)/lente
integ<-0
steplength <- step * 2
for(count in (2:(lente - 2))) {
  integ <- integ + (DC[count - 1] + 4 * DC[count] + DC[count + 1]) *
    steplength
}
integ <- integ/6
#
# Doing Monte Carlo simulation under the CTR
#
for(i in (2:nsim)) {
  if (!quiet)
    cat("Doing simulation ", i, "\n")
  tt<-floor(runif(lent,min(times),max(times)))
  xx<-runif(lent,minx,maxx)
  yy<-runif(lent,miny,maxy)
  pont<-as.points(cbind(xx,yy))
  tll<-cbind(min(tt),max(tt))
  kstmcc<-stkhath(pts=pont, times=tt, poly=poly, tlimits=tll, s=s, tm=tm)
  kl<-kstmcc$ks
  #
  #   Get statistics under CSR
  #
  tc <- (kl - pi*s^2)^2
  tk <- max(abs(kl-pi*s^2))
  step<-max(tm)/lente
  teg <- 0
  steplength <- step * 2
  for(count in (2:(lente - 2))) {
    teg <- teg + (tc[count - 1] + 4 * tc[count] + tc[count + 1]) *
      steplength
  }
}

```

```
teg <- teg/6
integ <- c(integ, teg)
DK<-c(DK,tk)
}
#
  rdatc<-rank(integ)[1]
  rdatk<-rank(DK)[1]
#
  pvc<-(nsim+1-rdatc)/nsim
  pvk<-(nsim+1-rdatk)/nsim
#
# Results: It uses P-values to see whether observed values
# are unusually large for the null distribution (CSR).
#
list(est_cramer=integ[1], p_valueC=pvc, Kolmogorov-Smirnov=DK[1],
p_valueK=pvk)
}
```



**ANEXO C** - Programa Ktenvelop: essa função implementa simulações Monte Carlo para criar e plotar envelopes

```
function(event,times, poly, s, tm, nsim = 10, quiet=FALSE){
  #
  # This function implements Monte Carlo simulations
  # to create and plot simulation envelopes to test
  # against the hypothesis of complete temporal randomness (CTR)
  # of an observed spatial-temporal point pattern.
  # It uses the temporal k-function and the cramer-vonMises measure of
distance.
  #
  # OBS: You need splancs loaded
  #
  # Where:
  # - event: A 2xn matrix that contains the coordinates of the events
  # - times: A 1xn vector of times
  # - poly:   A polygon enclosing the points
  # - s:      A vector of spatial distances for the analysis.
  # - tm:     A vector of times for the analysis
  # - nsim: Number of Monte Carlo simulations
  # - quiet: If quiet=TRUE then no output is produced,
  #          otherwise the function prints the number of simulations
  #
  # This is crude code by J.D.Scalon. No warranty! (16/11/12)
  #
  library(splancs)
  # Starting value
  maxx<-max(event[,1])
  maxy<-max(event[,2])
  minx<-min(event[,1])
  miny<-min(event[,2])
  lent<-length(event[,1])
  tl<-cbind(min(times),max(times))
  #
  # Get the K-function for space and time of the observed point pattern
  #
  ksto <- stkhat(pts=event, times=times, poly=poly, tlimits=tl, s=s, tm=tm)
  ko <- ksto$kt
  #
}
```

```

# Doing Monte Carlo simulation under the CSR
#
hold<-matrix(0,nsim,length(tm))
for(i in (1:nsim)) {
  if (!quiet)
    cat("Doing simulation ", i, "\n")
  tt<-floor(runif(lent,min(times),max(times)))
  xx<-runif(lent,minx,maxx)
  yy<-runif(lent,miny,maxy)
  pont<-cbind(xx,yy)
  tll<-cbind(min(tt),max(tt))
  kstmc<-stkhat(pts=pont, times=tt, poly=poly, tlimits=tll, s=s, tm=tm)
  kko<-kstmc$kt
  hold[i,]<-kko
}
#
up<- apply(hold,2,max)
dow<- apply(hold,2,min)
# Creating plot
par(pty = "s")
plot(tm, up, type = "n", xlab = "Tempo", ylab = "Função K temporal")
lines(tm, 2*tm)
lines(tm, ko, col="blue")
lines(tm, up, lty=2, col="red")
lines(tm, dow,lty=2,col="red")
}

```

**ANEXO D** - Programa Kttest: essa função implementa simulações Monte Carlo para testar a hipótese de CAT

```
function(event, times, poly, s, tm, nsim = 10){
  #
  # This function implements Monte Carlo simulations
  # for getting a test against the hypothesis of complete temporal
  # randomness (CTR) of an observed spatial-time point pattern.
  #
  # It uses the temporal K-function, Kolmogorov-Smirnov and the Cramer-
vonMises
  # measure of distances for getting statistic of test.
  #
  # Where:
  # - event: A 3xn matrix that contains the coordinates of the events
  # - times: A 1xn vector with times
  # - poly: A polygon enclosing the points
  # - s: A vector of spatial distances for the analysis.
  # - tm: A vector of times for the analysis
  # - nsim: Number of Monte Carlo simulations
  #
  # This is crude code by J.D.Scalon. No warranty! (16/11/12)
  #
  library(splancs)
  #
  # Starting values
  #
  maxx<-max(event[,1])
  maxy<-max(event[,2])
  minx<-min(event[,1])
  miny<-min(event[,2])
  lent<-length(event[,1])
  tl<-cbind(min(times),max(times))
  #
  # Get the K-function for space and time of the observed point pattern
  #
  ksto <- stkhat(pts=event, times=times, poly=poly, tlimits=tl, s=s, tm=tm)
  ko <- ksto$kt
  #
  # Get the statistics for the observed spatial point pattern
```

```

#
lente<- length(tm)
DC<-(ko-2*tm)^2
DK<-max(abs(ko-2*tm))
step<-max(tm)/lente
integ<-0
steplength <- step * 2
for(count in (2:(lente - 2))) {
  integ <- integ + (DC[count - 1] + 4 * DC[count] + DC[count + 1]) *
    steplength
}
integ <- integ/6
#
# Doing Monte Carlo simulation under the CTR
#
for(i in (2:nsim)) {
  tt<-floor(runif(lent,min(times),max(times)))
  xx<-floor(runif(lent,minx,maxx))
  yy<-floor(runif(lent,miny,maxy))
  pont<-cbind(xx,yy)
  tll<-cbind(min(tt),max(tt))
  kstmc<-stkhat(pts=pont, times=tt, poly=poly, tlimits=tll, s=s, tm=tm)
  kl<-kstmc$kt
  #
  # Get the statistics under CTR
  #
  tc <- (kl - 2*tm)^2
  tk <- max(abs(kl-2*tm))
  step<-max(tm)/lente
  teg <- 0
  steplength <- step * 2
  for(count in (2:(lente - 2))) {
    teg <- teg + (tc[count - 1] + 4 * tc[count] + tc[count + 1]) *
      steplength
  }
  teg <- teg/6
  integ <- c(integ, teg)
  DK<-c(DK,tk)
}
rdatc<-rank(integ)[1]
rdatk<-rank(DK)[1]
#

```

```
pvc<-(nsim+1-rdatc)/nsim
pvk<-(nsim+1-rdatk)/nsim
#
# Results: It uses P-values to see whether observed values
# are unusually large for the null distribution (CTR).
#
list(est_cramer=integ[1], p_valueC=pvc, est_kolm=DK[1], p_valueK=pvk)
}
```

**ANEXO E** - Programa Tempotest : essa função implementa simulações Monte Carlo para testar a hipótese de CAT e plota o gráfico

```
function(event, timein=c(0,365), nsim = 1000){
  #
  # This function uses the counting process theory and Monte Carlo
simulations
  # for getting a test against the hypothesis of complete temporal
  # randomness (CTR) of an observed time point pattern.
  #
  # It uses the empirical distribution function of time and
  # the Komogorov measure of distances for getting statistic of test.
  #
  # Where:
  # - event: A vector that contain the time occurrences of events
  # - timein: A vector that contain the time interval of the experiment -ex.
timein= c(0,b)
  # - nsim: Number of Monte Carlo simulations
  #
  # This is crude code by J.D.Scalon. No warranty! (25/12/12)
  #
  dimen<-max(timein[2])
  ct<-tabulate(event, nbins=dimen)
  tm<-seq(timein[1]:timein[2])
  #
  # Get the T-function of the observed point pattern
  #
  ko <- cumsum(ct)
  ko <-c(0,ko)
  #
  # Get the T-function of the theoretical point pattern under CRT
  #
  lamb<-length(event)/(timein[2]-timein[1])
  zz<- lamb*tm
  #
  # Get the kolmogorov distance for the observed temporal point pattern
  #
  D<-max(abs(ko-zz))
  #
  # Doing Monte Carlo simulation under the CTR
```

```

#
  for(i in (2:nsim)) {
    sp1<-floor(runif(length(event),min(event),max(event)))
    spts1<-tabulate(sp1, nbins=dimen)
    mxs<-max(cumsum(spts1))
    kl<-cumsum(spts1)
    kl<-c(0,kl)
#
# Get the kolmogorov under CTR
#
    t <- max(abs(kl - zz))
    D <- c(D, t)
  }
#
# Get the p-value
#
  pk <- rank(D)[1]
  pv <- (nsim + 1 - pk)/nsim
#
# Plot
#
  par(pty="s")
  plot(tm,zz,xlab="Tempo",ylab="Funcao      distribuicao      das
ocorrencias", type="n")
  lines(tm,ko,col="red")
  lines(tm,zz,lty=2)
#
# p-values
#
  list(est_test=D[1], p_value=pv)
}

```

**ANEXO F** - Programa Mctest: essa função implementa simulações Monte Carlo para criar e plotar envelopes

```
function(event, times, poly, tlimits, s, tt, nsim, quiet = FALSE,
         returnSims = FALSE)
#
# This function performs a Monte Carlo test of space-time clustering
# and returns p-value.
#
# Ho: There is NO space-time clustering
# H1: There is space-time clustering
#
# event:   A set of points as used by Splancs.
# times:   A vector of times, the same length as the number of points in
# pts.
# poly:    A polygon enclosing the points.
# s:       A vector of spatial distances for the analysis.
# tt:      A vector of times for the analysis.
# nsim:    The number of simulations to do.
# quiet:   If quiet=TRUE then no output is produced,
#          otherwise the function prints the number of simulations
#          completed so far, and also how the test statistic for
#          the data ranks with the simulations.
# returnSims: default FALSE, if TRUE, return the stkhat output
#             for the observed data and each simulation as
#             attributes obs and sims.
#             November 19nd, 2012
{
  stats <- NULL
  ntpts <- npts(event)
  ktsim <- stkhat(event, times, poly, tlimits, s, tt)
  resids <- ktsim$kst - outer(ktsim$ks, ktsim$kt)
  t0 <- sum(resids)
  if(returnSims) {
    obs <- ktsim
    sims <- vector(mode = "list", length = nsim)
  }
  for(isim in (2:nsim)) {
    if(!quiet)
      cat("Doing simulation ", isim, "\n")
  }
}
```



```
times <- sample(times)
ktsim <- stkhat(event, times, poly, tlimits, s, tt)
resids <- ktsim$kst - outer(ktsim$ks, ktsim$kt)
stats <- c(stats, sum(resids))
if (returnSims)
  sims[[isim]] <- ktsim
}
rdat <- rank(c(t0, stats))[1]
pv <- (nsim+1-rdat)/nsim
res <- list(t0 = t0, t = stats, pvalue=pv)
if (returnSims) {
  attr(res, "obs") <- obs
  attr(res, "sims") <- sims
}
res
}
```

**ANEXO G** - Programa Clustersatscan: essa função plota os círculos no mapa, por meio de saídas geradas pelo *software* SaTScan

```

library(RgoogleMaps)
#
## Leitura dos dados
#
gis <- read.table("Talhao-1.gis.txt", header=FALSE)
names(gis) <- c("LID", "Cl_Numb", "pvalue", "obs", "exp", "ratio",
               "Lobs", "Lexp", "ratioL", "V10", "V11")
#
geo <- read.table("Talhao-1.geo", header=FALSE)
names(geo) <- c("ID", "x", "y")
#
out <- read.table("Talhao-1.col.txt", header=FALSE)
names(out) <- c("ID", "geoID", "x", "y", "radius", "start", "end", "LID",
               "TStat", "pvalue", "obs", "expec", "ratio", "V14")
#
#
## Plota os círculos no mapa, no R
#
plot(geo$x, geo$y, xlab='x coordinate', ylab='y coordinate', pch=19, col='dark
green', cex=0.5)
#
vcolor <- c('red', 'orange', 'blue')
#
## Plota os três primeiros clusters
for(numCluster in 1:min(3, max(gis$Cl_Numb))){
#
  Xcenter <- geo$x[ out$geoID[numCluster] ]
  Ycenter <- geo$y[ out$geoID[numCluster] ]
  radius <- out$radius[numCluster]
#
  ## Plota os pontos com a cor do cluster a que pertencem
#
  auxID <- gis$LID[gis$Cl_Numb == numCluster]
  for(cont in 1:length(auxID)){
    coordX <- geo$x[ geo$ID == auxID[cont] ]
    coordY <- geo$y[ geo$ID == auxID[cont] ]
    points(coordX, coordY, pch=19, col=vcolor[numCluster], cex=1.0)
  }
}

```

```
}  
#  
  if(radius != 0){  
    angle <- seq(0, 2*pi, by=0.01)  
    coordX <- radius*sin(angle)  
    coordY <- radius*cos(angle)  
#  
    lines(Xcenter+coordX, Ycenter+coordY, col=vcolor[numCluster], lwd=3)  
  } else {  
    points(Xcenter, Ycenter, pch=19,col=vcolor[numCluster],cex=4)  
  }  
#  
  aux <- locator(1)  
  text(aux$x, aux$y, paste("pvalue: ",out$pvalue[numCluster]))  
}
```