

Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Modelo autologístico no estudo de padrões espaciais em doenças de citros

Luziane Franciscan

Dissertação apresentada para obtenção do título de Mestre em
Agronomia. Área de concentração: Estatística e Experimentação
Agronômica

Piracicaba

2008

Luziane Franciscon
Bacharel em Estatística

Modelo autologístico no estudo de padrões espaciais em doenças de citros

Orientador:

Prof. Dr. **PAULO JUSTINIANO RIBEIRO JÚNIOR**

Dissertação apresentada para obtenção do título de Mestre em
Agronomia. Área de concentração: Estatística e Experimentação
Agrônômica

Piracicaba

2008

**Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP**

Franciscon, Luziane
Modelo autolístico no estudo de padrões espaciais em doenças de citros / Luziane Franciscon. - - Piracicaba, 2008.
68 p.

Dissertação (Mestrado) - - Escola Superior de Agricultura Luiz de Queiroz, 2008.
Bibliografia.

1. Doenças de plantas 2. Estatísticas espaciais 3. Frutas cítricas 4. Verossimilhança
I. Título

CDD 634.3
F818m

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”

“Ao homem é dado um número limitado de experiências
- se ele não as desperdiçar, prolongará sua vida.”
George Ivanovitch Gurdjieff

AGRADECIMENTOS

Aos meus amáveis pais Altair e Lourdes Franciscan, por me ensinarem os primeiros passos, os grandes valores humanos, e por me deixarem voar.

Ao professor Paulo Justiniano Ribeiro Júnior, sua orientação, as incríveis conversas esclarecedoras, empolgantes e amigas, pelo entusiasmo acolhedor de sempre.

Ao meu esposo Dênis Weis Naressi, por me compreender e esperar com paciência, pela felicidade em compartilhar momentos da minha vida ao seu lado.

A minha professora do primário Odila, por me envolver nos estudos, com alegria, desde os primeiros anos de escola.

Aos companheiros de estudo do mestrado e do doutorado, pela amizade, as demonstrações matemáticas, e a convivência.

Ao querido amigo Elias, pela amizade, ensinamentos, a construção dos pacotes para as análises.

As amigas Aline e Fernanda por me acolherem em Piracicaba.

Aos professores do programa de Pós-graduação em Estatística e Experimentação Agronômica ESALQ/USP pelos ensinamentos.

Ao Fundecitros, na pessoa de Renato Bassanezi pela concessão dos dados.

À pesquisadora Ana Beatriz Czermainski da Embrapa Uva e Vinho, por suas contribuições.

À Embrapa Suínos e Aves pelo apoio.

À equipe do LEG, pelas boas conversas estatísticas e pelo espaço físico.

Aos funcionários do Departamento de Ciências Exatas da ESALQ/USP, sempre prontos a ajudar.

A todos, que contribuíram para a realização deste trabalho.

SUMÁRIO

RESUMO	7
ABSTRACT	8
LISTA DE FIGURAS	9
LISTA DE TABELAS	10
1 INTRODUÇÃO	11
2 DESENVOLVIMENTO	13
2.1 Revisão bibliográfica	13
2.1.1 Morte súbita dos citros	13
2.1.2 Leprose dos citros	15
2.1.3 Modelos lineares generalizados	16
2.1.4 Modelo de regressão logística	18
2.1.5 Modelo autológico	20
2.1.5.1 Estrutura de vizinhança	21
2.1.5.2 Especificação do modelo	22
2.1.5.3 Inferência	23
2.1.5.3.1 Método de máxima pseudo-verossimilhança	25
2.1.5.3.2 Método de <i>Bootstrap</i>	25
2.1.5.3.3 Método de Monte Carlo	26
2.1.5.3.4 Método de Monte Carlo com Cadeias de Markov (MCMC)	27
2.1.5.4 Covariáveis de vizinhança	28
2.1.5.5 Modelo autológico espaço-temporal	30
2.2 Material e métodos	32
2.2.1 Material	32
2.2.1.1 Dados de morte súbita dos citros	32
2.2.1.2 Dados de leprose dos citros	33
2.2.2 Métodos	35
2.2.2.1 Metodologia para os dados de MSC	35
2.2.2.2 Metodologia para os dados de leprose dos citros	38
2.3 Resultados e discussão	41
2.3.1 Dados de morte súbita em citros	41
2.3.2 Dados de leprose dos citros	44
3 CONCLUSÕES	51

REFERÊNCIAS 52

APÊNDICES 57

RESUMO

Modelo autológico no estudo de padrões espaciais em doenças de citros

A citricultura é uma das principais atividades agrícolas do Brasil e o estado de São Paulo concentra a maior área produtora de laranjas do mundo. O conhecimento de padrões da incidência de doenças cítricas no tempo e no espaço é relevante para o setor e permite a descrição da dinâmica dessas doenças, podendo indicar estratégias para controle de epidemias. Neste trabalho são consideradas duas doenças que afetam a cultura de citros, a leprose e a morte súbita dos citros utilizando dados provenientes do monitoramento de talhões. Um aspecto relevante para estudos de doenças como a leprose dos citros, considerada uma grave virose na citricultura brasileira, é a investigação do padrão espacial e dos efeitos temporais da sua incidência dentro do talhão. Métodos exploratórios para determinar se o padrão espacial é ou não agregado são frequentemente utilizados. Entretanto é possível explorar e descrever os dados adotando um modelo explícito, permitindo discriminar e quantificar os efeitos através de parâmetros para covariáveis que representam os aspectos de interesse. Uma das alternativas é a adoção de modelos autológicos, que estendem o modelo de regressão logística para acomodar efeitos espaciais. Para implementar esse modelo é necessário que se re-use os dados para extrair covariáveis espaciais, o que requer extensões na metodologia e algoritmos para avaliar adequadamente a variância das estimativas. Neste trabalho utiliza-se o modelo autológico na análise de dados de incidência de doenças em plantas cítricas coletados em pontos referenciados no espaço e no tempo em um talhão. É mostrado como o modelo autológico é apropriado para investigar doenças desse tipo, bem como é feita uma descrição do modelo e dos aspectos computacionais necessários para a estimação do modelo. São abordados métodos de seleção e avaliação de modelos autológicos que relacionam fatores que afetam a disseminação da doença com padrões espaciais e efeitos temporais. Desta forma é possível realizar avaliações objetivas dos efeitos dos fatores considerados sobre a incidência da doença através dos parâmetros estimados do modelo proposto e quantificar o efeito da presença da doença em diferentes estruturas de vizinhança. A modelagem dos dados identificou dependência espacial entre as plantas e o modelo adotado permitiu quantificar as variações na probabilidade de doença em função do *status* das plantas na vizinhança. A metodologia apresentada aqui não se restringe a cultura de citros pode ser usada na avaliação de padrões espaço-temporais e efeitos de fatores que afetem doenças de plantas em condições semelhantes.

Palavras-chave: Estatística espacial; Pseudo-verossimilhança; Doenças em citros

ABSTRACT

Autologistic model in the study of spatial patterns in citrus diseases

The citrus industry is a major agricultural activity in Brazil and the state of Sao Paulo concentrates the largest production area of oranges in the world. The knowledge of incidence patterns of citrus diseases in time and space is relevant to the industry and allows the description of the dynamics of these diseases and may indicate strategies for epidemics control. In this work are considered two diseases that affect the cultivation of citrus, leprosis and citrus sudden death using data from the tracking of stands. An important point when studying diseases such as the citrus leprosis, considered a serious viral disease in the Brazilian citrus industry, is the investigation of the spatial pattern and temporal effects of the disease incidence within a stand. Exploratory methods to determine if the spatial pattern is or not added are frequently used. However it is possible to explore and describe the data adopting an explicit model, allowing to discriminate and quantify the effects through parameters for covariates that represent aspects of interest. To implement this model is necessary to reuse the data in order to extract spatial covariates, which requires extensions in the methodology and algorithms to assess properly the variance of estimates. In this work, the autologistic model is used in the analysis of diseases incidence data in citrus plants collected in points referenced in space and time in a stand. It is shown how the autologistic model is appropriate to investigate such diseases, and there is a description of the model and computational aspects needed to estimate the model. Thus it is possible to achieve objective assessments of the effects of the factors considered on the incidence of the disease through the estimated parameters of the proposed model and quantify the disease presence effects in different neighborhood structures. The modeling of the data has identified a spatial dependence between the plants and the adopted model allowed to quantify the changes in the probability of disease according to the status of the plants in the neighbourhood. The methodology presented here is not restricted to the cultivation of citrus. It can be used in the assessment of spatial-temporal patterns and effects of factors that affect the diseases in plants under similar conditions.

Keywords: Spatial statistics; Pseudo-likelihood; Citrus diseases

LISTA DE FIGURAS

Figura 1 - Planta com sintomas de MSC	14
Figura 2 - Órgãos da planta com sintomas de leprose dos citros	15
Figura 3 - Estruturas de vizinhança: primeira ordem, segunda ordem e segunda ordem separadamente	22
Figura 4 - Incidência de MSC em talhão da Fazenda Vale Verde em 12 avaliações entre 05/11/2001 e 07/10/2002.	32
Figura 5 - Mapas com incidência de leprose dos citros em um talhão nas avaliações do ano de 2003	33
Figura 6 - Mapas com incidência de leprose dos citros em um talhão nas avaliações do ano de 2004	34
Figura 7 - Incidência de leprose em cada avaliação	44
Figura 8 - Coeficientes e erros-padrão para o modelo autológico	47

LISTA DE TABELAS

Tabela 1 - Equações dos modelos autologísticos espaço-temporais propostos para o ajuste dos dados de leprose	40
Tabela 2 - Estimativas dos parâmetros, erros-padrão e significância para o modelo $M1$ ajustado aos dados de MSC	42
Tabela 3 - Estimativas dos parâmetros, erros-padrão e significância para o modelo $M2$ ajustado aos dados de MSC	42
Tabela 4 - Estimativas dos parâmetros, erros-padrão e significância para o modelo $M3$ ajustado aos dados de MSC	43
Tabela 5 - Estimativas dos parâmetros, erros-padrão e significância para o modelo $M4$ ajustado aos dados de MSC	43
Tabela 6 - Estimativas dos parâmetros, erros-padrão e significância para o modelos $M1$, $M2$, $M3$ e $M4$ ajustados aos dados de MSC, considerando todas as avaliações conjuntamente.	44
Tabela 7 - Datas das avaliações, incidências de leprose, valores do teste de índice binomial de dispersão e teste da distância mínima média.	46
Tabela 8 - Estimativas e erros padrões dos parâmetros para os modelos ajustados aos dados de leprose dos citros	48
Tabela 9 - Valores de AIC para os modelos ajustados aos dados de leprose dos citros	49
Tabela 10 - Probabilidades da planta apresentar leprose dado que estava sadia no tempo anterior	49

1 INTRODUÇÃO

A produção citrícola brasileira tem notável importância na economia nacional, sendo a maior parte da produção destinada para a indústria de suco, a qual responde por cerca de 28% do suco de laranja produzido no mundo e por 80% do suco concentrado que transita pelo mercado internacional. Citricultores, indústrias e cientistas brasileiros criaram um setor de ponta na agroindústria nacional e trabalham em busca do aumento de produtividade e também controle e manutenção da capacidade produtiva agrícola. Contudo, a produção é limitada devido à presença de doenças nos pomares produtores (MARQUES et al., 2007). Patógenos transmitidos por vetores (insetos e ácaros) são especialmente difíceis de serem controlados e, quando o são, a custos financeiros e ambientais extremamente elevados (GUIRADO, 2000). Algumas doenças, como a leprose dos citros e a Morte Súbita dos Citros (MSC), podem causar a erradicação de talhões inteiros de plantas, na fazenda produtora, comprometendo a quantidade e a qualidade das frutas cítricas.

A MSC representa uma ameaça potencial para a citricultura paulista e nacional, uma vez que afeta todas as variedades comerciais enxertadas sobre limoeiro *cravo*, essa é a enxertia mais usada na produção. A doença provoca diminuição de tamanho, peso e quantidade de frutos, além de rápido definhamento e morte das plantas (BASSANEZI et al., 2003). A leprose dos citros foi considerada nos últimos tempos, uma das mais importantes viroses na citricultura nacional, pois reduz a produção e o período de vida das plantas de citros (RODRIGUES, 2000).

Os dados de plantas cítricas são, em geral, coletados ao longo do tempo nas inspeções periódicas, as plantas estão dispostas dentro de um talhão, em linhas e colunas com certa estrutura de vizinhança. A disposição das plantas, referenciadas por coordenadas locais, permite estudar o padrão espacial de doenças cítricas. Conhecer o movimento e a distribuição espacial de vetores que transmitem doenças e que afetam a disseminação nas plantas e o progresso da doença auxiliam na determinação de práticas de controle. Existem relativamente poucos estudos dos processos temporais e espaciais, relacionando a dinâmica de doenças cítricas na população de plantas hospedeiras e na dinâmica populacional do vetor (CZERMAINSKI, 2006; MAANEN; XU, 2003).

Em estudos para detectar o padrão espacial de plantas doentes são usadas técnicas predominantemente descritivas que visam diagnosticar se há agregação nos padrões da doença. Exemplos de tais técnicas são a análise por *quadrats*, variogramas, e o uso de diversos índices de associação (LIMA et al., 2006). Entretanto, a necessidade de quantificar padrões e estabelecer relações entre fatores que afetam a incidência, levou à proposição de modelos que permitem ir além da simples detecção da presença de padrões espaciais nos dados. São exemplos os modelos estatísticos com componentes espaciais tais

como o modelo autologístico, o modelo CAR (*Conditional Auto Regressive*) e modelos geoestatísticos espaciais (RIBEIRO JR. et al., 2008). Enquanto os dois últimos tratam o efeito espacial como aleatório e tipicamente requerem técnicas de Monte Carlo com Cadeias de Markov (MCMC) para ajuste e inferência, o modelo autologístico trata o componente espacial como efeito fixo. Nesse caso, o modelo é mais facilmente ajustado e de simples interpretabilidade, pois há coeficientes associados a cada termo que quantificam a relação com a incidência da doença, preservando ainda a incorporação explícita da estrutura de dependência espacial (VIOLA et al., 2008).

O modelo autologístico é flexível para descrever a incidência de doenças em plantas, considerando o *status* da planta como resposta binária e incorporando a dependência espacial ou autocorrelação espacial através de covariáveis de vizinhança (KRAINSKI et al., 2008). Os coeficientes de regressão estimados quantificam o acréscimo na probabilidade da presença da doença para cada fator do modelo incluído como covariável. Na agricultura, esse modelo foi estudado por Gumpertz, Graham e Ristiano (1997) para a análise da incidência de *Phytophthora* em pimentões, por Czermainski (2006) para análise de incidência de leprose dos citros, por Viola (2007) na análise de dados de um experimento com cebola e ainda por Krainski et al. (2008), na análise de dados de incidência de morte súbita em citros.

O interesse aqui é propor estratégias de modelagem contemplando questões que podem ser incorporadas no estudo quando ajustado o modelo autologístico: a estrutura de covariáveis de vizinhança, a dimensão temporal dos dados e o efeito defasado das covariáveis de vizinhança. São considerados dados de incidência de leprose dos citros de modo a selecionar um modelo adequado com fatores que descrevam a dinâmica da doença e que, de alguma forma, venham a auxiliar nas práticas de controle. Paralelamente são apresentadas as funcionalidades de dois pacotes, para o ajuste do modelo autologístico, *Rictrus* (KRAINSKI; RIBEIRO JR., 2007) e *stLattice* (KRAINSKI; RIBEIRO JR., 2008), do ambiente estatístico R (R DEVELOPMENT CORE TEAM, 2008). Para exemplificar as funções dos pacotes usa-se um conjunto de dados com incidência de morte súbita dos citros.

2 DESENVOLVIMENTO

2.1 Revisão bibliográfica

A revisão aborda, inicialmente, o contexto dos dados que serão analisados neste estudo. A descrição das duas doenças cítricas: a morte súbita dos citros e a leprose dos citros que causam grandes danos na produção de citros. É apresentada uma breve revisão dos modelos lineares generalizados e dos modelos logísticos. Por fim, descreve-se o modelo autológico e os métodos de inferência baseados em procedimentos computacionalmente intensivos.

2.1.1 Morte súbita dos citros

A MSC representa uma ameaça potencial para a citricultura nacional, uma vez que afeta diversas variedades comerciais de citros, causando o definhamento generalizado das plantas dentro da área produtora. Ela foi identificada no final da década de 90 em uma das principais áreas de produção no Brasil, no norte de São Paulo, região de Colômbia, e sul Minas Gerais, região de Comendador Gomes (GIRARDI, 2005). Desde então, esforços são empregados para entender a etiologia e os seus mecanismos de dispersão.

A doença causa a morte súbita de laranjeiras (*Citrus sinensis*), tangerinas (*C. reticulata*) enxertadas em limoeiros ‘Cravo’ (*C. limonia*), (BASSANEZI; FERNANDES; YAMAMOTO, 2003). O fato de ser transmitida por enxertia de tecidos e pela sua disseminação no pomar suspeita-se que a doença seja causada por um vírus, transmitido por vetor aéreo.

O sintoma característico da doença é o aparecimento de coloração amarela, tendendo para o alaranjado na parte interna da casca do porta-enxerto, logo abaixo da zona de enxertia (BASSANEZI et al., 2003). O tempo entre o aparecimento dos primeiros sintomas e a morte da planta é bastante variável em função da idade da planta, da época do ano e da condição de carga dos frutos. Seu diagnóstico é realizado com base nos sintomas apresentados pelas plantas doentes, pois não é possível diagnosticar a MSC antes do aparecimento deles na planta. De início, ocorre perda generalizada do brilho das folhas das plantas doentes, perdendo a coloração verde-escuro, tendendo para o amarelo-esverdeado. Com a evolução da doença, ocorre a desfolha parcial da planta e seu sistema radicular apresenta grande quantidade de raízes podres e mortas. A morte do sistema radicular é consequência do bloqueio dos vasos do floema, que transportam a seiva elaborada das folhas para as raízes. A Figura 1 mostra uma planta com sintomas de MSC.

Em 2004 dizia-se que nenhum agente causal havia sido associado a MSC (JESUS JR.; BASSANEZI, 2004). Em estudos de progresso espacial e temporal da doença suspeitava-se de uma doença de causa biótica associada a um vetor. Houve grande similaridade nos padrões espaciais da

morte súbita com os padrões espaciais da tristeza na presença do seu vetor, *Toxoptera citricida*. Sendo os sintomas similares aqueles do declínio rápido da tristeza, quanto à sintomatologia, velocidade de progresso da doença e distribuição espacial das plantas afetadas nos talhões, e por serem doenças de combinação copa/porta-enxerto. A suspeita era de que talvez houvesse o envolvimento de uma estirpe do vírus da tristeza com essa doença letal.



Figura 1 - Planta com sintomas de MSC

Em 2005, anunciou-se a descoberta de um novo vírus associado com a doença. Possivelmente, um novo tipo de vírus da família *Tymoviridae* pode ser o agente causador da MSC. Há 99,7% de associação entre o agora chamado *Citrus Sudden Death Virus* (CSDV, ou vírus associado à morte súbita dos citros) e o mal capaz de matar uma laranjeira ou uma tangerina em poucos meses. Mesmo assim, não se pode dizer que seja realmente esse o responsável pela morte das plantas. É preciso ainda demonstrar que existe uma relação clara de causa e efeito, o chamado postulado de Koch, que consiste em inocular o suposto agente causador da doença em organismos saudáveis, nesse caso as laranjeiras, e verificar se elas contraem ou não a enfermidade (MACCHERONI et al., 2005).

Suspeitou-se que se tratava de uma mutação do vírus da tristeza dos citros, uma doença que consumiu 90% dos laranjais paulistas entre 1939 e 1949. Depois, as diferenças se impuseram e foi demonstrado que o CSDV é um novo membro do gênero *Marafivirus*, integrante da família *Tymoviridae*, enquanto o vírus da tristeza pertence à família *Closteroviridae*. Mas ainda não se descarta a possibilidade de que os dois possam atuar em conjunto como causadores da morte súbita (MACCHERONI, 2005).

No entanto, o avanço da doença pelas regiões produtoras causa preocupação aos produtores e pesquisadores, que buscam conhecer o seu comportamento no tempo e espaço. O interesse está nas relações espaciais entre plantas com sintomas de MSC. Essas informações servirão de suporte para as inferências sobre o seu período de incubação, a velocidade e os mecanismos de disseminação

(BASSANEZI; FERNANDES; YAMAMOTO, 2003).

2.1.2 Leprose dos citros

A leprose dos citros foi considerada, nas últimas décadas, uma das mais importantes viroses na citricultura nacional, pois reduz a produção e o período de vida das plantas de citros (RODRIGUES, 2000). Causada por *Citrus leprosis virus* (CiLV), é uma doença endêmica nas regiões produtoras do estado de São Paulo. O vírus é transmitido exclusivamente pelo ácaro *Brevipalpus phoenicis*. O vírus da leprose pode ser transmitido de planta a planta somente mediante o ácaro vetor que se alimenta em planta infectada, adquirindo o vírus e transmitindo-o quando se alimenta em planta sadia. A alimentação e circulação do ácaro na planta e entre as plantas afetadas gera o mecanismo de dispersão da doença (BASSANEZI; LARANJEIRA, 2007).

Endêmica nas regiões produtoras do estado de São Paulo, a doença causa sérios prejuízos à produção. Os principais sintomas da doença na planta são manchas em frutos, folhas e ramos, provocando queda prematura, seca de ramos, e levando a planta ao definhamento (BASSANEZI; LARANJEIRA, 2007). Em geral, os sintomas são visíveis a partir de 17 a 60 dias após a infecção do tecido vegetal, sempre nos locais onde o ácaro se alimenta (BASSANEZI; SPÓSITO; YAMAMOTO, 2001).

Nas folhas, as lesões são inicialmente pouco salientes na face superior, apresentando coloração verde-pálida no centro e amarelada nas bordas. Nos frutos maduros, a lesão mostra-se na forma de uma depressão de coloração mais escura na casca, indo de verde a marrom-escura ou preta. Quando as lesões são abundantes, há queda intensa de folhas e frutos, reduzindo a capacidade fotossintética da planta. Nos ramos, a ocorrência de muitas lesões pode ocasionar a morte dos ponteiros. Os ramos mais grossos descascam transformando-se em porta de entrada para uma série de patógenos oportunistas. A Figura 2 mostra órgãos da planta com sintomas da doença.



Figura 2 - Órgãos da planta com sintomas de leprose dos citros

A principal prática de controle são os acaricidas que geram de 10 a 15% dos custos da produção (FUNDECITROS, 2008). Porém a utilização em grande escala de produtos químicos, além de causar problemas induzindo o aparecimento de populações de ácaros resistentes, gera preocupações

ecológicas e econômicas. Diante disso, há um aumento na busca de alternativas para controlar a leprose, entre eles, o uso comercial de diferentes tipos de agentes biológicos de controle e variedades resistentes ao vírus (BASTIANEL et al., 2006), visando um controle integrado, sobressaem-se o uso de cercas-vivas e/ou quebra-ventos adequados e o manejo correto das plantas daninhas hospedeiras do ácaro (MAIA; OLIVEIRA, 2005). O plantio de mudas isentas de ácaro e podas para eliminação dos ramos secos ou muito lesionados pela leprose. Além disso, conhecer a incidência da doença e do ácaro transmissor no tempo e espaço é de grande importância para descrever a dinâmica da doença e possíveis práticas de controle.

O movimento e a distribuição espacial de vetores que transmitem a virose afetam a disseminação nas plantas e o progresso da epidemia. Existem relativamente poucos estudos dos processos temporais e espaciais da dinâmica da doença nas plantas (CZERMAISNKI, 2006).

2.1.3 Modelos lineares generalizados

No contexto de doenças em plantas, tais como mencionadas anteriormente, os dados em geral, referem-se a indicação da presença da doença, contagem de indivíduos doentes, ou medidas de incidência. Para tais tipos de variáveis respostas não é adequado o ajuste de modelos lineares usuais.

Os Modelos Lineares Generalizados (MLG's), também denominados modelos exponenciais lineares, foram introduzidos por Nelder e Wedderburn (1972). A idéia básica consiste em estender os modelos lineares usuais para respostas gaussianas. Permite-se, então, diferentes opções para a distribuição da variável resposta, assumindo que a mesma pertença à família exponencial de distribuições e tenha maior flexibilidade para a relação funcional entre a média da variável resposta e o preditor linear.

A estrutura de um MLG é dada por três componentes:

- i) um componente aleatório composto de uma variável aleatória Y com n observações independentes de uma distribuição pertencente à família exponencial e um vetor de médias μ ;
- ii) um componente sistemático composto por variáveis explicativas x_1, \dots, x_p tais que, produzem um preditor linear;
- iii) e uma função $g(\cdot)$ monótona diferenciável, conhecida como função de ligação, que relaciona os dois componentes.

Sejam Y_1, \dots, Y_n variáveis aleatórias independentes, cada uma com função de densidade de probabilidade na família exponencial escrita como:

$$f(y_i; \theta_i, \phi) = \exp\{\phi^{-1}[y_i\theta_i - b(\theta_i) + c(y_i)] + a(y_i, \phi)\} , \quad (1)$$

em que $E[Y_i] = \mu_i$ e $\text{Var}[Y_i] = \phi V_i$, $V_i = d\mu_i/d\theta_i$ é a função de variância, sendo $i = 1, 2, \dots, n$, $\theta = \theta(\beta)$ é o parâmetro canônico e ϕ é o parâmetro de dispersão.

Os modelos lineares generalizados são definidos por (1) e pelo componente sistemático,

$$g(\mu_i) = \eta_i, \quad (2)$$

em que $\eta_i = x_i\beta$ é o preditor linear, $\beta = (\beta_1, \dots, \beta_p)^\top$, $p < n$, é o vetor dos parâmetros da regressão a serem estimados, $x_i = (x_{i1}, \dots, x_{ip})$ representa os valores de p variáveis explicativas e $g(\cdot)$ uma função monótona e diferenciável, denominada função de ligação.

As funções de ligação mais utilizadas são obtidas quando o parâmetro canônico coincide com o preditor linear, isto é, quando $\theta = \eta$ e a função de ligação, nessas situações, é chamada de ligação canônica. Nas ligações canônicas o preditor linear modela diretamente o parâmetro canônico, o que, geralmente, resulta em uma escala adequada para a modelagem com interpretação prática para os parâmetros de regressão.

Como exemplos ou casos particulares de distribuições que pertencem à família exponencial podem-se mencionar as distribuições normal, normal inversa, Poisson e binomial, dentre outras. A partir dessas distribuições podem-se obter os modelos de regressão normal inversa, regressão Poisson, regressão logística, entre outros.

A estimação dos parâmetros pode ser feita pelo método da máxima verossimilhança. Se considerar a variável aleatória Y_i com observações $\mathbf{y}^t = (y_1, y_2, \dots, y_n)$ de uma distribuição pertencente a família exponencial (1), a função de verossimilhança é dada por,

$$\begin{aligned} L(\boldsymbol{\theta}, \phi; \mathbf{y}) &= \prod_{i=1}^n f(y_i; \theta_i, \phi) \\ &= \exp \left[\sum_{i=1}^n \phi \{ y_i \theta_i - b(\theta_i) + c(y_i) \} + a(y_i, \phi) \right], \end{aligned}$$

cujo o logaritmo é dado por:

$$l(\boldsymbol{\theta}, \phi; \mathbf{y}) = \sum_{i=1}^n [\phi \{ y_i \theta_i - b(\theta_i) + c(y_i) \} + a(y_i, \phi)].$$

O ajuste adequado de um modelo aos dados, pode ser verificado através da estatística *scaled deviance* (NELDER; WEDDERBURN, 1972), dada por:

$$S_q = (\hat{\boldsymbol{\mu}}, \phi; \mathbf{y}) = -2\{l(\hat{\boldsymbol{\mu}}, \phi; \mathbf{y}) - l(\tilde{\mathbf{y}}, \phi; \mathbf{y})\},$$

em que, q é o número de parâmetros do modelo, $l(\hat{\boldsymbol{\mu}}, \phi; \mathbf{y})$ é o logaritmo da função de verossimilhança para o modelo sob ajuste e $l(\tilde{\mathbf{y}}, \phi; \mathbf{y})$ é o logaritmo da função de verossimilhança para o modelo saturado. No caso em que $\phi = 1$, esta estatística reduz-se à *deviance*,

$$D(\hat{\boldsymbol{\mu}}; \mathbf{y}) = -2\{l(\hat{\boldsymbol{\mu}}; \mathbf{y}) - l(\tilde{\mathbf{y}}; \mathbf{y})\} .$$

Portanto, para um modelo bem ajustado, espera-se que a *deviance* residual esteja próxima do número de graus de liberdade do resíduo do modelo. Uma maneira de obter uma *deviance* próxima do número de graus de liberdade é aumentar o número de parâmetros no modelo, mas com isso, aumenta-se também a complexidade na interpretação. O ideal é encontrar modelos simples com *deviance* moderada, procurar pela parcimônia (BORGATTO, 2004).

Dois critérios comumente usados para a seleção de modelos são o critério de informação de Akaike (AIC), proposto por Akaike (1973) e o critério de informação de Bayes (BIC), proposto por Schwarz (1978) expressos, respectivamente, por

$$AIC = -2l(\hat{\boldsymbol{\mu}}, \phi; \mathbf{y}) + 2q \quad (3)$$

e,

$$BIC = -2l(\hat{\boldsymbol{\mu}}, \phi; \mathbf{y}) + q \log(n) , \quad (4)$$

em que, q é o número de parâmetros do modelo e n é o número de observações. Enquanto que o AIC considera apenas q , o BIC considera ambos, q e n . Os dois critérios são usados tanto para modelos aninhados como não aninhados. Para ambos, seleciona-se como o modelo mais adequado aos dados aquele que apresentar menor valor.

Desde sua proposição os MLG's tem sido largamente usados em diversos contextos de modelagem estatística. Maiores detalhes sobre o assunto podem ser encontrados em Nelder e Wedderburn (1972), McCullagh e Nelder (1989) e Cordeiro e Demétrio (2007).

2.1.4 Modelo de regressão logística

Dentro da família dos modelos lineares generalizados está o modelo de regressão logística, adequado para a modelagem de variáveis com respostas binárias. Os dados assumem distribuição de probabilidade binomial, sob a forma de Y_i sucessos em m_i ensaios de Bernoulli, ($i = 1, \dots, n$). O modelo logístico é atrativo e comumente usado devido à facilidade de interpretação dos seus parâmetros. A partir dos parâmetros pode-se quantificar o aumento na chance de sucesso com o acréscimo de uma unidade da covariável associada enquanto as demais permanecem constantes.

Seja Y uma variável aleatória com distribuição binomial, escrita na forma da família exponencial por,

$$f(y; \theta) = \exp \left\{ \binom{m}{y} + y \ln \theta + (m - y) \ln(1 - \theta) \right\} I_A(y) .$$

O modelo de regressão logística pode ser expresso por:

$$\theta(x) = P(Y = 1|x) = \frac{\exp\{\beta_0 + \sum_{k=1}^p \beta_k x_k\}}{1 + \exp\{\beta_0 + \sum_{k=1}^p \beta_k x_k\}}, \quad (5)$$

em que x representa as covariáveis, $x = (x_1, \dots, x_p)$, o parâmetro β_0 é o intercepto, e $(\beta_1, \dots, \beta_p)$ são os p parâmetros de regressão associados às covariáveis. O modelo fornece uma estimativa da probabilidade do indivíduo ter a resposta, dado que o mesmo possui, ou não, determinados fatores de interesse. Observe, ainda, que

$$\ln \left(\frac{\theta(x)}{1 - \theta(x)} \right) = \beta_0 + \sum_{k=1}^p \beta_k x_k,$$

tem-se um modelo linear para o logito, isto é, o logaritmo neperiano da razão entre $\theta(x)$ e $1 - \theta(x)$. No contexto de MLG's, o logito é a função de ligação canônica para a regressão logística.

A estimação dos parâmetros do modelo é, usualmente, feita pelo método de máxima verossimilhança. Assumindo que as observações sejam independentes tem-se a seguinte expressão para a função de verossimilhança

$$L(\beta) = \prod_{i=1}^n \theta(x_i)^{y_i} (1 - \theta(x_i))^{1-y_i},$$

em que $\theta(x_i)$ é a contribuição para a função de verossimilhança dos pares (y_i, x_i) , $y_i = 1$ e $1 - \theta(x_i)$ a contribuição dos pares em que $y_i = 0$.

O princípio da máxima verossimilhança é estimar o valor de β que maximiza $L(\beta)$. Algebricamente, é mais fácil trabalhar com o logaritmo dessa função,

$$l(\beta) = \ln[L(\beta)] = \sum_{i=1}^n \left[y_i \ln \left(\frac{\theta(x_i)}{1 - \theta(x_i)} \right) + \ln(1 - \theta(x_i)) \right].$$

Para encontrar o valor que maximiza $l(\beta)$, basta diferenciar a função com respeito a cada parâmetro β_j ($j = 0, 1, \dots, p$), obtendo-se o sistema de $p + 1$ equações, que, igualadas a zero, produzem como solução as estimativas de máxima verossimilhança de β .

As $p + 1$ equações são chamadas de equações de verossimilhança e, por não serem lineares nos parâmetros β_j , requerem métodos numéricos para obtenção da solução. Os métodos iterativos de Newton-Raphson e o score de Fisher são algoritmos numéricos comumente utilizados para essa finalidade. Os valores ajustados para o modelo de regressão logística são, portanto, obtidos substituindo-se as estimativas de β em (5).

Após obter as estimativas dos coeficientes do modelo é preciso avaliar a adequação do modelo ajustado aos dados. Na regressão logística as estatísticas de teste para essa finalidade são denominadas estatísticas de qualidade do ajuste. As duas mais conhecidas são: a qui-quadrado de Pearson e a qui-quadrado da razão de verossimilhanças.

2.1.5 Modelo autológico

A chance de ocorrência da doença em determinada planta pode estar associada com a presença da doença em plantas vizinhas. Desta forma, espera-se que, em geral, plantas próximas tenham características similares fazendo com que a localização das plantas doentes apresente algum padrão espacial. A regressão logística é amplamente usada para análise de dados binários, como por exemplo, a presença de doenças em plantas. No contexto de plantas em um talhão, é razoável assumir que plantas vizinhas tendem a ter *status* similar, refletindo uma eventual agregação no padrão espacial da doença. O modelo autológico estende o modelo logístico usual, construindo covariáveis a partir da resposta observada na vizinhança de cada planta, uma vez que a configuração espacial dos dados está disponível.

O modelo autológico é uma das formas apresentadas na literatura para que a dependência espacial dos dados seja explicitamente considerada na modelagem. É um modelo auto-regressivo espacial para dados binários, considera a observação em determinado local dependente das observações em locais vizinhos na forma de covariáveis. Grande parte da teoria estatística do modelo autológico foi introduzida por Besag (1972), com alguns exemplos envolvendo doenças de plantas. Desde então, diversos autores como Gumpertz, Graham e Ristiano (1997), Huffer e Wu (1998), Gu e Zhu (2001), Zhu, Huang e Wu (2005) e Sherman, Apanasovich e Carrol (2006) desenvolveram suas pesquisas usando e fazendo generalizações do modelo.

O modelo é aplicável em diferentes contextos. Na modelagem de distribuições de espécies de plantas considerando covariáveis climáticas (WU; HUFFER, 1997), no estudo do efeito das informações de covariáveis de vizinhança na sobrevivência de árvores tropicais (HUBBEL; CONDIT; FOSTER, 2001) e em um conjunto de dados epidemiológicos espaciais envolvendo câncer de fígado (SHERMAN; APANASOVICH; CARROL, 2006).

Gumpertz, Graham e Ristiano (1997) usam o modelo para a análise da incidência de *Phytophthora* em pimentões, dispostos em um *lattice*, onde foi anotada a presença da doença em grupos de 2 a 3 plantas, denominados *quadrats*. O interesse em usar o modelo autológico, consistiu no estudo da probabilidade de um particular *quadrat* ter planta com a doença, a partir das informações de vizinhança e de covariáveis de solo. Em seus estudos sobre a distribuição de vegetação em duas diferentes florestas no Canadá He, Zhou e Zhu (2003) observam que o modelo autológico consegue explicar a dependência espacial dos dados, levando os autores a obter resultados satisfatórios. Os autores afirmam que a metodologia proposta no artigo pode ser usada nas áreas de geologia, ecologia, agricultura, epidemiologia, entre outras. Viola (2007) aplica o modelo a dados simulados e a dados provenientes de um experimento com pimentão.

Zhu, Huang e Wu (2005), em seu artigo, adicionam um componente temporal ao modelo autológico para analisar dados binários do tipo espaço-temporal. O modelo, então, captura a relação entre a variável binária e possíveis covariáveis, além de simultaneamente ajustar a dependência espacial e temporal para um campo aleatório Markoviano espaço-temporal. Para ilustrar o método, o modelo foi ajustado a dados de rupturas causadas por besouros em pinus na Carolina do Norte/USA. As observações foram feitas em diferentes períodos de tempo. Krainski et al. (2008) ajustam modelos autológicos para incidência de morte súbita dos citros de forma a analisar a presença da doença em plantas vizinhas no tempo contemporâneo à avaliação, no tempo anterior e um modelo com ambos os tempos. Czermainski (2006) ajusta modelos para dados de leprose dos citros observando a influência das covariáveis de vizinhança defasadas no tempo.

A essência do modelo autológico é utilizar covariáveis obtidas a partir dos valores da variável resposta em posições vizinhas da localização considerada. Portanto, o elemento chave do modelo é a definição e a estrutura de vizinhança que será discutida na próxima Seção.

2.1.5.1 Estrutura de vizinhança

Alguns tipos de dados estão dispostos regularmente em linhas e colunas em uma determinada região, ou seja, os dados formam uma malha regular de pontos, com as coordenadas de cada posição. A estrutura descrita permite construir covariáveis de vizinhança, que representam o *status* das observações vizinhas no modelo autológico.

A estrutura de vizinhança pode ser construída de diferentes formas, como ilustrado na Figura 3, considerando simplesmente a ordem de vizinhança, primeira ordem, segunda ou ordens superiores, ou ainda, considerando o efeito da planta nas linhas, colunas e diagonais separadamente (GUMPERTZ; GRAHAM; RISTIANO, 1997).

Considerando as ordens de vizinhança, uma possibilidade é construir uma covariável dentro de cada ordem de vizinhança. A covariável de primeira ordem é a soma das observações de quatro vizinhos, dois vizinhos na mesma linha em colunas adjacentes e dois vizinhos na mesma coluna em linhas adjacentes, e a de segunda ordem é a soma das observações da covariável de primeira ordem, com as informações dos quatro vizinhos nas duas diagonais. A construção das covariáveis é feita de forma análoga para ordens superiores.

Outra maneira de construir covariáveis de vizinhança é considerar linhas, colunas e diagonais separadamente, dentro de determinada ordem. Na primeira ordem, são construídas duas covariáveis. A primeira covariável inclui observações dos dois vizinhos na mesma linha em colunas adjacentes, posições $(k, l + 1)$ e $(k, l - 1)$, enquanto que a segunda covariável inclui informações dos dois vizinhos na mesma coluna em linhas adjacentes, posições $(k - 1, l)$ e $(k + 1, l)$. Na estrutura de segunda ordem

de vizinhança, são somadas as covariáveis de primeira ordem a uma terceira covariável, a qual inclui os dois vizinhos da diagonal $(-1, 1)$, posições $(k - 1, l + 1)$ e $(k + 1, l - 1)$ e a uma quarta covariável que inclui os dois vizinhos da diagonal $(1, -1)$, posições $(k - 1, l - 1)$ e $(k + 1, l + 1)$ (HE; ZHOU; ZHU, 2003). De forma análoga pode-se considerar vizinhança de maior ordem.

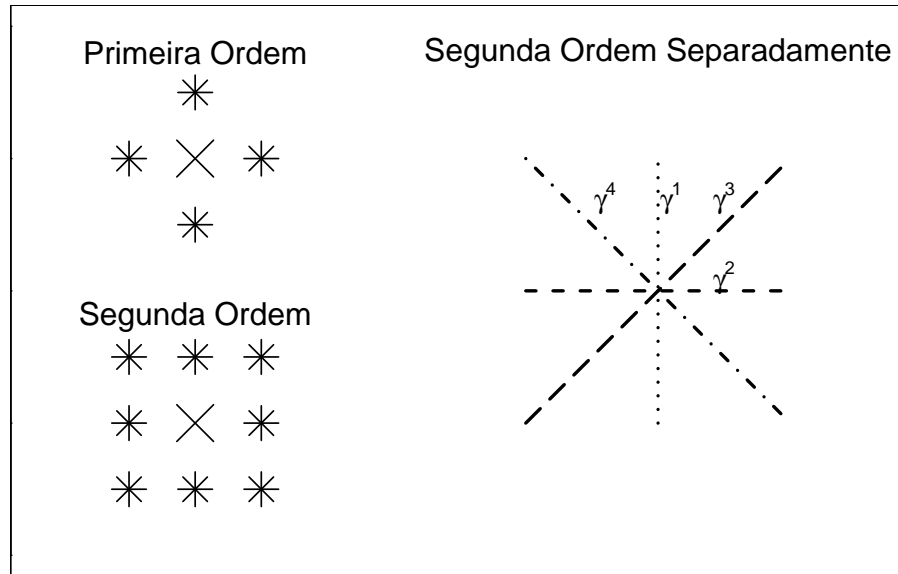


Figura 3 - Estruturas de vizinhança: primeira ordem, segunda ordem e segunda ordem separadamente

A estrutura que considera linhas, colunas e diagonais separadamente é mais flexível e particularmente útil no caso de doenças em plantas que possuem espaçamento diferente entre linhas e colunas. Permite medir o efeito das vizinhas dentro da linha, entre as linhas e efeitos direcionais das diagonais.

2.1.5.2 Especificação do modelo

No modelo autolístico com função de ligação logito, a probabilidade de sucesso é modelada como combinação linear de presença ou ausência de sucesso em locais vizinhos e de covariáveis que captam informações adicionais nesses locais, dado pela expressão,

$$P(Y = 1|x_{ij}, y_i) = \text{logit}(p_{ij}) = \sum_{j=1}^p \beta_j x_{ij} + \sum_{d=1}^q \gamma_d y_{id}, \quad (6)$$

em que β_j mede a influência da covariável x_{ij} , γ_d mede a influência de vizinhança de ordem k e y_{id} representam as covariáveis de vizinhança.

Quando $\gamma_d = 0$, o modelo (6) reduz-se ao modelo de regressão logística que é apropriado para dados binários independentes. Quando $\beta_j = 0$, o modelo torna-se o modelo autolístico proposto

Besag (1972, 1974).

Considere um *lattice* D uma região com n posições, cada uma descrita pelas coordenadas (k, l) especificando a linha e a coluna em que a observação é alocada. Em cada posição é observada uma resposta binária $y_{(k,l)}$, que tem valor 1 na presença da variável de interesse e zero, caso contrário, e um vetor $p \times 1$ de covariáveis $x_{(k,l)}$. As respostas binárias de n posições correspondem a $Y = (y_{(k,l)}, (k, l) \in D)$ que constituem um mapa da ocupação da variável de interesse (HE; ZHOU; ZHU, 2003). A probabilidade de sucesso do modelo autologístico passa a ser,

$$Pr(Y_{k,l} = 1 | x_{k,l}, y_{k,l}, (k, l) \in D) = \frac{\exp \left\{ \sum_{j=0}^p \beta_j x_{k,l,j} + \sum_{d=1}^q \gamma_d y_{k,l,d} \right\}}{1 + \exp \left\{ \sum_{j=0}^p \beta_j x_{k,l,j} + \sum_{d=1}^q \gamma_d y_{k,l,d} \right\}}, \quad (7)$$

em que, os β 's são parâmetros de regressão e os γ 's são os parâmetros de autocorrelação espacial, $x_{k,l,j}$ representam as covariáveis usuais e $y_{k,l,d}$ representam as covariáveis de vizinhança, com preditor linear $\eta = g(\mu) = \ln \left(\frac{\theta(x)}{1-\theta(x)} \right)$.

2.1.5.3 Inferência

No modelo autologístico não é possível obter a expressão da função de verossimilhança para a estimação dos parâmetros do modelo. A probabilidade de sucesso de uma observação $s_{k,l}$ é condicional ao *status* das outras observações $v_{k,l}$. Isso faz com que as observações sejam dependentes entre si, e ao ajustar um MLG usual, este não levará em conta a autocorrelação dos dados. Dessa forma, não é possível escrever a função de verossimilhança de forma fechada, sendo, em geral, desconhecida uma expressão analítica para a constante de normalização. Diversos métodos de estimação foram propostos na literatura.

O método de pseudo-verossimilhança proposto por Besag (1975) e utilizado por Gumpertz, Graham e Ristiano (1997), maximiza uma função de pseudo-verossimilhança para o modelo ajustado. A reestimação dos erros-padrão das estimativas de vizinhança é feita usando-se o método de *Bootstrap* com o amostrador de Gibbs utilizado na geração das amostras, pois, é preciso preservar a estrutura de vizinhança dos dados originais e o padrão de dependência espacial. Aplicações do método de pseudo-verossimilhança estão presentes em Strauss e Ikeda (1990), Arnold e Straus (1991) e Preisler (1991). Em Hubbell, Condit e Foster (2001), o método de estimação usado foi de máxima verossimilhança com simulação Monte Carlo com cadeias de Markov, utilizando o amostrador de Gibbs, semelhante ao procedimento utilizado por Gumpertz, Graham e Ristiano (1997).

Para a estimação do modelo autologístico Wu e Huffer (1997) usam três métodos de esti-

mação: *coding* (COD), máxima pseudo-verossimilhança (MPL) e Monte Carlo com cadeias de Markov (MCMC). A comparação entre os métodos mostrou que MCMC apresentou as melhores estimativas, porém requer muito mais tempo computacional que os outros métodos. A vantagem da precisão do método MCMC é substancial quando a correlação espacial é forte. Porém, quando a correlação é pequena, as estimativas de MPL são adequadas. Os autores recomendam usar MPL nos estágios preliminares da construção do modelo, quando ainda diferentes modelos e covariáveis estão sendo considerados e analisados. O MCMC pode então ser usado para um “ajuste fino” dos resultados preliminares.

He, Zhou e Zhu (2003) ajustam modelos autologísticos considerando a estrutura de vizinhança de segunda-ordem e comparam três métodos de estimação dos parâmetros: o método de máxima pseudo-verossimilhança, proposto por Besag (1975), o de máxima verossimilhança com simulação Monte Carlo, introduzido por Geyer e Thompson (1992), e a aproximação estocástica Monte Carlo com cadeias de Markov proposto por Gu e Zhu (2001). Estudos de simulação foram conduzidos para comparar a performance dos três métodos. O artigo relata que os dois últimos métodos dão resultados idênticos para as estimativas, enquanto que pseudo-verossimilhança apresenta valores um pouco diferentes. No entanto, afirmam que o método de pseudo-verossimilhança pode ser usado na maioria dos casos. Zhu, Huang e Wu (2005) estimam os parâmetros do modelo por máxima pseudo-verossimilhança e obtêm a predição de respostas futuras no *lattice* por meio do amostrador de Gibbs.

Sherman, Apanasovich e Carrol (2006) demonstram que o método de máxima verossimilhança não é adequado para estimação dos parâmetros no ajuste de dados binários espaciais usando o modelo autologístico. Discutem métodos alternativos de estimação: máxima pseudo-verossimilhança, máxima pseudo-verossimilhança generalizada e máxima verossimilhança com simulação Monte Carlo. Esses três métodos de estimação foram aplicados a um conjunto de dados reais e a dados simulados. O artigo discute os esforços computacionais para a implementação dos métodos. Os resultados indicam para o método de máxima verossimilhança com simulação Monte Carlo como mais eficiente se comparado ao de máxima pseudo-verossimilhança, por outro lado a implementação computacional é muito mais intensiva e complexa para o primeiro método. No caso específico os autores comentam que não existe um grande aumento da eficiência por usar o método de máxima verossimilhança com simulação Monte Carlo e devido ao esforço computacional desse método sugerem o uso do método de máxima pseudo-verossimilhança. Porém, eles afirmam que ao usar o método de máxima pseudo-verossimilhança é necessário que seja usado *Bootstrap* com amostrador de Gibbs para o cálculo dos erros-padrão.

No artigo de Viola et al. (2008) foi feito um estudo de simulação para verificar o comportamento dos estimadores de pseudo-verossimilhança dos parâmetros do modelo autologístico, considerando diferentes estruturas de covariáveis de vizinhança, três intensidades de infestação de uma praga e cinco valores para o parâmetro de correlação entre os vizinhos. Mostra-se que o método de estimação por

pseudo-verossimilhança pode ser usado, com certa cautela, quando o interesse está na contribuição das covariáveis, mas não deve ser usado quando o interesse está na estimação da correlação espacial.

2.1.5.3.1 Método de máxima pseudo-verossimilhança

O método de máxima pseudo-verossimilhança foi proposto por Besag (1975). Desde então, muitas pesquisas mostraram que sob condições adequadas o método é consistente e assintoticamente normal ao redor do verdadeiro valor do parâmetro. Isso para quando for usado em grandes amostras de processos com dependência espacial. Mais informações e detalhamentos em Jensen e Møller (1991), Guyon (1995), Huang e Ogata (2002), entre outros autores. O método de máxima pseudo-verossimilhança tem sua implementação baseada em algoritmos para o ajuste de MLG combinado com os procedimentos de reamostragem e, no ambiente R (R DEVELOPMENT CORE TEAM, 2008) está implementado nos pacotes Rcitrus (KRAINSKI; RIBEIRO JR., 2007) e stLattice (KRAINSKI; RIBEIRO JR., 2008).

Besag sugeriu um método de estimação para um vetor de parâmetros desconhecidos $\boldsymbol{\theta} = (\beta_0, \beta_1, \dots, \gamma)^T$, definido como um vetor $\hat{\boldsymbol{\theta}}$ que maximize a função de pseudo-verossimilhança,

$$\begin{aligned} l_{ps}(\boldsymbol{\theta}) &= \sum_{(k,l) \in D} \ln[P(Y_{k,l} = y_{k,l} | \text{todos os outros valores})] , \\ l_{ps}(\boldsymbol{\theta}) &= \sum_{(k,l) \in D} \{y_{k,l} f_{k,l}(\boldsymbol{\theta}) - \ln[1 + \exp(f_{k,l}(\boldsymbol{\theta}))]\} , \end{aligned} \quad (8)$$

em que $f_{k,l}(\boldsymbol{\theta}) = \beta_0 + \beta_1 x_{k,l}^T + \gamma_1 y_{(k,l)} + \gamma_2 y_{(k,l)} + \gamma_3 y_{(k,l)} + \gamma_4 y_{(k,l)}$, e $\boldsymbol{\theta} = (\beta_0, \beta_1^T, \gamma_1, \gamma_2, \gamma_3, \gamma_4)^T$

O estimador $\hat{\boldsymbol{\theta}}$, em geral, é consistente e assintoticamente normal. Entretanto, os erros-padrão do estimador obtido pela regressão logística usual são inválidos e precisam ser recalculados de outra maneira (ZHU; HUANG; WU, 2005).

Uma possível solução é usar métodos de reamostragem. No entanto, no contexto de padrões espaciais isso não é simples dada a necessidade de preservar a estrutura espacial. Isso pode ser alcançado através da reamostragem por bloco (CRESSIE, 1993). Uma forma de recalcular proposta por Gumpertz, Graham e Ristiano (1997) é usar o método de reamostragem *Bootstrap*.

2.1.5.3.2 Método de *Bootstrap*

A metodologia *Bootstrap* consiste em gerar amostras a partir dos dados originais para estimar a quantidade de interesse com essas amostras. O *Bootstrap* é utilizado quando se tem informação suficiente sobre a forma da distribuição dos dados. A amostra *Bootstrap* é formada realizando-se amostragem dessa distribuição com os parâmetros desconhecidos substituídos por estimativas paramétricas

obtidas dos dados. A distribuição da estatística de interesse aplicada aos valores da amostra *Bootstrap*, condicional aos dados observados, é definida como distribuição *Bootstrap* dessa estatística.

Devido à configuração espacial dos dados, o método de *Bootstrap* necessita usar o algoritmo do amostrador de Gibbs (GEMAN; GEMAN, 1984). O amostrador de Gibbs permite que o padrão espacial de vizinhança original seja preservado na reamostragem *Bootstrap*. O amostrador de Gibbs é particularmente útil para o modelo autológico, por ser um procedimento de amostragem baseado em distribuições condicionais, que permite retirar amostras sucessivas da distribuição de cada y_i , condicionado aos seus vizinhos.

A idéia básica é amostrar da distribuição de cada observação y_{ij} condicionando com o *status* atual das vizinhas, cuja as probabilidades são dadas pelo modelo autológico da equação (6). Esse procedimento é um algoritmo que funciona da seguinte forma: começa com valores observados $y^{(0)}$ dos quais obtém-se a estimativa do parâmetro $\gamma^{(0)}$ pela maximização da pseudo-verossimilhança da equação (8). Em seguida gera-se B amostras *Bootstrap* ($y^{(1)}, \dots, y^{(B)}$) obtendo estimativas ($\hat{\gamma}^{(1)}, \dots, \hat{\gamma}^{(B)}$) para cada uma delas. As amostras de *Bootstrap* são obtidas através dos seguintes passos: *i.* começando de uma localização arbitrária (planta), atualiza seu *status* amostrando da distribuição de Bernoulli $f(\hat{\gamma}^{(0)}, y^{(t)})$ com a probabilidade dada pelos parâmetros ajustados e pelo *status* atual das plantas, em uma sequência aleatória até que o ciclo seja completado, isto é, o *status* de todas as plantas sejam atualizadas gerando uma amostra *Bootstrap* com dados artificiais $y^{(t)}$; *ii.* quando o ciclo for completado obtém-se as estimativas dos parâmetros maximizando a função de pseudo-verossimilhança dada pela equação (8); *iii.* repita os passos *i* e *ii* até que o número B de amostras *Bootstrap* seja obtido (KRAINSKI et al., 2008).

Ao final do algoritmo de simulação, certifica-se que a cadeia das estimativas dos parâmetros converge para a distribuição alvo e então, a variância do estimador $\hat{\gamma}$ é dada simplesmente pela variância das estimativas ($\hat{\gamma}^{(1)}, \dots, \hat{\gamma}^{(B)}$). É também recomendável descartar um certo número M de reamostragens iniciais, o chamado período de aquecimento, período até a cadeia convergir para a distribuição estacionária, e utilizar as simulações a cada passo k , para reduzir o número de simulações armazenadas.

2.1.5.3.3 Método de Monte Carlo

Um método alternativo utiliza simulação Monte Carlo para aproximar a função de verossimilhança e gerar estimadores que são consistentes e assintoticamente normais (VIOLA, 2007). Geyer e Thompson (1992); He, Zhou e Zhu (2003) consideram o logaritmo da razão de verossimilhanças em relação a um ponto de referência $\psi \in \Theta$,

$$\ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\psi}) = (\boldsymbol{\theta} - \boldsymbol{\psi})^T T(\mathbf{Y}_0) - \log \frac{C(\boldsymbol{\theta})}{C(\boldsymbol{\psi})},$$

e dado que,

$$P_{\boldsymbol{\theta}}(\mathbf{Y}) = C(\boldsymbol{\theta})^{-1} \exp\{\boldsymbol{\theta}^T T(\mathbf{Y})\} \quad (9)$$

em que, $T(\mathbf{Y}) = \sum_{i \in D} y_i \tilde{x}_i$ e $\tilde{x}_i = (1, x_i^T, z_i^T)$, tem-se que,

$$\frac{C(\boldsymbol{\theta})}{C(\boldsymbol{\psi})} = \int \exp\{(\boldsymbol{\theta} - \boldsymbol{\psi})^T T(\mathbf{Y})\} P_{\boldsymbol{\psi}}(\mathbf{Y}) d\mathbf{Y} \quad (10)$$

Usando Monte Carlo a integral (10) pode ser aproximada com base em uma amostra aleatória Y_1, Y_2, \dots, Y_n de $P_{\boldsymbol{\psi}}(\mathbf{Y})$, isto é,

$$\frac{C(\boldsymbol{\theta})}{C(\boldsymbol{\psi})} \approx \frac{1}{n} \sum_{i=1}^n \exp\{(\boldsymbol{\theta} - \boldsymbol{\psi})^T T(\mathbf{Y}_i)\} \quad (11)$$

Substituindo (11) dentro de (9), obtém-se uma aproximação para o logaritmo da razão de verossimilhança em (9),

$$\ell(\boldsymbol{\theta}; \boldsymbol{\psi}) = (\boldsymbol{\theta} - \boldsymbol{\psi})^T T(\mathbf{Y}_0) - \ln \left[\frac{1}{n} \sum_{i=1}^n \exp\{(\boldsymbol{\theta} - \boldsymbol{\psi})^T T(\mathbf{Y}_i)\} \right] \quad (12)$$

Maximizando $\ell(\boldsymbol{\theta}; \boldsymbol{\psi})$ encontra-se a estimativa de $\hat{\boldsymbol{\theta}}_n$. O algoritmo de Newton-Raphson pode ser usado para calcular $\hat{\boldsymbol{\theta}}_n$, mais detalhes em He, Zhou e Zhu (2003). A eficiência do método depende da escolha de $\boldsymbol{\psi}$. Quando $\boldsymbol{\psi}$ está muito longe da estimativa de verossimilhança, o método pode deixar de ser uma solução. Huffer e Wu (1998) propõem usar a estimativa de pseudo-verossimilhança como valor inicial θ^0 para o valor de $\boldsymbol{\psi}$.

2.1.5.3.4 Método de Monte Carlo com Cadeias de Markov (MCMC)

Os métodos de Monte Carlo via cadeias de Markov (MCMC) são uma alternativa aos métodos não iterativos em problemas complexos. A idéia é obter uma amostra da distribuição a posteriori e calcular estimativas de características desta distribuição. As técnicas usadas são de simulação iterativa, baseadas em cadeias de Markov, e assim os valores gerados não serão mais independentes. Os métodos de MCMC mais usados são amostrador de Gibbs e o algoritmo de Metropolis-Hastings.

Huffer e Wu (1998) e, posteriormente, outros autores estudaram MCMC no contexto de modelos autolagísticos. Zhu, Huang e Wu (2005) usam para inferência estatística um algoritmo MCMC que permitiu obter a distribuição a posteriori dos parâmetros do modelo e as distribuições preditivas a posteriori. Huffer e Wu (1998) concluem que usar máxima verossimilhança com simulação MCMC é superior aos outros métodos usados para estimação dos parâmetros, como por exemplo, o

método de pseudo-verossimilhança. Em estudos de simulação as estimativas encontradas pelo método são aproximadamente normais e usando a informação de Fisher as estimativas de variância obtidas podem ser usadas para construir intervalos de confiança. Maiores detalhes em Geyer e Thompson (1992), Seymour e Ji (1996), Gu e Zhu (2001) e Sherman, Apanasovich e Carrol (2006).

Cadeias de markov: Uma cadeia de Markov é um processo estocástico $\{X_0, X_1, \dots\}$ tal que a distribuição de X_t dado todos os valores anteriores X_0, \dots, X_{t-1} depende apenas de X_{t-1} .

Os métodos MCMC requerem que a cadeia seja:

- i) homogênea, isto é, as probabilidades de transição de um estado para outro são invariantes;
- ii) irredutível, isto é, cada estado pode ser atingido a partir de qualquer outro número finito de interações;
- iii) aperiódica.

Uma questão de ordem prática é como os valores iniciais influenciam o comportamento da cadeia. A idéia é que conforme o número de iterações aumenta, a cadeia gradualmente esqueça os valores iniciais e, eventualmente, converge para uma distribuição de equilíbrio. Assim, em aplicações práticas, é comum que as iterações iniciais sejam descartadas, como se formassem uma amostra de “aquecimento”.

2.1.5.4 Covariáveis de vizinhança

A informação da covariável de vizinhança pode ser considerada de duas formas no modelo autolístico: de forma dicotômica com a informação de existência ou não de vizinhas doentes ou, o número/proporção de vizinhas doentes ou mesmo alguma medida do grau de infestação das vizinhas. No primeiro caso, plantas com uma vizinha doente serão consideradas de forma igual às plantas com mais de uma vizinha doente. Essas duas possibilidades de considerar a informação das vizinhas devem ser observadas na análise, pois, ambas contemplam diferentes formas de contágio. No primeiro caso assume-se que o efeito ou pressão de infecção de uma planta doente é o mesmo de duas plantas doentes.

Quando for considerado a contagem de plantas vizinhas doentes as expressões para o logito no caso de nenhuma vizinha doente, uma vizinha doente e duas vizinhas doentes são, respectivamente:

$$\begin{aligned} \text{logit}(p_{kl}) &= \beta_0 \\ \text{logit}(p_{kl}) &= \beta_0 + \gamma_1 y_{k,l,1} \\ \text{logit}(p_{kl}) &= \beta_0 + 2\gamma_1 y_{k,l,1} \end{aligned}$$

Quando for considerado a covariável de vizinhança na forma binária, presença ou ausência

de vizinhas doentes, as expressões são, respectivamente:

$$\text{logit}(p_{kl}) = \beta_0 + \gamma_1 y_{k,l,1}$$

$$\text{logit}(p_{kl}) = \beta_0$$

Covariáveis de vizinhança no período de tempo anterior: Nas situações em que as observações são medidas em diferentes períodos de tempo pode-se ajustar um modelo autologístico para as observações em cada um dos períodos. Uma proposta é modelar os dados de um período de tempo t , utilizando as covariáveis de vizinhança no período de tempo anterior $t - 1$. Czermainski (2006) modelou os dados com incidência de leprose, através do modelo autologístico com covariáveis de vizinhança defasadas no tempo.

Essa abordagem permite explorar o potencial preditivo do modelo. Por exemplo, ajusta-se o modelo em cada avaliação do tempo t , avaliando as covariáveis do tempo anterior $t - 1$, assim o modelo pode ser utilizado para prever a incidência no tempo futuro, baseado nas covariáveis de vizinhança do período atual. O modelo é dado por,

$$Pr(Y_{k,l} = y_{k,l,t} | x_{k,l}, y_{k,l,t-1}, (k,l) \in D) = \frac{\exp \left\{ \sum_{j=0}^p \beta_j x_{k,l,j} + \sum_{d=1}^q \gamma_d y_{k,l,t-1,d} \right\}}{1 + \exp \left\{ \sum_{j=0}^p \beta_j x_{k,l,j} + \sum_{d=1}^q \gamma_d y_{k,l,t-1,d} \right\}},$$

em que, os β 's são parâmetros de regressão e γ 's são parâmetros de autocorrelação espacial, $x_{k,l,j}$ representam as covariáveis e $y_{k,l,t-1,d}$ representam as covariáveis de vizinhança no tempo anterior, com preditor linear na forma $\eta = g(\mu) = \ln \left(\frac{\theta(x)}{1-\theta(x)} \right)$.

Nessa situação, a estimação dos parâmetros pode ser feito por máxima pseudo-verossimilhança, e as estimativas dos erros-padrão fornecidas pelo MLG usual estão corretas, pois não há autocorrelação das observações.

Algumas vezes o interesse do estudo está em medir a permanência do evento na observação, ou seja, dado que o evento de interesse estava presente no período anterior, qual a probabilidade do evento permanecer no tempo atual. A covariável de vizinhança nesse caso, é a própria unidade com a observação defasada no tempo.

Covariáveis de vizinhança no tempo atual e anterior: A partir do modelo ajustado com covariáveis do tempo anterior, pode-se testar se ainda restou algum efeito possível de ser explicado pelas covariáveis no tempo atual. Para isso, constrói-se o modelo autologístico considerando covariáveis no período de tempo atual e anterior em um mesmo modelo. Krainski et al. (2008) ajustam modelos com covariáveis de vizinhança no tempo anterior e modelos com covariáveis no tempo atual e anterior à avaliação, em um mesmo modelo. A expressão para a probabilidade de sucesso do modelo para esse caso,

$$Pr(Y_{k,l} = 1 | x_{k,l}, y_{k,l,t}, y_{k,l,t-1}, (k, l) \in D) = \frac{\exp\{\eta^*\}}{1 + \exp\{\eta^*\}},$$

sendo que,

$$\eta^* = \sum_{j=0}^p \beta_j x_{k,l,j} + \sum_{d=1}^q \gamma_d y_{k,l,t,d} + \sum_{d=1}^q \gamma_d y_{k,l,t-1,d}.$$

Ao fazer a estimação por pseudo-verossimilhança os erros-padrão das estimativas dos parâmetros no tempo t precisam ser estimados pelo procedimento de reamostragem, pois há a indução de autocorrelação e os erros-padrão usuais não são apropriados.

2.1.5.5 Modelo autologístico espaço-temporal

O modelo autologístico citado anteriormente é adequado para dados binários espaciais em *lattice* em determinado momento no tempo. Porém, muitas vezes as observações são repetidas ao longo do tempo e estão disponíveis dados do mesmo *lattice* em diversos pontos no tempo, como visto na Seção 2.1.5.4. Ou seja, para uma dada localização s_i e um determinado tempo t , a variável resposta é $Y_{i,t} \equiv Y_{(s_i,t)}$, em que $i = (1, \dots, n)$ e $t = 1, 2, \dots$

Fernández-Durán, Poiré e Rojas-Nandayapa (2004) estendem o modelo autologístico para incluir variáveis exploratórias contemporâneas e valores observados da covariável de vizinhança no período anterior. O artigo avalia a importância do efeito temporal de primeira ordem. Os autores referem-se ao modelo como modelo autologístico com efeitos temporais.

No artigo de Zhu, Huang e Wu (2005) os dados do *lattice* analisados estão disponíveis em diferentes períodos de tempo. Os autores destacam a necessidade do acréscimo da dependência temporal no modelo autologístico. O principal objetivo do artigo foi o de desenvolver um modelo autologístico espaço-temporal que, sistematicamente, modelasse a relação entre a variável resposta e potenciais variáveis explicativas, acomodando a dependência espacial e temporal simultaneamente. A formulação do modelo tem uma similaridade com Besag (1972) e com Preisler (1993). Em Besag (1972), o modelo autologístico espaço-temporal foi proposto com a suposição de estacionaridade. Preisler (1993) considera a regressão autologística espaço-temporal, mas supôs a independência entre as observações em diferentes tempos.

A contribuição de Zhu, Huang e Wu (2005) foi de melhorar o modelo de Preisler (1993) de modo que a dependência temporal esteja contida no modelo. Além disso, o método pode ser estendido para campos aleatórios markovianos com dependência espacial e temporal. O modelo captura ambas as dependências, espacial e temporal, simultaneamente, além das possíveis variáveis explicativas. Para a inferência, os autores usam máxima pseudo-verossimilhança, e comentam que para modelos autologísticos com dependência espacial mas sem variáveis explicativas, a eficiência do estimador de máxima pseudo-verossimilhança depende dos valores do coeficiente de autocorrelação espacial e às vezes pode

ser comparável com a eficiência das estimativas de máxima verossimilhança. No entanto, o erro-padrão das estimativas da regressão logística são inválidos e, portanto, precisam ser recalculados. Os autores usam *Bootstrap* de uma forma semelhante à Gumpertz, Graham e Ristiano (1997). Eles propõem para futuras pesquisas a extensão do modelo para coeficientes variando no tempo e termos com interação entre o espaço e tempo.

Um modelo autologístico espaço-temporal foi usado por Zheng e Zhu (2008) para relacionar uma variável resposta binária a potenciais covariáveis em um *lattice*. O modelo acomoda a dependência espacial ao longo do tempo de estudo. Essa abordagem é útil para a análise espaço-temporal dos dados binários. No entanto, o método de inferência estatística comumente usado, máxima pseudo-verossimilhança, em alguns casos pode ser estatisticamente ineficiente, principalmente quando a dependência espacial e temporal é forte. O artigo propõe uma abordagem bayesiana para a inferência e para previsão de respostas futuras. Um algoritmo de Metropolis-Hastings combinado com amostrador de Gibbs foi desenvolvido para a obtenção da distribuição a posteriori dos parâmetros, bem como para as distribuições preditivas a posteriori. Os autores demonstram a metodologia sugerida e comparam os resultados obtidos nesse método com resultados obtidos nos métodos de máxima pseudo-verossimilhança e máxima verossimilhança com MCMC através de um exemplo de dados de rupturas causadas por besouros em pinus.

A expressão do modelo autologístico espaço-temporal é dada por,

$$Pr(Y_{k,l} = y_{k,l} | x_{k,l}, y_{k,l,t}, (k, l) \in D) = \frac{\exp\{\sum_{j=0}^p \beta_j x_{k,l,j} + \Delta_t (\sum_{d=1}^q \gamma_d y_{k,l,d})\}}{1 + \exp\{\sum_{j=0}^p \beta_j x_{k,l,j} + \Delta_t (\sum_{d=1}^q \gamma_d y_{k,l,d})\}}. \quad (13)$$

A diferença desse modelo para o modelo autologístico atemporal é o acréscimo do Δ_t considerado *offset* no modelo, e que é o peso do valor da diferença entre os diferentes períodos de tempo (BESAG, 1972). Os métodos de estimação são os mesmos métodos usados no modelo autologístico (atemporal).

2.2 Material e métodos

2.2.1 Material

Neste trabalho serão analisados dois conjuntos de dados de doenças cítricas. O primeiro refere-se à incidências de MSC e o outro à incidências de leprose dos citros. Os dados são proveniente de talhões onde as plantas foram monitoradas regularmente durante um certo período de tempo e foram cedidos pelo Fundecitrus (Fundo de Defesa da Citricultura). O Fundecitrus é uma associação de citricultores e indústrias processadoras de frutas cítricas. Atua na sanidade dos pomares de cítricos, realiza e financia pesquisas científicas para a descoberta de formas de controle ou manejo de doenças e pragas que afetam a citricultura brasileira.

2.2.1.1 Dados de morte súbita dos citros

Os dados de MSC foram obtidos na Fazenda Vale Verde, localizada no município de Comendador Gomes, estado de Minas Gerais ($19^{\circ} 64' 68''$ S , $48^{\circ} 98' 04''$ O). O talhão possui 20 linhas com 48 plantas em cada uma. O espaçamento entre linhas é de 7,5 metros e entre as plantas na linha é de 4 metros. As plantas tinham 12 anos na época da coleta dos dados.

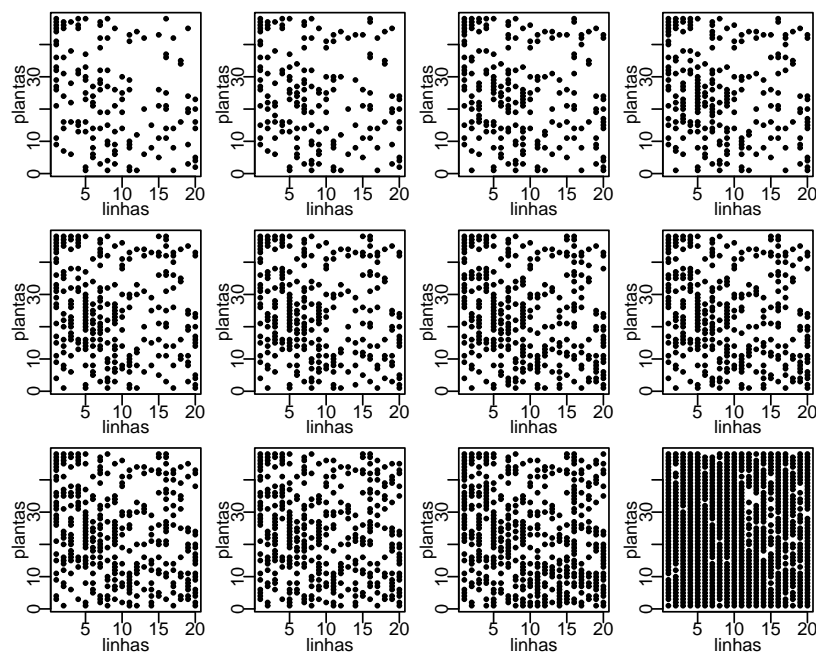


Figura 4 - Incidência de MSC em talhão da Fazenda Vale Verde em 12 avaliações entre 05/11/2001 e 07/10/2002.

Foram coletadas observações em 12 avaliações feitas entre os dias 05/11/2001 e 07/10/2002. A coleta dos dados resumia-se a diagnosticar o *status* de cada planta dentro do talhão. A cada avaliação, se a planta observada apresentasse sintomas da doença atribuíá-se o valor 1, se estivesse

sadia atribuía-se o valor 0.

O mapa da incidência da doença em cada avaliação está apresentado na Figura 4. A incidência da doença no talhão variou de 14,9% na primeira avaliação até 45,73% na 11ª avaliação. A 12ª avaliação foi descartada das análises pois apresentava incidência da doença em quase a totalidade das plantas.

2.2.1.2 Dados de leprose dos citros

Os dados com incidência de leprose de citros foram colatados em um talhão de citros da variedade ‘Valência’ enxertada sobre limoeiro ‘cravo’, plantado em 1996. O talhão está localizado em uma propriedade comercial no município de Santa Cruz do Rio Pardo/SP (22° 53' 56" S , 49° 37' 58" O). O espaçamento entre linhas é de 7,5 metros e entre plantas na linha é de 3,8 metros. Foram analisadas avaliações entre os períodos de 2002 a 2004. Nesse período não foram realizadas pulverizações com acaricidas no talhão, de modo a não afetar a população de ácaro da leprose.

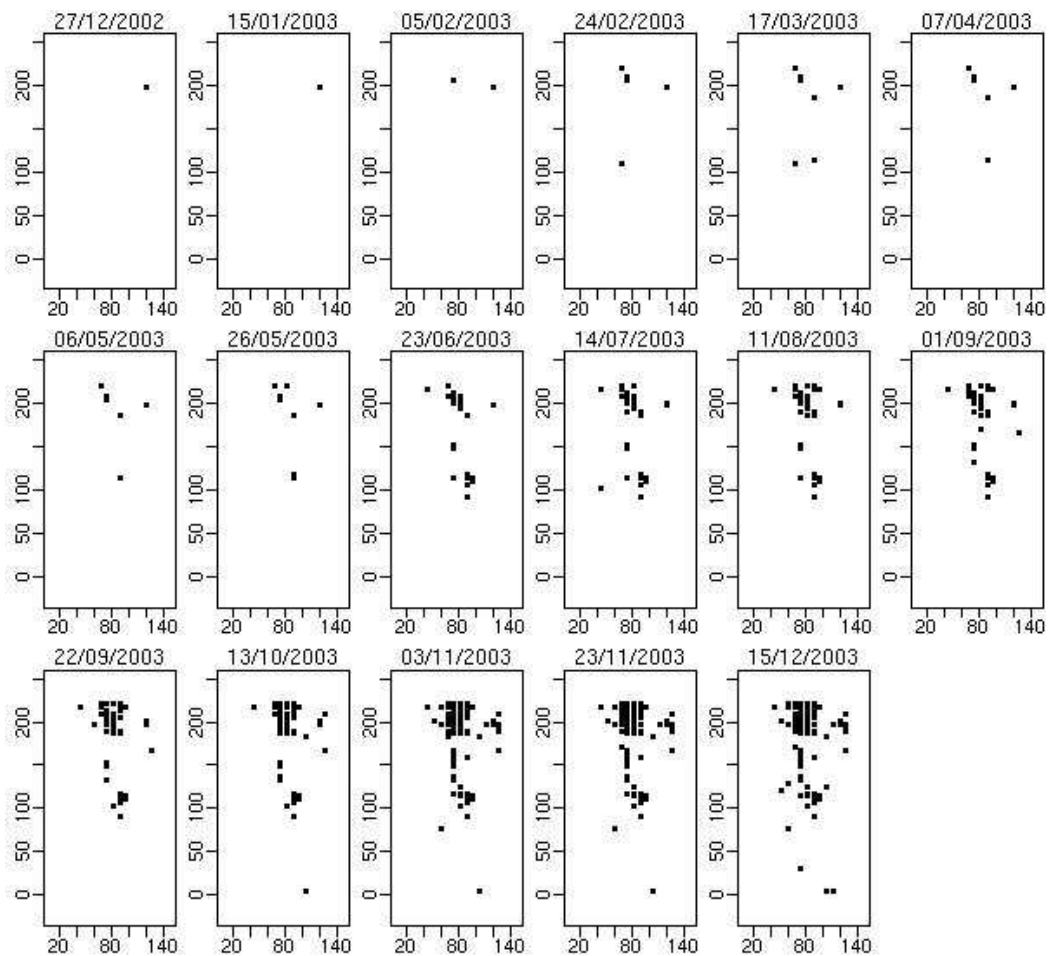


Figura 5 - Mapas com incidência de leprose dos citros em um talhão nas avaliações do ano de 2003

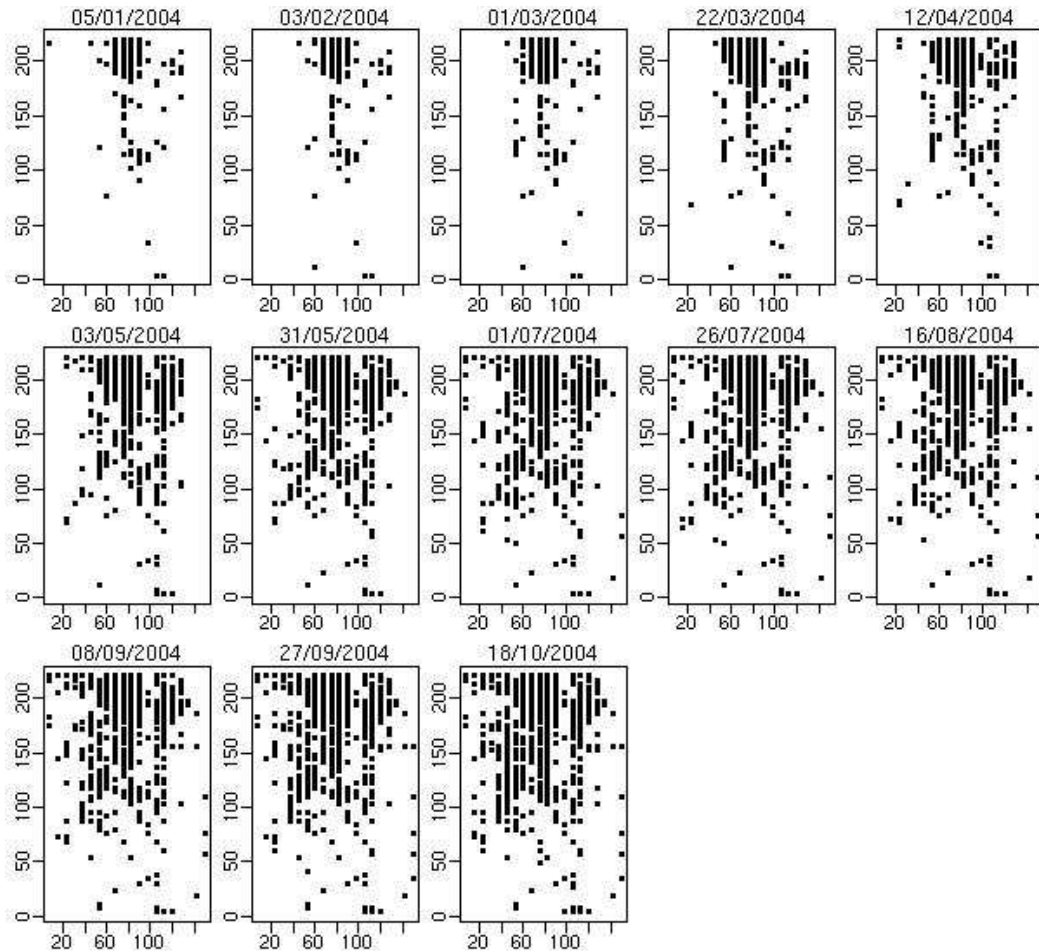


Figura 6 - Mapas com incidência de leprose dos citros em um talhão nas avaliações do ano de 2004

Foram registradas as incidências de doença e também a presença do ácaro transmissor em cada planta do talhão. O monitoramento do talhão foi iniciado em 20/01/2002 com avaliações aproximadamente mensais. Na avaliação de 27/05/2002 foi detectada a primeira planta com a doença, não tendo a doença sido detectada em outras plantas até a avaliação de 05/02/2003. As Figuras 5 e 6 mostram as avaliações dos anos 2003 e 2004, respectivamente onde observa-se a dispersão da doença pelo talhão ao longo do tempo de avaliação. As avaliações foram feitas em intervalos variáveis de tempo. Foram coletadas as seguintes informações a partir dos registros de infestação por *B. phoenicis* e de infecção por CiLV nas plantas: as incidências codificadas de forma binária sendo 0 para ausência do evento ácaro ou sintoma e 1 para presença do evento. A codificação foi feita separadamente para presença de sintomas em órgãos avaliados nas plantas e, para a planta como um todo, quanto a presença de sintomas e ácaros.

Os dados provenientes dos talhões de plantas cítricas a serem analisados neste trabalho, permitem desenvolver uma metodologia de análise que poderá ser estendida para avaliar os demais talhões monitorados espacialmente para a incidência de doenças cítricas.

2.2.2 Métodos

Considera-se que a presença de sintomas de determinada doença em uma planta pode estar associada com a presença da doença em plantas vizinhas. Desta forma, espera-se que plantas próximas tenham características similares, particularmente, que a localização das plantas doentes apresente algum padrão espacial. Na análise de dados binários geralmente é usada a regressão logística, como por exemplo, na análise de presença de doenças em plantas. No contexto de plantas em um talhão, é razoável assumir que plantas vizinhas tendem a ter *status* similar, refletindo uma eventual agregação no padrão espacial da doença. O modelo autológico (BESAG, 1972) estende a regressão logística usual para capturar a dependência espacial dos dados.

Para os dados de incidência de doenças em citros apresentados na Seção anterior a proposta de análise é de considerar a estrutura espacial através da modelagem da probabilidade condicional de uma planta ser infectada, dado o *status* das plantas vizinhas, por meio do modelo autológico.

2.2.2.1 Metodologia para os dados de MSC

O conjunto de dados de MSC consiste do registro da presença da doença em cada planta do talhão, para cada momento de coleta. Para a modelagem dos dados nesse caso não há covariáveis usuais e sugere-se utilizar a estrutura de vizinhança entre plantas na construção de covariáveis que possam auxiliar a descrição da probabilidade da ocorrência de doença nas plantas.

Os dados de MSC serão utilizados com intuito de ilustrar o uso das funções dos pacotes *Rcitrus* (KRAINSKI; RIBEIRO JR., 2007) e *stLattice* (KRAINSKI; RIBEIRO JR., 2008), para o ambiente estatístico R (R DEVELOPMENT CORE TEAM, 2008). Esses pacotes foram desenvolvidos motivados pela necessidade de implementar algoritmos para facilitar análises de dados de presença de doenças em monitoramento de talhões de plantas cítricas. Entretanto, as rotinas podem ser utilizadas para outras culturas com arranjos sistemáticos das plantas. Os pacotes implementam funções para manipulação e validação dos dados, métodos exploratórios para detecção de padrões espaciais e alguns modelos estatísticos de análise para dados de doenças em plantas, podendo ainda serem usados para análise de outros tipos de dados binários com arranjo espacial.

As funções *aglm()* do pacote *stLattice* e *autologistic.citrus()* do pacote *Rcitrus* possibilitam o ajuste do modelo autológico a dados binários com dependência espacial. Os argumentos dessas funções permitem diferentes especificações de modelo. Por exemplo, é possível combinar diferentes estruturas de vizinhança, diferentes métodos de estimação dos parâmetros e diferentes períodos de tempo de avaliação. Essa possibilidade de construir diversas combinações de estratégias de análise auxilia na busca por um modelo adequado e de fácil interpretabilidade.

A função *autologistic.citrus()* do pacote *Rcitrus* traz argumentos básicos para a especifi-

cação do modelo.

```
autologistic.citrus(formula=Y~R+C+dA+dB, bor=1, obj, size=NULL,
                    obj2=NULL, covariate=NULL, death=1,
                    healt=0, inf.method=c("pseudo","mc",
                    "bootstrap"), N, binary.covar=FALSE,
                    verbose=FALSE)
```

Se o interesse está em estudar diferentes estruturas de vizinhança utiliza-se o argumento *formula* que permite combinar covariáveis de linha, coluna, diagonais, de primeira ou segunda ordem, além de variar o tempo das covariáveis de vizinhança quando existirem diferentes períodos de avaliação. Caso existam covariáveis usuais estas são acrescentadas pelo argumento *covariate*. O argumento *obj* especifica o objeto contendo o conjunto de dados da classe *matrix* ou *data.frame*. A variável resposta entra com atributo de doente=1 e sadia=0, porém a codificação pode ser alterada pelos argumentos *death* e *health*. Os métodos de estimação implementados são o método de máxima pseudo-verossimilhança, *bootstrap* com amostrador de Gibbs para restimar as estimativas dos erros e o método de Monte Carlo. Existem outros argumentos pertencentes à função para situações mais específicas.

No pacote *stLattice* está disponível a função *aglm()* que implementa outras opções para o ajuste do modelo autologístico aos dados.

```
aglm(formula, data, sp = ~first, family = binomial,
      brlr=TRUE, bootstrap=TRUE, nboot=99,
      mctest = FALSE, nmc=99, bor = 1, lag = -1, times=30)
```

O argumento *formula* permite especificar o modelo de regressão logística com covariáveis usuais que possam existir no conjunto de dados. A classe *stLattice* define um formato específico para os dados a ser fornecido ao argumento *data*. A estrutura dessa classe é formada por três elementos: *coords* com as coordenadas (x,y) das localizações dos dados; *dates* com datas das avaliações e *data*, que é um *array* com quatro dimensões dadas pelo produto dos números de linhas, colunas, tempos e covariáveis. No argumento *sp* define-se as covariáveis de vizinhança do modelo, ou seja, a parte espacial do modelo. A vizinhança pode ser de primeira ordem (R + C), de segunda ordem (R+C+D), de segunda ordem com as diagonais separadas (R+C+dA+dB), ou ainda outras combinações de covariáveis. Em *lag* define-se o período do tempo em que as covariáveis de vizinhança são avaliadas, podem ser períodos anteriores ou contemporâneos. Define-se *bootstrap=TRUE* quando o procedimento de estimação usar pseudo-verossimilhança combinada com *bootstrap* via amostrador de Gibbs e *mctest=TRUE* quando o método de Monte Carlo é utilizado com o número de simulações dado por *N*. O argumento *brlr=TRUE* considera a correção de viés para ser usada no cálculo da função de pseudo-verossimilhança.

O procedimento de correção de viés (Firth, 1993) está implementado como uma opção no pacote *stLattice*. A aplicação desse método se dá nos casos em que as estimativas de parâmetros

não podem ser obtidas na maximização da função de pseudo-verossimilhança, fato que pode ocorrer na regressão logística. Isso é devido à separação completa ou quase-completa que ocorre quando há uma covariável que prediz perfeita ou quase perfeitamente a resposta. Supondo uma covariável com dois níveis e construindo com a resposta uma tabela de freqüências 2×2 , tem-se a separação completa no caso em que todos os elementos fora da diagonal desta tabela são nulos e quase completa quando um desses elementos é nulo. No caso onde a covariável é a presença da doença na vizinhança, a separação quase-completa pode ocorrer se, por exemplo, todas as plantas doentes do talhão tem vizinha na linha doente.

O que diferencia o uso dos dois pacotes no ajuste do modelo autolístico é o interesse do pesquisador e a estrutura dos dados. A função *aglm()* inclui o argumento *times*, para permitir o uso em situações onde além da estrutura espacial, a temporal também deva ser considerada. Esse argumento é o peso do valor da diferença entre os diferentes períodos de tempo e define o termo de *offset* do modelo linear generalizado. Em outras palavras, o pacote *stLattice* faz análises de dados espaço-temporais e incorpora os dados de todas as avaliações em um modelo único. Por outro lado, o pacote *Rcitrus* ajusta um modelo para cada avaliação, sem considerar o efeito temporal.

Segundo Krainsk et al. (2008) o modelo autolístico calcula a probabilidade p_{kl} de determinada planta na k -ésima linha e l -ésima coluna estar doente como uma função da combinação linear do *status* (doente/sadia) das plantas vizinhas. Considerando a função de ligação logística tem-se que:

$$\text{logit}(p_{kl}) = \beta_0 + \gamma_1 L_{kl1} + \gamma_2 C_{kl2} + \gamma_3 Da_{kl3} + \gamma_4 Db_{kl4} ,$$

em que $L_{kl1} = y_{k-1,l} + y_{k+1,l}$ é o *status* das plantas na linha adjacente e forma a covariável dentro da linha; $C_{kl2} = y_{k,l+1} + y_{k,l-1}$ o *status* das plantas em colunas adjacentes produzindo a covariável entre linhas, $Da_{kl3} = y_{k-1,l+1} + y_{k+1,l-1}$ e $Db_{kl4} = y_{k-1,l-1} + y_{k+1,l+1}$ o *status* das plantas das diagonais, formam as covariáveis de vizinhança das duas diagonais (decrecente e crescente, respectivamente). Os parâmetros γ 's medem o efeito de cada respectiva covariável de vizinhança. A separação dos efeitos de linhas e colunas de plantas, acomoda o fato de que o espaçamento dentro e entre linhas é diferente e permite estudar os efeitos direcionais da doença. A estimação do modelo é feita usando o método de pseudo-verossimilhança (BESAG, 1975).

Os modelos propostos para análise dos dados de MSC combinam diferentes estruturas de vizinhanças. São consideradas vizinhanças de primeira ordem, com efeito de linhas e colunas separadas e de segunda ordem, com efeito de linhas, colunas e diagonais. Considera-se ainda períodos de avaliação anteriores e contemporâneos à avaliação. Essas diferentes formas de combinar a estrutura das covariáveis de vizinhança com o tempo de avaliação auxiliam no estudo da estrutura de vizinhança que melhor se adapta ao ajuste dos dados e permite avaliar se a defasagem no tempo é fator importante na incidência

da doença.

Nas especificações de modelo utilizadas aqui tem-se que $M1$ considera como covariáveis de vizinhança as observações das plantas vizinhas dentro e entre linhas. O modelo $M2$ acrescenta as covariáveis das diagonais. Ambos consideram medidas no tempo de observação anterior,

$$\begin{aligned} M1 : \text{logit}(p_{klt}) &= \beta_0 + \gamma_1 L_{(k,l,t-1,1)} + \gamma_2 C_{(k,l,t-1,2)} \\ M2 : \text{logit}(p_{klt}) &= \beta_0 + \gamma_1 L_{(k,l,t-1,1)} + \gamma_2 C_{(k,l,t-1,2)} + \gamma_3 Da_{(k,l,t-1,3)} + \gamma_4 Db_{(k,l,t-1,4)} . \end{aligned}$$

Os modelos $M3$ e $M4$ consideram as mesmas covariáveis dos modelos $M1$ e $M2$, respectivamente, porém refletindo o *status* das plantas vizinhas no mesmo tempo de observação da resposta.

$$\begin{aligned} M3 : \text{logit}(p_{klt}) &= \beta_0 + \gamma_1 L_{(k,l,t,1)} + \gamma_2 C_{(k,l,t,2)} \\ M4 : \text{logit}(p_{klt}) &= \beta_0 + \gamma_1 L_{(k,l,t,1)} + \gamma_2 C_{(k,l,t,2)} + \gamma_3 Da_{(k,l,t,3)} + \gamma_4 Db_{(k,l,t,4)} . \end{aligned}$$

O teste de significância dos parâmetros de regressão são baseados na aproximação usual para modelos lineares generalizados, assumindo que $\frac{\hat{\gamma}}{\sqrt{\text{Var}(\hat{\gamma})}} \sim N(0, 1)$. Nos modelos apresentados, o teste da significância dos coeficientes estará testando a existência da dependência espacial e permitirá detectar se a agregação ocorre apenas nas linhas (curto alcance), entre as linhas (médio alcance) ou também nas diagonais (longo alcance). A escolha entre os diferentes modelos permite ainda verificar a importância de se considerar a defasagem no tempo para construção das covariáveis espaciais.

A análise geral da significância dos coeficientes dos modelos $M1$ e $M2$, estará testando a capacidade preditiva desses modelos. A análise da significância de cada coeficiente separadamente, permitirá estudar se a forma de propagação da doença possui efeitos direcionais. Os modelos $M3$ e $M4$ testam a significância das covariáveis de vizinhança no tempo contemporâneo e, desta forma, esses modelos dão um enfoque descritivo à análise dos padrões espaciais da doença.

2.2.2.2 Metodologia para os dados de leprose dos citros

Informações sobre o padrão espacial de plantas doentes auxiliam no conhecimento dos mecanismos de dispersão da doença, do patógeno e, conseqüentemente, das epidemias.

Medidas como a distância mínima entre plantas com incidência de leprose e a análise por *quadrats* são úteis para as análises exploratórias dos dados (SPÓSITO et al., 2007). O teste da distância mínima média consiste em calcular as distâncias mínimas entre plantas doentes. Aplica-se um teste Monte Carlo, comparando o valor observado de plantas com incidência da doença com valores obtidos sob hipótese nula de aleatoriedade espacial completa. O valor- p corresponde à proporção de distâncias médias mínimas calculadas nas simulações, menores ou iguais à média das distâncias mínimas observadas.

A análise por *quadrats* consiste em um procedimento simples para caracterizar o padrão espacial da doença. Para análise faz-se uma subdivisão do talhão em N regiões denominadas *quadrats*. O procedimento de seleção dos *quadrats*, a forma e o tamanho são flexíveis. Podem ser tomados de forma fixa, que consiste em uma subdivisão regular do talhão, ou de forma aleatória, em que a posição do *quadrat* é tomada por sorteio no talhão (KRAINSKI et al., 2005). A análise do padrão espacial é feita a partir do número de plantas doentes dentre as plantas de cada *quadrat*. Tipicamente, um padrão espacial é classificado como aleatório, agregado ou regular. O padrão espacial aleatório está relacionado a iguais oportunidades de infecção enquanto o agregado ocorre quando, em condições naturais, o patógeno dispersa-se predominantemente a curtas distâncias, como por exemplo, através de respingos da chuva. Nesse caso a probabilidade de que uma planta situada próxima à fonte de inóculo se torne infectada é maior do que a de plantas mais distantes. O padrão regular não ocorre de forma natural, mas pode ser provocado pelo plantio alternado de variedades resistentes e suscetíveis (LIMA et al., 2006). Para análise do padrão espacial, o teste do ajuste pode ser feito pelo índice de dispersão D , que é a razão entre a variância observada nos dados e a variância teórica, segundo uma distribuição de probabilidades assumida para as contagens. A hipótese de aleatoriedade espacial é testada considerando que, sob essa hipótese, $(N - 1)D$ tem distribuição $\chi^2_{(N-1)}$ onde N é o número de *quadrats*.

Após análises exploratórias os padrões espaciais podem ser investigados em mais detalhes através da modelagem dos dados. Dados de incidência de doenças em plantas coletados em diferentes períodos de tempo permitem que, além da dimensão espacial, a dimensão temporal dos dados também seja explorada. O modelo autológico espaço-temporal considera os diferentes períodos de tempo entre as avaliações e ajusta um modelo para o conjunto de todas as avaliações.

São considerados aqui diferentes modelos autológicos espaço-temporais para a probabilidade de doença nas plantas. As equações dos modelos são apresentadas na Tabela 1. Os modelos apresentam covariáveis usuais $x_{k,l,j}$ refletindo presença do ácaro em cada planta na k -ésima linha e l -ésima coluna. Considerou-se duas formas para tal covariável, a simples presença ou ausência do ácaro denotada por $x_{k,l,j,1}$ ou o número de ácaros na planta denotado por $x_{k,l,j,2}$. As seguintes covariáveis de vizinhança foram utilizadas: $L_{(k,l,t)} = (y_{k-1,l,t} + y_{k+1,l,t})$ é o *status* das plantas na linha adjacente e forma a covariável dentro da linha; $C_{(k,l,t)} = (y_{k,l+1,t} + y_{k,l-1,t})$ o *status* das plantas em colunas adjacentes produzindo a covariável entre linhas e $S_{(k,l,t-1)} = y_{k,l,t-1}$ que é a covariável definida pelo *status* da mesma planta no tempo anterior.

A primeira proposta é de modelar os dados de uma avaliação t , utilizando a mesma planta na avaliação anterior $t - 1$. Esse modelo ($M1$) permite verificar se o *status* da planta pode estar determinado pela condição dela mesma no tempo anterior. Na prática, o fato da planta estar ou não doente na avaliação anterior deve ser considerada sempre, ou seja, essa covariável permanece em todos

os modelos propostos, dada a persistência da leprose nas plantas. Um outro modelo considerado ($M2$), com covariáveis de vizinhança no período de tempo anterior $t - 1$, reflete o contágio da doença pela vizinhança do passado.

O terceiro modelo ($M3$) considera as covariáveis de vizinhança no período de tempo contemporâneo à avaliação. Esse modelo reflete um contágio por fontes adicionais que não foram capturadas em $M2$, atribuíveis a outros aspectos da propagação da doença que não só a influência das plantas vizinhas doentes na inspeção anterior.

Tabela 1 - Equações dos modelos autolísticos espaço-temporais propostos para o ajuste dos dados de leprose

Modelos	Equações
M1	$\beta_0 + \gamma_0 S_{(k,l,t-1,0)}$
M2	$\beta_0 + \gamma_0 S_{(k,l,t-1,0)} + \gamma_1 L_{(k,l,t-1,1)} + \gamma_2 C_{(k,l,t-1,2)}$
M3	$\beta_0 + \gamma_0 S_{(k,l,t-1,0)} + \gamma_1 L_{(k,l,t,1)} + \gamma_2 C_{(k,l,t,2)}$
M4	$\beta_0 + \beta_1 x_{(k,l,1,1)} + \gamma_0 S_{(k,l,t-1,0)}$
M5	$\beta_0 + \beta_1 x_{(k,l,1,1)} + \gamma_0 S_{(k,l,t-1,0)} + \gamma_1 L_{(k,l,t-1,1)} + \gamma_2 C_{(k,l,t-1,2)}$
M6	$\beta_0 + \beta_1 x_{(k,l,1,1)} + \gamma_0 S_{(k,l,t-1,0)} + \gamma_1 L_{(k,l,t,1)} + \gamma_2 C_{(k,l,t,2)}$
M7	$\beta_0 + \beta_1 x_{(k,l,2,1)} + \gamma_0 S_{(k,l,t-1,0)}$
M8	$\beta_0 + \beta_1 x_{(k,l,2,1)} + \gamma_0 S_{(k,l,t-1,0)} + \gamma_1 L_{(k,l,t-1,1)} + \gamma_2 C_{(i,j,t-1,2)}$
M9	$\beta_0 + \beta_1 x_{(k,l,2,1)} + \gamma_0 S_{(k,l,t-1,0)} + \gamma_1 L_{(k,l,t-1,1)} + \gamma_2 C_{(i,j,t-1,2)}$

A incidência de ácaro na própria planta, ou em suas vizinhas, pode influenciar a probabilidade de uma planta apresentar a doença. A fim de avaliar e quantificar este efeito, uma covariável é construída a partir da ocorrência do ácaro e acrescentada aos modelos anteriores, definindo-se três novos modelos - $M4$, $M5$ e $M6$. Avaliou-se ainda se, ao invés da simples ocorrência, o número de ácaros na planta poderia explicar melhor a probabilidade de ocorrer leprose. Os modelos $M7$, $M8$ e $M9$ avaliam esse efeito juntamente com as covariáveis de vizinhança.

Os modelos considerados sugerem diferentes mecanismos para explicar a dispersão da doença e o interesse é identificar o modelo que melhor reflete os padrões observados nos dados. Dois critérios de seleção de modelos comumente usados e adotados aqui são o critério de informação de Akaike (AIC) e o critério de informação de Bayes (BIC). Em ambas opções é considerado o melhor modelo aquele que apresenta o menor valor das tais medidas. Contudo, não deve substituir a interpretação e informação contextual, especialmente quando as diferenças entre os modelos são pequenas, em especial no caso específico dos modelos espaciais considerados aqui onde a verossimilhança é apenas uma aproximação (KRAINSKI et al., 2008).

O estudo da relação espacial e temporal da associação entre a incidência de plantas infestadas por *B. phoenicis* e da incidência de leprose dos citros, pode trazer estimativas mais precisas

sobre o crescimento da epidemia, no tempo e no espaço. A detecção de padrões espaço-temporais nos resultados sugerem métodos de controle do vetor e práticas de intervenção para a retirada de órgãos afetados concentrados nas áreas de localização dos focos da doença, ao invés da ação na área total (CZERMAINSKI, 2006).

A estimação do modelo será feita maximizando a função de pseudo-verossimilhança dada pela equação (8). Como dito na Seção 2.1 esse método traz estimativas consistentes dos parâmetros do modelo, porém as estimativas de variância são inacuradas no caso onde covariáveis são construídas a partir da própria variável resposta. É adotada aqui a proposta que utiliza a metodologia *Bootstrap* com amostrador de Gibbs, que se preserva o padrão espacial dos dados na reamostragem. Krainski et al. (2008) descreve o algoritmo de estimação implementado pelos autores para obter a reamostragem e as novas estimativas.

Os procedimentos de análise foram feitos no ambiente estatístico R (R DEVELOPMENT CORE TEAM, 2008), usando o pacote *stLattice* (KRAINSKI; RIBEIRO JR., 2008).

2.3 Resultados e discussão

2.3.1 Dados de morte súbita em citros

A Figura 4 mostra um rápido espalhamento da MSC no talhão durante o tempo de avaliação. Na primeira avaliação 15% das plantas dentro do talhão já apresentavam a doença, que atingiu 46% na 11ª avaliação, no período de 11 meses de observação. A simples inspeção visual sugere uma provável existência de dependência espacial da doença entre as plantas.

Os modelos apresentados na Seção 2.2.1.1 foram ajustados aos dados. Os procedimentos de análise com o uso de funções dos pacotes *Rcitrus* e *stLattice* estão descritos no Apêndice.

A Tabela 2 mostra as estimativas dos parâmetros, erros-padrão e significância para o modelo $M1$ ajustado aos dados de MSC. Ajustou-se um modelo para cada uma das avaliações, a partir da segunda. A significância das estimativas mostrou que a covariável de vizinhança dentro da linha, no instante de tempo anterior, influencia a probabilidade de presença da doença. Ao analisar as vizinhas de primeira ordem separadamente (dentro da linha e entre linhas de plantas) é possível medir o efeito do espaçamento entre as plantas. Neste caso, o modelo mostrou que o espaçamento menor dentro da linha tem maior impacto na probabilidade de ocorrência da doença.

A Tabela 3 apresenta as estimativas, erros-padrão e significância para o modelo $M2$ em cada uma das avaliações. Para as covariáveis de segunda ordem, defasadas no tempo $t - 1$, foram significativas ($p < 0,05$) apenas a covariável dentro da linha, a diagonal decrescente nas avaliações 5 e 6, e a diagonal crescente na avaliação 7.

A Tabela 4 mostra as estimativas dos parâmetros, erros-padrão e significância do modelo

Tabela 2 - Estimativas dos parâmetros, erros-padrão e significância para o modelo $M1$ ajustado aos dados de MSC

Avaliações	$\hat{\beta}_0 \pm EP(\hat{\beta}_0)$	$\hat{\gamma}_1 \pm EP(\hat{\gamma}_1)$	$\hat{\gamma}_2 \pm EP(\hat{\gamma}_2)$
2	$-1,806 \pm 0,163^{**}$	$0,362 \pm 0,181^*$	$0,120 \pm 0,192$
3	$-1,605 \pm 0,131^{**}$	$0,473 \pm 0,163^{**}$	$0,151 \pm 0,151$
4	$-1,567 \pm 0,141^{**}$	$0,650 \pm 0,137^{**}$	$0,026 \pm 0,147$
5	$-1,431 \pm 0,138^{**}$	$0,615 \pm 0,124^{**}$	$0,033 \pm 0,148$
6	$-1,388 \pm 0,166^{**}$	$0,611 \pm 0,118^{**}$	$0,058 \pm 0,140$
7	$-1,215 \pm 0,129^{**}$	$0,596 \pm 0,113^{**}$	$0,196 \pm 0,121$
8	$-1,276 \pm 0,130^{**}$	$0,621 \pm 0,117^{**}$	$0,143 \pm 0,119$
9	$-1,222 \pm 0,134^{**}$	$0,615 \pm 0,118^{**}$	$0,125 \pm 0,112$
10	$-0,931 \pm 0,129^{**}$	$0,503 \pm 0,108^{**}$	$0,021 \pm 0,110$
11	$-0,592 \pm 0,188^{**}$	$0,436 \pm 0,115^{**}$	$0,038 \pm 0,111$

** indica significância a 1% de probabilidade.

Tabela 3 - Estimativas dos parâmetros, erros-padrão e significância para o modelo $M2$ ajustado aos dados de MSC

Av.	$\hat{\beta}_0 \pm EP(\hat{\beta}_0)$	$\hat{\gamma}_1 \pm EP(\hat{\gamma}_1)$	$\hat{\gamma}_2 \pm EP(\hat{\gamma}_2)$	$\hat{\gamma}_3 \pm EP(\hat{\gamma}_3)$	$\hat{\gamma}_4 \pm EP(\hat{\gamma}_4)$
2	$-1,866 \pm 0,153^{**}$	$0,357 \pm 0,192^*$	$0,096 \pm 0,205$	$0,069 \pm 0,199$	$0,136 \pm 0,171$
3	$-1,726 \pm 0,136^{**}$	$0,441 \pm 0,153^{**}$	$0,116 \pm 0,161$	$0,165 \pm 0,167$	$0,245 \pm 0,166$
4	$-1,721 \pm 0,146^{**}$	$0,639 \pm 0,134^{**}$	$-0,047 \pm 0,137$	$0,205 \pm 0,141$	$0,224 \pm 0,144$
5	$-1,624 \pm 0,145^{**}$	$0,618 \pm 0,121^{**}$	$-0,078 \pm 0,137$	$0,292 \pm 0,137^*$	$0,207 \pm 0,151$
6	$-1,577 \pm 0,134^{**}$	$0,595 \pm 0,113^{**}$	$-0,028 \pm 0,124$	$0,242 \pm 0,131^*$	$0,209 \pm 0,137$
7	$-1,362 \pm 0,137^{**}$	$0,582 \pm 0,117^{**}$	$0,027 \pm 0,127$	$0,132 \pm 0,109$	$0,211 \pm 0,114^*$
8	$-1,392 \pm 0,161^{**}$	$0,611 \pm 0,111^{**}$	$0,101 \pm 0,120$	$0,075 \pm 0,122$	$0,155 \pm 0,126$
9	$-1,293 \pm 0,153^{**}$	$0,608 \pm 0,112^{**}$	$0,102 \pm 0,120$	$-0,016 \pm 0,121$	$0,152 \pm 0,112$
10	$-0,953 \pm 0,144^{**}$	$0,502 \pm 0,105^{**}$	$0,016 \pm 0,111$	$-0,089 \pm 0,118$	$0,127 \pm 0,104$
11	$-0,645 \pm 0,164^{**}$	$0,436 \pm 0,099^{**}$	$0,026 \pm 0,096$	$-0,043 \pm 0,110$	$0,126 \pm 0,106$

* e ** indicam significância a 5% e 1% de probabilidade, respectivamente.

$M3$. Esse modelo traz as covariáveis de vizinhança de primeira ordem, para o tempo contemporâneo a avaliação. Apenas a covariável dentro da linha mostra-se significativa ($p < 0,01$) a partir da segunda avaliação. Entre as linhas de plantas não houve significância das plantas vizinhas.

A Tabela 5 mostra as estimativas dos parâmetros, erros-padrão e significância do modelo $M4$. Esse modelo traz as covariáveis de vizinhança de segunda ordem, para o tempo contemporâneo a avaliação. A covariável dentro da linha mostra-se significativa ($p < 0,001$) a partir da segunda avaliação, além da diagonal decrescente na quinta avaliação.

O modelo autológico espaço-temporal utilizado aqui considera os diferentes períodos de tempo entre as avaliações e ajusta apenas um único modelo para todas as avaliações coletadas. Os mesmos modelos ajustados individualmente para cada avaliação foram agora ajustados de forma

Tabela 4 - Estimativas dos parâmetros, erros-padrão e significância para o modelo $M3$ ajustado aos dados de MSC

Avaliações	$\hat{\beta}_0 \pm EP(\hat{\beta}_0)$	$\hat{\gamma}_1 \pm EP(\hat{\gamma}_1)$	$\hat{\gamma}_2 \pm EP(\hat{\gamma}_2)$
1	$-1,965 \pm 0,134^{**}$	$0,327 \pm 0,199$	$0,004 \pm 0,210$
2	$-1,885 \pm 0,132^{**}$	$0,372 \pm 0,171^{**}$	$0,254 \pm 0,174$
3	$-1,703 \pm 0,129^{**}$	$0,639 \pm 0,141^{**}$	$0,043 \pm 0,149$
4	$-1,619 \pm 0,128^{**}$	$0,709 \pm 0,130^{**}$	$-0,005 \pm 0,138$
5	$-1,480 \pm 0,126^{**}$	$0,604 \pm 0,123^{**}$	$0,066 \pm 0,127$
6	$-1,448 \pm 0,128^{**}$	$0,647 \pm 0,120^{**}$	$0,080 \pm 0,122$
7	$-1,311 \pm 0,131^{**}$	$0,619 \pm 0,113^{**}$	$0,121 \pm 0,114$
8	$-1,312 \pm 0,133^{**}$	$0,633 \pm 0,112^{**}$	$0,162 \pm 0,112$
9	$-1,218 \pm 0,132^{**}$	$0,612 \pm 0,111^{**}$	$0,099 \pm 0,110$
10	$-0,930 \pm 0,135^{**}$	$0,474 \pm 0,107^{**}$	$0,003 \pm 0,106$
11	$-0,850 \pm 0,144^{**}$	$0,546 \pm 0,100^{**}$	$0,143 \pm 0,098$

** significativos a 1% de probabilidade.

Tabela 5 - Estimativas dos parâmetros, erros-padrão e significância para o modelo $M4$ ajustado aos dados de MSC

Av.	$\hat{\beta}_0 \pm EP(\hat{\beta}_0)$	$\hat{\gamma}_1 \pm EP(\hat{\gamma}_1)$	$\hat{\gamma}_2 \pm EP(\hat{\gamma}_2)$	$\hat{\gamma}_3 \pm EP(\hat{\gamma}_3)$	$\hat{\gamma}_4 \pm EP(\hat{\gamma}_4)$
1	$-2,021 \pm 0,155^{**}$	$0,321 \pm 0,200$	$-0,022 \pm 0,213$	$0,007 \pm 0,211$	$0,206 \pm 0,194$
2	$-1,973 \pm 0,153^{**}$	$0,349 \pm 0,173^{**}$	$0,229 \pm 0,175$	$0,139 \pm 0,177$	$0,161 \pm 0,171$
3	$-1,844 \pm 0,151^{**}$	$0,628 \pm 0,141^{**}$	$-0,025 \pm 0,154$	$0,163 \pm 0,151$	$0,233 \pm 0,146$
4	$-1,781 \pm 0,150^{**}$	$0,709 \pm 0,130^{**}$	$-0,098 \pm 0,145$	$0,212 \pm 0,141$	$0,208 \pm 0,140$
5	$-1,682 \pm 0,149^{**}$	$0,589 \pm 0,124^{**}$	$-0,026 \pm 0,132$	$0,309 \pm 0,130^*$	$0,167 \pm 0,130$
6	$-1,633 \pm 0,150^{**}$	$0,632 \pm 0,121^{**}$	$-0,007 \pm 0,128$	$0,187 \pm 0,127$	$0,239 \pm 0,125$
7	$-1,451 \pm 0,156^{**}$	$0,606 \pm 0,113^{**}$	$0,069 \pm 0,118$	$0,095 \pm 0,117$	$0,191 \pm 0,118$
8	$-1,391 \pm 0,155^{**}$	$0,624 \pm 0,112^{**}$	$0,133 \pm 0,116$	$0,027 \pm 0,117$	$0,131 \pm 0,115$
9	$-1,289 \pm 0,157^{**}$	$0,608 \pm 0,111^{**}$	$0,077 \pm 0,115$	$-0,059 \pm 0,115$	$0,187 \pm 0,113$
10	$-0,907 \pm 0,164^{**}$	$0,478 \pm 0,107^{**}$	$0,011 \pm 0,108$	$-0,117 \pm 0,111$	$0,071 \pm 0,108$
11	$-0,900 \pm 0,172^{**}$	$0,524 \pm 0,101^{**}$	$0,125 \pm 0,101$	$-0,078 \pm 0,103$	$0,163 \pm 0,103$

* e ** significativos a 5% e 1% de probabilidade, respectivamente.

conjunta para todas as avaliações. A Tabela 6 mostra as estimativas dos parâmetros, erros-padrão e significância para os quatro modelos. Observa-se que todas covariáveis de todos os modelos mostraram-se significativas ($p < 0,05$), possivelmente, devido ao fato de incluir o efeito de tempo de avaliação.

Ambos os pacotes ajustam adequadamente o modelo autológico aos dados de MSC. Os diversos argumentos dão maior flexibilidade, principalmente no estudo das estruturas de vizinhança e do tempo de avaliação. Quando há interesse em estudar cada avaliação separadamente, sem incluir o intervalo de tempo entre as avaliações, é indicado o uso da função *autologistic.citrus()* do pacote R*citrus*. Quando há interesse em estudar o enfoque espaço-temporal dos dados, em um estudo mais abrangente,

Tabela 6 - Estimativas dos parâmetros, erros-padrão e significância para o modelos $M1$, $M2$, $M3$ e $M4$ ajustados aos dados de MSC, considerando todas as avaliações conjuntamente.

Modelos	$\hat{\beta}_0 \pm EP(\hat{\beta}_0)$	$\hat{\gamma}_1 \pm EP(\hat{\gamma}_1)$	$\hat{\gamma}_2 \pm EP(\hat{\gamma}_2)$	$\hat{\gamma}_3 \pm EP(\hat{\gamma}_3)$	$\hat{\gamma}_4 \pm EP(\hat{\gamma}_4)$
M1	$-1,557 \pm 0,036^{**}$	$1,215 \pm 0,031^{**}$	$0,8471 \pm 0,030^{**}$
M2	$-1,842 \pm 0,038^{**}$	$0,941 \pm 0,033^{**}$	$0,403 \pm 0,035^{**}$	$0,512 \pm 0,034^{**}$	$0,571 \pm 0,034^{**}$
M3	$-2,139 \pm 0,039^{**}$	$1,336 \pm 0,031^{**}$	$0,894 \pm 0,029^{**}$
M4	$-2,395 \pm 0,041^{**}$	$1,031 \pm 0,033^{**}$	$0,407 \pm 0,035^{**}$	$0,494 \pm 0,035^{**}$	$0,589 \pm 0,035^{**}$

** significativos a 1% de probabilidade.

Nota : Sinal convencional utilizado:

.. Não se aplica dado numérico.

usa-se a função $aglm()$ do pacote `stLattice`.

2.3.2 Dados de leprose dos citros

A incidência de leprose dos citros foi monitorada, desde o estágio de ausência da doença, em 45 avaliações entre meados do ciclo de 2002 até o final do ciclo de 2004, em um talhão com 20 linhas de plantas de 58 plantas em cada linha. As Figuras 5 e 6 com os mapas da incidência da leprose durante o período de estudo mostram a leprose dos citros espalhando-se entre as plantas do talhão. Os valores da incidência de leprose nas plantas em cada avaliação estão presentes na Figura 7.

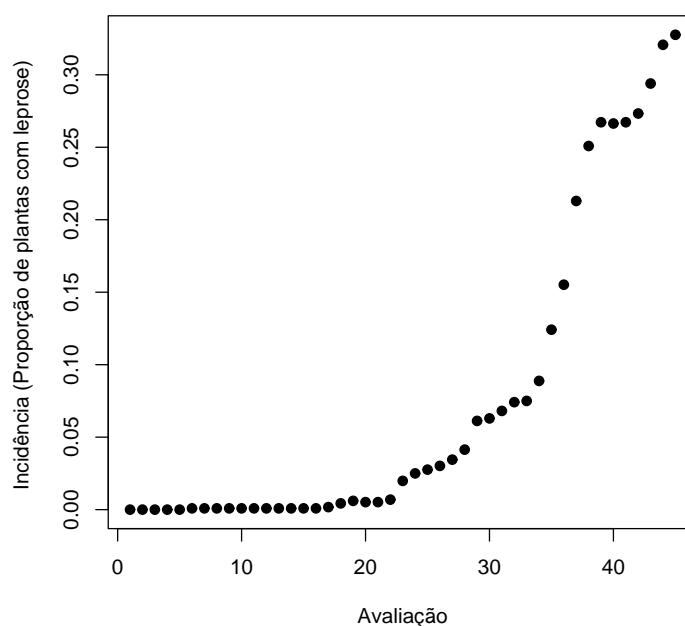


Figura 7 - Incidência de leprose em cada avaliação

Os mapas de incidência de leprose, Figuras 5 e 6, mostram incidências baixas nas avaliações iniciais. Observa-se na Figura 7 que a partir da avaliação 23 a incidência começa a crescer com maior rapidez. Em nosso estudo obteve-se por analisar as avaliações a partir da 23^a, quando 23 plantas apresentaram sintomas de leprose.

Foram feitas análises exploratórias através da medida de distância mínima e da análise por *quadrats*. A Tabela 7 apresenta as datas das avaliações, a incidência de leprose, os valores do índice binomial de dispersão e o teste da distância mínima média. Através da análise por *quadrats* verificou-se o padrão espacial agregado aos dados, ou seja, em condições naturais, o patógeno dispersa-se predominantemente a curtas distâncias. A probabilidade de uma planta situada próxima à fonte de inóculo, nesse caso outra planta doente, se tornar infectada é maior do que a de plantas mais distantes.

A distância mínima média entre plantas com incidência de leprose foi calculada e comparada com as médias obtidas em 199 simulações para obtenção do valor-*p*. A medida que cresce a incidência de leprose diminui a distância entre plantas doentes. Por exemplo, na última avaliação, quando a incidência atinge 32% das plantas no talhão, a média da distância mínima foi de 4,72 metros entre as plantas doentes.

Após as análises exploratórias indicarem a presença do padrão espacial entre as plantas. O interesse agora passa a ser a modelagem do padrão espacial através do modelo autológico. O modelo autológico possibilita o ajuste tanto de um modelo para cada uma das avaliações, quanto apenas um modelo único incorporando os dados de todas as avaliações. A Figura 8 mostra os coeficientes dos parâmetros e os erros-padrão de dois modelos ajustados para incidência de leprose em função apenas da covariável dentro da linha de plantas. Isso é apenas uma ilustração de como podem ser modelados os dados. No modelo *M1* tem-se apenas um coeficiente $\hat{\gamma}$ comum para todas as avaliações. No modelo *M2* há diferentes coeficientes $\hat{\gamma}$ para cada uma das avaliações consideradas.

A estrutura espaço-temporal dos dados permite agregar um estudo de padrões, associações de dependência espacial e de tempo de contágio da doença entre as plantas. Foram ajustados os diferentes modelos autológicos espaço-temporais descritos na Seção 2.2.2.2 e com expressões mostradas na Tabela 1 aos dados das avaliações 23 a 45. O valor da *deviance residual* mantevesse próxima ao valor dos graus-de-liberdade do resíduo para todos os modelos, indicando um ajuste adequado dos modelos aos dados. Os procedimentos de análise com o uso do pacote *stLattice* estão descritos no Apêndice.

As estimativas dos parâmetros e a correspondente significância de cada modelo são mostrados na Tabela 8. Foram obtidas estimativas significativas dos coeficientes em todos os modelos ajustados, o que confirma a existência de padrão espacial da doença. A covariável de vizinhança com maior efeito foi a que reflete o *status* das plantas dentro da linha onde o espaçamento entre plantas é menor comparado ao espaçamento entre linhas de plantas. Este resultado indica que plantas mais

Tabela 7 - Datas das avaliações, incidências de leprose, valores do teste de índice binomial de dispersão e teste da distância mínima média.

Número	Data	Incidências	Índice binomial	Teste da distância
	das avaliações	da doença	de dispersão	mínima média
			<i>Quadrat</i> (2×2)	Dist. Observada(m)
23	23/06/2003	0,0198	1,655 Agreg.	7,76**
24	14/07/2003	0,0250	1,636 Agreg.	6,85**
25	11/08/2003	0,0276	1,885 Agreg.	5,82**
26	01/09/2003	0,0302	1,975 Agreg.	6,65**
27	22/09/2003	0,0345	2,039 Agreg.	6,50**
28	13/10/2003	0,0414	2,313 Agreg.	7,36**
29	03/11/2003	0,0612	2,433 Agreg.	6,90**
30	23/11/2003	0,0629	2,532 Agreg.	6,87**
31	15/12/2003	0,0681	2,419 Agreg.	6,37**
32	05/01/2004	0,0741	2,401 Agreg.	6,57**
33	03/02/2004	0,0750	2,405 Agreg.	6,44**
34	01/03/2004	0,0888	2,335 Agreg.	6,39**
35	22/03/2004	0,1241	2,423 Agreg.	5,51**
36	12/04/2004	0,1552	2,311 Agreg.	4,84**
37	03/05/2004	0,2129	2,460 Agreg.	4,66**
38	31/05/2004	0,2509	2,426 Agreg.	4,67**
39	01/07/2004	0,2672	2,414 Agreg.	4,89**
40	26/07/2004	0,2664	2,327 Agreg.	4,97**
41	16/08/2004	0,2672	2,309 Agreg.	5,03**
42	08/09/2004	0,2733	2,311 Agreg.	5,03**
43	27/09/2004	0,2940	2,309 Agreg.	4,88**
44	18/10/2004	0,3207	2,347 Agreg.	4,80**
45	22/11/2004	0,3276	2,412 Agreg.	4,72**

** significativos a 1% de probabilidade.

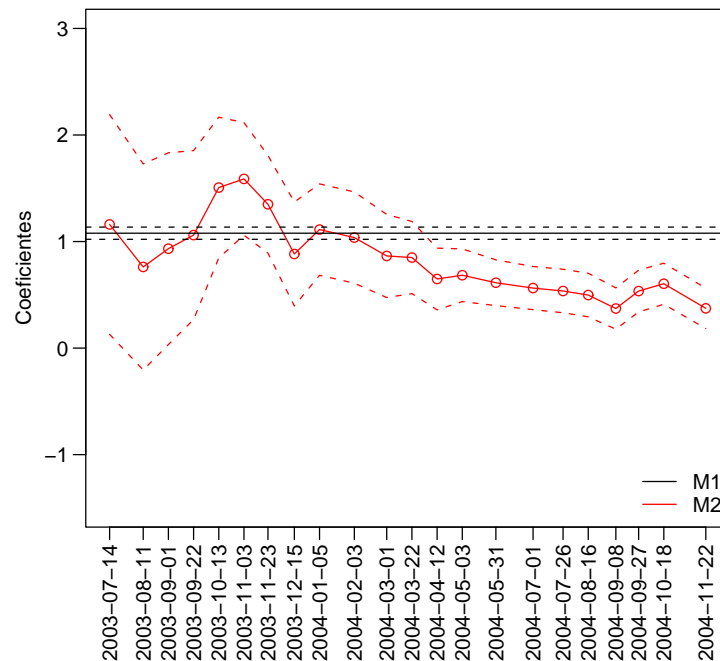


Figura 8 - Coeficientes e erros-padrão para o modelo autológico

próximas são mais susceptíveis ao contágio da doença por suas vizinhas doentes.

O fato de que a covariável de vizinhança entre as linhas de plantas é significativa ($p < 0,001$) em todos os modelos ajustados indica que a dependência não é de apenas curto alcance. Nos modelos em que foram incluídas as covariáveis de presença do ácaro, o efeito também foi significativo, tanto para incidência, quanto para número de ácaros na planta ($p < 0,001$). Este resultado confirma que a presença do ácaro em plantas vizinhas aumenta a probabilidade da doença, quantificando o efeito da mobilidade deste vetor. Os modelos *M4*, *M5* e *M6* apresentaram menores valores de AIC, que *M7*, *M8*, e *M9*, respectivamente. Portanto, considerar o número de ácaros por planta não implicou em vantagens na qualidade do ajuste do modelo em relação à simples incidência binária (presença/ausência) do ácaro na planta. Tal informação é relevante do ponto de vista prático, pois sugere que, nas condições deste estudo, o simples registro da presença de ácaros pode ser suficiente, sem a necessidade de quantificação dos ácaros. Entretanto, antes de recomendações definitivas, deve-se verificar se tal resultado se repete em dados coletados em outros talhões.

A escolha do modelo mais adequado aos dados de leprose de citros foi guiada pelo valor do AIC e BIC, selecionando como melhor, o modelo com menor valor para ambos. A Tabela 9 apresenta os valores de AIC e BIC para cada modelo ajustado. O modelo mais adequado aos dados, entre os ajustados, foi o modelo *M4*, uma vez que o modelo *M1* é considerado um modelo base por refletir a

Tabela 8 - Estimativas e erros padrões dos parâmetros para os modelos ajustados aos dados de leprose dos citros

Modelos	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
M1	-2,040±0,023**	..	1,625±0,043**
M2	-2,552±0,029**	..	1,064±0,047**	0,616±0,033**	0,959±0,032**
M3	-2,545±0,029**	..	1,043±0,048**	0,745±0,031**	0,716±0,032**
M4	-2,243±0,026**	0,801±0,041**	1,500±0,044**
M5	-2,661±0,031**	0,531±0,044**	1,012±0,047**	0,571±0,034**	0,919±0,032**
M6	-2,659±0,032**	0,556±0,043**	0,993±0,047**	0,702±0,032**	0,676±0,032**
M7	-2,341±0,023**	0,022±0,002**	1,572±0,043**
M8	-2,577±0,029**	0,013±0,001**	1,045±0,047**	0,597±0,033**	0,944±0,032**
M9	-2,567±0,029**	0,013±0,002**	1,028±0,047**	0,727±0,031**	0,699±0,032**

** significativos a 1% de probabilidade.

Nota : Sinal convencional utilizado:

.. Não se aplica dado numérico.

persistência da doença em plantas infectadas e portanto não é considerado na seleção dos modelos.

Uma vantagem do modelo autológico é a possibilidade de quantificar a probabilidade de doença em uma particular planta dado o *status* das plantas vizinhas ou da mesma planta no espaço e/ou no tempo. No presente estudo a escolha do modelo *M4* indica que o *status* da mesma planta em uma avaliação anterior juntamente com a incidência do ácaro é o que melhor explica a ocorrência da doença. Através dos coeficientes do modelo pode-se calcular a probabilidade da planta apresentar leprose, dado que ela estava ou não doente na avaliação anterior e da planta ter ou não presença de ácaro. Os resultados desses cálculos mostram que, a probabilidade da planta apresentar a doença dado que não estava doente na avaliação anterior e apresenta incidência de ácaro é de 0,191. Por outro lado, a probabilidade da planta apresentar a doença dado que não estava doente e não apresenta incidência de ácaro é de 0,096. Logo, em um período de tempo de aproximadamente 22 dias, intervalo aproximado entre as avaliações, uma planta sadia tem aproximadamente o dobro de chance de adquirir a doença se estiver hospedando o ácaro vetor.

Embora o modelo *M4* seja indicado pelo AIC e pelo BIC como de melhor ajuste dentre os modelos *M2* a *M9*, isso não exclui a importância e a possibilidade de extrair resultados práticos dos demais. A significância dos demais efeitos mostra a relevância do padrão espacial da doença nas plantas

Tabela 9 - Valores de AIC para os modelos ajustados aos dados de leprose dos citros

Modelos	AIC	BIC
M1	18052,37	18068,39
M2	33006,94	33038,96
M3	33419,41	33451,43
M4	26516,68	26540,70
M5	40868,39	40908,42
M6	41343,40	41383,44
M7	26828,33	26852,35
M8	41104,43	41144,47
M9	41629,12	41669,15

vizinhas bem como da presença de ácaros.

Tabela 10 - Probabilidades da planta apresentar leprose dado que estava sadia no tempo anterior

Vizinhas doentes dentro da linha	Vizinhas doentes entre linhas		
	0	1	2
0	0,073	0,126	0,212
1	0,169	0,274	0,411
2	0,347	0,496	0,645

Por exemplo, se o interesse for apenas de observar as covariáveis de vizinhança, o modelo mais adequado indicado pelo AIC e BIC será o *M2*, modelo com as covariáveis de vizinhança defasadas no tempo, $(t - 1)$. Com os valores dos coeficientes do modelo pode-se calcular a probabilidade estimada da planta apresentar leprose dado que ela estava sadia na avaliação anterior e do *status* das plantas vizinhas dentro da linha e entre linhas na avaliação anterior. Os resultados estão detalhados na Tabela 10. Observa-se que uma planta sadia na avaliação anterior tem probabilidade de 0,274 de apresentar leprose se uma vizinha na linha e outra vizinha entre as linhas de planta estivesse doente na avaliação anterior. Observa-se também que quando há duas plantas vizinhas dentro da linha doentes a probabilidade da planta apresentar doença é maior, se comparado com duas plantas doentes entre as

linhas, 0,347 e 0,212, respectivamente. Isso pode estar refletindo o espaçamento entre as plantas, que nesse caso é menor dentro da linha de plantas.

Portanto a escolha do modelo neste contexto não deve ser automática e guiada simplesmente por critérios como o AIC e o BIC, mas também pelas hipóteses fitopatológicas a serem avaliadas combinadas à estrutura de dados disponíveis.

3 CONCLUSÕES

O modelo autologístico descreve de forma explícita o padrão espacial de doenças em plantas, detectando a existência de padrões e quantificando o efeito da presença da doença em diferentes estruturas de vizinhança na probabilidade de uma planta tornar-se doente. Encontrar a probabilidade da planta estar doente dado o *status* das plantas na vizinhança contribui na compreensão da dinâmica da doença e pode auxiliar a definição de práticas de controle. Para isso é necessário propor, comparar, escolher e interpretar, à luz dos dados, modelos adequados.

Os dados de incidência de leprose dos citros no talhão estudado apresentaram dependência espacial entre as plantas. Os modelos considerados possuem a característica de tratar conjuntamente dados provenientes de diversas avaliações ao longo do tempo e considerar vizinhanças no espaço e no tempo, bem como a relação entre a presença do vetor e a probabilidade de ocorrência da doença. O estudo mostrou que quando a planta tem incidência de ácaro, ela tem maior chance de apresentar sintomas da doença, além de existir a dispersão entre plantas vizinhas, especialmente dentro das linhas onde o espaçamento é usualmente menor.

Os pacotes estatísticos R*citrus* e *stLattice* são flexíveis para trabalhar com o modelo autologístico, possibilitam combinar diferentes estruturas de vizinhança e tempos de avaliação. São pacotes ágeis para manipulação de grandes conjuntos de dados e eficientes para disponibilizar as análises estatísticas dos dados incorporando algoritmos que envolvem métodos computacionalmente intensivos.

A metodologia apresentada aqui não é restrita para análise de doenças de citros, podendo ser usada na avaliação de padrões espaço-temporais e efeitos de fatores que afetem doenças de plantas em condições análogas de configuração espacial e disponibilidade de dados. Uma proposta para continuidade deste estudo é de combinar dados de incidência de doenças cítricas de diferentes áreas. Isso permitirá uma investigação dos efeitos de diferentes espaçamento entre as plantas, idade das plantas, variedade de citros, os efeitos sazonais e de outros fatores que possam variar entre os talhões de plantas. Do ponto de vista da modelagem, pode-se pensar em flexibilizar o modelo permitindo, ainda numa análise conjunta, que os parâmetros variem ao longo do tempo.

REFERÊNCIAS

- AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In. INTERNACIONAL SYMPOSIUM OF INFORMATION THEORY 2., 1973, Budapest. **Proceedings...** Budapest: Akademiai Kiadó, 1973. p. 267-281.
- ARNOLD, B.C.; STRAUSS, D. Pseudolikelihood estimation: some examples. **Sankhya: The Indian Journal of Statistics**, Series B, Kolkata, v. 53, p. 233-243, 1991.
- BASSANEZI, R.B.; LARANJEIRA, F.F. Spatial patterns of leprosis and its mite vector in commercial citros groves in Brazil. **Plant Pathology**, Saint Paul, v. 56, p. 97-106, 2007.
- BASSANEZI, R.B.; FERNANDES, N.G.; YAMAMOTO, P.T. **Morte súbita do citros**. Araraquara: Fundecitros, São Paulo, 2003. 62 p.
- BASSANEZI, R.B.; SPÓSITO, M.B.; YAMAMOTO, P.T. Adeus à leprose. **Revista Cultivar Hortaliças e Frutas**, Pelotas, v. 10, p. 6-8, 2001.
- BASSANEZI, R.B.; BERGAMIN FILHO, A.; AMORIM, L.; GIMENES-FERNANDES, N.; GOTTWALD, T.R.; BOVÉ, J.M. Spatial and temporal analyses of citrus sudden death as a toll to generate hypotheses concerning its etiology. **Phytopathology**, Saint Paul, v. 93, n. 4, p. 502-512, 2003.
- BASTIANEL, M.; FREITAS-ASTUA, J.; KITAJIMA, E.W.; MACHADO, M.A. The citrus leprosis pathosystem. **Summa Phytopathologica**, Joboticabal, v. 32, p. 211-220, 2006.
- BESAG, J. Nearest-neighbour systems and the auto-logistic model for binary data. **Journal of the Royal Statistical Society**, Series B, London, v. 34, p. 75-83, 1972.
- _____. Spatial interaction and the statistical analysis of lattice systems. **Journal of the Royal Statistical Society**, Series B, London, n. 2, v. 23, p. 192-236, 1974.
- _____. Statistical analysis of non-lattice data. **The Statistician**, Alexandria, v. 24, n. 3, p. 179-195, 1975.
- BORGATTO, A.F. **Modelos para proporções com superdispersão e excessos de zeros: um procedimento bayesiano**. 2004. 104 p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2004.
- CORDEIRO, G.M.; DEMÉTRIO, C.G.B. Modelos lineares generalizados. In. SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 12., 2007, Santa Maria. **Minicurso**. Santa Maria: UFSM, 2007. 161 p.
- CRESSIE, N.A.C. **Statistics for spatial data**. New York: John Wiley, 1993. 928 p.

CZERMAISNKI, A.B.C. **Dinâmica espaço-temporal de populações do patossistema leprose dos citros em condições naturais de epidemia**. 2006. 91 p. Tese (Doutorado em Fitopatologia) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2006.

DEMÉTRIO, C.G.B. **Modelos lineares generalizados na experimentação agrônômica**. Piracicaba: ESALQ, Departamento de Ciências Exatas, 2001. 113 p.

FERNÁNDEZ-DURÁN, J.J.; POIRÉ, A.; ROJAS-NANDAYAPA, L. Spatial and temporal effects in Mexican direct elections for the chamber of deputies. **Political Geography**, Kidlington, v. 23, p. 529-548, 2004.

FIRTH, D. Bias reduction of maximum likelihood estimates. **Biometrika**, Cambridge, v. 80, n. 1, p. 27-38, 1993.

FRIEL, N.; PETTITT, A. N. Likelihood estimation and inference for the autologistic model. **Journal of Computational and Graphical Statistics**, Alexandria, v. 13, n. 1, p. 232-246, 2004.

FUNDECITRUS. **Manual da leprose dos citros**. Araraquara, 2006. 12 p. Disponível em: <<http://www.fundecitrus.com.br/manuais.html>>. Acesso em: 17 jan. 2008.

GELMAN, A.; MENG, X.L. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. **Statistical Science**, Beachwood, v. 13, n. 2, p. 163-185, 1998.

GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Washington, v. 6, p. 721-741, 1984.

GEYER, C.J.; THOMPSON, E.A. Constrained Monte Carlo maximum likelihood for dependent data. **Journal of the Royal Statistics Society. Series B**, London, v. 54, p. 657-699, 1992.

GIRARDI, E.A. **Métodos alternativos de produção de mudas cítricas em recipientes na prevenção da morte súbita em citros**. 2005. 89 p. Dissertação (Mestrado em Fitotecnia) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2005.

GU, M.G.; ZHU, H.T. Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation. **Journal of the Royal Statistics Society, Series B**, London, v. 63, p. 339-355, 2001.

GUIRADO, N. Defensivos naturais controlam a leprose dos citros. **O Agrônomo**, Campinas, v. 52, p. 11-12, 2000.

GUMPERTZ, M.L.; GRAHAM, J.M.; RISTAINO, J. B. Autologistic model of spatial pattern of phytophthora epidemic in bell pepper: effects of soil variables on disease presence. **Journal of Agricultural, Biological and Environmental Statistics**, Alexandria, v. 2, n. 2, p. 131-156, 1997.

GUYON, X. **Random fields on a network: modelling, statistics and applications**. New York: Springer, 1995. 255 p.

HE, F.; ZHOU, J.; ZHU, H. Autologistic regression model for the distribution of vegetation. **Journal of Agricultural, Biological, and Environmental Statistics**, Alexandria, v. 8, n. 3, p. 205-222, 2003.

HUANG, F.; OGATA, Y. Generalized pseudo-likelihood estimates for Markov random fields on lattice. **Annals of the Institute of Statistical Mathematics**, Tokyo, v. 54, n. 1, p. 1-18, 2002.

HUBBELL, S.P.; CONDIT, R.; FOSTER, R.B. Local neighborhood effects on long-term survival of individual trees in a neotropical forest. **Ecological Research**, New York, n. 16, p. 859-875, 2001.

HUFFER, F.W.; WU, H. Markov chain Monte Carlo for autologistic regression models with application to the distribution of plants species. **Biometrics**, Washington, v. 54, n. 2, p. 509-524, 1998.

JENSEN, J.L.; MOLLER, J. Pseudolikelihood for exponential family models of spatial point processes. **Annals of Applied Probability**, New York, v. 1, p. 445-461, 1991.

JESUS JR., W.C.; BASSANEZI, R.B. Análise da dinâmica e estrutura de focos da morte súbita dos citros. **Fitopatologia Brasileira**, Brasília, v. 29, n. 4, p. 399-405, 2004.

KRAINSKI, E.T.; RIBEIRO JR., P.J. **Introdução ao Rcitrus**. 2007. Disponível em: <<http://www.leg.ufpr.br/elias/Rcitrus/>>. Acesso em: 17 abr. 2008.

———. **Um pacote para modelos autologísticos**. 2008. Disponível em: <<http://www.leg.ufpr.br/elias/stLattice/>>. Acesso em: 15 abr. 2008.

KRAINSKI, E.T.; RIBEIRO JR., P.J.; ANDRADE NETO, P.R.; BASSANEZI, R.B. Um ambiente para monitoramento da morte súbita dos citros. In. SIMPÓSIO BRASILEIRO DE GEOINFORMÁTICA, 7., 2005, Campos do Jordão, **Anais...** Campos do Jordão: INPE, 2005. p. 223-235.

KRAINSKI, E.T.; RIBEIRO JR., P.J.; BASSANEZI, R.B.; FRANCISCON, L. Autologistic model with an application to the citrus sudden death disease. **Scientia Agricola**, Piracicaba, 2008. In press.

LIMA, R.R.; DEMETRIO, C.G.B.; RIBEIRO JR, P.J.; RIDOUT, M. Uma comparação de técnicas baseadas em quadrats para caracterização de padrões espaciais em doenças de plantas. **Revista Brasileira de Biometria**, São Paulo, v. 24, p. 7-26, 2006.

MAANEM, A; XU, X.M. Modelling plant disease epidemics. **European Journal of Plant Pathology**, Dordrecht, v. 109, n. 7, p. 669-682, 2003.

MACCHERONI, W. Sem segredos. **Pesquisa FAPESP**, São Paulo, n.109, 2005. Disponível em: <<http://www.revistapesquisa.fapesp.br/index.php?art=2704&bd=1&pg=1&lg=>>. Acesso em: 27 jan. 2008.

MACCHERONI, W.; ALEGRIA, M.C; GREGGIO, C.; PIAZZA, J.P.; KAMLA, R.F; ZACARIAS, P.R.A; BAR-JOSEPH, M.; KITAJIMA, E.W.; ASSUMPCÃO, L.C; CAMAROTTE, G.; CARDOZO, J.; CASAGRANDE, E.C.; FERRARI, F.; FRANCO, S.F.; GIACHETTO, P.F; et al. identification and genomic characterization of a new virus (*Tymoviridae* Family) associated with citrus sudden death disease. **Journal of Virology**, New York, v. 79, n. 5, p. 3028-3037, 2005.

MAIA, O.M.A.; OLIVEIRA, C.A.L. Transmissibilidade do vírus da leprose de cercas-vivas , quebra-ventos e plantas daninhas para laranjeiras através de *Brevipalpus phoenicis*(Geijskes). **Bragantia**, Campinas, v. 64, n. 3, p. 417-422, 2005.

MARQUES, J.P.R.; FREITAS-ASTÚA J.; KITAJIMA, E.W.; APPEZZATO-DA-GLÓRIA, B. Lesões foliares e de ramos de laranjeira-doce causadas pela leprose-dos-citros. **Pesquisa Agropecuária Brasileira**, Brasília, v. 42, p. 1531-1536, 2007.

McCULLAGH, P.; NELDER, J.A. **Generalized linear models**. 2nd ed. London: Chapman and Hall, 1989. 511 p.

NELDER, J.A.; WEDDERBURN, R.W.M. Generalized linear models. **Journal of the Royal Statistical Society**, Series A, London, v. 135, p. 370-384, 1972.

PREISLER, H.K. Spatial patterns of trees attacked by beetles: pseudolikelihood estimation and iterative simulations, **Proceedings Computer Science and Statistics**, Washington, v. 23, p. 494-494, 1991.

_____. Modelling spatial patterns of trees attacked by bark beetles. **Applied Statistics**, Berkeley, v. 42, p. 501-514, 1993.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna. Disponível em: <<http://www.R-project.org>>. Acesso em: 27 maio 2008.

RIBEIRO JR., P.J.; VIOLA, D.N.; DEMETRIO, C.G.B; MANLY, B.F. Spatial pattern detection modeling of onion thrips (*Thrips tabaci*) on onion fields. **Scientia Agricola**, Piracicaba, 2008. In press.

RODRIGUES, J.C.V. **Relações patógeno-vetor-planta no sistema leprose dos citros**. 2000. 168 p. Tese (Doutorado em Energia Nuclear na Agricultura) - Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2000.

SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, Philadelphia, v. 6, n. 2, p. 461-464, 1978.

SEYMOUR, L.; JI, C. Approximate Bayes model selection procedures for GibbsMarkov random fields. **Journal of Statistical Planning and Inference**, Amsterdam, v. 51, p. 75-97, 1996.

SHERMAN, M.; APANASOVICH, T. V.; CARROL, R. J. On estimation in binary autologistic spatial models. **Journal of Statistical Computation and Simulation**, Philadelphia, v. 76, n. 2, p. 167-179, 2006.

SPÓSITO, M. B.; AMORIM, L.; RIBEIRO JR, P. J.; BASSANEZI, R. B.; KRAINSKI, E.T. Spatial pattern of trees affected by black spot in citrus groves in Brazil. **Plant Disease**, Saint Paul, v. 91, p. 36-40, 2007.

STRAUSS, D.; IKEDA, M. Pseudolikelihood estimation for social networks. **Journal of the American Statistical Association**, Alexandria, v. 85, p. 204-212, 1990.

VIOLA, D.N. **Detecção e modelagem de padrão espacial em dados binários e de contagem**. 2007. 119 p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2007.

VIOLA, D.N.; DEMÉTRIO, C.G.B.; RIBEIRO, J.R.; MANLY, B.F. Uma avaliação do estimador de pseudo-verossimilhança para modelos autolísticos espaciais, **Revista Brasileira de Biometria**, São Paulo, v. 26, p. 61-68, 2008.

WU, H.; HUFFER, F.W. Modelling the distribution of plant species using the autologistic regression model, **Environmental and Ecological Statistics**, London, v. 4, p. 49-64, 1997.

ZHENG, Y.; ZHU, J. Markov chain Monte Carlo for a spatial-temporal autologistic regression model. **Journal of Computational & Graphical Statistics**, Alexandria, v. 17, n. 1, p. 123-137, 2008.

ZHU, J.; HUANG, H.C.; WU, J.; Modeling spatial-temporal binary data using Markov random fields. **Journal of Agricultural, Biological, and Environmental Statistics**, Alexandria, v. 10, n. 2, p. 212-225, 2005.

APÊNDICES

APÊNDICE A - Códigos em R para a obtenção da análise dos dados

Após inicializar o R, carregue o pacote `Rcitrus` e o pacote `stLattice` com o comando `library` (ou `require`).

```
>require(Rcitrus)
```

```
>require(stLattice)
```

Na função `autologistic.citrus` do pacote `Rcitrus` está implementado o modelo autologístico, com os seguintes argumentos:

```
> args(autologistic.citrus)
```

```
function (formula = Y ~ R + C + dA + dB, bor = 1, obj, size = NULL,
  obj2 = NULL, covariate = NULL, death = 1, healt = 0,
  inf.method = c("pseudo",
  "mc", "bootstrap"), N, fit.method = c("glm", "brlr"),
  binary.covar = FALSE, verbose = FALSE)
```

```
#####
#####
### DADOS DE MORTE SUBITA DOS CITROS #####
```

A função `autologistic.citrus` será usada para o ajuste dos dados de MSC. Define-se a entrada dos dados em um `data.frame` ou matriz, o passo seguinte é o ajuste do modelo autologístico.

```
>data(vv202)
```

Observe que para o ajuste dos modelos foram implementados diversos argumentos na função. O primeiro deles é `formula`, aqui define-se a estrutura de vizinhança, observe o modelo `m1`, `Y` é a resposta, que está em função das covariáveis de vizinhança dentro da linha de plantas (`R2`) e entre linhas (`C2`) no tempo anterior a avaliação. O modelo `m2` acrescenta as diagonais (`dA2` e `dB2`) ao modelo anterior. Os modelos `m3` e `m4` ajustam as covariáveis no tempo contemporâneo a avaliação.

O próximo argumento dos modelos é o conjunto de dados, `obj`. O `obj2` é utilizado para construção das covariáveis, quando estão defasadas. A estimação dos parâmetros por reamostragem é feita através do argumento `inf.method=bootstrap` e `N=199` para o número de simulações da reamostragem.

```
###modelos
```

```
>fit <- list()
```

```

fit$m1 <- lapply(avals, function(i)
  autologistic.citrus(Y~R2+C2, obj=vvi[, ,i], obj2=vvi[, ,i-1],
    inf.method="bootstrap", N=199))

>fit$m2 <- lapply(avals, function(i)
  autologistic.citrus(Y~R2+C2+dA2+dB2, obj=vvi[, ,i], obj2=vvi[, ,i-1],
    inf.method="bootstrap", N=199))

>fit$m3 <- lapply(avals, function(i)
  autologistic.citrus(Y~R+C, obj=vvi[, ,i],
    inf.method="bootstrap", N=199))

>fit$m4 <- lapply(avals, function(i)
  autologistic.citrus(Y~R+C+dA+dB, obj=vvi[, ,i],
    inf.method="bootstrap", N=199))

```

Para observar o sumário de algumas avaliações de alguns modelos usam-se os seguintes comandos:

```

>summary(fit$m1$av2)
>summary(fit$m3$av3)
>summary(fit$m2$av4)

```

```

#####
###modelos usando o pacote stLattice

```

Quando faz-se uso do pacote `stLattice` para ajuste dos dados de MSC, a função `aglm` tem os seguintes argumentos implementados:

```

> args(aglm)

function (formula, data, sp = ~first, family = binomial, brlr = TRUE,
  bootstrap = TRUE, nboot = 99, mctest = FALSE, nmc = 99, bor = 1,
  lag = -1, times = 30, only.naive = FALSE)

```

Os dados,

```

>data(vv202,package="Rcitrus")

```

O formato de `vv202` é um *array* com três dimensões. A primeira dimensão indica o número de linhas de plantas, a segunda indica o número de plantas na linha e a terceira indica o número de avaliações.

Para criar as coordenadas na classe `stLattice`,

```

> cvv <- new("rlcoords", x = 1:nrow(vv202), y = 1:ncol(vv202))

```

O novo conjunto de dados da classe stLattice,

```
> dvv <- new("strldata", value = array(ifelse(vv202, 1, 0),
  c(dim(vv202), 1), dimnames = list(NULL, NULL, NULL,
  "resp")))

> vv202str <- new("mstr1", coords=cvv,dates=as.ISOdate(as.numeric
  (dimnames( vv202)[[3]]), "numeric"), data=dvv)
```

Pronto o conjunto de dados, ajustam-se os quatro modelos aos dados utilizando-se os seguintes argumentos: *formula* que é a parte do modelo de regressão logística usual, como não há covariáveis usuais usa-se 1, *data* para o conjunto de dados, *sp* para a parte espacial do modelo e *lag* para o tempo em que as covariáveis foram medidas, *lag=1* indica covariável no tempo anterior, *lag=0* indica covariável no tempo contemporâneo.

```
>m <- list()

>m$avv1 <- aglm(formula=resp~1, data=vv202str, sp=~R+C, lag=-1)
>summary(m$avv1)

>m$avv2 <- aglm(formula=resp~1, data=vv202str, sp=~R+C+dA+dB, lag=-1)
>summary(m$avv2)

>m$avv3 <- aglm(formula=resp~1, data=vv202str, sp=~R+C, lag=0)
>summary(m$avv3)

>m$avv4 <- aglm(formula=resp~1, data=vv202str, sp=~R+C+dA+dB, lag=0)
>summary(m$avv4)
```

```
#####
#####
###DADOS DE LEPROSE DOS CITROS#####
```

Aos dados de leprose dos citros são ajustados modelos autológicos espaço-temporais utilizando o pacote stLattice.

Lendo o arquivo de dados,

```
>popVal <- read.table('ValenciaR.txt', header=TRUE)
```

Converter os dados para o formato da classe stLattice,

```
>id.av <- which(as.numeric(substring(as.character(popVal$mapa), 5))>22)

>print(names(popVal)[c(7,9,13,15,17)], quote=F)
```

```
>ilep <- df2mstr1(popVal[id.av, c(4,5,2,19,10,11)])
>dim(ilep@data@value)
>str(ilep)
```

Agora serão ajustados os nove modelos aos dados.

```
>dimnames(ilep@data@value)[[4]]
>dftime <- as.numeric(mean(diff(ilep@dates)))
>m <- list()
```

Para ajuste dos modelos são usados os argumentos: *formula* para o modelo linear usual, *data* para o conjunto dos dados, *sp* para as covariáveis espaciais, os *lags* e *times* para o peso do valor da diferença entre os diferentes períodos de tempo.

```
>m$m1 <- aglm(formula=inplsin~1, data=ilep, sp=~S, times=dftime)
>summary(m$m1);AIC(m$m1)

>m$m2 <- aglm(formula=inplsin~1, data=ilep, sp=~S+R+C, lag=c(-1,-1,-1),
              times=dftime)
>summary(m$m2);AIC(m$m2);logLik(m$m2)

>m$m3 <- aglm(inplsin~1, data=ilep, sp=~S+R+C, lag=c(-1,0,0), times=dftime)
>summary(m$m3);AIC(m$m3);logLik(m$m3)

>m$m4 <- aglm(inplsin~inplaaca, data=ilep, sp=~S, lag=-1, times=dftime)
>summary(m$m4);AIC(m$m4)

>m$m5 <- aglm(inplsin~inplaaca, data=ilep, sp=~S+R+C, lag=-1, times=dftime)
>summary(m$m5);AIC(m$m5);logLik(m$m5)

>m$m6 <- aglm(inplsin~inplaaca, data=ilep, sp=~S+R+C, lag=c(-1,0,0),
              times=dftime)
>summary(m$m6);AIC(m$m6);logLik(m$m6)
```

```
>m7 <- aglm(inplsin~nacaplan, data=ilep, sp=~S, times=dftime)
>summary(m7);AIC(m7)

>m8 <- aglm(inplsin~nacaplan, data=ilep, sp=~S+R+C, times=dftime)
>summary(m8);AIC(m8);logLik(m8)

>m9 <- aglm(inplsin~nacaplan, data=ilep, sp=~S+R+C, lag=c(-1,0,0),
            times=dftime)
>summary(m9);AIC(m9);logLik(m9)
```

APÊNDICE B - Códigos em R para a obtenção das análises dos dados de MSC e para leprose dos citros

```
#####
### DADOS DE MORTE SUBITA DOS CITROS #####

#####
###modelos usando o pacote Rcitrus

###dados

require(Rcitrus)
data(vv202)
vvi <- vv202
vvi[vv202>0] <- 1
apply(vvi, 3, mean)

###modelos

avals <- c(av2=2, av3=3,av4=4,av5=5,av6=6,av7=7,
           av8=8,av9=9,av10=10,av11=11)

fit <- list()

fit$m1 <- lapply(avals, function(i)
  autologistic.citrus(Y~R2+C2, obj=vvi[, ,i], obj2=vvi[, ,i-1],
    inf.method="bootstrap", N=199))

fit$m2 <- lapply(avals, function(i)
  autologistic.citrus(Y~R2+C2+dA2+dB2, obj=vvi[, ,i],
    obj2=vvi[, ,i-1], inf.method="bootstrap",
    N=199))

fit$m3 <- lapply(avals, function(i)
  autologistic.citrus(Y~R+C, obj=vvi[, ,i],
    inf.method="bootstrap",N=199))

fit$m4 <- lapply(avals, function(i)
  autologistic.citrus(Y~R+C+dA+dB, obj=vvi[, ,i],
    inf.method="bootstrap",N=199))

###coeficientes

col <- t(rbind(sapply(fit$m1, function(x) coef(x$ps))))

require(xtable)
```



```

co2 <- t(rbind(sapply(fit$m2, function(x) coef(x$ps))))
co2
xtable(round(co2,3))

xtable(round(t(rbind(sapply(fit$m3, function(x) summary(x)$coef[,1]))),4))

xtable(round(t(rbind(sapply(fit$m4, function(x) summary(x)$coef[,1]))),4))

xtable(round(t(rbind(sapply(fit$m5, function(x) summary(x)$coef[,1]))),4))
summary(fit$m5$av2)

### extraindo os erros

summary(fit$m1$av2)
summary(fit$m1$av3)
summary(fit$m1$av4)
summary(fit$m1$av5)
...

summary(fit$m2$av2)
summary(fit$m2$av3)
summary(fit$m2$av4)
summary(fit$m2$av5)
...

t(rbind(sapply(fit$m3, function(x) summary(x)$coef[,2])))

t(rbind(sapply(fit$m4, function(x) summary(x)$coef[,2])))

t(rbind(sapply(fit$m5, function(x) summary(x)$coef[,2])))

summary(fit$m5$av2)

### extraindo os valores-p

round(pv.res, 4)

t(rbind(sapply(fit$m3, function(x) summary(x)$coef[,4])))

t(rbind(sapply(fit$m4, function(x) summary(x)$coef[,4])))

t(rbind(sapply(fit$m5, function(x) summary(x)$coef[,4])))

summary(fit$m5$av2)

#####
###modelos usando o pacote stLattice

```

```

require(stLattice)

####dados

data(vv202,package="Rcitrus")

cvv <- new("rlcoords", x=1:nrow(vv202), y=1:ncol(vv202))

dvv <- new("strldata", value=array(ifelse(vv202,1,0), c(dim(vv202),1),
                                   dimnames=list(NULL, NULL, NULL, "resp")))

vv202str <- new("mstr1", coords=cvv,dates=as.ISOdate(as.numeric(dimnames(
  vv202)[[3]]), "numeric"), data=dvv)

##dump("vv202", "../data/vv202")

str(vv202str)

##modelos

m <- list()

m$avv1 <- aglm(resp~1,vv202str, ~R+C)
summary(m$avv1)

m$avv2 <- aglm(resp~1,vv202str, ~R+C+dA+dB)
summary(m$avv2)

m$avv3 <- aglm(resp~1,vv202str, ~R+C,lag=0)
summary(m$avv3)

m$avv4 <- aglm(resp~1,vv202str, ~R+C+dA+dB,lag=0)
summary(m$avv4)

#### tabelas

coefs <- sapply(m, function(x) summary(x)$fit[,1])

ep <- sapply(m, function(x) summary(x)$fit[,2])

pvals <- sapply(m, function(x) summary(x)$fit[,4])

aics <- sapply(m, AIC)

bics <- sapply(m, function(x) AIC(x, k=log(length(x$resid))))

coefs

ep

```

```

pvals

aics

bics

#####
### DADOS DE LEPROSE DOS CITROS #####

###Análise descritiva

uniConv <- function(dat, x, y, dates, find.form='array',
                    attrib=NULL) {
  obj <- tapply(dat, list(x, y, dates), unique)
  attributes(obj) <- c(attributes(obj)[1], attrib)
  dimnames(obj) <- list(Rows=unique(x), Cols=unique(y),
                       Evals=unique(dates))
  class(obj) <- c('citrus', 'array')
  if (find.form!='array')
    obj <- citrus.conv(obj, find.form=find.form)
  return(obj)
}

### Funcao para conversao de varias variaveis
multConv <- function(obj, data.col=6:19, coords.col=21:20,
                    dates.col=2, find.form='array',
                    attrib=list(Municipio=NULL,
                                Propriedade=NULL, Proprietario=NULL,
                                Talhao='T257', Variedade=NULL, PortaEnxerto=NULL,
                                EntreLinhas=7.5, DentroLinhas=3.8, Plantio=NULL,
                                Latitude=NULL, Longitude=NULL))
  lapply(obj[, data.col], uniConv, obj[, coords.col[1]],
         obj[, coords.col[2]], obj[, dates.col], find.form, attrib)

### lendo o arquivo de dados
popVal <- read.table('ValenciaR.txt', header=TRUE)
dim(popVal)
names(popVal)
summary(popVal)
popVal$x <- (popVal$x - min(popVal$x))
popVal$y <- (popVal$y - min(popVal$y))
max(popVal$x)
max(popVal$y)

aVal <- multConv(popVal, find.form='array')
names(aVal)

###Análise por quadrats
quad <- disp.quadrats(aVal$inplsin,2)

```

```

quad

quad <- disp.quadrats(aVal$inplsin,3)
quad

###Dsitância mínima média
md <- mmdist.test(aVal$inplsin, eval=1:dim(aVal$inplsin)[3])

md
summary(md)

#####
### arrumando datas
### colocando datas em formato ISOdate
dchar <- as.character(popVal$datalev)
ano <- as.numeric(substr(dchar, 7, 8))+2000
isodt <- ISOdate(ano, substr(dchar,4,5), substr(dchar,1,2))
popVal$datalev <- substr(as.character(isodt), 1, 10)
dts.avs <- unique(isodt)[23:45]

### convertendo avaliacoess 23:45 formato mstr1
### com as variaveis: inplacfr, inplacra, inplsifr, inplsira, inplsifo
id.av <- which(as.numeric(substring(as.character(popVal$mapa), 5))>22)
print(names(popVal)[c(7,9,13,15,17)], quote=F)

require(stLattice)
names(popVal)

ilep <- df2mstr1(popVal[id.av, c(4,5,2,19,10,11)])
dim(ilep@data@value)
str(ilep)

#####
### ajustando os modelos
dimnames(ilep@data@value)[[4]]

dftime <- as.numeric(mean(diff(ilep@dates)))

m <- list()

m$m1 <- aglm(formula=inplsin~1, data=ilep, sp=~S, times=dftime)
summary(m$m1); AIC(m$m1); deviance(m$m1); df.residual(m$m1)

m$m2 <- aglm(formula=inplsin~1, data=ilep, sp=~S+R+C, lag=c(-1,-1,-1),
             times=dftime)
summary(m$m2);AIC(m$m2);logLik(m$m2);deviance(m$m2); df.residual(m$m2)

m$m3 <- aglm(inplsin~1, data=ilep, sp=~S+R+C, lag=c(-1,0,0), times=dftime)
summary(m$m3);AIC(m$m3);logLik(m$m3);deviance(m$m3); df.residual(m$m3)

```

```

m$m4 <- aglm(inplsin~inplaaca, data=ilep, sp=~S, lag=-1, times=dftime)
summary(m$m4);AIC(m$m4);deviance(m$m4); df.residual(m$m4)

m$m5 <- aglm(inplsin~inplaaca, data=ilep, sp=~S+R+C, lag=-1, times=dftime)
summary(m$m5);AIC(m$m5);logLik(m$m5); deviance(m$m5); df.residual(m$m5)

m$m6 <- aglm(inplsin~inplaaca, data=ilep, sp=~S+R+C, lag=c(-1,0,0),
             times=dftime)
summary(m$m6);AIC(m$m6);logLik(m$m6); deviance(m$m6); df.residual(m$m6)

m$m7 <- aglm(inplsin~nacaplan, data=ilep, sp=~S, times=dftime)
summary(m$m7);AIC(m$m7); deviance(m$m7); df.residual(m$m7)

m$m8 <- aglm(inplsin~nacaplan, data=ilep, sp=~S+R+C, times=dftime)
summary(m$m8);AIC(m$m8);logLik(m$m8);deviance(m$m8); df.residual(m$m8)

m$m9 <- aglm(inplsin~nacaplan, data=ilep, sp=~S+R+C, lag=c(-1,0,0),
             times=dftime)
summary(m$m9);AIC(m$m9);logLik(m$m9); deviance(m$m9); df.residual(m$m9)

#### tabelas

coefs <- sapply(m, function(x) summary(x)$fit[,1])
ep <- sapply(m, function(x) summary(x)$fit[,2])
pvals <- sapply(m, function(x) summary(x)$fit[,4])

aics <- sapply(m, AIC)
bics <- sapply(m, function(x) AIC(x, k=log(length(x$resid))))

coefs
ep
pvals
aics
bics

```