



**Laboratório de Geoprocessamento da Escola de
Arquitetura da UFMG
Profa Ana Clara M. Moura**

**A escolha de interpoladores e recursos de visualização na estruturação de
bases de dados para produção de informações espaciais apoiadas por
geoprocessamento**

Ana Clara Mourão Moura

Escola de Arquitetura, Depto Urbanismo, UFMG

Rua Paraíba 697, Savassi, Belo Horizonte – MG, CEP 30130-140, anaclara@ufmg.br

Resumo:

O artigo apresenta inicialmente uma discussão sobre o estado da arte na estruturação de bases de dados para a construção de análises espaciais apoiadas por geoprocessamento. Defende as estruturas raster e a construção de superfícies potenciais de distribuição de ocorrências e fenômenos territoriais. Apresenta explicações sobre critérios para a escolha de interpoladores e recursos de visualização segundo os objetivos da análise espacial, discutindo aplicações, restrições e potencialidades. Para desenvolver suas discussões, elege dados a partir de estudo de caso do emprego dos interpoladores de Buffer e de Kernel na distribuição de atividades de comércio e indústria na RMBH, com vistas a demonstrar que uma mesma coleção de dados pode ser tratada de diferentes maneiras, segundo os objetivos da análise espacial.

Palavras-Chave: Interpoladores, Análise Espacial, Geoprocessamento

Abstract:

The article presents a discussion about the state of the art in the structuring of data bases for the construction of spatial analysis supported by GIS. It defends raster structures and construction of surface potential distribution of occurrences and territorial phenomena. Presents explanations of criteria for the choice of interpolation and visualization features according to the objectives of spatial analysis, discussing applications, limitations and potential. The data collection to develop the discussions are chosen from a case study of the use of the interpolators Kernel and Buffer in a study of distribution activities of commerce and industry in the RMBH territory, in order to demonstrate that a single data collection can be treated in different ways, according to the objectives of spatial analysis.

Key-Words: Interpolation, Spatial Analysis, GIS

1. Introdução – o Estado da Arte na construção de bases de dados para análise espacial apoiada por geoprocessamento:

Moura, Ana Clara M. A escolha de interpoladores e recursos de visualização na estruturação de bases de dados para produção de informações espaciais apoiadas por geoprocessamento. Laboratório de Geoprocessamento da Escola de Arquitetura da UFMG. Belo Horizonte, 2012. 21 p. Apostilas. In.: <http://www.arq.ufmg.br/>

O presente artigo visa dar apoio às aplicações em geoprocessamento que prescindem da construção de bases de dados que caracterizem a distribuição territorial de atributos e fenômenos para quem a partir desta distribuição, possam ser realizadas combinações de variáveis que respondam por objetivos de análises espaciais.

A utilização dos recursos de geoprocessamento para a análise espacial requer a adoção de um roteiro metodológico que parta das discussões sobre os motivos da investigação proposta, a escolha das variáveis que respondem pelo tema, à captura e tratamento dessas variáveis que devem ser transformadas em planos de informação, a escolha dos modelos de análise espacial que irão promover a combinação das variáveis e, finalmente, aplicados os modelos de análise espacial, a verificação do grau de acerto da análise espacial.

O roteiro metodológico sugerido em nossos estudos de análise espacial apoiada por geoprocessamento segue as etapas de trabalho:

1. Clara definição dos objetivos e aplicações no uso do sistema de análise apoiada por geoprocessamento. O pesquisador deve sobre a sua proposta de investigação e verificar se possui uma pergunta norteadora que deverá ser respondida em sua investigação. A falta de clareza na definição dos objetivos pode resultar em coleta de dados desnecessários, a erros no tratamento e representação desses dados, assim como erros na escolha dos modelos de análise espacial.

2. Definição da base de dados cartográfica e alfanumérica. O pesquisador deve ter amplo domínio sobre as características do elemento ou fenômeno investigado para selecionar as variáveis que irá coletar, mensurar e representar em bases cartográficas (mapas) e alfanuméricas (tabelas). Esta escolha de variáveis pode ser apoiada por revisão bibliográfica, entrevista a diferentes especialistas ou por processos heurísticos, nos quais se aplica método intuitivo de tentativas para abordar um problema e chegar a uma solução final mediante aproximações sucessivas. (Moura, 2004).

3. Organização do SIG – Sistema de Informações Geográfico. Estruturação de conjunto de dados cartográficos e alfanuméricos para consultas e análises espaciais. Hoje temos à disposição softwares gratuitos e em português que permitem o desenvolvimento de análises espaciais complexas.

4. Conversão de escalas de medição. Os dados podem ser organizados, segundos suas próprias naturezas, em escala nominal ou seletiva, ordinal ou de razão. Moura (2003, p. 42), apresenta exemplos para se compreender esta divisão:

- a) Nominal - é qualitativa ou seletiva. Usando como exemplo o uso do solo, observa-se que os componentes não podem ser hierarquizados, a não ser segundo algum julgamento de valor. Exemplo: Mapa de Uso do Solo com as legendas divididas em cobertura vegetal, área urbana, solo exposto e cursos d água. Não se pode dizer que a cobertura vegetal seja mais ou menos importante que a área urbana, a não ser segundo um recorte classificatório como, por exemplo, avaliação do valor da terra para fins de implantação de linha de transmissão de energia elétrica. Essa escala não permite a

aplicação de operações aritméticas, pois não é possível dizer, por exemplo, que "cobertura vegetal" seja o dobro de "área urbana".

- b) Ordinal - é possível ordenar os componentes segundo grandeza ou preferência, mas, como na classe anterior, não é possível a aplicação de operações aritméticas, mas somente avaliar a frequência e a classe modal. Baseia-se na hierarquização de posições. Usando, como no exemplo, a classificação de estradas *em municipal, estadual e federal*, observa-se que não é possível dizer que uma seja o dobro da outra.
- c) Intervalo - os intervalos são conhecidos e cada observação pode receber valor numérico preciso. O ponto zero é arbitrário e não indica ausência da característica medida, como o exemplo de temperatura. É infinita em extensão e densidade entre duas posições. A utilização de operações aritméticas é limitada a transformações lineares. Podem-se aplicar técnicas de estatística paramétrica.

A transformação das escalas nominal e ordinal para intervalo ou razão possibilita a aplicação de operações aritméticas e de técnicas de estatística paramétrica. Componentes são colocados em um *ranking* segundo um olhar, ou recorte, e permitem a atribuição de valores numéricos. Assim, os dados passam por um processo de pré-processamento, visando à padronização que irá permitir os estudos de correlações de variáveis nas avaliações de fenômenos espaciais.

5. Conversão de dados vetoriais em matriciais – aplicação de processos que irão simplesmente converter as formas de representação dos dados, ou escolha de interpoladores que irão promover a geração de “superfícies potenciais” do fenômeno ou ocorrência investigada. Nos exemplos a seguir, um trecho de mapa de Lei de Uso e Ocupação do Solo (Figura 1), dados nominais ou seletivos, é simplesmente transformado de vetor para raster através da escolha da resolução (tamanho do pixel). Em outras situações é necessário planejar esta conversão para que os dados sejam representados em superfície potencial, como é o caso de dados pontuais (Figura 2) que quando têm atributos nominais podem ser representados segundo a área de influência ou o grau de concentração territorial, e quando possuem atributos em escala ordinal ou de razão (em ranking ou valores numéricos) podem ter suas distribuições espaciais considerando, também, a massa ou força do ponto.



Figura 1 – Mapa de Lei de Uso e Ocupação do Solo - Mapa de dados nominais ou seletivos, representação vetorial e matricial (raster).

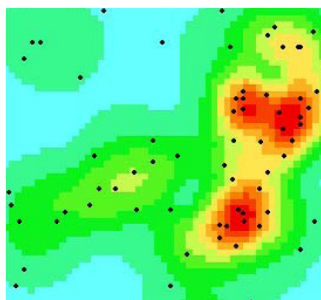


Figura 2 – Dados pontuais transformados em superfícies de distribuição espacial.

A maioria das coleções cartográficas para estudos urbanos e regionais é estruturada em formato vetorial, em pontos, linhas e polígonos, o que é geralmente denominado “geo-objetos” aos quais são associadas tabelas alfanuméricas contendo informações qualitativas e quantitativas da ocorrência (Câmara et al., 2005). Ainda são poucas as informações que se apresentam em formato raster ou matricial, mais comum em dados ambientais oriundos de imagens de satélite ou de representações da topografia do terreno.

Na estrutura *raster*, a base é uma matriz de pontos (“*grid*”), sendo cada ponto um *pixel*. Um *pixel*, ou “*picture element*” é a menor unidade de representação da grelha. A relação entre a dimensão da célula de representação e a área que ela representa na realidade dá a *resolução* empregada. A resolução é a capacidade de distinção entre entidades e eventos contíguos nas dimensões espacial, temporal ou taxonômica. DPI (“*dots per inch*”) é uma medida de densidade de rastreamento, enquanto que a resolução espacial é uma medida de área. O termo resolução pode ser aplicado ao espaço geográfico, ao espaço taxonômico (de atributos) e ao tempo.

Por trabalhar com uma superfície discretizada, a estrutura *raster* utiliza a geometria digital. Tanto o ponto, como a linha, como uma área são representados por um conjunto de *pixels*. Sobre as informações contidas em cada quadrícula, “*pixel*”, ou unidade mínima, Teixeira *et al.* (1992, p.17) mostram que “...ocorre um processo de generalização onde os vários elementos que podem constituir uma quadrícula deixam de ser individualizados”. A aceitação dessa generalização irá depender dos objetivos na utilização dos dados espaciais. A resolução empregada deve responder às expectativas de posicionamento e de dimensão dos elementos retratados.

Contudo, desde a década de 90 autores importantes do geoprocessamento já anunciavam que o futuro da análise espacial caminharia para os trabalhos em matrizes, em formato raster TOMLIN (1990:44): “*yes, raster is faster, but raster is vaster, and vector just seems more correcter*”.

Esta afirmação de Dana Tomlin merece explicações, pois é uma frase emblemática que, ao ser defendida ainda no século passado, causou questionamentos, e hoje vemos como afirmativa totalmente pertinente. Começemos pela afirmativa do meio: o raster é mais vasto – disso não há dúvida, pois o armazenamento da informação em formato de imagens (*tif* ou equivalentes) ou de matrizes (*grids* com informações pontuais onde cada posição recebe necessariamente uma informação, ainda que seja a de “não há ocorrência do fenômeno ou fato nessa posição”) é bem vasto que o armazenamento de informações espaciais na forma de vetores que, por sua vez, são compostos por pontos (linha de comprimento zero), linhas (conjunto de segmentos de

reta interligados por vértices) e polígonos (conjunto de segmentos de rede onde o vértice final de um dos elementos do conjunto coincide com o vértice inicial do segmento mais próximo, configurando superfície). Em formatos vetoriais para cada primitiva gráfica é associado um registro no banco de dados. Por outro lado, no formato matricial para cada célula do conjunto é armazenada uma informação.

Sobre a afirmativa da autora de que o raster é mais rápido, é comum se perguntar os motivos, uma vez que só para visualizar uma imagem o tempo necessário já é maior. Sim, por inspeção visual seria complexo, mas quem analisa, combina, e coteja os dados é o processo de varredura do computador. E sendo assim, varrer matrizes e promover combinações de valores contidos nos pixels ou células é bem mais facilitado que promover as relações de análises topológicas (interseção, união, diferença, soma) entre formas geométricas vetoriais. Observe a Figura 3. É muito mais lógico para um processo matemático computadorizado combinar os valores dos pixels entre as camadas das informações, representadas cada uma delas como uma matriz.

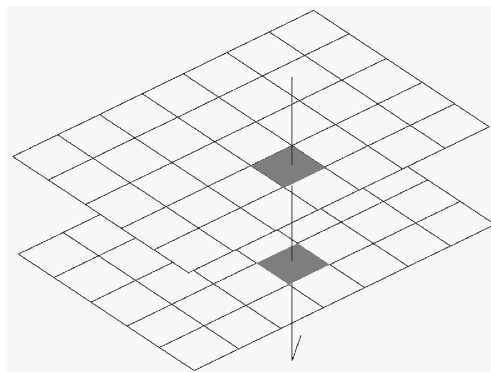


Figura 3 – Combinação e cotejo pixel a pixel - estudo e sobreposição de planos de informação

Sobre a questão de que o vetor “só se parece mais correto”, há sempre muita reação a esta afirmativa, pois é comum pensarmos que a linha vetorial, por ser mais exata, é mais correta. Verifica-se um forte apego às referências da cartografia analógica ou tradicional, pois se espera do produto final um "desenho" que se assemelhe a um produto antes elaborado pelo homem. Moura (1999, p. 60-61), ao abordar a questão da comunicação visual e das linguagens utilizadas pela cartografia digital, faz uma relação entre a representação *raster* e o modo Impressionista na pintura, particularmente o Pontilhismo (Figura 4). Segundo a autora, as pessoas se sentem mais confortáveis com o desenho vetorial porque ele se aproxima da linguagem de desenho que utilizam para representar uma forma qualquer. Quando vamos desenhar uma simples flor, o fazemos através de suas linhas de contorno, e não através dos micro pontos que compõem sua imagem. Um momento da linguagem de desenho em que essa forma de olhar o mundo foi alterada ocorreu no Impressionismo, quando os artistas passaram a representar as imagens através da justaposição de pequenos pontos, que deveriam ser vistos a certa distância para que conformasse o todo. Pode-se dizer que os impressionistas discretizavam a imagem para recompô-la através da pintura. O processo exige capacidade de análise e síntese, mais complexas para a linguagem de desenho, mas absolutamente possível para a tecnologia do geoprocessamento.

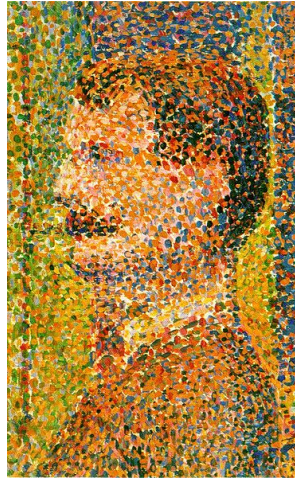


Figura 4 - Exemplo de Potilismo – discretização na representação gráfica – Quadro “La Parade” de Georges Seurat (1889). Fonte: Wikipedia.

O vetorial é, muitas vezes, considerado mais correto porque se aproxima da linguagem convencional de desenho, baseada em linhas. Contudo, essas linhas são generalizações cartográficas que têm sua correção geométrica limitada a diversos fatores, tais como a escala, os processos de interpolação, resolução taxonômica, entre outras. A estrutura *raster*, por sua vez, apresenta a correção definida pela resolução adotada pelo *pixel*.

A opção por sistema *raster* ou sistema vetorial significa a opção por diferentes *métodos* para a estruturação de dados. Muito já foi discutido sobre vantagens e desvantagens de uma estrutura ou outra. A questão principal refere-se à capacidade de armazenamento das informações. A memorização de dados *raster* é feita como se o espaço cartografado fosse "quebrado" em unidades discretas, e cada uma dessas unidades é registrada. No caso dos dados vetorizados são memorizadas as coordenadas dos nós (ou vértices) dos elementos gráficos e as conexões e informações para reconstruir objetos complexos. Na estrutura vetorial, como são memorizados apenas os extremos do segmento, a ocupação de memória é inferior. Observa-se, contudo, uma rápida evolução nos algoritmos de compactação, bem como do suporte *hardware* para o armazenamento de dados.

Segundo Câmara (2005), o trabalho em matrizes é denominado representação na forma de “geo-campos”, pois para cada célula representada existe, necessariamente, uma informação sobre o grau de existência espacial do fenômeno ou ocorrência mapeado naquela posição. O mesmo autor apresenta os conceitos de geo-objetos para estruturas vetoriais e geo-campos para estruturas matriciais.

Dentro desse mesmo raciocínio, pode-se dizer que as estruturas *raster* são mais orientadas para a posição, enquanto as estruturas vetoriais são mais orientadas para o tema. No *raster*, armazenam-se características associadas a localizações, enquanto no vetorial registram-se localizações associadas a características. Defendemos o conceito de que trabalhar em formato raster para promover, por diferentes modelos de análise espacial, a combinação de matrizes, significa desenvolver formas de espacializar as variáveis em “superfícies potenciais” de distribuição da ocorrência ou fenômeno de interesse.

Xavier da Silva (2001) explica a mudança de paradigma de uso de dados vetoriais, utilizados a partir de inspeções pontuais e generalizações a partir da investigação, para se trabalhar com o processo de varredura e integração locacional, que ele denomina “vail”, que promove a análise de dados pixel a pixel e integração das camadas ou matrizes de dados.

Citamos um exemplo que sustenta o nosso argumento: a apresentação de um mapa de distribuição de supermercados em um município por si só é apenas uma coleção de dados georreferenciados. Para os dados se transformarem em informação, que dê apoio à tomada de decisões, nos interessa saber não apenas que em um determinado local “tem” ou “não tem” supermercado, mas sim até onde vai a área de influência de um supermercado, se uma determinada posição no território é bem atendida por supermercados, se existem sobreposições de atuações de supermercados diferentes em uma região da cidade gerando concorrência, entre outras questões. Assim, cada posição de uma matriz da região de estudo deve representar um valor qualitativo ou quantitativo, na forma de uma superfície potencial, indicando o grau de relação com a forma de distribuição da variável no território. Isto significa trabalhar em formato raster, em geo-campos, e não apenas em formato vetorial, em geo-objetos.

Xavier-da-Silva (2009) também defende este conceito, que ele denomina “geoinclusão”, quando argumenta: “No Geoprocessamento são tratados enormes volumes de dados, exigindo “técnicas computacionais” e disponibilizados atributos geotopológicos para fins de análises, sínteses e utilização imediata no planejamento ambiental e na gestão territorial. Tornam-se reveláveis atributos espaciais dos fenômenos, tais como: localizações sistemáticas ou eventuais; extensões de ocorrência e respectivos níveis diversos de intensidade; formas e padrões de distribuição espacial; níveis de proximidades geográficas, de tempo e de custo; relacionamentos hierárquicos e funcionais de inúmeras naturezas, a serem usados em classificações ambientais, em simulações sinérgicas e na elaboração de cenários prospectivos.”

Oliveira *et al.* (1994, p. 35-36) também defendem a adoção das estruturas *raster*, alegando: *"One of the most important operation on a GIS is overlay, which is facilitated if the maps are on a raster format. The raster represents the data in a more realistic (although less aesthetically pleasing) than its vector counterpart. Zooming operations in raster show the data at most basic unit: the pixel. In opposition, a vector system zoom maintains the topology at any scale, although some approximation and manipulation is always executed but under a sort of hidden form."*

Assim, significa etapa importante do preparo dos dados, para que sejam tratados de modo a representarem informações. Isto significa transformá-los de formato tabular (dados alfanuméricos) ou vetoriais (pontos, linhas e polígonos) para o formato matricial (raster). Em algumas situações o processo é a simples conversão de formatos, escolhendo a delimitação da área de trabalho e a resolução adequada (dimensão do pixel), mas em outros casos cabe a escolha de recursos de gradeamentos e interpoladores, utilizados quando o tratamento de dados requer uma atribuição de valores para todos as posições do território, ou seja, a construção de uma superfície potencial.

Bonham-Carter (1994) e Xavier-da-Silva (1999a) citam como exemplos de gradeamentos e interpoladores:

Moura, Ana Clara M. A escolha de interpoladores e recursos de visualização na estruturação de bases de dados para produção de informações espaciais apoiadas por geoprocessamento. Laboratório de Geoprocessamento da Escola de Arquitetura da UFMG. Belo Horizonte, 2012. 21 p. Apostilas. In.: <http://www.arq.ufmg.br/>

- Triangulação de Delaunay.
- Raio de busca para uso da influência dos pontos na área.
- Geoestatística utilizando a krigagem para gerar superfícies de tendência e levar em consideração os arranjos espaciais.
- Trabalho com escala ordinal possibilitando identificar os níveis de correlação entre diversos pares dos planos de informação por meio do coeficiente de correlação de Spearman.
- Trabalho com escalas de intervalo e razão possibilitando utilizar o coeficiente paramétrico de Pearson, definindo índices de sua validade (covariância em relação à variância total), a partir de distribuições de frequência consideradas normais ou normalizadas.
- Nas escalas nominal e ordinal, nas quais as operações numéricas são restritas, o ideal é utilizar "o vizinho mais próximo".

Cabe, então, com o objetivo de dar apoio a esta importante etapa de preparo da coleção de dados e produção das primeiras informações da análise espacial, discutir as condições de construção das estruturas raster e a escolha de interpoladores.

2. A escolha de interpoladores

A escolha de interpoladores ocorre no momento em que se decide transformar dados vetoriais ou alfanuméricos em dados matriciais para composição da superfície potencial de distribuição do fenômeno ou ocorrência espacial. Ela parte da seguinte pergunta: Como transformar dados cuja captura ou representação se localiza em pontos específicos do território para sua representação em estudos de espacialização/territorialização dos fenômenos e ocorrências, representados por fatores ou objetos de nosso interesse? A resposta está na construção de superfícies potenciais que indicam o grau de concentração ou o modo de distribuição dos fenômenos/ocorrências, através da aplicação de interpoladores.

Apresentaremos aspectos de alguns interpoladores com o objetivo de demonstrar que um usuário mal informado sobre suas aplicações pode fazer a escolha errada para espacializar as suas informações. Os exemplos foram elaborados a partir de uma mesma coleção de dados, coleção simples de pontos apenas para a realização da demonstração, e é importante observar que os resultados da aplicação de cada interpolador são bastante diferentes.

2.1. Kernel

O Kernel é indicado para se representar o grau de concentração de um fenômeno ou tipo de ocorrência territorial. Segundo ESRI (2008), *"Kernel Density calculates the density of features in a neighborhood around those features."*

O interpolador considera o arranjo espacial: o agrupamento de pontos ou ocorrências por área e pode ser ponderado por algum dado da tabela. Citando como exemplo o caso de interesse de distribuição de escolas em um determinado município, se o pesquisador tiver apenas os dados de localização espacial das escolas, o interpolador irá gerar a superfície potencial que indica as áreas onde há maior concentração de escolas e as áreas onde não há ocorrências de escolas. Por outro lado, se o pesquisador tiver uma tabela associada aos pontos das escolas contendo informações sobre número de vagas, por exemplo, o interpolador pondera e atribui mais peso

para as áreas onde, além da concentração de escolas, há também maior concentração de vagas. (Exemplo Figura 5).

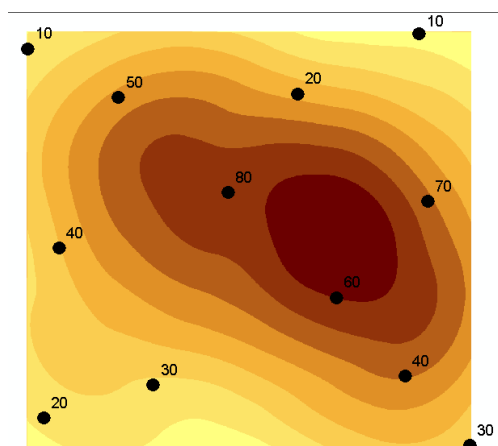


Figura 5 – Exemplo de aplicação de Kernel considerando a concentração de pontos e a massa ou valor do atributo no ponto.

Inicialmente, cabe perguntar: o que eu quero responder com a investigação? Muda se forem escolas, centros de saúde ou bancos? Sim, muda. Se objetivo for verificar se estou bem atendida por serviços de saúde, basta investigar se estou na área de influência de algum deles, pois basta um (não vou tomar a vacina mais de uma vez...). Nesse caso, o aplicativo para espacialização é um simples Buffer (que desenha geometricamente um raio de ação nas redondezas do elemento de localização) ou um Voronoi (que considera a força da massa do ponto, tal como o número de atendentes ou de especialidades atendidas no posto de saúde, e os atritos ou impedâncias espaciais para se chegar àquele ponto). (Figura 6). Para saber um pouco mais sobre os modelos de área de influência na forma de Buffer ou Voronoi, indicamos o artigo publicado no ENABER de 2009 (Moura, 2009).

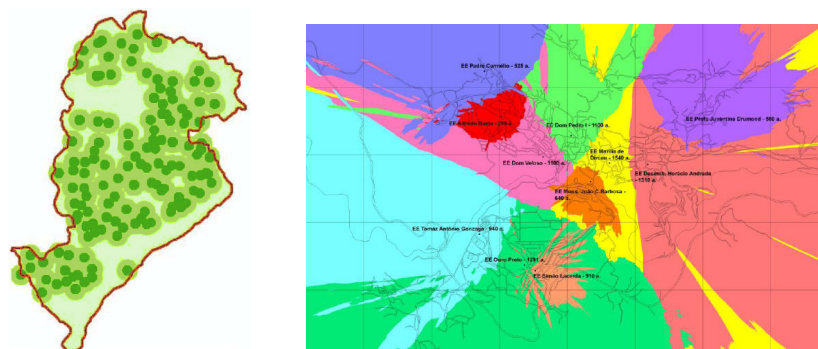


Figura 6 – Exemplo de Buffer nas proximidades de centros de saúde me Belo Horizonte e exemplo de Voronoi no estudo de área de influência de escolas em Ouro Preto considerando o número de vagas como massa e as impedâncias territoriais de acessibilidade como atrito.

Contudo, se o objetivo for investigar a facilidade de acesso a serviços de bancos, por exemplo, não cabe realizar o estudo por Buffer ou Voronoi, porque de nada adianta eu estar bem próxima de um determinado banco se eu não sou correntista dele. Nesse caso, me interessa mais estar em área onde há concentração de bancos, pois aumentam as chances de eu conseguir o serviço. Quando a opção for pela concentração da ocorrência, indica-se o Kernel.

Os dados pontuais de distribuição de serviços urbanos, tais como escolas, centros culturais ou outros de mesma natureza, em geral são trabalhados apenas segundo a distribuição e concentração territorial. Se o pesquisador estruturar também tabelas associadas a esses elementos contendo suas características ou atributos, além do interpolador indicar a concentração do serviço por agrupamentos territoriais, a superfície de distribuição gerada responde também pela qualidade do serviço, ponderada por algum atributo da tabela. Por exemplo, além da concentração de bancos, será levado em consideração o número de caixas de auto-serviço.

Cabe ainda refletir sobre o seguinte exemplo: pode ocorrer de existir um agrupamento de escolas em uma dada região, mas todas com poucas vagas, ao passo que em outra região existe uma única escola com muitas vagas. Nesse caso, se existirem tabelas associadas com a informação sobre a massa (número de vagas), a superfície de distribuição indicaria que tanto uma região como a outra é importante. Por outro lado, sem a referida informação sobre o número de vagas, a superfície potencial destacaria apenas onde há concentração de escolas.

Além de decidir sobre a resposta que atende aos objetivos da investigação, é importante o conhecimento específico para se decidir sobre as métricas. Do que se trata: o Kernel faz a superfície de distribuição mensurando ocorrências a partir de um raio de busca. No exemplo citado, poderíamos fazer uma primeira simulação indicando um raio de busca de 500 metros, arbitrando que esta é uma distância ideal máxima para crianças caminharem a pé para a escola. Segundo a Organização Mundial de Saúde, indicam-se quatro quadras para uma criança carregar mochila, mas cabe investigação, a partir do conhecimento especialista sobre as condições do local. Os valores seriam diferentes, por exemplo, para o deslocamento de um adolescente ou de uma criança? Do deslocamento para uma atividade cultural ou para um posto de saúde? A cada investigação é necessária escolher a métrica mediante amplo conhecimento de fenômeno estudado.

2.2. Inverso do Quadrado da Distância (IDW – Inverse Distance Weighted)

O interpolador IDW atribui valores para áreas não mensuradas a partir da captura de dados de suas vizinhanças, de modo que os valores mais próximos de sua área de interesse têm mais influência sobre a decisão. Isto significa que se for utilizada a distribuição de escolas no território e se o objetivo for saber como estão as condições de atendimento em relação a este serviço em uma determinada localidade no território, a resposta é influenciada pelos valores dos vizinhos mais próximos.

Segundo ESRI (2008): *“IDW interpolation explicitly implements the assumption that things that are close to one another are more alike than those that are farther apart. To predict a value for any unmeasured location, IDW will use the measured values surrounding the prediction location. Those measured values closest to the prediction location will have more influence on the predicted value than those farther away. Thus, IDW assumes that each measured point has a local influence that diminishes with distance. It weights the points closer to the prediction location greater than those farther away, hence the name inverse distance weighted.”* Explicam ainda: *“IDW assumes that the surface is being driven by the local variation, which can be captured through the neighborhood.”*

A aplicação deste interpolador faz sentido quando existe uma tabela de dados associada aos pontos, contendo, por exemplo, o número de vagas na escola. Assim o interpolador responde, a partir desta tabela, qual é o grau de atendimento que cada porção do território apresenta. Não pode ser utilizado quando não há atributos em tabela para se ponderar, como é o caso da localização simples de escolas. (Figura 7).

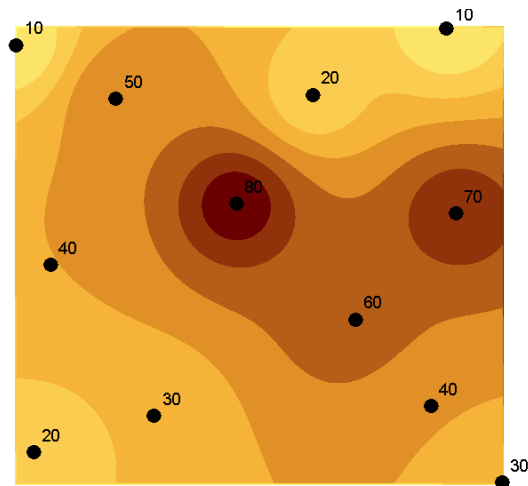


Figura 7 – Exemplo de aplicação do Inverso do Quadrado da Distância considerando a posição de cada pontos e a massa ou valor do atributo no ponto para responder sobre valor do serviço em posições do território.

É muito comum, em estudos econômicos urbanos e regionais, o emprego de dados agrupados por setores, como é o caso de setores censitários ou áreas homogêneas. A dificuldade em trabalhar com estes recortes poligonais é que eles não têm justificativa espacial para sua definição, mas seguem decisões tais como o número de questionários a serem aplicados. Observa-se, por exemplo, setores muito grandes territorialmente e setores muito pequenos, e ao se fazer um mapa temático de distribuição de uma variável, como exemplo a renda, ambos podem ser classificados em uma mesma faixa, mas o resultado visual irá induzir a compreensão de que a superfície territorial maior é mais importante. Na verdade, a superfície maior é uma área menos densa na qual o somatório de valores a equiparou com uma área mais densa e menor territorialmente.

O interpolador IDW se aplica bem a esses casos, pois ele faz um estudo das vizinhanças e atribui, para cada célula da região (ou pixel de saída da matriz ou mapa raster), um valor de potencial de distribuição da variável escolhida em função dos valores dos vizinhos mais próximos. É um modelo que trabalha tanto com a organização espacial dos pontos como com os valores de atributos nesses pontos.

É importante destacar que para aplicar o procedimento para dados armazenados em superfícies (polígonos), como é o caso de setores censitários ou outros, antes é necessário tratar o dado para que cada polígono seja representado pelo seu centróide e a este centróide sejam associados os dados contidos na tabela, antes associada aos polígonos (Figura 8).

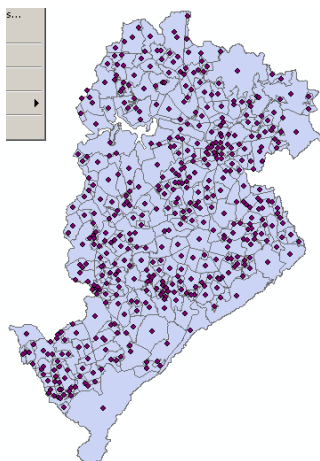


Figura 8 – Dados associados a polígonos (exemplo de setores censitários) representados por seus centróides, para aplicação de interpolador.

Os motivos de se usar o IDW e não o Kernel Ponderado está no fato que o Kernel é indicado quando o interesse é densidade de distribuição de alguma ocorrência, enquanto o IDW é indicado quando o interesse é a área de influência dos valores e pontos, e a indicação de qual é o valor resultante em cada ponto do território. Mais uma vez, segundo o conhecimento especialista do pesquisador e dos objetivos da análise, define-se o interpolador mais indicado.

2.3. Splining

Apresenta a lógica da triangulação de Delauney. Força a criação de uma superfície regular de distribuição das ocorrências. É muito útil em dados ambientais tais como temperatura e topografia, mas *não* é indicado para dados sócio-econômicos, uma vez que parte do pressuposto que entre dois valores necessariamente existirão intermediários regularmente distribuídos (entre uma escola com 50 vagas e outra escola com 30 vagas existiria uma com 40 vagas, o que não se sustenta).

Além de não se aplicar para dados sócio-econômicos, precisa necessariamente de tabela de valores associada. No exemplo apresentado na Figura 9 observa-se que o aplicativo criou isolinhas de distribuição regular de áreas de influência a partir das informações de quantitativas existentes na tabela, como se o fenômeno possuísse distribuição regular no território.

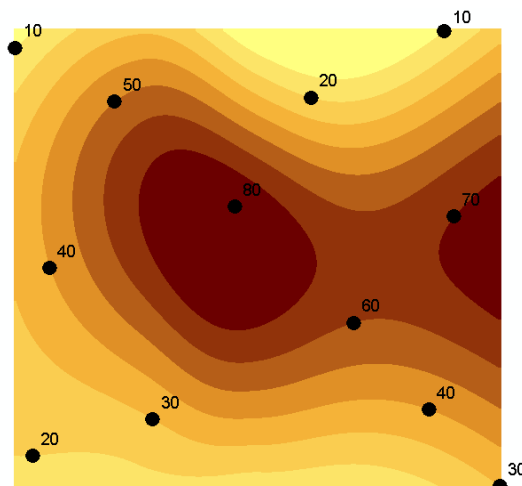


Figura 9 – Exemplo de aplicação de Splining – baseado na triangulação de Delauney. Observa-se a distribuição regular de áreas de influência.

O interpolador não se aplica aos dados sócio-econômicos de interesse dos estudos econômicos urbanos e regionais. Assim, o pesquisador deve procurar saber, ao escolher um software para trabalho e ao solicitar a aplicação de um interpolador, qual é o procedimento que o aplicativo utiliza.

2.4. Krigagem

Usado para quando se quer atribuir valores para posições do mapa não mensurados, a partir do comportamento espacial da distribuição do dado. Ele apresenta uma “direção” ou “espacialização” provável para a distribuição de um fenômeno a partir do comportamento dos dados já mensurados. É muito usado, por exemplo, em estudos de mineração e geologia para, a partir de um conjunto de amostras, identificarem a provável espacialização de um veio de um material de interesse. (Figura 10).

Destaca-se que é uma ferramenta para investigação de possíveis respostas, mais do que para se obter uma resposta definitiva e exata a partir da primeira interpolação.

Segundo ESRI (2008): “*Kriging is an advanced geostatistical procedure that generates an estimated surface from a scattered set of points with z-values. Unlike other interpolation methods, Kriging involves an interactive investigation of the spatial behavior of the phenomenon represented by the z-values before you select the best estimation method for generating the output surface.*”

The following is an introduction to the procedure of using Kriging to produce an estimated surface. Kriging is a complex procedure that requires greater knowledge about spatial statistics than can be conveyed in this command reference. Before using Kriging, you should have a thorough understanding of its fundamentals and have assessed the appropriateness of your data for modeling with this technique. If you do not have a good understanding of this procedure, it is strongly recommended that you review some of the references listed at the end of this topic.

Kriging is based on the regionalized variable theory that assumes that the spatial variation in the phenomenon represented by the z-values is statistically homogeneous throughout the surface (for example, the same pattern of variation can be observed at all locations on the surface). This hypothesis of spatial homogeneity is fundamental to the regionalized variable theory.”

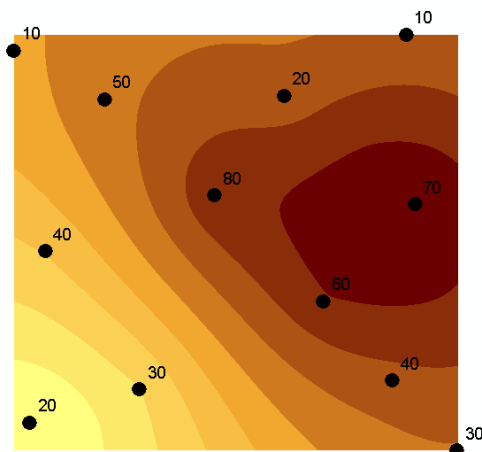


Figura 10 – Aplicação da Krigagem na mesma coleção de dados apresentada nas figuras anteriores.

No caso de aplicações em dados sócio-econômicos indica-se profundo conhecimento de suas características, variabilidades e comportamentos, pois a Krigagem irá atuar como ferramenta de apoio à investigação, mas será o especialista quem conduzirá as diferentes simulações e será capaz de responder se, de fato, o modelo representa a distribuição que ele está estudando. Requer amplo processo de calibração e validação, em investigação de aproximações com a realidade representada, segundo diferentes olhares e recortes.

É importante lembrar que qualquer modelo é um recorte da realidade conduzida por um objetivo, um olhar, uma definição espacial e temporal. O modelo não é a realidade, mas apenas uma de suas facetas, escolhida pelo investigador, segundo os objetivos de sua busca. Se o processo de modelagem for assim compreendido, não como uma resposta cabal e conclusiva, mas apenas como uma escolha segundo o recorte metodológico do investigar, a Krigagem pode ser, sim, uma ferramenta muito útil como interpolar.

2.5 – O fatiamento das legendas

Uma vez aplicados os interpoladores, cabe ainda ao pesquisador decidir pela forma de visualização dos dados em componentes de legenda, de modo a escolher os limites que comporão as faixas de representação. Em geoprocessamento se chama este processo de “simbolização da informação por fatiamento”. Os aplicativos já apresentam sugestões de como realizar esses recortes, sendo os mais conhecidos os valores únicos, os intervalos iguais, as quebras naturais e os quantis.

Para que se entenda a importância de se escolher corretamente o modo de fatiamento, em função dos objetivos da investigação, apresentamos na Figura 11 apresenta a mesma coleção de dados, a distribuição de renda em Belo Horizonte segundo dados do IBGE de 2000, por setor censitário. Escolhemos, a título de exemplo, apenas os fatiamentos de intervalos iguais, Moura, Ana Clara M. A escolha de interpoladores e recursos de visualização na estruturação de bases de dados para produção de informações espaciais apoiadas por geoprocessamento. Laboratório de Geoprocessamento da Escola de Arquitetura da UFMG. Belo Horizonte, 2012. 21 p. Apostilas. In.: <http://www.arq.ufmg.br/>

quebras-naturais e quantil. Observa-se que a divisão em intervalos iguais, no caso em específico, retrata o que é a realidade de grande concentração de ocorrências na baixa renda, assim como de baixa ocorrência na alta renda – o histograma demonstra que a primeira faixa reserva quase todo o conjunto. No fatiamento por quebras-naturais (*natural breaks*) a separação das faixas identifica onde há mudanças de padrão e agrupa sub-conjuntos que possuem características semelhantes, sendo aplicável quando este for o objetivo da investigação. O terceiro fatiamento, por quantil, realiza a distribuição de ocorrências no número de faixas indicado procurando colocar número semelhante de indivíduos os registros em cada faixa. No caso de aplicação na variável renda, só se destinaria a, por exemplo, a necessidade de atender à população separada por rendas diferentes e conseguir o atendimento de um número equivalente de registros por vez.

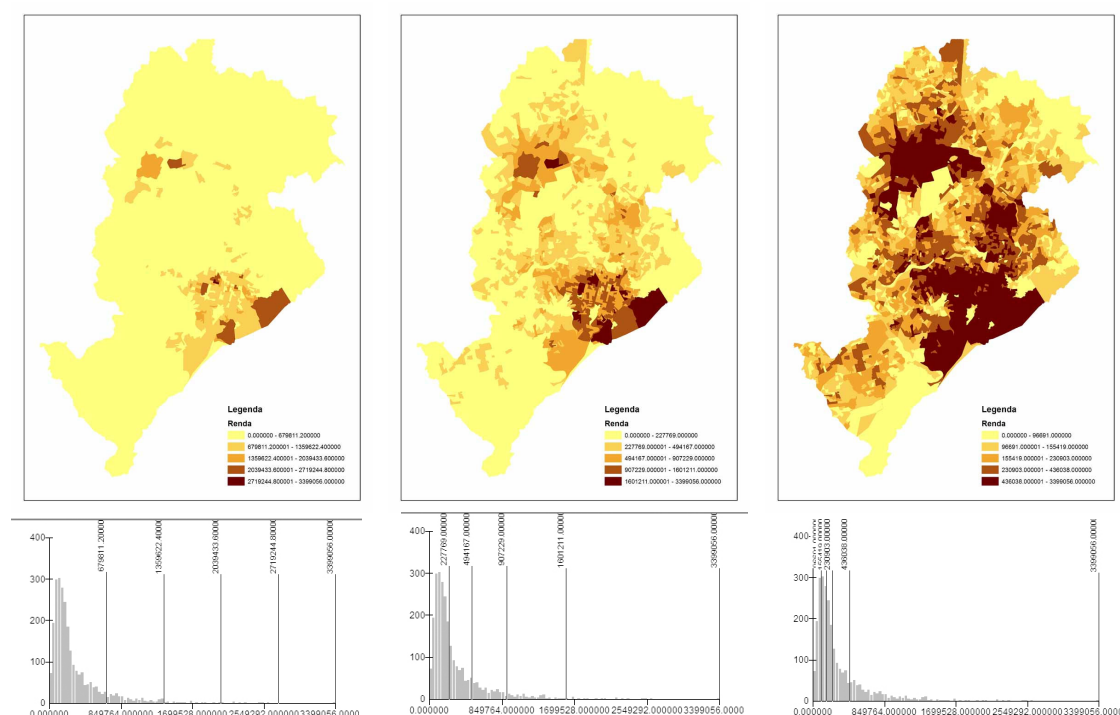


Figura 11 – Coleção de dados de renda por setor censitário apresentada por fatiamentos de intervalos iguais (a) e seu respectivo histograma, por quebras-naturais (b) e seu respectivo histograma e por quantil (c) e seu respectivo histograma.

Sobre esta questão de como manusear os dados e sobre as diferentes respostas visuais que se pode obter pelo tratamento da informação, indicamos a leitura do livro “How to lie with maps” (Monmonier, 1996), que nos chama a atenção para ter claro o objetivo de investigação para a escolha da visualização, assim como se pautar pela ética na escolha do modo de difusão da informação.

3. Estudo de caso de espacialização de postes de energia elétrica para indicação de concentração do uso antrópico do território – RMBH

A elaboração do mapeamento de manchas urbanas em escala regional, como o necessário para o trabalho aqui apresentado, é classicamente realizado através de processamento digital de

Moura, Ana Clara M. A escolha de interpoladores e recursos de visualização na estruturação de bases de dados para produção de informações espaciais apoiadas por geoprocessamento. Laboratório de Geoprocessamento da Escola de Arquitetura da UFMG. Belo Horizonte, 2012. 21 p. Apostilas. In.: <http://www.arq.ufmg.br/>

imagens de satélite de média resolução ou de vetorização a partir de imagens de melhor resolução. Os estudos foram iniciados testando essas fontes e os resultados que poderiam ser obtidos. Contudo, para estudo de uso e ocupação do solo, crescimento das manchas urbanas, cálculo de métricas das formas da ocupação, entre outros, seria necessário obter representações espaciais capazes de identificar as nuances de suas morfologias, tais como os braços de eixos, as formas conformadas pelos agrupamentos, enfim: um nível mais bem detalhado que o promovido pela generalização das classificações de imagens de satélite.

3. 1. Criação das manchas de ocupação antrópica no território

A proposta foi a de criar uma metodologia de aproveitamento de dados pontuais a partir de postes de energia elétrica da CEMIG (Companhia Energética de Minas Gerais) e explorar suas potencialidades para agregar condições de interpretação da ocupação humana no espaço a partir dos dados. A coleção de dados pontuais de postes e seus atributos de uso fazem parte do Projeto GEMINI da CEMIG, aos quais aplicamos modelos de transformação de dados pontuais em manchas de agrupamento do serviço, o que conforma a mancha atropizada na área de estudo, ou seja: a representação mais próxima da real mancha de ocupação urbana que se pode obter (Figura 12 e Figura 13).



Figura 12 – Nuvem de pontos de postes de fornecimento de energia elétrica na RMBH e Detalhe da nuvem de postes de fornecimento de energia elétrica na RMBH Projeto GEMINI – CEMIG.

FID	Shape	FUSO	ABSCISSA	ORDENADA	CLASSE	QUANTIDADE
65514	Point	23	590330	7783229	Residencial	72
65515	Point	23	593380	7757760	Rural	1
65516	Point	23	610084	7815627	Residencial	63
65517	Point	23	611550	7795886	Comercial	938
65518	Point	23	604779	7762675	Residencial	3
65519	Point	23	605228	7810910	Residencial	72
65520	Point	23	611739	7825931	Residencial	1
65521	Point	23	604120	7761870	Residencial	6
65522	Point	23	609967	7793612	Residencial	14
65523	Point	23	605693	7811316	Residencial	18
65524	Point	23	610255	7795845	Industrial	1
65525	Point	23	559287	7788892	Industrial	3
65526	Point	23	606567	7792050	Comercial	1
65527	Point	23	615399	7802717	Residencial	16
65528	Point	23	630391	7810605	Residencial	16
65529	Point	23	585151	7793516	Residencial	4
65530	Point	23	609942	7806020	Residencial	42
65531	Point	23	601528	7794615	Comercial	4

Figura 13 – Detalhe da base de dados alfanumérica associada aos postes, contendo informações relativas ao tipo de uso (uso rural, residencial, comercial, industrial, público, entre outros); assim como a quantidade de ligações, por tipo de ligação, por poste.

O argumento se baseia no fato de que o serviço de energia elétrica em Minas Gerais é muito bem distribuído, chegando onde há ocupação humana do território, independente de ser área rural ou urbana. Destaca-se que a área de estudo se localiza na Região Metropolitana de Belo Horizonte, aonde chega eletrificação em praticamente 100% de onde há ocupação humana. O mesmo não pode ser dito sobre o mapeamento de qualquer outra infra-estrutura, pois não há este rígido controle da informação por geoprocessamento de qualquer outra variável, seja ela rede de água, esgotamento sanitário, georreferenciamento de unidades domiciliares, entre outras.

O banco de dados de poste separa o serviço em área rural e urbana, o que nos permitiu fazer a mancha da área antropizada total, a área antropizada em espaço rural e a área antropizada em espaço urbano. Cada poste recebe também o registro de número de ligações por atividade de comércio, indústria, uso institucional e residencial.

De acordo com diferentes objetivos da investigação foram aplicados dois interpoladores: o Buffer para construção da mancha de ocupação antrópica e o Kernel para construção da superfície potencial de distribuição e concentração de atividades de comércio e de atividades industriais na região.

Assim, com o objetivo de gerar a mancha de ocupação antrópica, a partir da nuvem de postes foram aplicados *buffers* para definição das áreas de influência de cada poste, definindo assim uma mancha de distribuição do serviço. Foram utilizados buffers de 50 e de 100 metros, definindo assim uma área de alcance máximo para a influência do poste. Como a escala de trabalho era a região metropolitana, observou-se que a mancha resultante da soma dos buffers de 100 metros de cada poste era bastante adequada às investigações. Destaca-se que a definição deste buffer seguiu indicação técnica da empresa responsável pelo fornecimento de energia elétrica de que a infra-estrutura implantada pode atender até esta distância territorial. Mapeada a mancha, realizou-se então procedimento de filtragem do resultado, para se trabalhar apenas com áreas maiores que a influência de um único poste isolado, e concentrar os cálculos, métricas e observações nos aglomerados da ocupação antrópica.

O estudo de distribuição dos postes de energia elétrica resultou no mapeamento da mancha da área antropizada na RMBH, partindo-se do pressuposto de que nesta região onde há ocupação

há, certamente, energia elétrica. Foram mapeados separadamente e conjuntamente ocupação antrópica rural, urbana, e na ocupação urbana a distribuição dos tipos de uso. Na Figura 14 está representada a mancha de postes que servem à área urbana, com todos os tipos de uso.

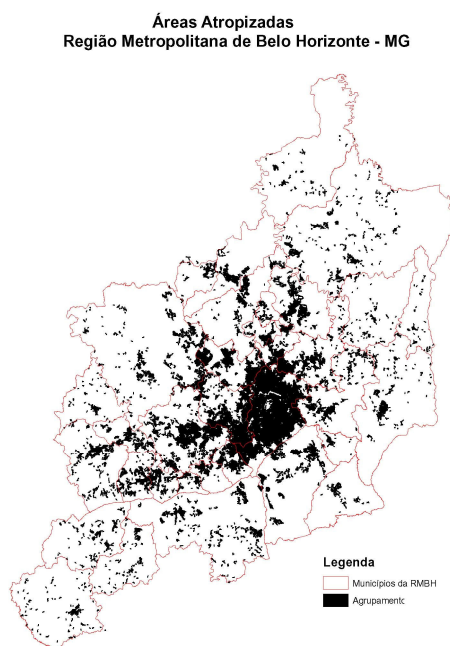


Figura 14 – Manchas urbanas na área de estudo - RMBH

Muitas foram as aplicações dos resultados obtidos, entre os quais citamos o trabalho do grupo MOM (Morar de Outras Maneiras) da Escola de Arquitetura, que partiu dessa espacialização para o estudo mais detalhado da vacância imobiliária na região. A equipe do Laboratório de Geoprocessamento da mesma escola elaborou o estudo de vacância da ocupação territorial, em escala menos detalhada que a domiciliar.

Uma forma que o estudo permite a identificação da vacância territorial é o comparativo da mancha antropizada a partir dos postes com a mancha de área antropizada obtida por classificação de imagens de satélite. Onde a classificação da imagem indica área urbana ou antropizada e não há manchas de poste, isto significa que houve alteração da paisagem através de implantação de novas glebas ou de solo exposto, mas isto não foi seguido de implantação de energia elétrica. Na RMBH este tipo de combinação significa vacância na ocupação, uma vez que onde há pessoas há luz (Figura 15).

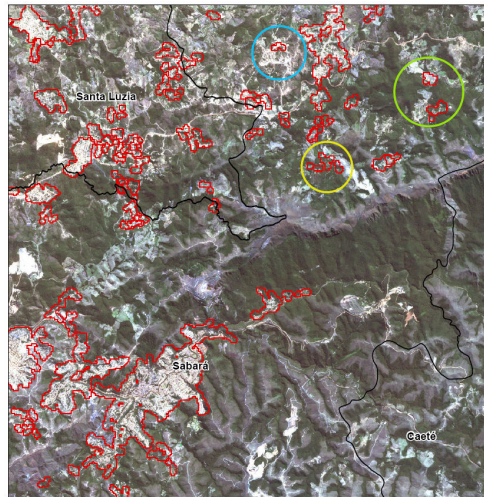


Figura 15 – Imagem Rapid Eye e manchas de áreas antropizadas geradas pelo agrupamento de postes (em vermelho).

3.2. A construção da superfície potencial de concentração das ocorrências territoriais

Com o objetivo de identificar as áreas de maior concentração de atividades de comércio e indústria e, através desta superfície de concentração extrair valores que seriam utilizados como massa para se avaliar o potencial de interação entre os vários pontos da rede viária da RMBH, foi utilizado o interpolador de Kernel. Como o banco de dados apresentava não somente a distribuição de pontos no território, assim como tabelas contendo o número de ligações por grupo de atividade por poste, foi possível gerar a superfície de Kernel ponderada, tendo como resultado a identificação das áreas de maior concentração de comércio e de maior concentração de indústrias no território (Figura 16).

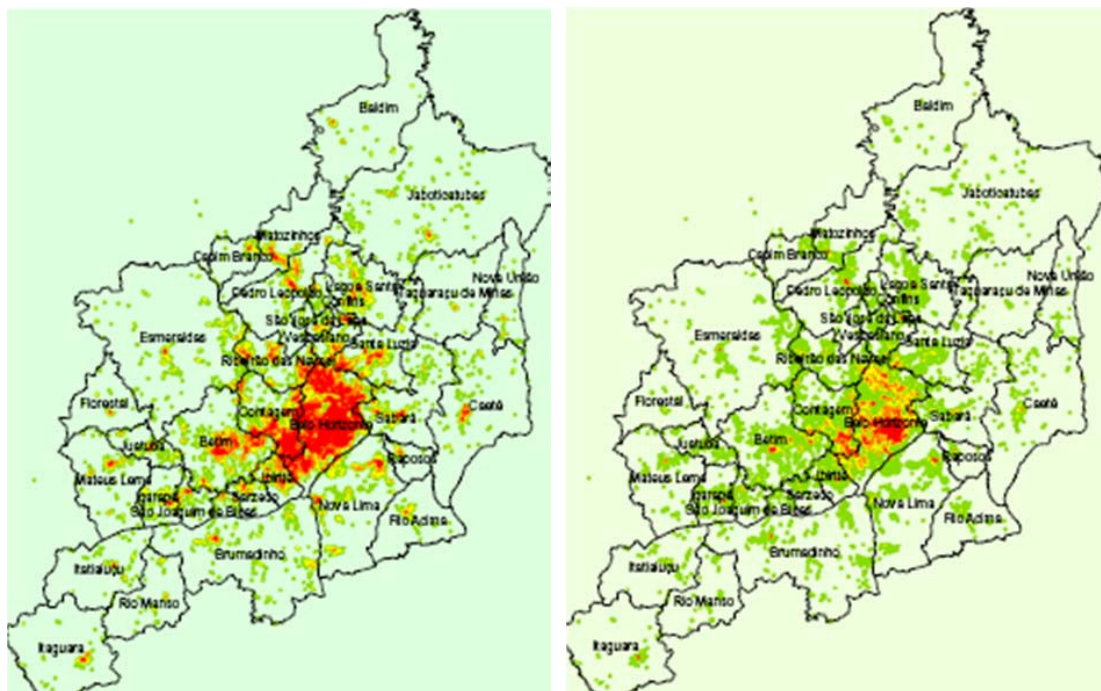


Figura 16 – Kernel por concentração de postes e Kernel ponderado por número de ligações no poste

A construção de modelo de Potencial de Interação, que calcula o quanto cada nó da rede viária da região é influenciado e influencia os demais pontos, foi elaborado utilizando como massa de atração de cada ponto o número de ligações de atividades de comércio e indústria. O referido modelo foi objeto de outros artigos, que podem ser consultados nas publicações de 2011 do Laboratório de Geoprocessamento da Escola de Arquitetura da UFMG (http://www.arq.ufmg.br/Laboratorio_Geo/).

4. Conclusões:

O presente artigo visou apresentar uma abordagem conceitual e prática sobre o modo de organização de dados de distribuição territorial para a construção de análise espacial apoiada por geoprocessamento, com vistas a esclarecer sobre o uso de estruturas raster ou matriciais, sobre a escolha de interpoladores e sobre a escolha da visualização das informações geradas.

Apresenta estudo de caso de exploração de tabela de dados contendo a informação locacional (coordenadas de localização no globo) associada a dados sobre a distribuição de serviços de energia elétrica, coleção de dados aparentemente muito singela, mas que, se bem explorada pelos recursos de geoprocessamento e análise espacial, podem resultar em complexas informações para o apoio à tomada de decisões em planejamento e gestão do território urbano e regional.

Relacionado ao presente artigo estão outras investigações desenvolvidas quando de nosso envolvimento no Plano Diretor de Desenvolvimento Integrado da RMBH, através do Laboratório de Geoprocessamento da Escola de Arquitetura da UFMG. Assim, indica-se a leitura dos demais estudos realizados, disponíveis no site mencionado.

A nosso ver, a principal contribuição do artigo é a apresentação didática de conceitos, aplicações e restrições sobre a construção de bases de dados e produção de informações que utilizem interpoladores, assim como a defesa de estruturas matriciais de coleções de dados na estruturação de modelos de análise espacial apoiados por geoprocessamento.

5. Agradecimentos

Agradecemos à valiosa colaboração da CEMIG, pela concessão de dados do projeto GEMINI para as investigações do PDDI – Plano de Desenvolvimento Integrado da Região Metropolitana de Belo Horizonte.

Agradecemos à FAPEMIG pelo apoio à participação no evento e consequente publicação do presente artigo.

6. Referências Bibliográficas:

BONHAM-CARTER, Graeme F. *Geographic Information Systems for Geoscientists; modelling with GIS*. Ottawa: Pergamon, 1994. 398 p.

Moura, Ana Clara M. A escolha de interpoladores e recursos de visualização na estruturação de bases de dados para produção de informações espaciais apoiadas por geoprocessamento. Laboratório de Geoprocessamento da Escola de Arquitetura da UFMG. Belo Horizonte, 2012. 21 p. Apostilas. In.: <http://www.arq.ufmg.br/>

- CÂMARA, G. et al. **Banco de Dados Geográficos**. São Paulo: MundoGEO, 2005. 506p
- ESRI. **ArcGis Desktop Help** 9.3. In.: webhelp.esri.com/arcgisdesktop/9.2/(2008).
- MONMONIER, Mark. **How to Lie With Maps**. Chicago, University of Chicago Press, 1996.
- MOURA, Ana Clara M. Discussões metodológicas para aplicação do modelo de Polígonos de Voronoi em estudos de áreas de influência fenômenos em ocupações urbanas - estudo de caso em Ouro Preto –MG. In: VII Encontro Nacional da Associação Brasileira de Estudos Regionais e Urbanos, 2009, São Paulo. **Anais do VII Encontro Nacional da Associação Brasileira de Estudos Regionais e Urbanos**. São Paulo: Associação Brasileira de Estudos Regionais e Urbanos - USP, 2009. p.1-16. In.: http://www.arq.ufmg.br/Laboratorio_Geo/Artigos/AnaClara-ENABER-2009.pdf
- MOURA, Ana Clara M. Geoprocessamento Aplicado à Caracterização e Planejamento Urbano de Ouro Preto – MG. In.: Xavier-da-Silva, Jorge, Zaidan, Ricardo (Org). **Geoprocessamento e Análise Ambiental – Aplicações**. Rio de Janeiro, Bertrand Brasil. 2004. p. 227.
- MOURA, Ana Clara M. **Geoprocessamento na Gestão e Planejamento Urbano**. Belo Horizonte, Edição da Autora, 2003. 294p.
- MOURA, Marcela M. **As representações gráficas aplicadas a variáveis ambientais em mineração de erro a céu aberto**. Orientadora: Janine Le Sann. Belo Horizonte: IGC-UFGM, 1999. 122 p. Dissertação. (Mestrado em Geografia).
- OLIVEIRA, Osmar Moreira, XAVIER-DA-SILVA, Jorge, ALMEIDA, Luís Fernando. Methodology for associating a conventional database to a geographical information system: The Saga/UFRJ case study. **1º Congresso Brasileiro de Cadastro Técnico Multifinalitário**, Florianópolis, ago. 1994. p. 35-39.
- TEIXEIRA, Amandio, CHRISTOFOLETTI, Antônio, MORETTI, Edmar. **Introdução aos Sistemas de Informação Geográfica**. Rio Claro: Ed. dos Autores, 1992. 80 p.
- TOMLIN, Dana. **Geographic Information Systems and cartographic modeling**. New Jersey: Prentice Hall, Englewood Cliffs. 1990. 249p.
- XAVIER-DA-SILVA, Jorge. **Acesso a dados e transformações preparatórias à análise ambiental**. Rio de Janeiro: Lageop, 1999a. 12 p. (apostila do Curso de Especialização em Geoprocessamento - Midia CD-rom).
- XAVIER-DA-SILVA, Jorge. **Geoprocessamento para análise ambiental**. Rio de Janeiro: J. Xavier da Silva, 2001. 227 p.
- XAVIER-DA-SILVA, Jorge. O que é Geoprocessamento? Conceito não pode ser confundido com todo o conjunto das geotecnologias, como o Sensoriamento Remoto, a Cartografia e os Sistemas de Posicionamento Global (GPS). **Revista do CREA-RJ**, Outubro/Novembro de 2009. p. 42-44.