

GEOPROCESSAMENTO NOS DIAGNÓSTICOS E PROGNÓSTICOS DE ÁREAS DE INTERESSE ESPECIAL AO NA ÁREA DE INFLUÊNCIA DAS LINHAS DE TRANSMISSÃO DA CEMIG – ESTUDO DE CASO DA RMBH

Ana Clara Mourão Moura

Universidade Federal de Minas Gerais

Escola de Arquitetura, Depto Urbanismo

Rua Paraíba 697, Savassi, Belo Horizonte – MG, CEP 30130-140

anaclara@ufmg.br

Carlos Alberto Moura

CEMIG – Companhia Energética de Minas Gerais

CEMIG-GE

Av. Barbacena 1200, Santo Agostinho, Belo Horizonte-MG, CEP 30123-970

camoura@cemig.com.br

Sheyla Aguilar de Santana

Coffey Information

Av. Afonso Pena, 3924 - cj 206 - Cruzeiro - Belo Horizonte - MG - Cep 30130-009

sheyla_santana@coffey.com

Lucas de Souza Lanna

CEMIG – Companhia Energética de Minas Gerais

CEMIG-GE

Av. Barbacena 1200, Santo Agostinho, Belo Horizonte-MG, CEP 30123-970

llanna@cemig.com.br

Úrsula Caroline de S G Azevedo

CEMIG – Companhia Energética de Minas Gerais

Av. Barbacena 1200, Santo Agostinho, Belo Horizonte-MG, CEP 30123-970

ursula.azevedo@cemig.com.br

Paula Moreira B Lourenço

CEMIG – Companhia Energética de Minas Gerais

CEMIG-GE

Av. Barbacena 1200, Santo Agostinho, Belo Horizonte-MG, CEP 30123-970

llanna@cemig.com.br

RESUMO

O estudo visa à exploração e otimização de técnicas de geoprocessamento na caracterização e análise do uso e ocupação do solo nas proximidades de linhas de transmissão de energia elétrica da CEMIG, com vistas a identificar áreas de interesse prioritário para o acompanhamento da dinâmica de ocupação, e para a construção de estudos preditivos das possibilidades de transformação nessas áreas. Faz parte de ações do projeto Geomape, que engloba construção de bases de dados detalhados sobre a região das linhas de transmissão. Promove a estruturação de ampla base de dados, com variáveis que caracterizam as condições ambientais e antrópicas da ocupação do solo nas áreas de influência das linhas de transmissão na Região Metropolitana de Belo Horizonte, entre as quais citamos infra-estrutura, densidade e tipo de ocupação, condições topográficas, fatores de acessibilidade, dados sócio-econômicos, áreas de expansão urbana, entre outros. Estrutura Sistema de Informações Geográficas para consultas e análises espaciais. A partir da coleção de dados aplica modelos de área de influência de ocorrência espacial de atividades, de estudos de correlações de variáveis e de combinação de variáveis. Aplica modelos de análise espacial baseada na metodologia de Multicritérios abordando procedimentos guiados por especialistas (*knowledge driven evaluation*) e procedimentos estatísticos segundo o comportamento dos dados (*data driven evaluation*), entre os quais o Expert System, a Assinatura

e a Mineração de Dados (*data mining*). Por se basear em estudo de exploração metodológica de análise multicritérios objetivando ampliar a sua aplicação para todo o estado de Minas Gerais, gera quatro sínteses segundo quatro procedimentos metodológicos diferentes, com vistas a validar resultados e identificar os procedimentos de menores investimentos em tempo e custo operacional. Os resultados obtidos são mapas de distribuição potencial de áreas de interesse para as observações e atenções da empresa, assim como a caracterização de suas condições quanto ao uso.

Palavras chaves: Geoprocessamento, Análise multicritérios, estudos diagnósticos e prognósticos, LTs

ABSTRACT

The study aims at the exploration and optimization of geospatial technologies in the characterization and analysis of the use and occupation of land near transmission lines Cemig's electricity, with a view to identifying areas of priority interest for monitoring the dynamics of occupation and as for the construction of predictive studies of the possibilities of transformation in these areas. Part of the project actions Geomap, which includes other actions to build detailed databases about the area of transmission lines, the actions reported in this article refer to planning actions on the region of transmission lines. It promotes broad-based structuring of data, with variables that characterize environmental conditions and the anthropogenic land cover in the areas of transmission lines in the metropolitan region of Belo Horizonte, among which we mention the infrastructure, density and type of occupation, topographical conditions, accessibility factors, socio-economic areas of urban expansion, among others. It structures Geographic Information System for queries and spatial analysis. From the collection of data the models identify areas of occurrence of space activities, studies of correlations of variables and combining variables. Apply models of spatial analysis based on Multicriteria methodology for addressing procedures guided by experts (knowledge driven evaluation) and statistical procedures according to the behavior of the data (data driven evaluation), including the Expert System, the Signature and Data Mining (Data Mining). By relying on the study of methodological exploration of multi-criteria analysis aiming to expand its application to the entire state of Minas Gerais, generates four summaries in four different methodological procedures, in order to validate results and identify the procedures less investment in time and operating cost. The results are distribution maps of potential areas of interest to the comments and attention from the company, as well as the characterization of conditions for the use.

Keywords: Geoprocessing, Multicriteria analysis, diagnostic and prognostic studies, LTs

1 INTRODUÇÃO

O estudo visa à exploração e otimização de técnicas de geoprocessamento na caracterização e análise do uso e ocupação do solo nas proximidades de linhas de transmissão de energia elétrica da CEMIG, com vistas a identificar áreas de interesse prioritário para o acompanhamento da dinâmica de ocupação, assim como para a construção de estudos preditivos das possibilidades de transformação nessas áreas.

O projeto, denominado Geomape, engloba outras ações de construção de bases de dados detalhados sobre a região das linhas de transmissão, com o apoio de geotecnologias de ponta para a elaboração de cadastros de alta precisão. Como parte do projeto maior, as ações relatadas pelo presente artigo se referem a ações de planejamento sobre a região das linhas de transmissão.

O projeto Geomape tem como objetivos a geração de relatórios e caracterizações das áreas das linhas de transmissão com para ações de planejamento e gestão. Entende-se gestão como o conjunto de ações em menores dimensões territoriais e temporais: é o cotidiano, baseado em cadastros detalhados e em procedimentos de intervenções. Por outro lado, o planejamento visa ações em maiores dimensões territoriais e temporais, para definição de objetivos e metas. Entende-se que ambas as ações, de

planejamento e gestão, são fundamentais para a empresa, sendo que a primeira define as grandes metas e objetivos, e a segunda estabelece como atuar localmente e de imediato para se obter os resultados esperados.

As ações de planejamento abordam, sobretudo, o estudo das condições das faixas de domínio das linhas de transmissão e seus entornos imediatos, mais especificamente de 250 metros de distância para cada lado do eixo. Os procedimentos incluem a identificação das áreas invadidas e das áreas passíveis de invasão por terceiros, definindo áreas de atenção especial. A caracterização das áreas foi obtida pelo mapeamento de variáveis, com o apoio do geoprocessamento, relativa aos temas de uso do solo, relevo, densidade de ocupação, sistema viário, condições topográficas, cobertura vegetal, condições sócio-econômicas, entre outras.

A metodologia se baseou na construção de bases cartográficas e na integração das informações através da álgebra de mapas conhecida como "Análise de Multicritérios". Como região de estudo, foi inicialmente selecionada uma área-piloto dentro de Belo Horizonte, para a qual foi construído o roteiro metodológico. Uma vez realizado o primeiro estudo, o projeto se estendeu para a Região Metropolitana, com os ajustes necessários para o incremento da complexidade espacial da área.

O piloto foi executado em dois traçados por um total de 12 km aproximadamente, na Região Metropolitana de Belo Horizonte, nas linhas LT1 e LT2 BH Bonsucesso – Sion e LT1 e LT2 BH Bonsucesso – Gutierrez. O projeto na Região Metropolitana de Belo Horizonte complexa rede de LTs (linhas de transmissão) que atende à área, totalizando 1.173 km de LTs de circuito simples e de 147 km de LTs de circuito duplo, em uma área de trabalho de 532 km² ao longo de 31 municípios.

O resultado do projeto maior é avaliado tanto em termos técnico-metodológicos, como pela ampla caracterização das áreas segundo as variáveis que as compõem e a produção de sínteses que configuram perfis sobre a área. As sínteses estabelecem uma classificação de áreas prioritárias para a observação, através dos graus de prioridade alto, médio a alto, médio, médio a baixo e baixo interesse de cuidados espaciais e atenção por parte da empresa. Quanto ao resultado técnico- metodológico, o projeto permitiu a exploração de métodos e técnicas de geoprocessamento com o objetivo de confirmar pareceres, cotejar diferentes propostas de roteiro metodológico, calibrar e validar resultados.

2 BASE CONCEITUAL - ANÁLISE DE MULTICRITÉRIOS

Segundo Moura (2003) a Análise de *Multicritérios* é um procedimento metodológico de cruzamento de variáveis amplamente aceito nas análises espaciais. Ela é também conhecida como *Análise Hierárquica de Pesos*. O procedimento baseia-se no mapeamento de variáveis por plano de informação e na definição do grau de pertinência de cada plano de informação e de cada um de seus componentes de legenda para a construção do resultado final. A matemática empregada é a simples *Média Ponderada*.

O emprego da Média Ponderada cria um espaço classificatório, ordinal, que pode ser também entendido como uma escala de intervalo. Esse processo pode também ser utilizado em escala nominal, desde que os eventos sejam hierarquizados segundo algum critério de valor. A ponderação deve ser feita por conhecedores dos fenômenos e das variáveis da situação avaliada, ou pelo conhecimento prévio de situações semelhantes. Nesse processo, a possibilidade de se ponderar de modo inadequado uma situação é o inverso do número de ponderações atribuídas.

O método *Delphi* (ou a consulta a especialistas) na obtenção dos pesos e notas baseia-se na escolha de um grupo multidisciplinar de especialistas, que conheçam bem o fenômeno e melhor ainda se conhecerem bem a realidade espacial onde ele se localiza. A esses especialistas é solicitado que hierarquizem ou coloquem as variáveis (ou planos de informação) em ordem de importância para a manifestação ou ocorrência de fenômeno estudado.

Outra forma de atribuição de pesos é a construção de análises baseadas por estatística, de modo a identificar as situações nas quais exista baixa correlação entre as variáveis, e alta correlação com o fenômeno ou ocorrência a ser explicado. Isto porque se existe altas correlações entre as variáveis, ambas estariam contribuindo da mesma forma para a explicação do fenômeno (é como se uma substituísse a outra na análise, e a incorporação de ambas não faria diferença).

Contudo, é importante observar que embora existam métodos e técnicas para tirar do especialista a responsabilidade de atribuir todos os valores numéricos da análise, representados pelos pesos e notas dos planos de informação e de seus respectivos componentes de legenda, sempre haverá um nível de subjetividade. Sempre haverá a necessidade de indicar uma hierarquia entre variáveis (que se somam em 100%) ou os pesos de suas subdivisões (que recebem grau de pertinência de 0 a 10). Não existe pesquisa sem um pequeno traço de subjetividade, pois a própria escolha de um modelo já é a opinião de um especialista segundo um olhar sobre a realidade espacial.

2.1 Os modelos de Análise Espacial na aplicação da metodologia de Análise de Multicritérios

A metodologia de análise espacial baseada em cruzamento de planos de informação em formato matricial começa com a montagem da base de dados e a espacialização destes em mapas georreferenciados. Uma vez montados os mapas, ou planos de informação, há aplicativos que disponibilizam ferramentas para procedimentos diagnósticos e procedimentos prognósticos. Os procedimentos diagnósticos caracterizam-se pela análise da situação espacial vigente, enquanto os prognósticos, somando-se aos conhecimentos dos diagnósticos, permitem antever situações e construir propostas de intervenção ambiental (XAVIER-DA-SILVA, 2001).

Ao longo dos processos de análise, sempre que são produzidas sínteses ou avaliações segundo algum critério, os mapas obtidos devem ser verificados segundo algum dos procedimentos:

- Identificação de áreas cujas características são bastante conhecidas e verificação da classificação por elas obtida no mapa.

- Uso de outras classificações ou mapeamentos de situações específicas, elaboradas por outros pesquisadores, como modo de comparação com os resultados obtidos.

- Realização de trabalho de campo para verificação de certas classificações e ocorrências surgidas nas análises.

- Consultas aos conhecedores da realidade espacial analisada.

Quando necessário, diante das verificações citadas, as avaliações devem ser refeitas com nova escolha de variáveis, ou mudanças em seus pesos e ou atribuição de novas notas aos componentes de legenda.

3. MAPEAMENTO DAS CARACTERÍSTICAS ESPECIAIS DAS FAIXAS DE DOMÍNIO DE LINHAS DE TRANSMISSÃO DE ENERGIA ELÉTRICA NA RMBH

A montagem da base de dados segue o seguinte roteiro:

- a) Definição das variáveis de mapeamento, fontes e dados disponíveis e novos dados a serem gerados;
- b) Definição das escalas de mapeamento de cada variável;
- c) Escolha da unidade territorial de integração dos dados, o que significa a dimensão do pixel e definição da matriz de representação.
- d) Reamostragem dos planos de informação para a resolução definida no item anterior;
- e) Montagem da base de dados cartográfica e alfanumérica;

Uma vez composta a base de dados, que são planos de informação na forma de mapas, tabelas e imagens, é estruturado o SIG (Sistema de Informações Geográficas), segundo o roteiro:

- a) Associação de dados alfanuméricos e composição de consultas temáticas de interesse para o estudo e para o processo de modelagem;
- b) Montagem do SIG contendo planos de informação em diferentes formatos de arquivamento (vetorial ou *raster*) mas, sobretudo, com planos que serão utilizados nos modelos de análise espacial em formato *raster* (matricial);
- c) Construção de mapas temáticos relativos a consultas e distribuição dos fenômenos no território.

Uma vez estruturado o SIG, é aplicado o modelo de análise espacial denominado “Análise de Multicritérios” segundo o roteiro:

- a) Verificação da escolha de variáveis de mapeamento – conferência junto a especialistas e conhecedores do fenômeno sobre a possibilidade de inserção de novos planos de informação;
- b) Aplicação de método Delphi para hierarquização dos pesos das variáveis de análise;
- c) Aplicação de método estatístico de Análise Hierárquica de Pesos, *Expert System* (Sistema Especialista) e *Data Mining* (mineração de dados) para atribuição de pesos às variáveis.

O resultado da análise de multicritérios é a geração de um mapa de distribuição hierárquica dos interesses, que define cada posição do território segundo o seu grau de importância para as ações de observação e acompanhamento pela empresa. Contudo, o processo exige validação dos resultados, segundo os procedimentos:

- a) Verificação e validação dos resultados obtidos através de conferência dos dados por verificação na imagem de alta resolução ou por comparação a situações conhecidas.
- b) Os produtos gerados caracterizaram *zoneamentos* segundo diferentes variáveis ambientais,

com identificação de situações especiais que caracterizam a área analisada, segundo conflitos, potenciais, riscos e prioridades de intervenção.

3.1 Composição do Banco de Dados

Foram trabalhadas bases de dados vetoriais, alfanuméricos (tabelas) e de imagens de alta resolução com vistas ao cruzamento das variáveis e construção dos mapas de potencial de interesse espacial. A primeira etapa de todo projeto na área de geoprocessamento é reunir a base de dados disponíveis e compatibilizá-las. Como estamos trabalhando com dados produzidos por diversas instituições, foi necessário padronizar as unidades territoriais de integração, a projeção cartográfica e a escala dos dados trabalhados. A projeção escolhida foi a UTM, Sad 69, fuso 23. Algumas informações foram trabalhadas por setores censitários, dados pontuais, dados lineares (infra-estrutura), pixel (imagens) e outras a partir do modelo digital de elevação resultante do trabalho de extração do Modelo Digital de Terreno a partir de imagens do satélite ASTER. Camadas de informação geradas:

a) Distribuição de infra-estrutura: Serviço de Abastecimento de Água por Rede Geral, Serviço de Esgotamento Sanitário por Rede Geral, Serviço de Coleta de Lixo por Serviço Público, Serviço de Fornecimento de Energia Elétrica. As informações foram representadas por setor censitário ou por superfície de distribuição dos serviços (mapas de Kernel) identificando as áreas mais bem servidas até as mais mal servidas de infra-estrutura na região.

b) Acessibilidade Viária: mapa construído a partir da classificação da concentração de vias urbanas e estradas, sendo as últimas separadas em pavimentadas e não-pavimentadas. As informações foram construídas a partir da vetorização de imagens de satélite de alta resolução, a Quickbird. A partir dos eixos viários foram traçados buffers, que são faixas de distância, de modo que a superfície gerada permite avaliar o grau de acessibilidade em cada ponto de território.

c) Serviços Urbanos: distribuição de comércio, prestação de serviços, indústria e serviços de uso coletivo a partir de dados associados aos postes da CEMIG. A densidade de Kernel foi utilizada para gerar a superfície de proximidade à distribuição de um ou mais serviços urbanos. Para classificar os ramos de atividades dentro desses quatro grupos foi utilizado como referência uma tabela apresentada pelo Plano Diretor e Lei de Parcelamento, Ocupação e Uso do Solo de Belo Horizonte.

d) Densidade Populacional e de Implantação de Edificações: distribuição da população e das edificações na área. O dado de densidade populacional foi trabalhado a partir dos setores censitários e da tabela do Censo de 2000. Já a densidade de implantação das edificações foi obtida a partir da classificação das imagens de alta resolução.

e) Declividades (Condições Topográficas): A declividade foi extraída das imagens de satélite ASTER. O procedimento promoveu a construção de modelo digital de elevação e posterior elaboração de mapa temático nas faixas de declividade indicadas para a caracterização das condições topográficas em faixas que seguem a Lei Federal 6766/79, que define áreas de restrição à ocupação urbana.

f) Condições Socioeconômicas foram mapeadas as distribuições espaciais de renda média familiar e de escolaridade, ambas a partir de setores censitários e dados do IBGE.

g) Proximidade a Vilas e Favelas: variável mapeada através da classificação de imagens de alta resolução e pela construção de buffers ou faixas de domínio destas ocorrências.

h) Padrão de Vegetação: mapeamento foi feito através da separação por tipologias de vegetação arbórea, arbustiva, rasteira, cultivo e solo exposto a partir da classificação das imagens QuickBird.

i) Uso e Cobertura do Solo: mapeada através da vetorização das imagens de alta resolução (QuickBird, Ortofotos e WordView), segundo as chaves de classificação: favelas, alta densidade urbana, média densidade urbana, baixa densidade urbana, grandes estruturas da paisagem, solo exposto, vegetação rasteira, vegetação arbórea, vegetação arbustiva, cultivo, mineração, subestações da Cemig, cursos d'água e vias.

j) Eixo de Expansão Urbana: obtido através da classificação e comparação de imagens do satélite Cbers de 2001 e 2007.

k) Proximidade ao perímetro urbano: construída a partir da classificação das imagens de satélite Cbers de 2007. É importante lembrar que o conceito de perímetro urbano é a divisão político-administrativa arbitrada pelas Câmaras Municipais, de modo que o objetivo no trabalho era de identificação dos contatos de concentração de ocupações antrópicas.

l) Proximidade às subestações: classificação do uso e cobertura do solo nas imagens QuickBird.

m) Proximidade às Grandes Estruturas da Paisagem: vetorizadas a partir das imagens de alta resolução. Elas são compostas por grandes galpões, estacionamentos, campos de futebol, clubes, indústrias, escolas, hospitais e outras estruturas que se destacam na paisagem pelo seu poder de atratividade.

n) Fiscalização: mapeia as rotas de fiscalização das linhas de transmissão e suas periodicidades.

3.2. Análise de Multicritérios

A análise de multicritérios consiste na combinação de variáveis, por processo de álgebra de mapas, visando à composição de uma síntese que traduza algum potencial que se deseje investigar. No presente estudo a investigação é sobre a hierarquização de áreas de interesse especial para a CEMIG.

Uma vez estruturada a coleção de dados são construídos planos de informação raster (matrizes compostas por “n” linhas e “m” colunas definidas pelo retângulo de envolvimento da área de trabalho e pelo “pixel” ou célula de análise, conhecido também como unidade territorial de integração de dados.

Na sequência estes planos de informação devem ser combinados segundo média ponderada de pesos e notas. A atribuição destes pesos e nota é etapa fundamental do processo e foi motivo de muitas investigações, que apresentamos a seguir. Foram atribuídos valores pelos métodos de *Knowledge Driven Evaluation* e pelo método de *Data Driven Evaluation* (Bonham-Carter, 1994).

3.2.1 Aplicação do método Delphi para integração das variáveis – Knowledge Driven Evaluation

O método Delphi na obtenção dos pesos e notas baseia-se na escolha de um grupo multidisciplinar de especialistas, que conheçam bem o fenômeno e melhor ainda se conhecerem bem a realidade espacial onde ele se localiza. A esses especialistas é solicitado que hierarquizem ou coloquem as variáveis (ou planos de informação) em ordem de importância para a manifestação ou ocorrência de fenômeno estudado. Exemplo: para a geração do mapa de Potencial de Riscos, qual é a ordem de importância das variáveis escolhidas? Uma vez recebidas respostas do grupo, realiza-se a seleção da média e a indicação do predomínio nas manifestações. O especialista então recebe o resultado da consulta e é solicitado a rever suas posições – caso ele tenha firmeza das suas escolhas, mantém suas respostas, mas caso ele decida ajustar suas avaliações diante da resposta do grupo, ele manifesta nova opinião. Assim é feito por 2 rodadas, mas há situações em que se aplicam 3 rodadas. (Moura, 2009).

Inicialmente foi realizada palestra explicativa sobre o método, seus procedimentos e objetivos, ilustrados com o exemplo do piloto desenvolvido no projeto anterior. A partir desta palestra a CEMIG estruturou a lista de pessoas que iriam participar do questionário Delphi, cujo objetivo é a maximização de consenso e contemplar diferentes visões sobre o tema. Estes participantes receberam um email informando a relação de variáveis contempladas no projeto piloto, a partir da qual eles deveriam responder sobre a manutenção daquelas variáveis, assim como a inserção de novas variáveis.

Foi organizada uma lista de variáveis a partir da justificativa de seus usos e verificações de condições para a realização deste mapeamento. Estabelecida a lista, a segunda etapa consistiu em pedir aos participantes opinarem sobre a importância de cada variável para a composição de uma área de interesse especial para observação pela empresa, indicando valores de 0 a 10, sendo 0 a indicação de que a variável teria importância mínima e 10 que a variável teria alta

relação com as áreas que devem ser mais bem observadas.

Os participantes receberam a tabela resultante da etapa 2, sem identificação dos demais participantes, e contendo a média e a moda das opiniões. A terceira etapa do processo consistiu em verificar se eles mantinham as opiniões ou gostariam de rever suas sugestões a partir da informação da média e da moda. Cerca de metade dos participantes ajustaram suas opiniões. O objetivo é evitar que sejam dadas notas ao acaso e que a pessoa tenha a oportunidade de conhecer o pensamento do grupo para se posicionar em relação à maioria.

Uma vez recebidas as opiniões de revisão dos participante, nova média e moda são computados e é elaborada a tabela final de valores atribuídos a cada variável. Realizada esta síntese, é aplicado o fatiamento dos valores pela a estatística de “Quebra-Natural” (*Natural Breaks*) para que o conjunto de variáveis seja dividido entre aquelas que têm alta, média a alta, média, média a baixa ou baixa importância, segundo os especialistas, quanto ao interesse de cuidados especiais pela empresa

As variáveis indicadas pelo grupo foram estruturadas em 20 camadas de informação, e transformadas em matrizes de dados configuradas como superfícies potenciais de distribuição de cada variável. Segundo Moura (2009) o primeiro passo é a composição de uma base de dados cartográficos, composta na forma de planos de informação e que deverão ser conjugados nas aplicações de modelos de análise espacial, pode ser realizada em formato vetorial ou matricial, mas há fortes tendências para o predomínio das operações dos modelos em formatos matriciais (raster). A questão se justifica pela relação de topologia implícita ao processo matricial, o que não só otimiza o cruzamento de dados, como também é condição sine qua non em alguns modelos.

É importante definir a unidade territorial de integração das análises, o que significa a escolha da resolução dos planos de informação e, conseqüentemente, a resolução ou precisão espacial das análises geradas. No presente estudo a resolução espacial adotada foi de 5 metros. Os valores de cada posição da matriz, que correspondem a uma posição territorial, foram reclassificados segundo o grau de importância. Uma vez construídas as superfícies potenciais ou camadas de informação elas são sintetizadas pela média ponderada em processo de álgebra de mapas, conforme o procedimento:

$$A_{ij} = \sum_{k=1}^n (P_k \times N_k)$$

Sendo:

A_{ij} – a posição na matriz de análise (linha/coluna), ou do pixel no mapa;

n – número de mapas ou camadas de variáveis cruzadas;

P_k – pontos percentuais ou peso atribuído ao mapa ou camada de variável k ;

N_k – graus de influência (de 0 a 10) da tipologia da variável para o risco final avaliado

O emprego da Média Ponderada cria um espaço classificatório, ordinal, que pode ser também entendido como uma escala de intervalo. Foi estruturada tabela que relaciona as variáveis combinadas e seus respectivos pesos, e o produto deste cruzamento foi um mapa geral de classificação da área, apresentado na figura 1.

3.2.2 Aplicação do método de Mineração de Dados e Assinatura – Data Driven Evaluation

O método de combinação de variáveis por multicritérios indica duas formas de obtenção de pesos e notas para indicação do grau de pertinência de cada uma delas no conjunto da síntese: pelo *Knowledge Driven Evaluation* (pela opinião dos especialistas) e pelo *Data Driven Evaluation* (pelo comportamento dos dados). Os processos não são excludentes, mas sim complementares, pois resultam em visões diferentes sobre o mesmo estudo.

No *Knowledge Driven Evaluation* o objetivo é extrair da experiência dos especialistas a indicação da lista de variáveis de interesse para o fenômeno, assim como o comportamento das variáveis. Há diferentes métodos para se obter a síntese das visões dos especialistas, com o objetivo da busca de maximização de consenso. No presente trabalho optou-se pelo método Delphi para a realização da consulta e síntese dos dados a partir das entrevistas aos especialistas.

No *Data Driven Evaluation* o objetivo é extrair dos dados a relação das variáveis que tornam as áreas de interesse para a observação devido ao risco de invasão. No presente estudo este processo foi realizado em três abordagens: *Assinatura* nas áreas invadidas e análise dos resultados e seleção das variáveis mais significativas no surgimento do fenômeno pelo *Expert Choice*, e uma última abordagem por *Data Mining*, visando a validação final dos resultados.

a) Assinatura:

Foram identificadas 254 localizações de invasões confirmadas, e ali foram marcados pontos destas ocorrências. Da mesma forma, foram selecionadas 254 outras localidades onde foi confirmada a não-invasão, e marcados pontos destas ocorrências. Uma vez construída a coleção de pontos, foram construídos buffers (áreas de influência) de 15 metros de raio ou 30 metros de diâmetro a partir deles, e realizadas a Assinatura em cada camada para identificação do comportamento das variáveis naquelas localizações. O resultado é o predomínio, por área, de cada componente de legenda das variáveis mapeadas.

A interpretação da tabela de assinaturas nos permite identificar quais variáveis são mais relevantes para a ocorrência do fenômeno, assim como quais

componentes de legenda (caracterização das variáveis) são mais relevantes para a ocorrência do fenômeno. A interpretação dos resultados de uma Assinatura deve seguir os procedimentos:

- Assinar pontos ou localidades onde o fenômeno que nos interessa foi confirmado, mas é importante assinar também toda a área de estudo para conhecer o comportamento das variáveis de um modo geral.

- Analisar os resultados sempre de fenômeno confirmado frente às características gerais do conjunto.

- Hierarquizar a importância da variável ou mapa para caracterização de fenômeno. Observar se o fenômeno acontece em qualquer situação (se há valores semelhantes em todas as legendas do mapa) daquela variável, ou se há o predomínio de uma ou mais legendas. Quando a distribuição é homogênea, aquela variável ou mapa não é útil para destacar áreas potenciais para o fenômeno, mas se predominam componentes de legenda na distribuição, aquela variável ou mapa deve ser considerado.

- Uma vez escolhidas as variáveis ou mapas que serão empregados, os valores obtidos na assinatura devem nos ajudar a atribuir pesos e notas. É feita a ordenação de importância observando estes valores.

É construída nova tabela de relação de pesos para as variáveis e realizada nova síntese, seguida da representação espacial em mapa. Observa-se uma expressiva semelhança entre os mapas produzidos na primeira etapa: o de *knowledge driven evaluation* (que seguiu os valores indicados pelo Delphi) e o de *data driven evaluation* (que seguiu os valores indicados pela assinatura, que traduz o comportamento dos dados na realidade territorial estudada). A semelhança expressa que os olhares especialistas estão muito calibrados com a realidade, cabendo então uma investigação mais detalhada de onde acontecem as diferenças nas hierarquizações e participações das variáveis (Figura 1)

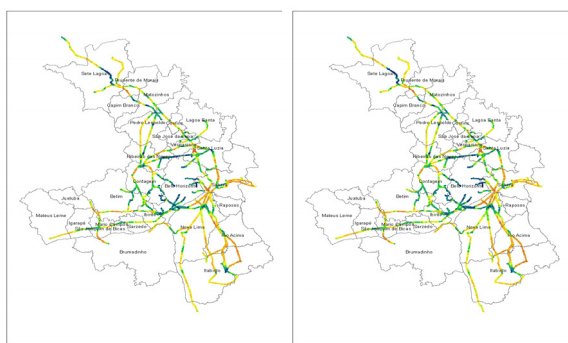


Fig 1: à esquerda o mapa obtido pelo processo Delphi e à direita o mapa obtido pelo de Assinatura

b) Expert Choice

Outro estudo realizado é a busca da seleção das variáveis que seriam os principais componentes na definição das áreas propícias à ocorrência do evento, o que funciona como um filtro no grupo de variáveis trabalhadas na Assinatura. Trata-se da aplicação de um

mecanismo de consulta especialista, ao mesmo tempo em que seleciona e hierarquiza um conjunto de variáveis. Para esta avaliação utilizamos um aplicativo de *Expert Choice*, baseado na lógica de hierarquização pelo procedimento que se chama "AHP - Análise Hierárquica de Peso", desenvolvido pelo Prof. Thomas Saaty em 1978, na Universidade da Pensilvânia. Ele auxilia na atribuição dos pesos dos planos de informação, para determinar a contribuição relativa de cada um, mas ainda assim o especialista ainda deve definir a hierarquia entre as variáveis e os pesos de cada componente de legenda das variáveis. O método propõe a comparação de variáveis par a par e é atribuído um critério de importância relativa entre eles.

Utilizamos como suporte técnico o software "Expert Choice" em sua versão 11.5, que disponibiliza um demo para testes por 15 dias. Através deste software carregamos as variáveis par a par e o critério de decisão sobre o grau de importância de uma sobre a outra foi a observação dos resultados obtidos na assinatura. O procedimento resultou na redução do conjunto de variáveis em 9 camadas, compostas por 15 variáveis de nosso mapeamento. Empregamos a tabela de pesos e notas e obtivemos novo mapa síntese, que também é baseado no *Data Driven Evaluation*, o que nos deu uma primeira resposta sobre a possibilidade de reduzir o número de variáveis na análise.

c) Data Mining

Como o objetivo de apurar o processo de decisão na seleção das principais variáveis que interferem mais ativamente na ocorrência do fenômeno estudado, nos propusemos a empregar a lógica da heurística para a mineração dos dados.

Segundo Xavier-da-Silva, em comunicação oral em 24 de outubro de 2007: "A heurística é uma forma de contemplar a visão qualitativa na combinação de variáveis que respondem por um determinado fenômeno, uma vez que permite que diferentes profissionais opinem sobre uma situação ambiental e ajustem suas opiniões a partir da compreensão do grau de pertinência de cada componente ambiental de acordo com o contexto onde ele está inserido, assim como do grau de pertinência de suas observações frente ao pensamento do grupo de outros especialistas."

Entre os procedimentos heurísticos, adotamos a mineração de dados como apoio à seleção das variáveis, com vistas a identificar padrões de relação entre elas que justificassem a identificação de associações que melhor respondessem pelo fenômeno. O *Data Mining*, em particular, é o passo onde são aplicados algoritmos voltados para atingir objetivos específicos, produzindo uma enumeração particular de padrões nos dados (Goebel e Gruenwald, 1999). Portanto, em um processo completo de descoberta de conhecimento, podem ser utilizados diversos algoritmos de *Data Mining*.

A mineração de dados espaciais é uma extensão de *Data Mining* voltada para domínios de

aplicação onde a consideração da dimensão espacial é essencial na extração de conhecimento. De forma similar, existem desenvolvimentos de trabalhos considerando o contexto temporal (*Temporal Data Mining*) e ainda, considerando ambas as dimensões (*Spatio-Temporal Data Mining*).

É organizada tabela que separa registros de áreas invadidas e registros de áreas não invadidas, juntamente com suas caracterizações segundo as variáveis de mapeamento. A tabela é tratada com o software livre WEKA (Waikato Environment for Knowledge Analysis), “Ambiente Waikato para Análises de Conhecimento”, desenvolvido pela Universidade de Waikato, na Nova Zelândia.

Foi selecionado o classificador de Bayes (*Naive Bayes*) e a ação de validação cruzada (*cross-validation*). Classificador Naive Bayes sugere que a presença (ou ausência) de uma característica particular de uma classe não tem relação com a presença (ou ausência) de qualquer outra característica. O classificador considera que todas as propriedades contribuem independentemente para a probabilidade de uma ocorrência (Zhang, 2001). O resultado apresenta a matriz de confusão:

Tabela 1– Matriz de Confusão

	a	b
a	5423	1943
b	566	6800

A interpretação da matriz indica que foram inseridos dados de tipo “a” – área invadida e dados de tipo “b” – área não invadida. Do conjunto do “a” 5423 registros foram identificados como tendo as características predominantes de “a” e 566 foram identificados como tendo as características predominantes de “b”. Do conjunto de dados de tipo “b” 6800 foram classificados como tendo características predominantes de “b” e 1943 como tendo características predominantes de “a”. Ou seja: o sistema realiza uma classificação e identifica padrões de comportamento, e em seguida ele verifica a quantidade de ocorrências que ele esperava que respondessem, por semelhança de padrão, com o tipo “a” ou “b”.

No caso em específico, consideramos que a confusão de 10% na classificação de “invadida” é um índice de erro muito baixo. A confusão maior ficou com a “não invadida” foi maior, cerca de 29%, o que significa que as áreas não propícias à invasão têm uma maior complexidade em sua conformação. Foi um ótimo resultado o das áreas invadidas, pois sinaliza que existe, sim, um conjunto de variáveis semelhantes que formam um padrão que favorece a ocorrência.

Outro índice que sustenta esta observação é o índice Kappa que o sistema gera, muito utilizado para dar idéia de quanto as observações se afastam daquelas esperadas, frutos do acaso, indicando assim quão legítimas as interpretações são (Jensen, 2005). Calcula-se o índice Kappa para cada matriz de confusão.

Tabela 2 – Interpretação do índice Kappa

Valor de Kappa	Concordância
0	Pobre
0-0,20	Ligeira
0,21-0,40	Considerável
0,41-0,60	Moderada
0,61-0,80	Substancial
0,81-1	Excelente

No estudo o índice Kappa deu 0.66, o que o coloca, segundo a tabela acima, na categoria de “substancial”, só perdendo para o excelente. A partir do estudo das relações entre as variáveis par a par e pela observação o desvio padrão obtido por cada uma foi então proposta nova relação de seleção de variáveis e seus valores, para realizar novo cruzamento de multicritérios. O objetivo desta etapa era identificar os principais componentes, ou seja, as principais variáveis que respondem pelo fenômeno e, uma vez recortado o grupo, ter avaliação do grau de assertividade que se pode obter trabalhando com apenas aquelas variáveis. Foi proposto o novo cruzamento, com a redução de número de variáveis.

4 CONCLUSÕES

O mapa resultante pelo *Data Mining* (*data driven evaluation*) foi comparado com o mapa resultante pela Assinatura e o resultado foi a validação dos resultados e a indicação de que seria sim possível reduzir o número de variáveis às principais, sem grande perda de informação. Assim, em projetos futuros, é possível reduzir para 10 as camadas de variáveis e chegar a resultados bastante semelhantes, como pode ser observado na Figura 2.

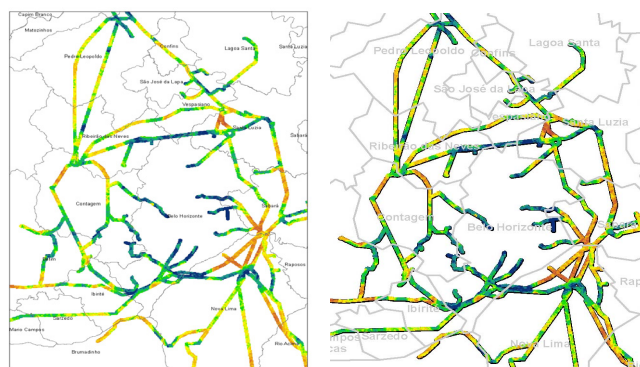


Fig 2: A esquerda trecho do mapa obtido pela Assinatura e à direita o obtido pelo Data Mining

É importante destacar que o mapa resultante do *Data Mining* atua como validação dos resultados frente à realidade, pois são aplicados índices de correlação espacial e do grau de reconhecimentos de

padrões comportamentais para o fenômeno a partir de suas variáveis principais. O procedimento é também uma forma de ajustar dados para se chegar a um índice de legitimidade dos procedimentos.

Diante do exposto, consideramos o trabalho satisfatório no sentido que contemplou diferentes procedimentos metodológicos para comparação de resultados, e apresentou validação dos resultados obtidos.

A partir da validação da metodologia, foi promovido o cruzamento entre o mapa síntese final, que hierarquiza o território segundo o grau de interesse para cuidados especiais e observações na fiscalização, e o mapa de ocupação do solo. O objetivo era realizar medições sobre os tipos de ocorrência que se encontram nas áreas identificadas como de alto a médio interesse, tais como ocupação urbana densa, cobertura vegetal rasteira, solo exposto, favelas, entre outras.

Os resultados obtidos foram validados por diferentes procedimentos metodológicos e confirmam a potencialidade da metodologia e das técnicas empregadas para apoio à tomada de decisões no planejamento das áreas de influência das linhas de transmissão de energia elétrica da CEMIG. A coleção de dados sobre os municípios da RMBH é um acervo expressivo que pode e deve ser empregado em outros projetos da empresa, uma vez que aborda caracterizações de condições ambientais e antrópicas, de infra-estrutura e condições de uso do solo.

5 REFERENCIAS BIBLIOGRÁFICAS

Bonham-Carter, G. Geographic information systems for geoscientists: modelling with GIS. New York: Pergamon, 1994.

Goebel, M.; Gruenwald, L. A survey of data mining and knowledge discovery software tools. SIGKDD Explorations, p. 20-33, 1999. In.: Neves, Marcos Corrêa, Freitas, Corina Costa e Câmara, Gilberto. Mineração de Dados em Grandes Bancos de Dados Geográficos. INPE, Relatório Técnico, Novembro, 2001.

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. Anais VII ENABER, São Paulo, Brasil, 9-11 setembro 2009, FEA/USP.

Moura, Ana Clara M. Geoprocessamento na Gestão e Planejamento Urbano. Belo Horizonte, Ed da autora, 2003.

Xavier-da-Silva, J. Geoprocessamento para análise ambiental. Rio de Janeiro: J. Xavier da Silva, 2001. 227 p.

6 Agradecimentos

Agradecemos à CEMIG, pela oportunidade de testar nossos saberes e capacidades.

Agradecemos à Fapemig – Fundação de Amparo à Pesquisa no Estado de Minas Gerais - pelo apoio financeiro para a apresentação do trabalho no evento.