

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE VETERINÁRIA

REVISÃO DE MÉTODOS DE ANÁLISE ESPACIAL DE DADOS APLICADOS EM
EPIDEMIOLOGIA VETERINÁRIA

Autora: Ana Paula Serafini Poeta Silva

PORTO ALEGRE

2015/1

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE VETERINÁRIA**

**REVISÃO DE MÉTODOS DE ANÁLISE ESPACIAL DE DADOS APLICADOS EM
EPIDEMIOLOGIA VETERINÁRIA**

Autora: Ana Paula Serafini Poeta Silva

Orientador: Luís Gustavo Corbellini

Co-orientador: Mauro Riegert Borba

**Trabalho de conclusão de curso
apresentado à Faculdade de Veterinária
como requisito parcial para a obtenção de
graduação em Medicina Veterinária**

PORTO ALEGRE

2015/1

AGRADECIMENTOS

Agradeço imensamente aos meus pais, Paulo Poeta Silva e Angela Maria Serafini, pelo amor e confiança. Ao meu pai, especialmente, um agradecimento especial pelo apoio nessa minha primeira monografia.

Ao meu orientador, Luís Gustavo Corbellini, pelos grandes ensinamentos e oportunidades durante esses três últimos anos de graduação.

Ao meu co-orientador, Mauro Riegert Borba, pela paciência e contribuição para essa monografia.

A todos os colegas do Laboratório de Epidemiologia Veterinária (EpiLab) pela troca de conhecimentos.

Às minhas amigas e amigos, que divertem a minha vida e me apoiam nas ideias mais absurdas. Um agradecimento especial para as minhas amigas Rochelle e Marina pelo convívio diário.

A duas pessoas essenciais na minha vida, Sueli e Elisa Rosa, pelo grande apoio nesse último ano.

Minha profunda gratidão a todos que estão comigo!

RESUMO

O ambiente é cada vez mais reconhecido como um fator influente na ocorrência das doenças infecciosas, dada a atual realidade na forma em que ocorre a produção dos animais domésticos. As criações estão mais populosas e com restrições de espaços, fatos que associados a falhas nas medidas de segurança sanitária dessas produções podem determinar a presença de diversas doenças. Para a epidemiologia veterinária contemporânea, a ocorrência de doenças infecciosas nos animais é determinada pela interação de três principais fatores: o agente infeccioso, o hospedeiro e o ambiente; também conhecida como tríade epidemiológica. Logo, entender a forma de distribuição das doenças no espaço é um elemento chave para a investigação epidemiológica. A epidemiologia espacial é uma subdisciplina da epidemiologia com o enfoque em descrever e explicar os padrões espaciais de ocorrência das enfermidades. Existem inúmeras ferramentas e técnicas para análise dos dados de episódios sanitários vinculados com o espaço, os quais vêm auxiliando as atuais formas de combate a diversas enfermidades. Além do mais, o surgimento dos Sistemas de Informação Geográfica (SIG) facilitou a coleta desses dados para análises e previsões, podendo ser utilizado por diferentes programas de informática, de maneira rápida e viável. O presente trabalho objetiva revisar os principais métodos de análise espacial de dados utilizados em estudos epidemiológicos na Medicina Veterinária. Além disso, a título de demonstração prática, o método espacial de interpolação por krigagem será aplicado para estimação da localização de javalis de vida livre no estado do Rio Grande do Sul.

Palavras-chave: epidemiologia espacial, estatística espacial, ambiente, mapeamento.

ABSTRACT

The environment is increasingly recognized as an influential factor in the occurrence of infectious diseases, given the current reality in the way is the production of livestock. The breedings are more populated and space constraints, facts associated with failures in health and safety measures of these productions can determine the presence of various diseases. For contemporary veterinary epidemiology, the occurrence of infectious diseases in animals is determined by the interaction of three main factors: the infectious agent, the host and the environment; also known as epidemiological triad. So understand the form of distribution of diseases in space is a key element for epidemiological research. Spatial epidemiology is a sub-discipline of epidemiology with a focus on describing and explaining the spatial patterns of occurrence of diseases. There are numerous tools and techniques for analysis of health data episodes linked with the space, which have been helping the current forms of fighting various diseases. Moreover, the emergence of Geographic Information Systems (GIS) facilitated the collection of such data for analysis and predictions that can be used by different computer programs quickly and viable way. This document aims to review the main spatial data analysis methods used in epidemiological studies in veterinary medicine. In addition, as a practical demonstration, the kriging interpolation will be applied to estimate the location of free-living wild boars in the state of Rio Grande do Sul.

Keywords: spatial epidemiology, spatial statistics, environment, mapping.

LISTA DE FIGURAS

Figura 1 - Mapa da localização das mortes devido à Cólera (pontos) e os locais de suprimento de água (cruzes) na cidade de Londres feito por John Snow em 1854. Fonte: Pfeiffer, 2012.	11
Figura 2 - Diagrama de uma análise espacial em epidemiologia.	15
Figura 3 - Exemplo de um conjunto de células no formato de <i>raster</i>	16
Figura 4 - Representação de um mapa com a rede de triângulos irregulares.	17
Figura 5: Demonstração da sobreposição de quatro camadas diferentes, com diferentes tipos de representação cartográfica dos SIGs.....	17
Figura 6 - Divisão do globo terrestre em meridianos e paralelos utilizados pelo sistema.	18
Figura 7 - Demonstração do sistema UTM.	19
Figura 8 - Mapa de visualização de pontos. Os pontos representam focos positivos para tuberculose bovina em abatedouros na Inglaterra durante os anos de 1985 a 1997. Destaque para a aglomeração de pontos no sul do país. Fonte: Pfeiffer, 2012.	21
Figura 9 - Exemplos de mapas coropléticos demonstrando a prevalência de tuberculose bovina (TB) em abatedouros na Inglaterra nos anos de 1985 a 1997. A prevalência de TB é expressa em número de abatedouros positivos a cada 100.000 abatedouros. O mapa (a) tem sua distribuição somente pelas fronteiras político-administrativas da Inglaterra. O mapa (b) tem a correção de um cartograma, onde as fronteiras foram modificadas de acordo com o tamanho da população de bovinos. Fonte: Pfeiffer, 2012.	23
Figura 10 - Mapas isopléticos elaborados a partir do método de Kernel. Os mapas demonstram a evolução da prevalência de BSE na Inglaterra durante os anos de 1987 a 1997. A prevalência de BSE é expressa pelo número de focos positivos a cada 100 propriedades. Fonte: Stevenson, 2000.	25

Figura 11 - Demonstração da página inicial do programa SatScan para utilização na análise espacial <i>scan</i> . Fonte: http://www.satscan.org/	31
Figura 12 - Mapa indicando as áreas com maior risco para HPAI (círculos vermelhos) na Europa. Os <i>clusters</i> foram elaborados a partir de dados de surtos de HPAI em aves silvestres detectados de 2005 a 2008 (pontos). Fonte: Iglesias, 2010.	32
Figura 13 - Gráfico de dispersão de Moran.	35
Figura 14 - Mapas com as probabilidades de surtos de HPAI no Vietnã, e estimadas para cada um dos três períodos de epidemia, com base em efeitos aleatórios regressão logística multivariável. Fonte: Pfeiffer, 2007.	36
Figura 15 - Exemplo de variograma e seus elementos. Fonte: Deutsch & Journel, 1998 (adaptado).	39
Figura 16 - Demonstração gráfica dos parâmetros necessários para construção do variograma e escolha do ângulo onde se tem o maior alcance. Fonte: Deutsch & Journel, 1998 (adaptado).	40
Figura 17 - Visualização da distribuição dos pontos para a construção dos semivariogramas no <i>software</i> SGeMS.	43
Figura 18 - Determinação dos parâmetros número de <i>lags</i> e direções angulares para construção dos semivariogramas.	44
Figura 19 - Semivariogramas construídos para verificação da dependência espacial.	44
Figura 20 - Mapa demonstrando o resultado da interpolação de pontos derivado da análise por krigagem. Pela escala de cor, quando mais escuro, maior o risco de presença de javali.	45
Figura 21 - Mapa com a localização dos <i>clusters</i> primário e secundário e suas abrangências.	46

LISTA DE ABREVIATURAS

BSE – Encefalopatia Espongiforme bovina

CDC – Centro de Controle de Doenças

GPS – Sistema de Posicionamento Global

HPAI – Influenza Aviária Alta Patogenicidade

IBAMA - Instituto de Meio Ambiente e dos Recursos Naturais Renováveis

LISA – Indicador local de agregação espacial

MAUP – Problema de unidade de área modificável

PSC – Peste Suína Clássica

RS – Rio Grande do Sul

SIG – Sistema de Informação Geográfica

TB – Tuberculose bovina

TIN - Rede Triangular Irregular

UTM - Universal Transverso de Mercator

SUMÁRIO

1. INTRODUÇÃO	10
2. REVISÃO BIBLIOGRÁFICA	12
2.1 Epidemiologia Espacial	12
2.1.1 Análise espacial	12
2.1.1.1 Sistemas de Informação Geográfica (SIG)	12
2.1.1.2 Geoprocessamento	13
2.2 Análise espacial de dados em epidemiologia espacial	14
2.2.1 Dados espaciais	15
2.2.1.1 Dados espaciais tipo “entidades”	15
2.2.2 Coleta de dados espaciais	17
2.2.3 Conceitos básicos para a análise espacial	19
2.2.3.1 Dependência espacial	19
2.2.3.2 Heterogeneidade espacial e isotropia	20
2.3 Visualização espacial	20
2.3.1 Visualização espacial de dados por padrão de pontos	20
2.3.2 Visualização espacial de dados agregados por áreas	22
2.3.3 Visualização espacial de dados contínuos	23
2.3.3.1 Kernel	24
2.3.3.2 Krigagem	25
2.4 Exploração espacial	26
2.4.1 Análise exploratória de pontos	26
2.4.1.1 Testes globais para análise de cluster em pontos	27
2.4.1.1.1 Cuzick e Edwards’ k-nearest neighbour test	27
2.4.1.1.2 Ripley’s K-function	28
2.4.1.2 Testes locais para análise de clusters em pontos	29
2.4.1.2.1 Kulldorff’s spatial scan statistic	30
2.4.1.3 Testes globais e locais para análise de clusters em áreas	32
2.4.1.3.1 I de Moran Global	33
2.4.1.3.2 LISA	34
2.5 Modelagem espacial	35
2.6 Softwares para análises espaciais	36
3. METODOLOGIA	38
3.1 Krigagem	38
3.2 Javali	40
3.3 Banco de dados	41
3.4 Metodologia estatística da análise estatística	42
4. RESULTADOS	43
5. DISCUSSÃO	47
6. CONCLUSÃO	49
REFERÊNCIAS BIBLIOGRÁFICAS	50

1. INTRODUÇÃO

A transmissão e o delineamento de uma doença infecciosa estão estritamente ligados ao espaço e a sua proximidade espaço-temporal. Há muito tempo atrás, estudos epidemiológicos trabalham com os fatores relacionados ao espaço, ao tempo e ao indivíduo (MacMahon & Pugh, 1970; Lilienfeld & Lilienfeld, 1980). No século 19, John Snow (1990) utilizou técnicas de mapeamento para analisar a grande epidemia de cólera ocorrida no ano de 1854 em Londres (Figura 1). O estudo vinculou as mortes por cólera com a localização do suprimento de água das diferentes bombas públicas de abastecimento, identificando assim a origem da epidemia, mesmo sem conhecer o agente etiológico, utilizando o fator espaço para a resolução dos casos. A análise espacial em eventos sanitários refere-se ao estudo qualitativo e quantitativo da distribuição de doenças ou fatores que podem determinar sua ocorrência, no qual o objeto principal do estudo é a sua localização geográfica e sua relação com o ambiente (Gesler, 1986). O campo da análise de dados espaciais envolve dados que representam o desfecho de um processo operando no espaço. Diversos métodos estatísticos são utilizados para descrever e explicar esses dados. Os avanços das técnicas de geoprocessamento e geostatística contribuíram para que as análises espaciais de dados se tornassem uma ferramenta indispensável para a epidemiologia e investigações epidemiológicas. Entretanto, a incorporação efetiva dessas técnicas na área da saúde animal ainda ocorre de forma muito pontual e esparsa. Com isso, perante essa necessidade, o presente trabalho de conclusão de curso em Medicina Veterinária elabora uma análise dos principais aspectos de análise espacial de dados e propõe uma revisão dos principais métodos de estatística espacial conhecidos e aplicados em estudos em saúde animal.



Figura 1: Mapa da localização das mortes devido à Cólera (pontos) e os locais de suprimento de água (cruzes) na cidade de Londres feito por John Snow em 1854. Fonte: Pfeiffer, 2012.

2. REVISÃO BIBLIOGRÁFICA

2.2 Epidemiologia Espacial

O termo Epidemiologia Espacial é utilizado para definir estudos nos quais a localização dos eventos é o principal componente a ser analisado (Saez & Saurina, 2007). É considerada uma subdisciplina da Epidemiologia, onde a distribuição espacial de eventos possibilita determinar padrões da atual situação sanitária de uma área, caracterizar a dinâmica e explicar padrões de ocorrência de uma enfermidade, formular hipóteses e identificar possíveis fatores de risco, além de conduzir e direcionar o delineamento de programas sanitários, sejam eles de controle, monitoramento ou ações de vigilância.

Na medicina humana, existem estudos de epidemiologia espacial antes do século XIX (Lawson & Williams, 2001). Já em medicina veterinária encontram-se poucos exemplos, na literatura, de estudos espaciais antes do anos 80. Tratam-se basicamente de trabalhos elaborados por parasitologistas que buscavam a relação do espaço e clima com vetores de algumas doenças. Um dos primeiros trabalhos foi conduzido na Tanzânia (Robson, 1961), no qual foi demonstrado que a febre da Costa Leste em bovinos estava confinada a áreas onde existiam moscas tsé-tsé.

2.1.1 Análise espacial

A análise espacial é o estudo de fenômenos que são vinculados ao espaço no campo da epidemiologia espacial, na qual são incluídas uma série de métodos estatísticos. Pode se dizer que a análise espacial faz a ligação entre um valor de natureza espacial e as diversas ferramentas aplicadas em estatística e modelagem matemática, permitindo assim, combinar variáveis georreferenciadas. A partir delas, criar e analisar novas variáveis de interesse (Rosa, 2011). Os objetivos da aplicação dessas análises em epidemiologia espacial são descrever os padrões espaciais, identificar agrupamentos geográficos e explicar os riscos para ocorrência dos eventos sanitários. Para isso, é fundamental o acesso a dados com informações claras e concisas. Atualmente, isso é feito através do uso de ferramentas de geoprocessamento, aos quais se associam os sistemas de informação geográfica (SIG), o processamento digital de imagens e a cartografia digital.

2.1.1.1 Sistemas de Informação Geográfica (SIG)

O Sistema de Informação Geográfica (SIG) é a ferramenta de coleta de dados do geoprocessamento. É caracterizado por um conjunto de ferramentas computacionais, composta por equipamentos e programas que realizam a integração dos dados coletados para posterior análise do geoprocessamento (Rosa, 2011). Esse sistema é composto pela entrada de

dados (*input*), processamento, saída (*output*) e retroalimentação (*feedback*). Logo, o SIG caracteriza a relação entre a tecnologia da informação com a formação e o gerenciamento de bancos de dados georreferenciados, e sua posterior visualização através da cartografia digital em *softwares* específicos para essa finalidade. Ainda, o SIG é uma ferramenta acessível, prática e economicamente viável, sendo assim, é a principal ferramenta usada no geoprocessamento e na coleta de dados espaciais.

2.1.1.2 Geoprocessamento

Os dados tratados pelo geoprocessamento incluem aqueles de natureza essencialmente geográfica, como a localização (pontos geográficos) ou uma determinada extensão de área e, também, aqueles sem natureza geográfica que podem ser representados cartograficamente. Um exemplo de dados sem natureza geográfica são os nomes dos proprietários e as características socioeconômicas de suas propriedades rurais. Esses dados podem ser georreferenciados por um endereço ou CEP, por exemplo.

A visualização ocorre por meio dos mapas temáticos, os quais são um veículo de comunicação visual valioso que demonstram fenômenos geográficos que estão ocorrendo naquele momento. Ainda, existe a possibilidade de modificar esses mapas de acordo com a intenção da pesquisa, proporcionando diversas características visuais distintas.

O uso das técnicas de geoprocessamento vem aumentando cada vez mais no planejamento, no monitoramento e nas avaliações das ações em saúde animal e pública (Barcellos, 2004). Especificamente, no campo da Epidemiologia, tanto animal quanto humana, o geoprocessamento faz parte da análise da dinâmica espacial das doenças em suas relações com o ambiente e espaço. Na avaliação da situação real da doença e na identificação de regiões de alto risco, é um instrumento importante de apoio às atividades de vigilância epidemiológica, planejamento de ações de prevenção e controle de enfermidades, especialmente no contexto atual, devido ao surgimento de doenças emergentes. Segundo SMOLINSKI ET AL 2003, 75% das doenças emergentes que surgiram nas últimas décadas são zoonoses, estreitando cada vez mais a relação entre a saúde humana e animal com o ambiente.

Além disso, com os recentes avanços na área da computação relacionados aos equipamentos de coleta de informações geográficas e *softwares* de análises de dados espaciais, foi possível intensificar a interação das grandes bases de dados espaciais com o desenvolvimento da epidemiologia espacial. Os métodos de captura, armazenamento e demonstração de dados geográficos (SIG) criaram uma demanda para novas ferramentas de análise de estatística espacial. Contudo, as técnicas de análise espacial foram desenvolvidas

paralelamente, mas de forma independente. Como resultado disso, *softwares* SIG ainda tem algumas funcionalidades limitadas para essas análises espaciais (Pfeiffer, 2004).

2.2 Análise espacial de dados em epidemiologia espacial

De acordo com os autores BAILEY & GATRELL (1995) a análise espacial dos dados em epidemiologia segue um padrão cronológico de trabalho e é dividida em três grupos de métodos de análise: visualização, exploração e modelagem matemática (Figura 2). Os dois primeiros grupos são métodos que se concentram exclusivamente em analisar a dimensão espacial dos dados.

A visualização é, provavelmente, o método de análise espacial mais utilizada, pois resulta em mapas que descrevem os padrões espaciais e tem a finalidade de estimular e apontar uma análise mais complexa.

A exploração de dados espaciais envolve o uso de métodos estatísticos quantitativos para determinar se os padrões observados são aleatórios no espaço.

Já a modelagem matemática introduz o conceito de relação causa-efeito usando os dados e seus respectivos atributos, tanto espaciais quanto não espaciais para explicar ou prever os padrões espaciais.

Entretanto, nenhuma destas abordagens permite uma inferência independente, ou seja, existe uma ligação e uma sobreposição entre os três grupos, particularmente entre a visualização e a exploração, uma vez que a apresentação visual pode exigir a utilização de métodos de análise quantitativos. Ainda, esse processo não segue uma linearidade restrita, a apresentação dos resultados de exploração e modelagem requer um retorno ou *feedback* aos métodos de visualização.

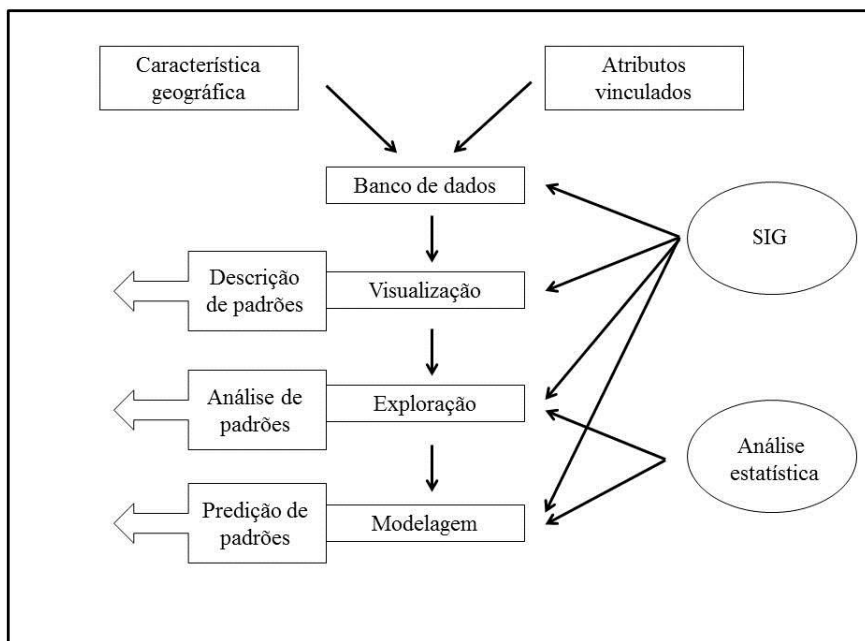


Figura 2: Diagrama de uma análise espacial em epidemiologia.

2.2.1 Dados espaciais

Os dados coletados para a investigação epidemiológica de um evento sanitário têm como objetivo a localização geográfica e o conhecimento dos atributos dessas observações. A presença de um referencial geográfico para cada observação permite em primeiro lugar, para a análise espacial, a determinação das relações da localização geográfica entre todas as observações coletadas e seus atributos e, em segundo lugar, os dados dos atributos adicionais podem ser obtidas através da ligação espacial com outros dados georreferenciados.

Existem vários modelos conceituais usados para representar um espaço geográfico específico em uma imagem cartográfica ou mapa. Os mais utilizados são os conceitos de “entidades” e “campos”. A primeira abordagem vê o espaço sendo ocupado por entidades com seus atributos específicos ou características intrínsecas. Sua posição pode ser mapeada através de um sistema de coordenadas geográficas. A segunda abordagem descreve a variação de determinado atributo no espaço, com uma função matemática vinculada e com certa continuidade no espaço. Por exemplo, as mensurações de precipitações ou temperaturas em uma determinada área.

2.2.1.1 Dados espaciais tipo “entidades”

Os dados espaciais do tipo entidades utilizados pelos SIGs são divididos em pontos, linhas e polígonos.

A representação de uma entidade pontual é feita por um par de coordenadas (X, Y). As linhas e os polígonos são um conjunto de pares de pontos. As linhas podem representar características lineares contínuas, como rodovias ou rios. Os polígonos são utilizados para definir áreas contínuas com uma característica comum, como a área de uma cidade ou a abrangência de uma floresta. Todas essas entidades possuem seus atributos, ou seja, possuem as informações vinculadas a essa entidade. Por exemplo, um ponto, além de determinar a localização de um rebanho infectado em um espaço, contém informações sobre a espécie animal afetada e a doença em questão.

Na linguagem espacial e cartográfica, os SIGs projetam essas entidades em mapas através de três maneiras: vetores, *rasters* e redes triangulares irregulares (TIN, do termo em inglês *triangulated irregular network*).

Os vetores são a união de pontos por retas que representam formas com características espaciais baseadas em um conjunto ordenado de coordenadas, como as fronteiras dos municípios, que são uma série de pontos unidos com linhas retas.

Conjuntos de dados *rasters* representam características geográficas dividindo o mundo em forma de células quadrangulares ou retangulares dispostas em uma grade (Figura 3). Cada célula tem um único valor de atributo que é utilizado para representar o fenômeno espacial presente, tal como a variação da temperatura ou a altitude, por exemplo. Esse atributo é vinculado a uma função matemática que determinar a variação desse atributo ao longo da célula. São utilizados para representar características de pontos, linhas e polígonos.

Redes triangulares irregulares (TIN) são usadas para representar superfícies em três dimensões, como a morfologia geográfica de uma área, com suas elevações e distribuição no espaço (Figura 4). TINs são uma forma de conjunto de dados geográficos baseados em vetores e são construídos por triangulação de um conjunto de pontos. Os pontos estão conectados com uma série de arestas ou retas para formar uma rede de triângulos.

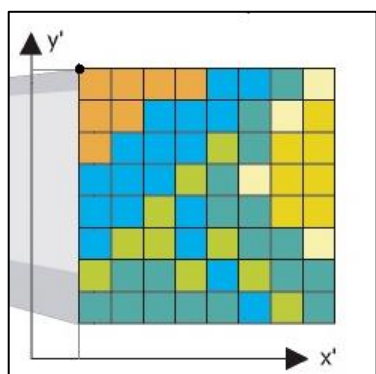


Figura 3: Exemplo de um conjunto de células no formato de *raster*.

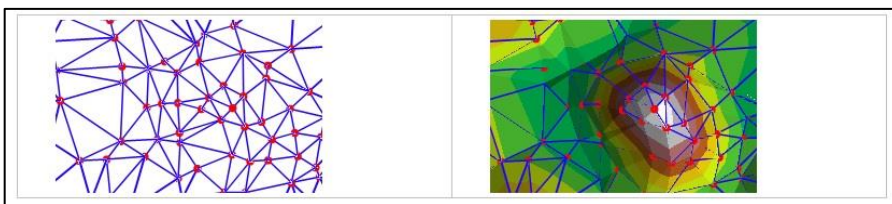


Figura 4: Representação de um mapa com a rede de triângulos irregulares.

Um conjunto de representações cartográficas representando uma área específica forma um banco de dados, o qual representa uma camada quando projetada espacialmente. Como consequência de uma união de diversas camadas distintas, uma nova região é elaborada e é vinculada a um novo banco de dados, que pode ser denominado *geodatabase*. A Figura 5 faz uma representação da união de camadas com tipos de representações diferentes formando uma nova região.

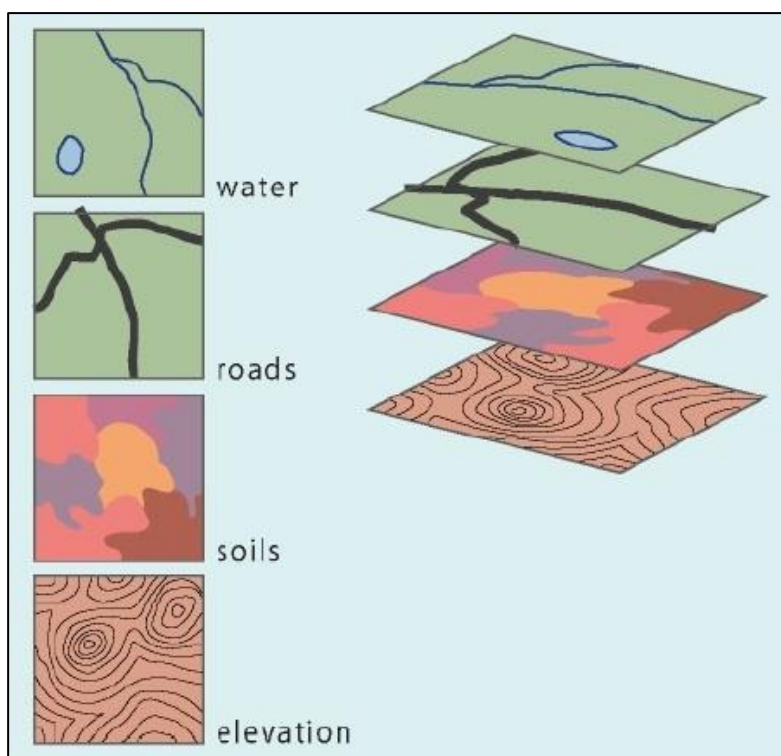


Figura 5: Demonstração da sobreposição de quatro camadas diferentes, com diferentes tipos de representação cartográficas dos SIGs.

2.2.2 Coleta de dados espaciais

A coleta de dados espaciais envolve a captura, verificação e uma estruturação do processo (Burrough & McDonnell, 1998). Atualmente, com a tecnologia de aparelhos móveis

como o GPS (*global positioning system*), que é um sistema de navegação regido por satélites, proporcionou a captura da informação da posição no espaço representado por uma coordenada geográfica. Para projetar esses dados coletados em um mapa é preciso definir o sistema de coordenadas geográficas, pois as camadas de dados que são processadas conjuntamente em um SIG, devem ser representadas pelo mesmo sistema de projeção cartográfica (Aronoff, 1989). Existem dois tipos principais de projeções: sistema de coordenadas geográficas e sistema de projeção Universal Transverso de Mercator (UTM).

O sistema de coordenadas geográfica expressa qualquer posição horizontal em uma superfície plana através de duas das três coordenadas existentes num sistema esférico de coordenadas, alinhadas com o eixo de rotação da Terra (Figura 6). Nesse caso, o globo está dividido em paralelos e meridianos e sua localização é representada por meio do sistema de latitude e longitude, as quais são orientadas conforme os pontos cardeais e referenciadas de acordo com o meridiano de Greenwich e a Linha do Equador. A longitude (eixo X) varia na direção Leste-Oeste, a latitude (eixo Y) varia na direção Norte-Sul e são medidas em graus.

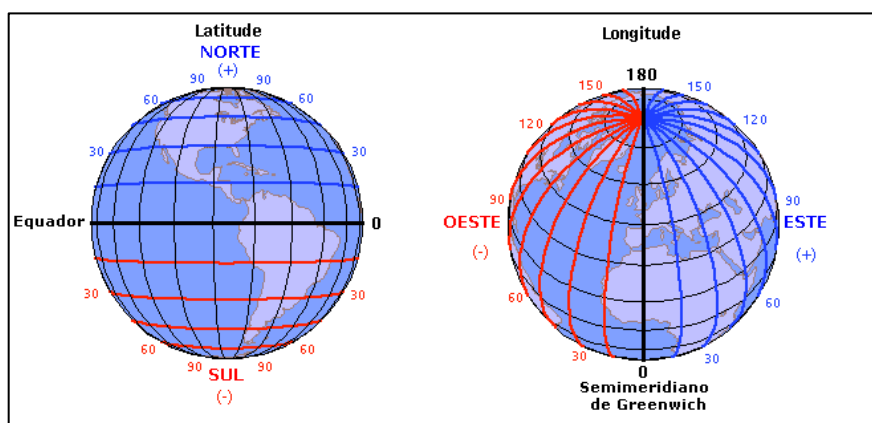


Figura 6: Divisão do globo terrestre em meridianos e paralelos utilizados pelo sistema de coordenadas geográfico.

Já no sistema de projeção UTM (Figura 7), o globo é dividido em 60 fusos de 6 graus de longitude (eixo X), numerados a partir do antimeridiano de Greenwich, seguindo de oeste para leste até o encontro com o ponto de origem. A extensão latitudinal está compreendida entre 80° Sul e 84° Norte. O eixo central do fuso, denominado como meridiano central estabelece, junto à linha do Equador, a origem do sistema de coordenadas de cada fuso. Cada fuso também é chamado de Zona UTM e a unidade de medida é o metro.

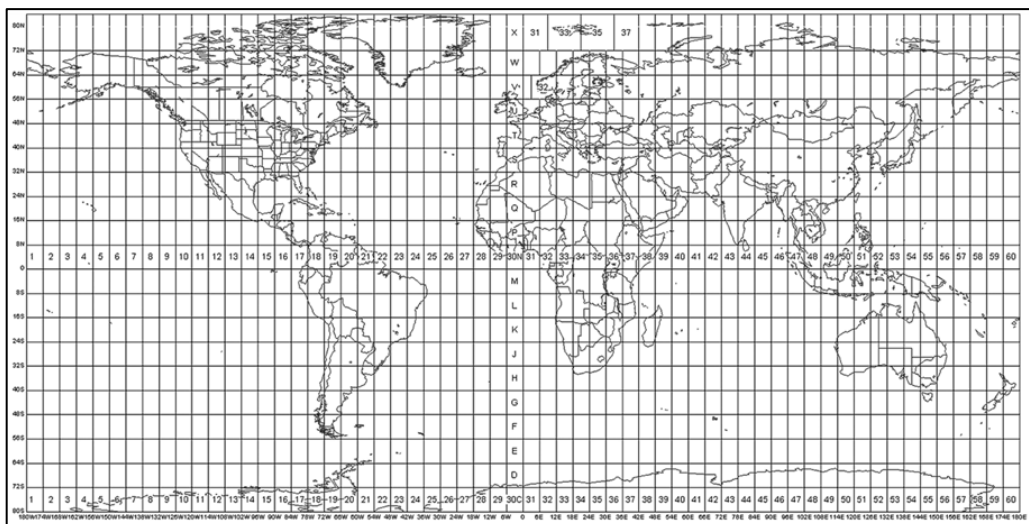


Figura 7: Demonstração do sistema UTM.

2.2.3 Conceitos básicos para a análise espacial

Antes de apresentar os principais conceitos e parâmetros para a análise espacial é necessário entender que a denominação “estatística espacial” surgiu em oposição às diversas técnicas da estatística nos estudos em saúde humana e animal especificamente (Santos & Souza, 2007). Isso porque um dos pressupostos mais básicos e gerais da estatística é o da independência entre as observações – o que se mede em um indivíduo não está associado ao que está sendo medido no outro indivíduo, exceto, quando existem possíveis fatores comuns que podemos estimar por meio dos diversos testes estatísticos existentes, como teste do qui-quadrado e regressão múltipla. Todavia, quando a localização espacial é uma característica da observação, esse pressuposto perde poder explicativo, pois segundo TOBLER (1979) “todas as coisas são parecidas, mas coisas mais próximas se parecem mais do que as coisas mais distantes”, portanto, a informação da localização das observações contribui para as análises. Uma potencial consequência de ignorar o efeito espacial nas análises estatísticas é subestimar os erros e sobrestimar a significância dos testes, com isso, aumentando os riscos de erros do Tipo 1 estarem enviesando os resultados (Pfeiffer, 2008). O erro Tipo 1 superestima o resultado, devido à maior proporção de falsos positivos.

2.2.3.1 Dependência espacial

Um conceito primordial na compreensão e análise correta de dados com natureza espacial é a dependência espacial (Druck, 2004). A dependência espacial surge sempre que o atributo de uma observação de algum lugar no espaço está relacionado com outro atributo ou em outros lugares no espaço. Esse fenômeno ocorre devido à similaridade das observações que estão próximas, isto é, como foi dito anteriormente, observações com maior proximidade

no espaço possuem uma correlação maior do que com as observações mais distantes (Krempi, 2004). A quantificação da dependência espacial é denominada autocorrelação. A preposição “auto” está indicando que a medida de relação de dependência é realizada em vários locais distintos do espaço com a mesma variável. Dois aspectos induzem a ocorrência da autocorrelação espacial: erros de medidas e fenômenos de interação espaciais externos, como por exemplo, o aspecto topológico do espaço.

2.2.3.2 Heterogeneidade espacial e isotropia

A heterogeneidade espacial surge devido ao fato que os dados utilizados para explicar determinados fenômenos espaciais são de unidades distintas na grande parte das vezes (Krempi, 2004). Essa heterogeneidade espacial está relacionada com a variação da dependência espacial através da distância. Caso a dependência espacial não se altere com a mudança da distância, o fenômeno é denominado estacionário, e o contrário, portanto, não-estacionário ou heterogêneo.

A isotropia acontece quando a dependência espacial de um processo estacionário é a mesma em todas as suas direções, entretanto, é dito anisotrópico, caso essa dependência espacial varie em direções diferentes.

As descrições e conceitos apresentados até aqui neste trabalho descreveram um conjunto básico de propriedades dos dados espaciais e devem ser levados em consideração para as futuras análises espaciais em estudos de epidemiologia humana e veterinária.

2.3 Visualização espacial

Como já foi salientado anteriormente nesse trabalho, o primeiro passo para qualquer análise e investigação epidemiológica é a visualização das características das observações coletadas (Bailey & Gatrell, 1995). O objetivo básico da visualização de dados espaciais é demonstrar os padrões e possíveis tendências de distribuição das observações por meio dos mapas temáticos. Existem três maneiras diferentes de representar os eventos em mapas de acordo com os métodos de visualização: representação de dados por padrão de pontos, dados agregados por áreas e dados contínuos.

2.3.1 Visualização espacial de dados por padrão de pontos

A visualização por meio de pontos é a maneira mais tradicional e mais utilizada pelos estudos epidemiológicos, como é o caso no estudo da cólera na Inglaterra conduzido por John Snow. Os pontos determinam uma localização geográfica representada por um par de coordenadas geográficas. Mapas de pontos são considerados a maneira mais simples de visualização de eventos sanitários com localizações conhecidas. No entanto, apresentam

problemas onde existem uma grande quantidade de eventos ou múltiplos eventos numa mesma localização. Em ambas as situações, o resultado são mapas apresentando aglomerações desses pontos, o que dificulta a sua visualização e identificação dos atributos das observações (Figura 8). Há a possibilidade de usar diferentes símbolos para diferenciar esses atributos, mas ainda assim, é prejudicado pelo grande número de dados. Portanto, os pontos são escolhidos quando existem poucos dados e com o objetivo simples de localização dos eventos (Stevenson, 2003). Como alternativa a essa formação de aglomerados, que resultam numa superfície praticamente contínua de pontos, existe os métodos de suavização de imagens cartográficas. Como a interpolação de dados da geoestatística (Lawson & Williams, 2001) e o método de *Kernel*. Esses métodos são caracterizados pela construção de mapas isopléticos. Isoplético quer dizer “mesmo valor”. O método isoplético é aplicável para fenômenos geográficos de natureza contínua, tais como: médias, razões, proporções e medidas de dispersão de algum atributo, atribuído à área envolvida. Além disso, mapas isopléticos são caracterizados por apontarem as direções e as intensidades que os valores dos atributos crescem ou decrescem ao longo da área. Na epidemiologia espacial, por exemplo, esse método é aplicado em mapas de densidade populacional.

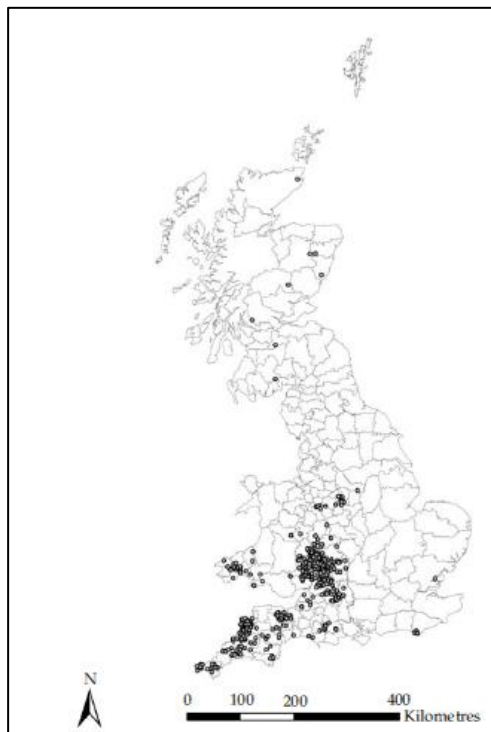


Figura 8: Mapa de visualização de pontos. Os pontos representam focos positivos para tuberculose bovina em abatedouros na Inglaterra durante os anos de 1985 a 1997. Destaque para a aglomeração de pontos no sul do país. Fonte: Pfeiffer, 2012.

2.3.2 Visualização espacial de dados agregados por áreas

A visualização de dados agregados por área geográfica é a mais utilizada para manipular dados de população. O processo de agregação de dados espaciais envolve a sumarização de um determinado atributo pertencente a um grupo de pontos, assim, produzindo um valor único que se refere a uma área específica. A sumarização pode ser resultado de uma soma, uma média ou uma proporção, por exemplo. As áreas que frequentemente delimitam a agregação de dados são as divisões político-administrativas, como um país, um estado ou um município.

Em estudos epidemiológicos, a representação do número de casos por área é a forma mais usual de dados agregados por áreas (Pfeiffer, 2012). A contagem desses casos pode ser expressa em função do tamanho da população, sob a forma de prevalência ou incidência por unidade de área. Os mapas coropléticos são a melhor forma de representação nesse caso, pois têm como finalidade representar em imagens os valores calculados na área delimitada. Contudo, quando se analisa mapas coropléticos é necessário levar em consideração pelo menos dois fatores (Pfeiffer & Jones, 2002):

- a) Devido à escolha de áreas delimitadas por razões políticas, isso pode ser um fator irrelevante ou ter um impacto direto sobre o espalhamento do evento sanitário.
- b) Áreas geograficamente muito extensas podem ter resultados superestimados, pois a população alvo não possui uma distribuição homogênea por toda a extensão da área demarcada. Esse fenômeno é chamado de MAUP (*modifiable area unit problem*) – “problema de unidade de área modificável”. Os efeitos do MAUP são inerentes aos dados coletados através de agregação por áreas, pois os pontos de coleta de dados são aleatórios e não podem ser diretamente associados à distribuição espacial do atributo analisado (Openshaw, 1984).

O cartograma é um método que corrige o problema derivado das agregações em extensas áreas (Dorling, 1995). O cartograma é um tipo de mapa que representa a densidade de um atributo ou variável escolhida sobre a área de sua ocorrência, ao invés da simples apresentação em uma determinada área (Dorling, 1995). Os contornos originais da área são deformados em função do peso desta variável. A conformidade com os limites originais da superfície são preservados ao máximo.

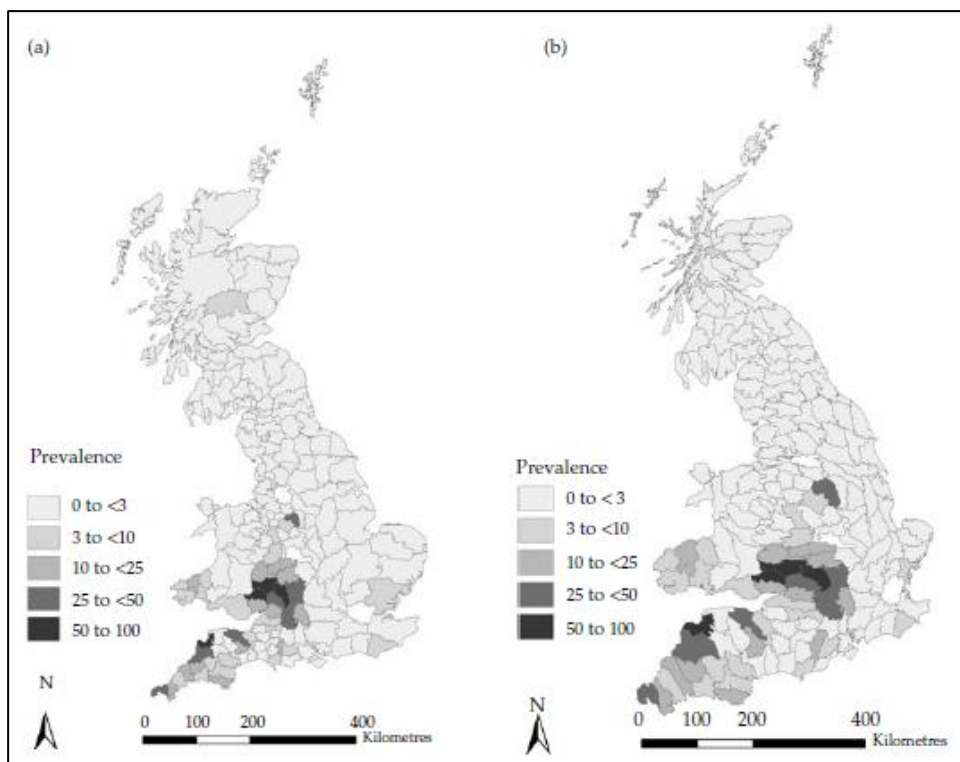


Figura 9: Exemplos de mapas coropléticos demonstrando a prevalência de tuberculose bovina (TB) em abatedouros na Inglaterra nos anos de 1985 a 1997. A prevalência de TB é expressa em número de abatedouros positivos a cada 100.000 abatedouros. O mapa (a) tem sua distribuição somente pelas fronteiras político-administrativas da Inglaterra. O mapa (b) tem a correção de um cartograma, onde as fronteiras foram modificadas de acordo com o tamanho da população de bovinos. Fonte: Pfeiffer, 2012.

2.3.3 Visualização espacial de dados contínuos

Dados de natureza contínua no espaço têm sua localização estimada ao longo de uma determinada área, como a distribuição das chuvas, humidade, poluição do ar ou concentração de minerais no solo. Nesse caso, o dado é uma medida obtida em um ponto amostrado, logo, apesar de ser atribuído ao ponto, o atributo do dado é referente à área amostrada, a qual é, geralmente, um espaço contínuo e sem delimitação explícita. Portanto, esse ponto possui uma coordenada geográfica e um atributo na forma de valor medido, sendo um dado contínuo ou discreto.

As técnicas mais usuais para a visualização desses tipos de dados são voltadas para a predição espacial e interpolação, pois os valores são estimados a partir de valores já mensurados (Berke, 2004). As principais técnicas utilizadas em estudos epidemiológicos são a krigagem e o método de Kernel.

2.3.3.1 Kernel

O método de Kernel pode ser usado para estimativa de densidade de dados categóricos e contínuos. É uma técnica de interpolação exploratória que gera uma superfície de visualização de “áreas quentes”. Uma área quente é uma concentração de eventos que pode indicar alguma forma de aglomeração espacial.

É um método analítico não paramétrico, em que uma distribuição de pontos ou eventos é transformada em uma superfície contínua. Uma das maiores vantagens é a rápida visualização de áreas que merecem atenção, além de não ser afetada por divisões político-administrativas. Entretanto, a análise por Kernel é estritamente subjetiva e depende do conhecimento prévio da área estudada.

A estimativa de densidade Kernel é uma função de distribuição de probabilidade bivariada aplicada para determinar a intensidade de um processo espacial (Bailey & Gatrell, 1995). Para aplicação é necessária a definição de dois parâmetros básicos, que podem ser visualizados pela fórmula:

$$\hat{\lambda}_{\tau}(s) = \frac{1}{\delta_{\tau}(s)} \sum_{i=1}^n \frac{1}{\tau^2} k \left(\frac{(s - s_i)}{\tau} \right)$$

Onde τ é raio de influência que define a vizinhança do ponto a ser interpolado, e k é a função de estimação de Kernel que fará o alisamento da superfície.

Uma aplicação muito comum de Kernel é a estimativa de densidade populacional de animais e propriedades em grandes áreas. STEVENSON ET AL. (2000) conduziram um estudo descritivo espacial da ocorrência da Encefalopatia Espongiforme Bovina (BSE) na Inglaterra usando o método de Kernel (Figura 10).

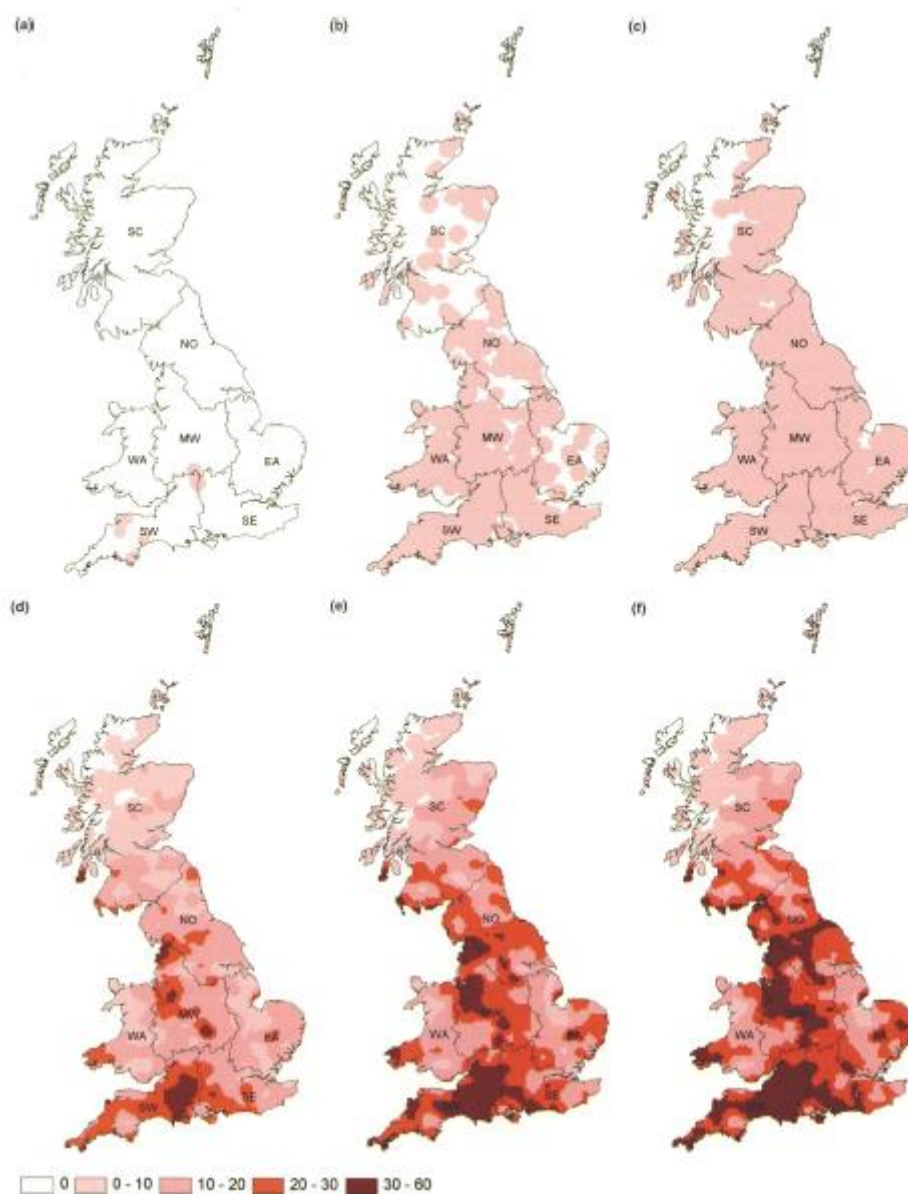


Figura 10: Mapas isopléticos elaborados a partir do método de Kernel. Os mapas demonstram a evolução da prevalência de BSE na Inglaterra durante os anos de 1987 a 1997. A prevalência de BSE é expressa pelo número de focos positivos a cada 100 propriedades. Fonte: Stevenson, 2000.

2.3.3.2 Krigagem

A krigagem é um método de interpolação espacial da geoestatística baseada na suavização de uma superfície, a partir de poucos pontos amostrados vinculados com uma variável contínua de interesse. Diferente da estimação de Kernel, a krigagem é fundamentada num modelo paramétrico espacial (Berke, 2004). A predição da krigagem é construída por

meio de variogramas empíricos. O variograma é uma ferramenta que mede quantitativamente a relação de dependência de um evento no espaço. GRAHAM ET AL. (2005) realizou uma pesquisa da distribuição espacial, no Texas, de anticorpos contra *Salmonella* entérica serovar *Typhimurium* em rebanhos leiteiros.

2.4 Exploração espacial

A análise exploratória de dados espaciais implica na busca da presença de algum tipo de padrão na ocorrência dos eventos sanitários, baseada na formulação de hipóteses estatísticas.

A diferenciação entre visualizar e explorar dados é uma questão importante, pois suas delimitações não estão bem definidas, uma vez que existe uma ligação entre alguns dos seus métodos. Portanto, essa diferenciação depende basicamente do grau de manipulação dos dados que o método aplicado envolve. Essas abordagens são normalmente acompanhadas de uma interpretação intuitiva e tem como objetivo estimar uma impressão inicial da área de estudo. Segundo os autores que propuseram essa diferenciação dos métodos de análise espacial (Bailey & Gatrell, 1995), a tendência é realmente mesclar esses dois métodos devido à semelhança e inter-relação entre as suas técnicas. Nesse contexto, o presente trabalho utiliza-se dessa integração dos métodos de visualização com os de exploração. A partir da revisão dos principais aspectos das análises ditas exploratórias, as técnicas de interpolação de dados citadas anteriormente serão analisadas e demonstradas em estudos epidemiológicos de medicina veterinária.

Neste capítulo, os métodos de análise exploratória de dados espaciais também serão divididos em: análise exploratória de pontos, de dados agregados em áreas e de dados contínuos.

2.4.1 Análise exploratória de pontos

A primeira etapa da análise exploratória de pontos é a verificação se existe um padrão espacial de distribuição de pontos. Pensando em um teste estatístico, as hipóteses seriam formuladas com base na pergunta: o padrão de distribuição dos pontos é aleatório? Está distribuído ao acaso? Em estudos epidemiológicos, o esperado é que a distribuição de eventos sanitários esteja vinculada com determinados fatores ligados ao espaço e ao tempo (Wakefield et al., 2000). Padrões de distribuição espacial podem ser classificados como regular, aleatório ou agregado.

No padrão de distribuição de pontos regular, a distância média entre os pontos amostrados tende a ser constante, essa situação ocorre em estudos com amostragens

planejadas. O padrão aleatório não tem nenhum modelo de distribuição, está ocorrendo por um completo ao acaso.

O termo *cluster* é utilizado para descrever eventos sanitários que estão agregados ou aglomerados no espaço e tempo de forma inesperada (CDC, 1990). Essa agregação espacial de uma doença pode ser atribuída a fatores demográficos, genéticos, ambientais ou sócio-culturais, por exemplo (Werneck & Struchiner, 1997).

O objetivo principal da investigação de *clusters* no espaço em epidemiologia é o esclarecimento do mecanismo responsável pela sua formação e, se essa formação é ou não devido ao acaso. São basicamente dois mecanismos explicativos de interesse epidemiológico que podem estar contribuindo para a formação de *clusters* de uma doença (Marshall, 1991). O primeiro pode ser explicado por fatores de risco transitórios que expõem os indivíduos a um maior risco de adoecer de maneira independente, por exemplo, algum manejo de criação. Já o segundo, processos infecciosos ou heranças genéticas são responsáveis pela formação dos agregados, na qual a independência entre eventos não é observada.

Segundo os autores BESAG & NEWELL (1991), os métodos estatísticos de análise espacial para investigação de clusters podem ser categorizados em dois tipos: globais ou locais. Os testes estatísticos globais objetivam identificar um padrão geral de distribuição da doença em uma grande região, sem identificar a sua localização geográfica.

Já os testes locais são utilizados quando se tem a localização dos dados, ou seja, verificam a existência de *clusters* em regiões pré-definidas, antes mesmo da observação da sua ocorrência (Santos & Souza, 2007). A sua hipótese alternativa é que o número de eventos sanitários observados são maiores do que o esperado para aquela região. Um exemplo prático do uso desses testes focais é para determinar a localização de um foco suspeito, com o risco aumentado para determinada doença.

Muitas dessas técnicas exploratórias de identificação de aglomerados utilizam o método de Monte Carlo para avaliar a significância estatística dos clusters encontrados (Pfeiffer, 2012). Após o cálculo da estatística da técnica espacial, o método de Monte Carlo faz diversas simulações aleatórias com os dados observados, acumulando distribuições e a partir destas, será calculado um nível descritivo (valor de p).

2.4.1.1 Testes globais para análise de cluster em pontos

2.4.1.1.1 Cuzick e Edwards' k-nearest neighbour test

CUZICK & EDWARDS (1990) desenvolveram um método para detecção espacial de *clusters* para populações expostas ao risco com distribuição heterogênea ou não-estacionária numa determinada área. O método é baseado na localização dos casos (positivos para a

doença) e numa seleção aleatória dos controles (negativos), presumindo que controles vizinhos também podem ser outros casos.

O teste estatístico (T_k) conta quantos controles vizinhos são casos. Para o cálculo estatístico, é incluído um parâmetro “k”, que é determinado empiricamente pelo usuário. Esse parâmetro refere-se ao número de controles vizinhos dos casos que será utilizado. Para cada caso, o teste conta quantos controles vizinhos, ou quantos parâmetros k vizinhos, também são casos. Pela fórmula:

$$T_k = \sum_{i=1}^{n_i} m_i(k)$$

Hipótese nula: Casos e controles forma amostrados a partir de um ponto de distribuição espacial comum.

Hipótese alternativa: Os casos estão espacialmente agrupados em relação aos controles.

Onde T_k é o total sobre todos os casos no conjunto de dados e n é número da população amostral. Por exemplo, T_1 é o número de casos no conjunto de dados cujos vizinhos mais próximos também são casos ($k = 1$).

Se os casos estão agrupados, os vizinhos próximos tendem a serem outros casos o valor T_k é alto. Entretanto, quando todos os casos possuem pontos vizinhos somente como controles, T será igual a zero.

CUZICK & EDWARDS (1990) utilizaram o método para a triagem de aglomerados de casos de leucemia infantil e linfoma diagnosticado no Reino Unido entre 1974 e 1986. Os dados mostraram uma aglomeração espacial significativa de casos de leucemia em relação à distribuição espacial dos controles. Existem muitos exemplos na literatura em saúde animal utilizando também esse método. CARPENTER ET AL (2006) procuraram detectar agrupamentos espaciais de abortos de rebanhos leiteiros na Dinamarca. PEREZ ET AL (2002) investigaram *clusters* de tuberculose bovina na Argentina, para o qual os dados foram agregados pelos municípios. Os municípios com uma prevalência maior do que a mediana foram definidos como casos, e aqueles com uma prevalência mais baixa do que ou igual à mediana como controles. Importância global do agrupamento é avaliada usando valores de k variam de um a dez.

2.4.1.1.2 Ripley's K-function

A função K de Ripley utiliza uma medida de primeira ordem, assume o pressuposto que não há variação da intensidade e na direção na ocorrência de um evento, portanto, é

utilizada para detectar *clusters* em populações com distribuição estacionária em um processo isotrópico.

Pela fórmula, a função K é calculada através do número de eventos dentro de uma distância s. Para ser um *cluster*, cada caso tem maior risco de estar cercado por outros casos em uma mesma área de interesse (R), portanto, para distâncias s menores, a função K será maior. Por outro lado, se os casos estão espaçados e distribuídos de forma regular, para valores pequenos de s, a função de K também será pequena.

$$K(s) = \frac{1}{\lambda^2 R} \sum_{i \neq j} \sum I_s(d_{ij})$$

Todavia, devido às variações na distribuição espacial da população em risco, uma única função K calculada para os casos pode não conter a informação completa. Assim, a função K dos casos é comparada com a função K dos controles.

$$D(S) = K_{case}(s) - K_{control}(s)$$

Hipótese nula: A distribuição dos casos é uniforme.

Hipótese alternativa: A distribuição dos casos está aglomerada.

A diferença observada é interpretada como a medida em si de agregação de casos, e maior do que foi observado nos controles (Stevenson, 2003).

O 'BRIEN (2000) usa a função K de Ripley para investigar agrupamentos de diferentes tipos de neoplasias em humanos e em cães em Michigan, Estados Unidos, entre os anos 1964 e 1994, concluindo que os processos de determinação de agregação espacial dos casos nas duas espécies não são independentes uma da outra. ABERNETHY ET AL. (2000) utilizam a função K para identificar agrupamentos espaciais grupos de aves afetadas pela doença de Newcastle na Irlanda do Norte entre 1996 e 1997.

2.4.1.2 Testes locais para análise de clusters em pontos

Os métodos espaciais ditos globais apresentados anteriormente, permitem avaliar se existe uma diferença estatística significativa entre as observações dentro e fora de uma área de influência. Mas na grande parte das vezes, principalmente em estudos de surtos sanitários em grandes áreas, é necessário identificar o local onde estão localizadas as possíveis fontes de risco. Para isso, existe uma técnica que avalia o número de casos por meio de um círculo que caminha no espaço, mudando o seu raio e a posição do seu centro. Essa técnica faz uma varredura (*scan* em inglês) da área estudada, com objetivo de localizar espacialmente a área com

uma incidência de casos elevada dentro, e muito baixa, fora do círculo. Logo, a estatística espacial *scan* possui a habilidade de detectar e localizar geograficamente o *cluster*, já que se possui a informação da localização dos casos e controles.

2.4.1.2.1 Kulldorff's spatial scan statistic

O método proposto por Kulldorff (1997) leva em consideração a distribuição populacional irregular que conseqüentemente influenciará na distribuição dos pontos sob a hipótese nula. A estatística espacial *scan* é definida por uma janela geográfica circular que se move através da área de interesse, ou seja, é uma estatística de varredura onde, o círculo inclui diferentes conjuntos de áreas vizinhas (Barcellos, 2007). Somente considera-se uma outra área, se o seu ponto central está contido na janela. O raio de cobertura do círculo pode variar de zero a um valor máximo, entretanto, a janela nunca pode incluir mais do que 50% da população total (Kulldorff, 1997). Para cada localização em relação ao ponto central do círculo, é verificado se existe uma elevação na proporção do número de eventos quando comparado com a região externa do círculo.

A técnica cria um conjunto infinito de círculos distintos. Para cada um desses círculos, é testada a hipótese alternativa que existe um risco elevado de ocorrência de eventos dentro deste círculo em comparação com o exterior, gerando em um valor de probabilidade (*p-valor*). É comparado o número de casos dentro do círculo com o número de casos esperados, se os casos estivessem aleatoriamente distribuídos no espaço.

O teste é calculado por meio de uma função de verossimilhança. Na estatística, função de verossimilhança é uma relação matemática por meio de uma probabilidade condicionada, isto é, a ocorrência de um evento está condicionada a outro evento. No caso da estatística *scan*, a função de verossimilhança é maximizada para todas as janelas e, aquela com a maior verossimilhança é considerada como o *cluster* mais provável de estar ocorrendo. Em outras palavras, o *cluster* que tem a menor probabilidade de ter ocorrido ao acaso.

Além de identificar o aglomerado mais verossímil, o método também identifica outros aglomerados secundários que também são significativos. Existe a possibilidade de sobreposição desses aglomerados.

Dois tipos diferentes de modelos probabilísticos devem ser considerados quando se aplica a estatística espacial *scan*: modelo de distribuição Bernoulli e modelo de distribuição de Poisson.

O modelo de distribuição de Bernoulli é utilizado para dados categóricos binários, os eventos são classificados como casos ou controles, positivos ou negativos. Os eventos ou

indivíduos são identificados como pontos e sua localização (coordenadas geográficas) constitui um processo pontual.

Já no modelo de Poisson, o valor do evento em cada ponto ou área é considerado como uma distribuição normal de acordo com a população sob risco conhecida. Sob a hipótese nula que o número esperado de casos em cada área é proporcional ao tamanho de sua população. O modelo de Poisson requer a contagem de casos e controles para cada localização dos eventos.

Para ambas os modelos, a estatística *scan* ajusta a densidade populacional irregular presente e sua análise é condicionada ao número total de casos observados. A determinação do modelo que deve ser utilizado depende estritamente da variável a ser respondida, como dito anteriormente, para dados binários, modelo de Bernoulli, dados em forma de contagens, modelo de Poisson (Kulldorff, 2003).

Para realizar a estatística *scan*, utiliza-se o software SatScan (Figura 11). O SatScan é um aplicativo gratuito desenvolvido pelo professor Martin Kulldorff, do Instituto de Câncer dos Estados Unidos (<http://www.satscan.org>). A entrada de dados é constituída da identificação dos casos e controles e suas coordenadas geográficas correspondentes. O resultado dessa análise é a identificação dos centroides dos *clusters* significativos existentes e os seus raios de abrangência. Esses dados podem ser convertidos em um mapa de visualização utilizando softwares para essa finalidade.

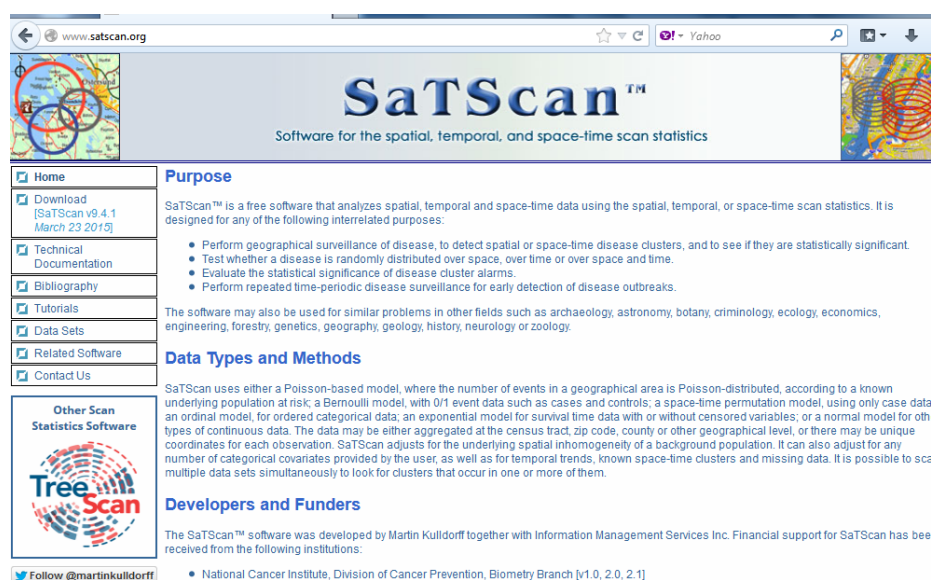


Figura 11: Demonstração da página inicial do programa SatScan para utilização na análise espacial *scan*. Fonte: <http://www.satscan.org/>.

Atualmente, a estatística *scan* é a técnica exploratória espacial mais utilizada em estudos relacionados à saúde animal, devido a essa acessibilidade e disponibilidade gratuita do programa SatScan. Através dessa técnica, foram identificados *clusters* de Encefalopatia Espongiforme bovina na Inglaterra (Stevenson et al., 2000), na Suíça (Doherr et al., 2002) e na França (Abrial et al., 2003). IGLESIAS ET AL. (2010) identificaram áreas de maior risco para Influenza Aviária de Alta Patogenicidade (HPAI) na Europa, utilizando análises de cluster sem conhecer o tamanho da população em risco (Figura 12).

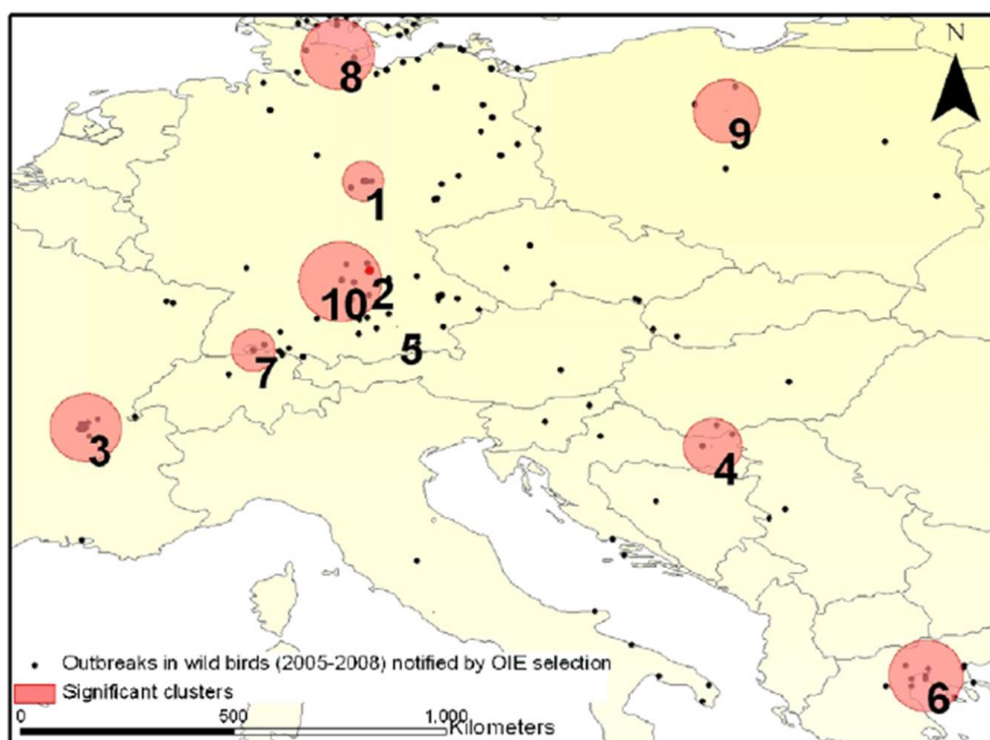


Figura 12: Mapa indicando as áreas com maior risco para HPAI (círculos vermelhos) na Europa. Os *clusters* foram elaborados a partir de dados de surtos de HPAI em aves silvestres detectados de 2005 a 2008 (pontos). Fonte: Iglesias, 2010.

2.4.1.3 Testes globais e locais para análise de clusters em áreas

As análises estatísticas dos testes exploratórios em áreas, tanto os testes globais, quanto os locais, são baseadas na relação de distância e vizinhanças perante a uma variável nas áreas que constituem o estudo. Essa relação é denominada dependência espacial.

Os métodos globais constituem uma aproximação mais conservadora do efeito da dependência espacial, em que a relação de dependência poder ser resumido em um único valor. Pode ser obtido pelo coeficiente I de Moran e G de Geary, ou ainda, através de uma variograma (para dados pontuais). Já os métodos locais, também conhecidos como LISA

(*Local Indicator of Spatial Association*, ou Indicador Local de Associação Espacial) são valores associados a cada área específica. Fornecem assim uma informação acerca da relevância de um agrupamento espacial de valores ao redor de cada uma destas áreas e a soma destes valores é proporcional ao indicador global de associação espacial (Krempf, 2004).

A estrutura de dependência espacial é calculada por uma função de autocorrelação espacial, que mede a correlação de uma única variável em diferentes lugares no espaço. Assim, altas correlações indicam que os dados não estão distribuídos de forma aleatória no espaço, logo, possuem uma relação de dependência e indicam a formação de *cluster*.

Um aspecto fundamental para o cálculo da autocorrelação espacial é a construção de uma matriz de pesos de vizinhança, que é usada para definir a relação espacial das áreas, por exemplo, se essas áreas possuem a mesma fronteira, ou não (Pfeiffer, 2012). Visto que existe uma dependência entre a autocorrelação espacial e a distância, portanto, áreas mais próximas ou com limites vizinhos devem possuir maior peso no cálculo, do que as mais distantes (Moran, 1950).

Os dados utilizados nessas análises são geralmente taxas, pois estão representando um grupo ou uma população em uma área delimitada, como as taxas de incidência de doença ou taxas de mortalidade.

2.4.1.3.1 I de Moran Global

O coeficiente de autocorrelação I de Moran é essencialmente uma alteração do coeficiente de correlação comum da estatística analítica, o coeficiente de correlação de Pearson, mas com adição de um parâmetro que mede a proximidade espacial (Durr & Gatrell, 2004). Essa proximidade espacial é a quantificação da similaridade de uma variável entre as áreas determinadas (Moran, 1950). Pela fórmula do coeficiente I de Moran global:

$$I = \frac{n \sum_i \sum_j W_{ij} (Z_i - \bar{Z})(Z_j - \bar{Z})}{(\sum_i \sum_j W_{ij}) \sum_k (Z_k - \bar{Z})^2}$$

Onde n é o número de áreas do estudo, Z é a diferença dos eventos observados menos o esperado, W é a medida de proximidade espacial entre um par de áreas. Com dados espaciais por áreas, uma definição usada para proximidade espacial é que as unidades de cada área deve ter um limite comum. As ponderações para os valores de vizinhança podem ser especificadas como binário (sim ou não), por exemplo, caracterizando se é uma área com limite adjacente ou não. Ou ainda, utilizar uma matriz com ponderações das distâncias, como a distância entre um par de pontos (Saez & Saurina, 2007).

Conforme o cálculo do coeficiente I de Moran, a autocorrelação de uma variável vai variar de +1 a -1. A correlação da variável com ela mesma, medida no mesmo local, será sempre 1. Quanto mais próximo do 1, maior é a autocorrelação, maior a semelhança entre as áreas, portanto, indica a aglomeração espacial dos eventos. O valor 0 indica inexistência de correlação, e os valores negativos indicam correlação espacial negativa.

O teste de Moran deve ser usado para detecção de *clusters* em eventos com distribuição estacionária, ou seja, a autocorrelação espacial é constante em todas as distâncias.

Como exemplos do uso do coeficiente de Moran para detecção de aglomerados em medicina veterinária, KITRON & KAZMIERCZAK (1997) investigaram a distribuição espacial da incidência da enfermidade de Lyme por distritos no Estado de Wisconsin nos Estados Unidos e identificaram uma autocorrelação com a distribuição do vetor (*Ixodes scapularis*) que transmite essa doença. PEREZ ET AL. (2000) também usou esse método em seu trabalho sobre a distribuição espacial da tuberculose bovina na Argentina.

Existem outros testes para detecção de agrupamentos por áreas, como o coeficiente C de Geary (1954) e teste de Tango (1995).

2.4.1.3.2 LISA

O LISA pode ser calculado a partir do método I de Moran Local (Anselin, 1995), que constitui uma decomposição do coeficiente global I de Moran. O teste determina coeficientes específicos de autocorrelação espacial para cada área da região do estudo, permitindo assim, a identificação das áreas que estão agrupadas. A fórmula segue o mesmo desenho do teste global:

$$I_i = Z_i \sum_{j, j \neq i}^n w_{ij} Z_j$$

Onde Z_i e Z_j são os valores observados e w_{ij} é a matriz dos pesos das relações espaciais. A soma de todos os diferentes coeficientes I é proporcional ao índice global de Moran.

Existem dois usos para o LISA. Ou como indicador de autocorrelação local, ou como testes para valores extremos nos padrões espaciais globais, sob a forma de um gráfico de dispersão de Moran (Anselin, 1996). Em um gráfico de dispersão de Moran, o eixo horizontal (Z) representa o vetor dos valores observados e os específicos de eixo vertical (W), como a média ponderada dos valores vizinhos (Figura 12). A extensão da união dos pares dentre os quatro tipos de associação nos quadrantes (baixo-elevado, maior alta, alta-baixo e baixo-baixo) fornece uma indicação da estabilidade da associação espacial ao longo dos dados.

Também pode sugerir a existência de diferentes tipos de associação em diferentes subconjuntos de dados, por exemplo, associação positiva em uma área e associação negativa em outro (Pfeiffer, 2012).

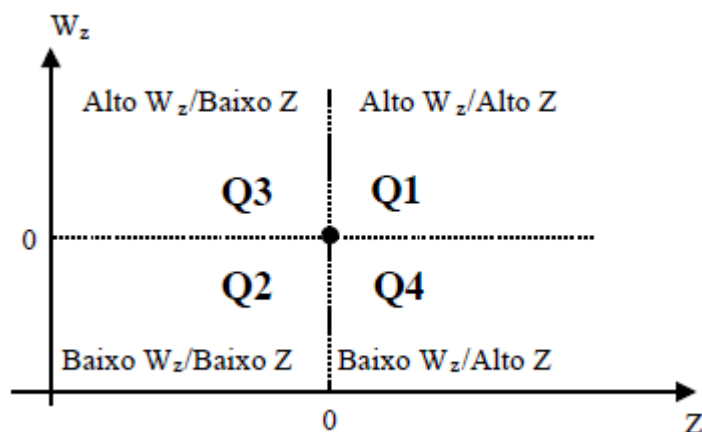


Figura 13: Gráfico de dispersão de Moran.

NODTVEDT ET AL. (2007) utilizou método de LISA para analisar a distribuição espacial de casos de dermatite atópica em cães na Suécia nos anos de 1995 a 2000.

2.5 Modelagem espacial

A modelagem de dados espaciais são modelos estatísticos que objetivam estimar parâmetros a partir de eventos estocásticos. Ou seja, aqueles eventos que estão sujeitos a incertezas e são influenciados pelas leis de probabilidade. Desse modo, são usados métodos de análise que consideram a probabilidade de ocorrência de um evento em relação a outro conjunto de eventos localizados em um espaço específico.

Modelos de regressão da estatística básica não são apropriados para análise de dados espacialmente dependentes, visto que assumem a independência das observações (Bailey & Gatrell, 1995). A modelagem espacial exige uma avaliação iterativa da força de autocorrelação espacial e suas covariáveis.

Modelos autoregressivos espaciais são métodos que incorporam os pesos de autocorrelação espacial em seus desenhos, a partir de matrizes de vizinhança que especificam as relações espaciais entre pontos vizinhos (Auchincloss et al., 2012).

Métodos bayesianos também são aplicados para analisar possíveis riscos relativos à localização espacial, utilizando outras covariáveis preditoras em seu delineamento (Auchincloss et al., 2012).

Na Inglaterra, um estudo realizado por STEVENSON ET AL. (2005) ilustraram, a partir de um modelo hierárquico bayesiano, a evolução da epidemia de Encefalopatia espongiforme em bovinos nascidos antes e depois da proibição do uso de alimentos com proteína animal de julho de 1988.

PFEIFFER ET AL. (2007) usou um modelo de regressão logística multivariável para analisar a associação entre fatores de riscos para HPAI com o risco de um surto no Vietnã. O risco de o surto ocorrer aumentava em granjas próximas a plantações de arroz (Figura 13).

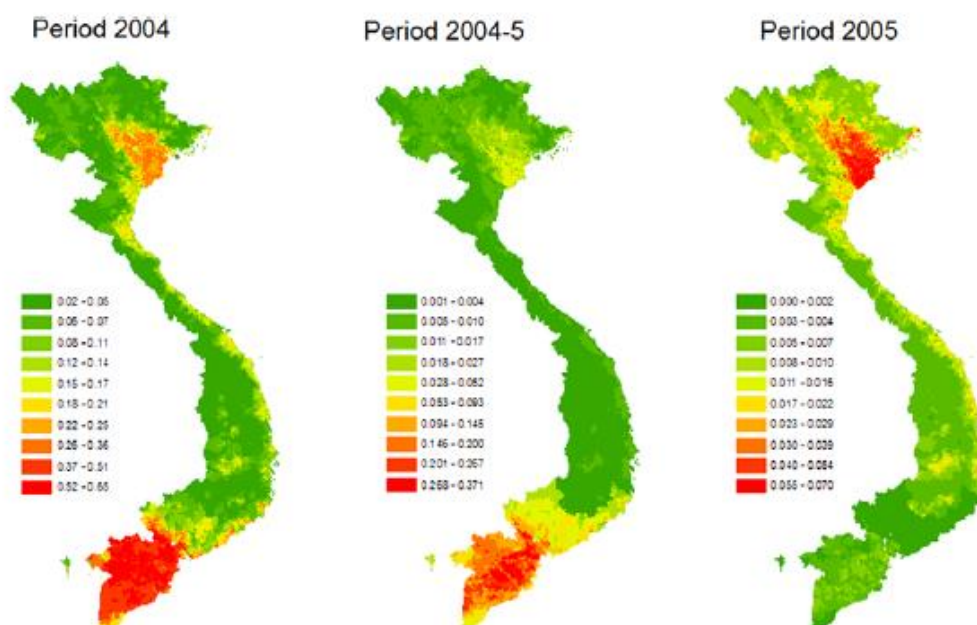


Figura 14: Mapas com as probabilidades de surtos de HPAI no Vietnã, e estimadas para cada um dos três períodos de epidemia, com base em efeitos aleatórios regressão logística multivariável. Fonte: Pfeiffer, 2007

2.6 Softwares para análises espaciais

Há uma grande disponibilidade de programas de gerenciamento SIG e softwares de análise espacial. O seu uso aumentou devido aos avanços que ocorreram em relação à funcionalidade de programas SIG. Entretanto, para algumas análises estatísticas espaciais mais complexas, há a necessidade da utilização de pacotes estatísticos que não são ferramentas SIG. Esses pacotes estatísticos requerem certo nível de conhecimentos de programação computacional. Exemplos desses softwares estatísticos são o SAS, o R e SPSS.

É útil distinguir entre mapeamento e *software* de análises SIG. O primeiro só produz mapas e, geralmente, tem a entrada de dados limitada. Enquanto que o segundo fornece uma

ampla gama de funções espaciais, que podem ser classificados em entrada de dados, gerenciamento e análises (Pfeiffer & Hugh-Jones, 2002). Exemplos de pacotes de software de mapeamento incluem os software livres ArcExplorer e EpiMap, que é parte do pacote do *software* de domínio público EpiInfo, elabora do pelo CDC. Softwares de análises SIG incluem, por exemplo, o ArcGIS da ESRI, QGIS e IDRISI.

A maioria dos softwares de análises SIG pode lidar com fontes de dados tanto vetoriais como *raster*. Também são capazes de relacionar bancos de dados espaciais com não espaciais. Softwares de análise espacial especializada inclui o ClusterSeer e o software de domínio público SaTScan, que permitem analisar os agrupamentos, tanto espacial, quanto espaço-temporal.

3. METODOLOGIA

Este capítulo dividiu-se em quatro partes: a primeira é uma breve introdução sobre o método de análise espacial escolhido para demonstração prática. A segunda parte é referente à problemática gerada por javalis de vida livre; a terceira, ao banco de dados utilizado para aplicação da técnica, e a última, os métodos de análise estatística.

3.1 Krigagem

A krigagem é um método de interpolação espacial pertencente à Geoestatística. A geoestatística é um ramo da estatística espacial que usa o conceito da função de aleatoriedade para incorporar a autocorrelação espacial. A geoestatística tem como objetivo principal a análise exploratória e modelagem da variabilidade espacial de um fenômeno ou variável.

A krigagem utiliza um atributo específico e sua posição geográfica para calcular as interpolações, baseada na Primeira Lei da Geografia de Tobler (Tobler, 1979). Conforme essa lei, valores de uma variável em posições mais próximas são mais correlacionadas do que as mais afastadas. A krigagem utiliza funções matemáticas para acrescentar pesos maiores nas posições mais próximas aos pontos amostrais e pesos menores nas mais distantes, e assim criar, novas inferências em novos pontos com bases em combinações lineares dos dados (Jakob, 2002).

A verificação da dependência espacial é realizada a partir de variogramas. O variograma é a descrição matemática da relação entre a variância de pares de observações (dois pontos) e a distância que separa essas observações (h).

Ainda, a krigagem faz inferências a partir de uma função aleatória $Z(x)$, originando os pontos $Z(x_1)$, $Z(x_2)$, ..., $Z(x_n)$. A função:

$$Z(x) = m(x) + \gamma(h) + \varepsilon$$

Apresenta a média constante, a correlação espacial e o erro residual. A correlação espacial é dada pelo variograma:

$$\gamma(h) = \frac{1}{2} N(h) \sum_i [Z(x_i) - Z(x_i+h)]^2$$

Onde $N(h)$ é o número total de pares de observações separadas pela distância h .

A Figura 15 demonstra os elementos que compõe o variograma. O efeito pepita (*nugget*) é ponto inicial da curva (C_0), onde a curva toca o eixo γ , quando $h=0$. O alcance (*range*) é o ponto máximo onde existe autocorrelação espacial das variáveis (a).

O patamar (*sill*) é o valor do variograma correspondente ao seu alcance, é o ponto em que não existe mais correlação entre as variáveis (C). A contribuição (C_1) é a diferença entre o patamar e o efeito pepita.

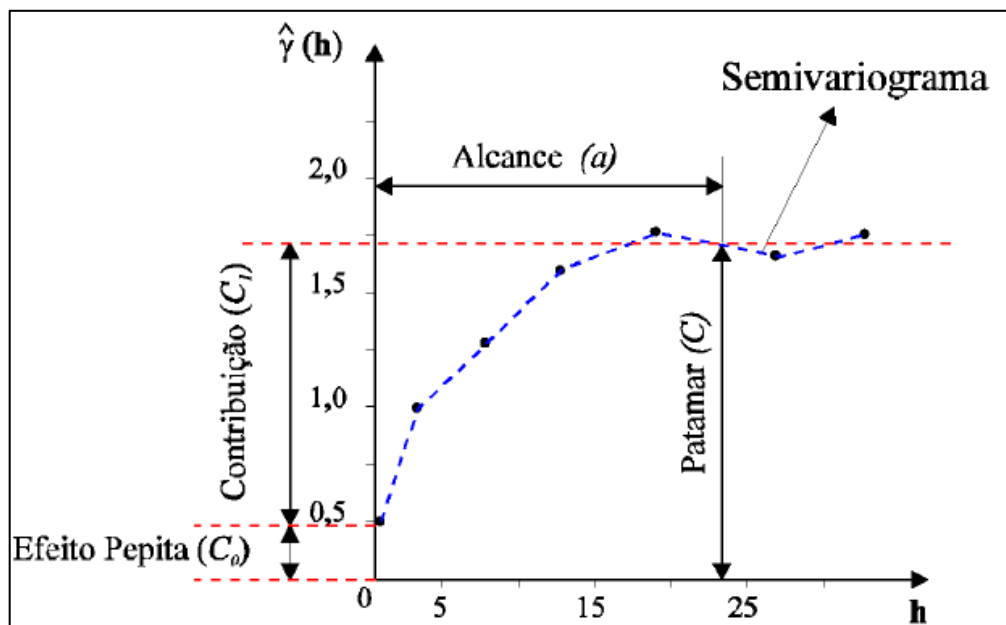


Figura 15: Exemplo de variograma e seus elementos. Fonte: Deutsch & Journel, 1998 (adaptado).

Existem três tipos básicos de variogramas, o esférico, o exponencial e o gaussiano. A parte mais importante do variograma é sua forma próximo à origem, uma vez que são os pontos mais próximos os que possuem maior peso no processo de interpolação (Jakob, 2002).

Para realizar a estimativa dos valores dos componentes do variograma é necessário definir dois parâmetros. Inicialmente deve-se determinar o tamanho da área. A determinação do tamanho da área foi proposta por JOURNAL & HUIJBERGTS (1978), a construção do variograma deve considerar distâncias iguais ou menores do que a metade da distância máxima da região de estudo. Depois, essa distância é dividida em tamanhos iguais, denominados *lags*.

Outro parâmetro fundamental é a definição da direção da maior continuidade espacial, ou alcance máximo. Variogramas são construídos para diferentes direções angulares, como 0° , 45° , 90° e 135° . Para que o ângulo de busca dos pares não se abra indefinidamente, é necessário restringi-lo através de uma largura de banda ou *bandwidth* (Deutsch & Journel, 1998). A figura 16 exemplifica os parâmetros para a construção do variograma.

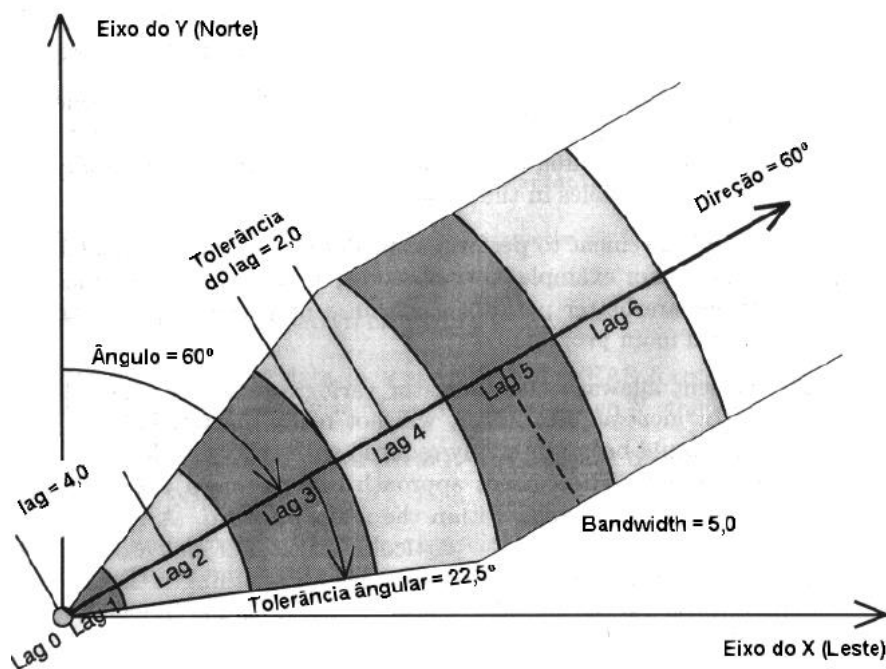


Figura 16: Demonstração gráfica dos parâmetros necessários para construção do variograma e escolha do ângulo onde se tem o maior alcance. Fonte: Deutsch & Journel, 1998 (adaptado).

A krigagem é um método de estimativa ou predição espacial que visa estimar o valor de um fenômeno em posições não amostradas, a partir de dados amostrados da variável em posições vizinhas. Os pesos das distâncias das posições vizinhas são baseados nos semivariogramas.

3.2 Javali

O javali (*Sus scrofa*, Linnaeus 1758) é a principal espécie dos porcos silvestres e tem sua origem na Europa, Ásia e Norte da África. Está amplamente distribuído, inclusive em áreas distintas da sua origem. Esses animais foram introduzidos pelo homem com o objetivo de criação e consumo, entretanto, devido ao seu temperamento selvagem e instável, acabaram fugindo das criações, dispersando-se em diferentes ambientes. Essa dispersão sem controle ocasionou diversos cruzamentos com o porco doméstico (*Sus suis*), resultando em animais híbridos de vida livre, também conhecidos pelos termos “javali selvagem”, “porco selvagem” ou “porco monteiro” (Sweeney et al., 2003). Nesse trabalho, o termo “javali” possui um caráter geral, referindo-se a todas as formas fenotípicas possíveis de *Sus scrofa*, diferentes do porco doméstico, que se asselvajaram devido a solturas ou escapes de criações autorizadas e

estão presentes em praticamente todos os continentes, principalmente nas Américas (SBCF, 2010).

Os impactos negativos da distribuição de javalis incluem a degradação da vegetação nativa e água de superfície (Cushman et al., 2004; Kaller & Kelso, 2006), a predação sobre a fauna e pecuária (Seward et al., 2004), e bem como a transmissão de patógenos para humanos e animais (Jay et al., 2007; Hall et al., 2008; Hermoso de Mendonza et al., 2006; Montagnaro et al., 2010; Mohamed et al., 2011).

O javali introduzido em diversas regiões do mundo é classificado pela *International Union for Conservation of Nature* (IUCN) como uma das 100 piores espécies exóticas invasoras (Lowe et al., 2000).

Em todo o mundo, há relatos de aumento da densidade populacional dos javalis (Saez-Royuela & Telleria, 1986; Gortazar et al., 2000; Acevedo et al., 2006), juntamente com o problemas acarretados por essa espécie. Nos Estados Unidos, os javalis causam cerca de US \$ 800 milhões em danos a produtores de gado e a agricultores a cada ano (Pimental et al., 2005). No Brasil, as populações de javalis vêm crescendo devido à falta de predadores naturais e à grande disponibilidade de alimentos presente nas culturas agrícolas e no meio ambiente, além da falta de uma estratégia adequada para o controle destes animais na natureza.

Outro ponto importante sobre o javali é a sua alta produtividade em comparação com outros ungulados silvestres (Taylor et al., 1998), também por consequência dos cruzamentos e hibridização da espécie com o porco doméstico, como consequência, é extremamente resistente à programas de controle de espécies (Giles, 1980), capazes de superpopular uma área em um curto espaço de tempo (Barrette Birmingham, 1994).

Devido a essa problemática dos javalis, o presente trabalho objetivou aplicar o método de estatística espacial escolhido, a krigagem, para estimar uma localização em todo o estado do Rio Grande do Sul (RS), Brasil. A título de comparação desse método, foi utilizada uma técnica de análise exploratória local de pontos.

3.3 Banco de dados

Os dados foram obtidos a partir de um inquérito soroepidemiológico, realizado no ano de 2014 no RS. O estudo foi delineado para verificar a ausência de Peste Suína Clássica (PSC) em suínos de criatórios não comerciais no RS. A amostragem foi calculada para uma prevalência esperada de 1% para PSC por unidade amostral (propriedade) e nível de confiança de 95%. Ao todo, foram amostradas 320 criatórios de suínos.

Um questionário epidemiológico foi aplicado por 27 médicos veterinários do Serviço Oficial do RS. Através desse questionário, foram coletadas informações sobre a localização geográfica (latitude e longitude) e dados relacionados à possível interação entre os javalis de vida livre e os criatórios de suínos domésticos, como: visualização de javalis, presença de rastros de javalis, presença de ataque na propriedade e/ou propriedade vizinha. Esses indicadores foram agrupados em uma única variável resposta: a presença de javalis. Dentre todas as 320 propriedades, 44 criatórios (14%) apresentaram um ou mais indicadores da presença de javalis de vida livre. Com as informações dos questionários, um banco de dados foi montado em Excel 2010®.

3.4 Metodologia da análise estatística

Para a análise estatística espacial foi construído um banco de dados com a identificação de cada criatório, as coordenadas geográficas (latitude e longitude) e a variável resposta (negativo ou positivo para a presença de javali). As coordenadas geográficas foram transformadas em projeções UTM para homogeneizar as unidades de medidas das análises espaciais.

A verificação da relação espacial entre a variável resposta e as projeções UTM foi realizada utilizando semivariogramas, que a metade do variograma. Semivariogramas empíricos foram construídos para as direções 0°, 45°, 90° e 135°, a fim de identificar o maior *range* e, assim, determinar os elementos indicadores da dependência espacial (*nugget* e *sill*). Foi utilizado o *software* livre SGeMS para análise da distribuição da variável resposta e construção dos semivariogramas.

A comparação da análise espacial por krigagem foi feita através de uma análise exploratória de pontos por *cluster* com o mesmo banco de dados. A estatística *scan* foi calculada usando o *software* SatScan, no qual foi determinado o centroide e raio de influência do *cluster*. O modelo probabilístico utilizado foi Bernoulli, com a variável categoria presença de javali (sim ou não) e as coordenadas geográficas.

Os mapas foram elaborados, tanto as interpolações dos pontos por krigagem, quanto os círculos de análise por *cluster*, no *software* ArcGis Map 10 (ESRI®), utilizando a ferramenta Spatial Analysis Tools.

4. RESULTADOS

A Figura 17 demonstra a distribuição dos pontos para verificação da dependência espacial. Os pontos vermelhos representam os criatórios positivos para a presença de javali e os pontos azuis, os negativos para presença de javali.

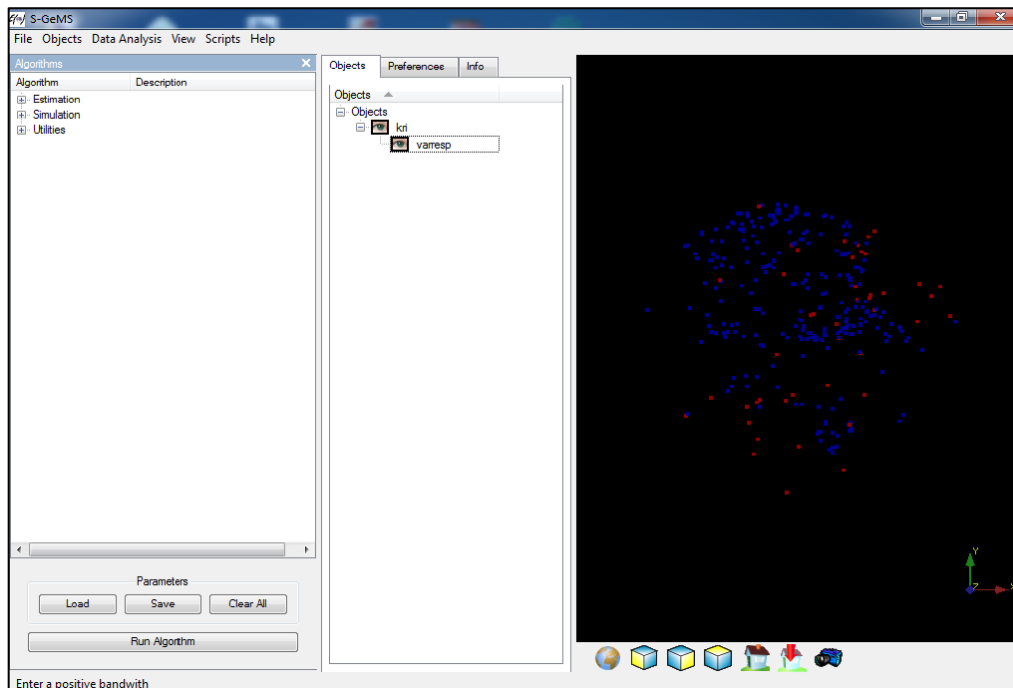


Figura 17: Visualização da distribuição dos pontos para a construção dos semivariogramas no *software* SGeMS.

Os semivariogramas empíricos foram construídos com o número de *lags* igual a 15, com uma separação de 50.000 metros (50 km) e tolerância de 25.000 metros (25 km). Foram calculados para os ângulos 0° , 45° , 90° e 135° , com uma tolerância angular de $22,5^\circ$ (Figura 19 e 20).

Input the parameters for variogram computation

Load Parameters... Save

Lags

Number of lags: 15
 Lag separation: 50000
 Lag tolerance: 25000

Directions

Number of directions: 4

Angles are in degrees. Use a tolerance $tol > 90$ to indicate an omni-directional variogram

	azimuth	dip	tolerance	bandwidth	measure type	head indic. cutc	tail indic. cutoff
1	45	0	22.5	10000	variogram		
2	135	0	22.5	10000	variogram		
3	90	0	22.5	10000	variogram		
4	0	0	22.5	10000	variogram		

Figura 18: Determinação dos parâmetros número de *lags* e direções angulares para construção dos semivariogramas.

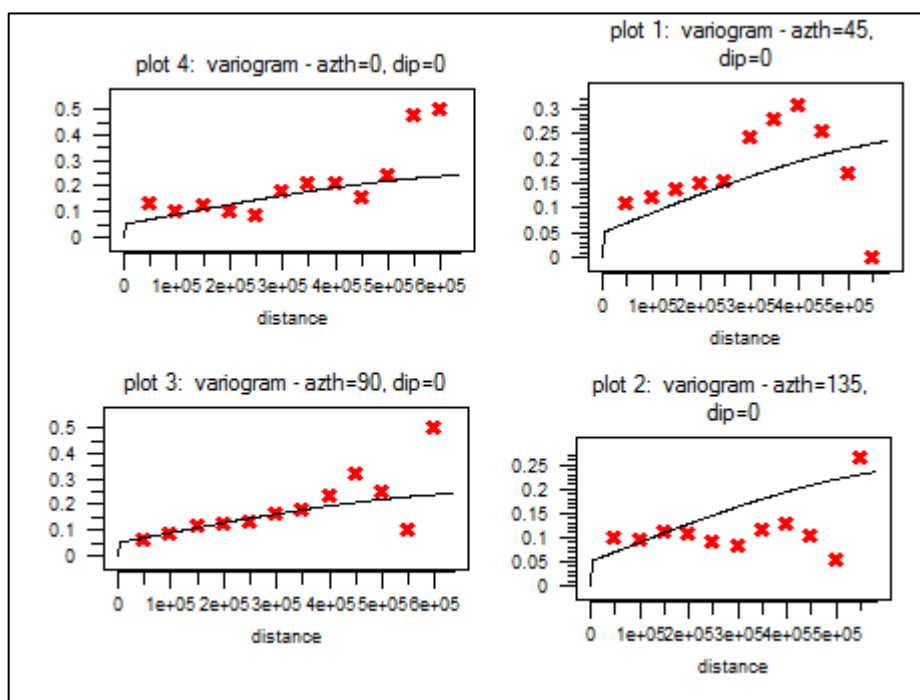


Figura 19: Semivariogramas construídos para verificação da dependência espacial.

Segundo a Figura 19, o semivariograma utilizado para definição dos parâmetros da krigagem foi o de 90°, onde se verifica o melhor ajuste de relação de continuidade espacial. Os parâmetros encontrados a partir desse ângulo:

- Efeito pepita (*nugget*): 0,05

- Patamar (*sill*): 0,2
- Alcance máximo (*range*): 750 km

A Figura 20 demonstra o mapa elaborado em que se utilizam os parâmetros do semivariograma e a ferramenta interpoladora krigagem. O resultado foi uma superfície suavizada, na qual foi feita uma estimativa de onde pode estar ocorrendo um risco maior da presença de javali a partir de 320 pontos georreferenciados no RS. Sendo um mapa isoplético, a intensidade do evento é representada por uma escala de cor. No caso, nos pontos onde não existe a informação da presença de javali, quanto mais escuro, maior o risco.

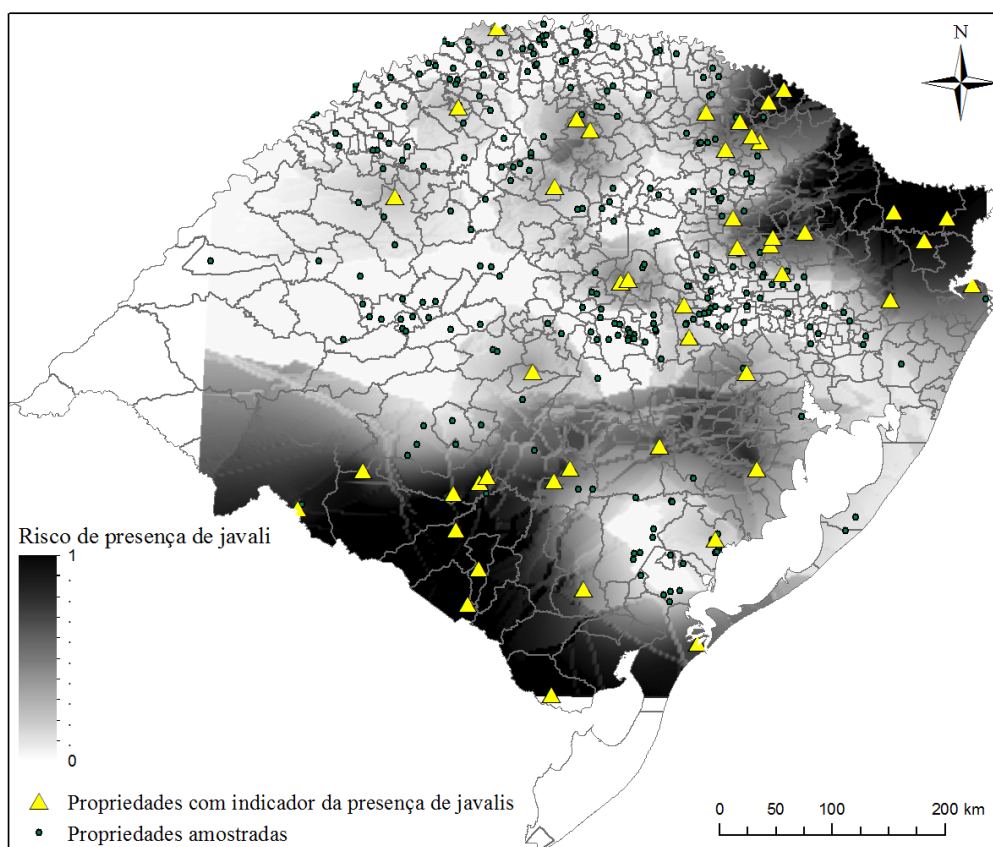


Figura 20: Mapa demonstrando o resultado da interpolação de pontos derivado da análise por krigagem. Pela escala de cor, quando mais escuro, maior o risco de presença de javali.

A análise estatística *scan* mostrou que existem dois *clusters* significativos ($p < 0,05$) para a presença de javali. O raio do *cluster* primário é de 127 km, do *cluster* secundário é de 88 km. A Figura 21 demonstra a localização de duas possíveis aglomerações de javalis.

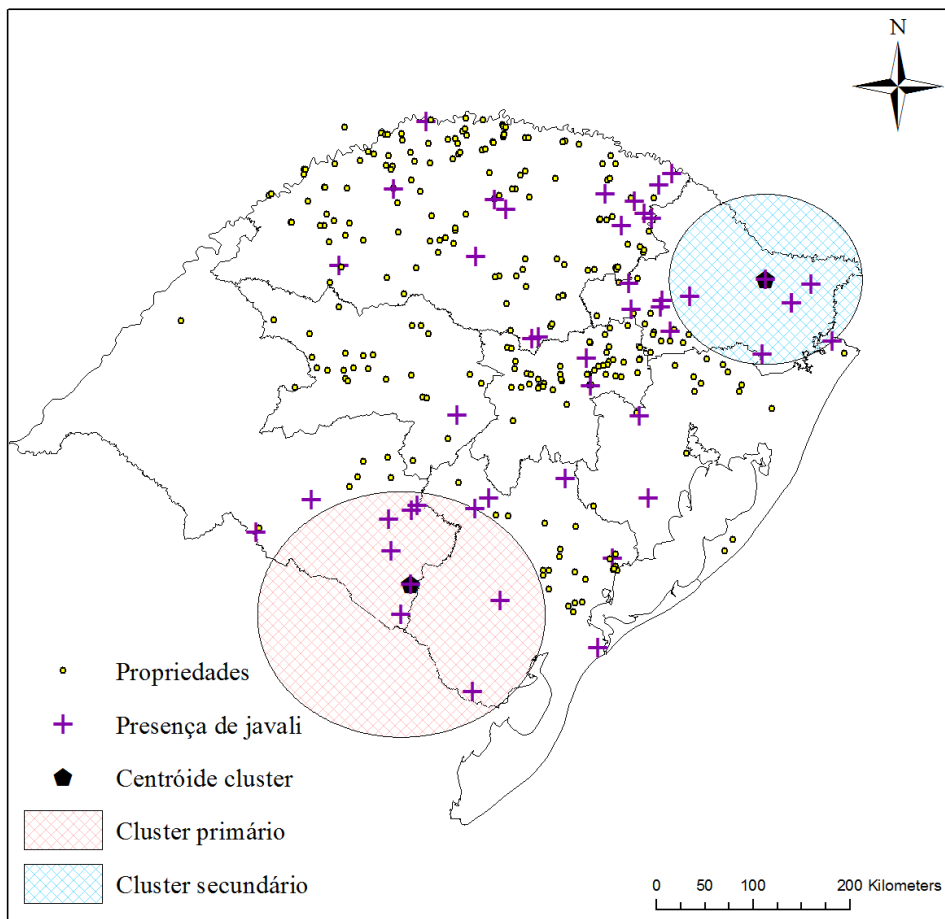


Figura 21: Mapa com a localização dos *clusters* primário e secundário e suas abrangências.

Conforme visto pelas figuras 20 e 21, os javalis de vida livre estão localizados principalmente nas regiões nordeste e sul/sudoeste do RS.

5. DISCUSSÃO

A análise por krigagem resulta em um mapa suavizado, isto é, o resultado é uma superfície contínua, onde os atributos de áreas antes desconhecidos, agora possuem um resultado estimado. De acordo com o mapa isoplético realizado pela krigagem (Figura 21), as áreas próximas a pontos positivos para presença de javali sempre possuem algum tipo de coloração. Entretanto, é necessário notar que a coloração está mais intensa, ou mais escura, nas áreas onde há maior concentração de pontos positivos para presença de javali, como as regiões nordeste e sudoeste do RS. Isso pode ser reflexo de uma maior dependência espacial nessas áreas, ou seja, essas áreas com a mesma intensidade de cor são mais semelhantes entre si.

A krigagem é um método geoestatístico que leva em consideração a autocorrelação de variáveis regionalizadas. Logo, valores de uma variável em pontos próximos são correlacionados, e essa correlação diminui com a distância. Para a análise por krigagem, deve ser constada a continuidade espacial da variável, que é verificada pelo variograma. A construção do variograma é de forma empírica, ou seja, o usuário determina alguns dos parâmetros para a sua construção. Perante a esse aspecto, a krigagem depende de um conhecimento prévio e racionalidade na interpretação dos resultados. Pode não ser possível a construção de um variograma adequado devido à natureza da variação espacial da variável analisada. Isto pode ocorrer devido a erros de amostragem e de análise (Krajewski & Gibbs, 1966).

Objetivo da visualização pela interpolação de pontos é evidenciar tendências das variáveis na região estudada. No caso, estimar áreas mais abrangentes no RS onde o javali possa estar circulando.

Para comparação do método interpolador, a análise exploratória de *cluster* foi utilizada. A análise, por meio de um círculo de varredura, identificou uma área onde a ocorrência da presença do javali é maior que o esperado. Ou seja, perante todos os outros pontos identificados, o círculo determinou a área onde ocorre uma maior incidência de pontos positivos para presença de javali do que os pontos fora desse círculo.

Ambas as análises apontaram para as regiões nordeste e sudoeste. Segundo a história da trajetória do javali no RS, uma estiagem ocorrida no Uruguai na década de 90, acarretou na invasão desses animais pela fronteira sudoeste do RS (Frankenberg, 2005; Valério, 1999). Somado a isso, em 1996, as importações de javalis originários da Europa foram autorizadas para criação no RS. Em 1998, o IBAMA proibiu essas importações e abertura de novos

criatórios. Devido a esse fato, muitos produtores resolveram soltar os seus animais com medo de multas e infrações dos órgãos ambientais (Debredt et al., 2005). Muitos desses criatórios existentes estavam localizados principalmente na região norte do Estado, concomitante com a criação de suínos domésticos. Nos dias atuais, a situação do javali está diferente, o Instituto de Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA) determinou que o javali-europeu (*Sus scrofa*), em todas as suas formas, linhagens, raças e diferentes graus de cruzamento com o porco doméstico, são animais exóticos invasores, permitindo assim, a sua caça em todo território nacional. Com isso, a preocupação com a localização e o comportamento dessa espécie vem aumentando significativamente pelos órgãos públicos, principalmente, porque ainda não se tem um programa de controle apropriado para a espécie.

Logo, a interpolação de pontos por krigagem e a análise exploratória de pontos por estatística *scan* podem ser usadas para localização dessa espécie por programas de controle e de defesa animal.

6. CONCLUSÃO

Nos últimos anos, devido aos avanços das tecnologias da informação, as análises espaciais vêm ganhando espaços nas investigações epidemiológicas, muito das vezes, servem como ferramentas chaves para resolução efetiva de um surto sanitário. A melhora na coleta de dados e acessibilidade dos SIG contribuiu para que essas ferramentas incorporem programas de monitoramento e vigilância na defesa sanitária animal na atualidade.

Não existe o melhor método de análise espacial, ou o mais correto, mas sim, a melhor maneira de usá-los. É primordial compreender a natureza dos dados coletados, criar um banco de dados organizado para assim, aplicar a técnica espacial que melhor explica a ocorrência de possíveis fenômenos sanitários vinculados ao espaço.

REFERÊNCIAS BIBLIOGRÁFICAS

ABERNETHY, D. A. et al. Evaluating airborne spread in a Newcastle disease epidemic in Northern Ireland. In: Salman, M., Morley, P., and Ruch-Gallie, R. (Eds.), *Proceedings of the 9th Symposium of the International Society for Veterinary Epidemiology and Economics*. Breckenridge, Colorado, USA, August 7–11, 2000.

ABRIAL, D. et al. Descriptive spatial analysis of BSE in western France. **Veterinary Research**, v. 34, p. 749–760, 2003.

ACEVEDO, P. et al. Factors affecting wild boar abundance across an environmental gradient in Spain. **Acta Theriologica**, v. 51, p. 327–336, 2006.

ACEVEDO, P. et al. Estimation of European wild boar relative abundance and aggregation: a novel method in epidemiological risk assessment. **Epidemiology and Infection**, v. 135, n. 3, p. 519 – 527, Aug 2007.

ANSELIN, L. Local indicators of spatial association – LISA. **Geographical Analysis**, v. 27, p. 93–115, 1995.

ANSELIN, L. The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In: Fischer, M., Scholten, H., Unwin, D. (Eds.), *Spatial Analytical Perspectives on GIS*, pp. 111–125. Taylor and Francis, London. 1996.

ARONOF, S. **Geographic Information Systems: a management perspective**. Canada: WDL Publications, 1989.

AUCHINCLOSS, A. H. et al. A review of spatial methods in epidemiology, 2000 – 2001. **Annual Review of Public Health**, v.33, p. 107 – 122, Apr. 2012.

BAILEY, N. T. J. & GATRELL, A. **Interactive Spatial Data Analysis**. Essex: Longman. 1995.

BARCELLOS, C. A. Saúde nos Sistemas de Informações Geográficas: Apenas uma camada a mais?. **Caderno Prudentino de Geografia**, v. 25, p. 29 – 43, 2004.

BARRETT, R. H. & BIRMINGHAM, G. H. Wild pigs. In: HYGNESTROM, S. E. ;TIMM R. M. & LARSON, G. E. *Prevention and Control of Wildlife Damage*. Lincoln: Cooperative Extension Service, University of Nebraska, USA. 1994. P. 65–70.

BERKE, O. Exploratory disease mapping: kriging the spatial risk function from regional count data. **International Journal of Health Geographics**, v. 3, p. 1-18, Aug. 2004.

BESAG, J.; YORK, J.; MOLLIE, A. Bayesian image restoration with two applications in spatial statistics. **Annals of the Institute of Statistics and Mathematics**, v. 43, p. 1–59, 1991.

BURROUGH, P. A. & McDONNEL, R. A. **Principles of geographic information systems**. New York: Oxford University Press, 1998. 333p.

CARPENTER, T. E. et al. An epidemiologic study of late-term abortions in dairy cattle in Denmark, July 2000-August 2003. **Preventive Veterinary Medicine**, v. 77, p. 215–229, 2006.

CDC, 1990. Guidelines for investigating clusters of health events.

CUSHMAN, J. H.; TIERNEY, T. A.; HINDS, J. M.; Variable effects of feral pig disturbances on native and exotic plants in a California grassland. **Ecological Applications**, v. 14, p. 1746–1756, 2004.

CUZICK, J. & EDWARDS, R. Spatial clustering for inhomogeneous populations (with discussion). **Journal of the Royal Statistical Society Series**, v. 52, p. 73–104, 1990.

DEBERDT, A. J. & SCHERER, S. B. O javali asselvajado: ocorrência e manejo da espécie no Brasil. **Natureza & Conservação**, v. 5, n. 2, p. 31 – 44, Out 2005.

DEUTSCH, C. V. & JOURNEL, A. G. **GSLIB: Geostatistical Software Library and user's guide**. New York, Oxford University Press, 1998. 369p.

DOHERR, M. G. et al. Geographical clustering of cases of bovine spongiform encephalopathy (BSE) born in Switzerland after the feed ban. **The Veterinary Record**, v. 151, p. 467–472, 2002.

DORLING, D. The visualization of local urban change across Britain. **Environment and Planning B: Planning and Design**, v. 22, p. 269–290, 1995.

DRUCK, S.; CARVALHO, M. S.; CÂMARA, G.; MONTEIRO, A. V. M. **Análise Espacial de Dados Geográficos**. Brasília: Embrapa, 2004. 190 p.

DURR, P. & GATRELL, AGIS and Spatial Analysis in Veterinary Science. CABI Publishing, Wallingford. 2004.

GESLER, W. The uses of spatial analysis in medical geography: a review. **Social Science and Medicine**, v. 23, p. 963-973, 1986.

GEARY, R. The contiguity ratio and statistical mapping. **The Incorporated Statistician**, v. 5, p. 115–145, 1954.

GORTAZAR, C. et al. Historical examination of the status of large mammals in Aragon, Spain. **Mammalia**, v. 64, p. 411–422, 2000.

GRAHAM, S. L. et al. Spatial distribution of antibodies to Salmonella enterica serovar Typhimurium O antigens in bulk milk from Texas dairy herds. **Preventive Veterinary Medicine**, v. 69, p. 53 – 61, Jan 2005.

HALL, J. S. et al. Influenza exposure in feral swine from the United States. **Journal Wildlife Disease**, v. 44, 362–368. 2008.

HERMOSO de MENDOZA, J. et al. Bovine tuberculosis in wild boar (*Sus scrofa*), red deer (*Cervus elaphus*) and cattle (*Bos taurus*) in a Mediterranean ecosystem (1992–2004). **Preventive Veterinary Medicine**, v. 74, p. 239-247, 2006.

IGLESIAS, I. et al. Identifying areas for infectious animal disease surveillance in the absence of population data: highly pathogenic avian influenza in wild bird populations of Europe. **Preventive Veterinary Medicine**, v. 96, p. 1-8, Aug 2010.

ISAAKS, E. H. & SRIVASTAVA, R. M. **Applied geostatistics**. New York, Oxford University Press, 1989. 561p.

JAKOB, A. E. Kriging demographic variables in order to look for trends in the spatial distribution of population. Tese (Mestrado em Spatial Modeling and Analysis) - Department of Crop and Soil Sciences, Cornell University, Ithaca, NY, 2002.

JAY, M. T. et al. Escherichia coli O157:H7 in feral swine near spinach fields and cattle, central California coast. **Emerging Infectious Diseases Journal**, v. 13, p. 1908–1911. 2007.

JOURNEL, A. G. Geoestatics: Tools for advanced spatial modeling in GIS. In: Applications of GIS to the modeling of non-point source pollutants in the vadose zone. Soil Sci. Soc. Am. J., Special Publication:39-55, 1996.

JOURNEL, A. G. & HUIJBERGTS, C. J. **Mining geostatistics**. London: Academic Press, 1978.

KALLER, M. D.; KELSO, W. E. Swine activity alters invertebrate and microbial communities in a Coastal Plain watershed. **The American Midland Naturalist Journal**, v. 156, p. 163–177, 2006.

KITRON, U. & KAZMIERCZAK, J. J., Spatial analysis of the distribution of Lyme disease in Wisconsin. **American Journal of Epidemiology**, v. 145, p. 558–566, 1997.

KREMPI, A. Explorando recursos de estatística espacial para análise da acessibilidade da cidade de Bauru. 2004. 94 f. Tese (Doutorado em estatística) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Paulo. 2004.

LAWSON, A. B. & WILLIAMS, F. L. R. (2001). An Introductory Guide to Disease Mapping. John Wiley & Sons.

LILIENFELD, A. M. & LILIENFELD, D. E. **Foundations of Epidemiology**. 2a ed, New York: Oxford, 1980.

LOWE, S., BROWNE, M., BOUDEJELAS, S. **100 of the world's most invasive species: a selection from the global invasive species database**. ISSG: Auckland. 2000.

MACMAHON, B. & PUGH, T., **Epidemiology: Principles and Methods**. Boston: Little, Brown & Co. 1970.

MARSHALL, R. J. A review of methods for the statistical analysis of spatial patterns of disease. **Journal of the Royal Statistical Society Series**, v. 154, p. 421–441, 1991.

MOHAMED, F. et al. Foot-and-mouth disease in feral swine: susceptibility and transmission. **Transboundary Emerging Diseases**, v. 58, p. 358-371. 2011.

MONTAGNARO, S. et al. Prevalence of antibodies to selected viral and bacterial pathogens in wild boar (*Sus scrofa*) in Campania Region, Italy. **Journal Wildlife Disease**, v. 46, p. 316-319, 2010.

MORAN, P. A. P. Notes on continuous stochastic phenomena. **Biometrika**, v. 37, p. 17–23, 1950.

NODTVEDT, A. et al. The spatial distribution of atopic dermatitis cases in a population of insured Swedish dogs. **Preventive Veterinary Medicine**, v. 78, p. 210–222, Mar 2007.

O'BRIEN, D. J. et al. Spatial and temporal comparison of selected cancers in dogs and humans, Michigan, USA, 1964-1994. **Preventive Veterinary Medicine**, v. 47, p. 187–204, 2000.

OPENSHAW, S. *The Modifiable Areal Unit Problem (Concepts and Techniques in Modern Geography)*. Norwich: Geo Books. 1984.

PEREZ, A. M. Use of spatial statistics and monitoring data to identify clustering of bovine tuberculosis in Argentina. **Preventive Veterinary Medicine**, v. 56, p. 63–74, 2002.

PFEIFFER, D. U. Geographical information science and spatial analysis in animal health. In: Durr, P., Gatrell, A. (Eds.), *GIS and Spatial Analysis in Veterinary Science*, pp. 119–144. CABI Publishing: Wallingford. 2004.

PFEIFFER, D. et al. **Spatial analysis in epidemiology**. Oxford: Oxford University Press, 2012. 142 p.

PFEIFFER, D. U. & HUGH-JONES, M. Geographical information systems as a tool in epidemiological assessment and wildlife disease management. **Revue Scientifique et Technique de l'Office International des Epizooties**, v. 21, p. 91–102, 2002.

PFEIFFER, D. U. et al. An analysis of the spatial and temporal patterns of highly pathogenic avian influenza occurrence in Vietnam using national surveillance data. **Veterinary Journal**, v. 174, p. 302 – 309. Jul 2007.

PIMENTAL, D.; LACH, L.; ZUNIGA, R.; MORRISSON, D. Update on the environmental and economic costs associated with alien-invasive species in the United States. **Ecological Economics**, v. 52, p. 273–288, 2005.

ROBSON, J.; YEOMAN, G. H.; ROSS, J. P. J. *Rhipicephalus appendiculatus* and East Coast Fever in Taganyika. **East African Medicine Journal**, v. 38, p. 206-214, 1961.

ROSA, R. Análise especial em geografia. Associação Nacional de Pesquisa e Pós-Graduação em Geografia, v. 7, n. 1, p. 275 – 289, Out. 2011.

SAEZ-ROYUELA, C. & TELLERIA, J. L. The increased population of the wild boar (*Sus scrofa* L.) in Europe. **Mammal Review**, v. 16, p. 97–101, 1986.

SAEZ, M. & SAURINA, C. Estadística y epidemiología espacial. Girona: Edicions a petició. Documenta universitària. Capítulo 1: Introducció a la estadística espacial. 2007.

SANTOS, S. M. & SOUZA, W. V. Introdução à estatística espacial para a saúde pública/ Ministério da Saúde, Fundação Oswaldo Cruz. Brasília, Ministério da Saúde, 2007. 120p.

SBCF, 2010. O Javali no Brasil. 27p.

SEWARD, N. W. et al. Feral swine impacts on agricultural and the environment. **Sheep Goat Reserch Journal**, v. 19, p. 24–40, 2004

SNOW, J. *On the Mode of Communication of Cholera*. John Churchill, London. 1855.

STEVENSON, M. A. et al. Descriptive spatial analysis of the epidemic of bovine spongiform encephalopathy in Great Britain to June 1997. **Veterinary Record**, v. 147, p. 379 – 384, Set 2000.

STEVENSON, M. A. et al. Area-level risks for BSE in British cattle before and after the July 1988 meat and bone meal feed ban. **Preventive Veterinary Medicine**, v. 69, p. 129–144, Jan 2005.

SWEENEY, J. R., SWEENEY, J. M. & SWEENEY, S. W. Feral pigs. In: FELDHAMER, G.; THOMPSON, B. & CHAPMAN, D. J. *Wild mammals of North America.*, Baltimore: John Hopkins University Press Maryland, USA. 2003. P. 1164–1179.

TAYLOR, R. B. et al. Reproduction of feral pigs in southern Texas. **Journal of Mammalogy**, v. 79, p. 1325–1331, 1998.

TOBLER, W. R. Cellular geography. **Philosophy in geography**, p. 379 – 386. 1979.

VALÉRIO, L. A. J. 1999. Ocorrência e alimentação da linhagem javali (*Sus scrofa*, Mammalia, Artiodactyla) em estado silvestre no sudoeste do Rio Grande do Sul, Brasil. Tese de Mestrado Universidade Federal do Rio Grande do Sul, Porto Alegre, 56 p.

WAKEFIELD, J. C.; KELSALL, J. E.; MORRIS, S. E. Clustering, cluster detection and spatial variation in risk. In: Elliott, P., Wakefield, J.C., Best, N.G., Briggs, D.J. (Eds.), *Spatial Epidemiology – Methods and Applications*. pp. 128–152. Oxford University Press, Oxford, 2000.

WERNECK, G. L.; & STRUCHINER, C. J. Estudos de agregados de doenças no espaço-tempo: conceitos, técnicas e desafios. **Cadernos de Saúde Pública**, v. 13, p. 611 - 624. Dez 1997.