

CONTRIBUIÇÕES AOS ESTUDOS DE ANÁLISES DE INCERTEZAS COMO COMPLEMENTAÇÃO ÀS ANÁLISES MULTICRITÉRIOS - “SENSITIVITY ANALYSIS TO SUITABILITY EVALUATION”

Ana Clara Mourão Moura¹
Piotr Jankowski²
Chiara Cocco³

¹Universidade Federal de Minas Gerais
Escola de Arquitetura, Laboratório de Geoprocessamento
anaclara@ufmg.br, geoproea.arq.ufmg.br

²San Diego State University
Department of Geography
pjankows@mail.sdsu.edu

³Università Degli Stud di Cagliari
Facoltà di Ingegneria e Architettura
kiara-mg@hotmail.it

RESUMO

A Análise Multicritérios tem sido popular entre os usuários de GIS quando o objetivo é a construção de modelos descritivos ou preditivos de um território, em diferentes escalas e para diferentes aplicações. Uma vez que o método tem sido amplamente utilizado durante os últimos vinte anos, e que já existe uma vasta literatura sobre o assunto, cabe desenvolver estudos que levam a questionar o impacto das incertezas nos resultados, e por isto este artigo tem como objetivo a discussão da análise de sensibilidade espacial como uma abordagem para lidar com a incerteza no processo de Análise Multicritérios. O trabalho visa apresentar propostas de como descobrir e representar a incerteza oriunda das características dos dados iniciais, e visa propor critérios para a definição de faixas de pesos para cada critério utilizado na função de agregação ponderada da análise multicritério e demonstra essas propostas como subsídio para o emprego de análise de sensibilidade nos estudos espaciais – SASE (Sensitivity Analysis to Suitability Evaluation).

Palavras chaves: Análise de Sensibilidade, Análise de Incertezas, Análise Multicritérios

ABSTRACT

Multicriteria Analysis has been popular among GIS users, as an integrative method of spatial analysis when the objective is the construction of descriptive or prescriptive model of a territory, at different scales and for different applications. Given that the method has been widely used for the past twenty years, and that there is already an extensive literature on the subject, it's important to develop studies to discuss spatial sensitivity analysis as an explicit approach to dealing with uncertainty as part of Multicriteria Analysis process. It presents a methodological guide for discovering, capturing and representing uncertainty characteristics in input data, followed by defining ranges of weights for each criterion used in the weighted aggregation function of Multicriteria Analysis and explains this proposals to give support to the use of sensitivity analysis in spatial analysis – SASE (Sensitivity Analysis to Suitability Evaluation).

Keywords: Sensitivity Analysis, Uncertainty Analysis, Multicriteria Analysis

1. INTRODUÇÃO

Análise multicritério tem sido popular entre os usuários de GIS, um método integrado da análise espacial quando o objetivo é a construção de modelos descritivos ou preditivos de um território, em diferentes escalas e para diferentes aplicações. Há duas razões para esta popularidade do método: é um método simples para integrar dados segundo os objetivos de análise segundo múltiplos critérios, e a existência de ferramentas prontamente disponíveis em softwares de geoprocessamento, que podem ser utilizadas para implementar as etapas do processo de análise espacial por combinação de variáveis.

A função principal da análise multicritério é facilitar a integração variáveis espaciais, sob a forma de mapas temáticos, a fim de encontrar locais adequados para as alocações de recursos espaciais ou atividades humanas. A análise multicritério pode ser usada, por exemplo, para a identificação de lugares adequados para crescimento e expansão urbana; lugares onde é necessária uma proteção ambiental; locais com propensões para certa mudança de uso da terra, entre outros. Em sua forma mais simples envolve a apenas a combinação de variáveis espaciais sob a forma de camadas de mapa por média ponderada.

Uma vez que o método tem sido amplamente utilizado durante os últimos vinte anos, e que já existe uma vasta literatura sobre o assunto, o conhecimento prático e os exemplos na utilização de análise multicritério para suporte à tomada de decisões levam a questionar o impacto das incertezas inerentes em dados utilizados na análise, que são uma parte inerente ao método. Assim, este artigo tem como objetivo a discussão da análise de sensibilidade espacial como uma abordagem para lidar com a incerteza no processo de análise multicritério.

A Análise da Sensibilidade Espacial complementa a análise multicritério para medir e promover a visualização do grau de certeza e incerteza que existem nos resultados produzidos pela análise multicritério. É possível categorizar os resultados da análise da sensibilidade espacial: - local adequado para um determinado evento ou fenômeno, acompanhado do baixo grau de incerteza sobre o resultado; - local adequado é um determinado evento ou fenômeno, mas com um alto grau de incerteza sobre os resultados; - local inadequado para o determinado evento ou fenômeno, acompanhado do baixo grau de incerteza sobre os resultados; - local inadequado para o determinado evento ou fenômeno, com um alto grau de incerteza sobre os resultados.

Este trabalho complementa uma abordagem "Análise de sensibilidade para avaliação da adequabilidade" (SASE) proposta por Jankowski e Ligmann-Zielinska (2010, 2012, 2014). Visa apresentar propostas de como descobrir e representar a incerteza oriunda das características dos dados iniciais, e visa propor critérios para a definição de faixas de pesos para cada critério utilizado na função de agregação ponderada da análise multicritério.

O procedimento proposto, que serve como complemento para o SASE proposto por Jankowski e Ligmann-Zielinska, introduz o usuário à lógica da incerteza em variáveis espaciais (critério de avaliação) e mostra como chegar à escolha de valores para as faixas de peso de variáveis utilizadas na análise de Monte Carlo.

O procedimento é inicialmente apresentado através de dados de uma pequena região na fronteira de Belo Horizonte, em área limítrofe com Nova Lima, apenas para ilustrar a proposta metodológica. Em seguida, é desenvolvido estudo de caso na área maior, relativo a toda a região da zona norte de Belo Horizonte, região da Pampulha. É importante ressaltar que Análises de Sensibilidade é uma técnica que complementa e traz maior credibilidade ao processo de Análise Multicritérios.

2. ANÁLISE MULTICRITÉRIOS

Análise multicritério se popularizou significativamente entre os usuários de SIG, sendo o modelo de análise espacial mais utilizado quando o objetivo é a construção de diagnósticos e prognósticos do território, em escalas diferentes e com diferentes objetivos. Há duas razões para essa ampla divulgação do modelo: a clareza de traduzir os objetivos baseada na estruturação lógica do modelo, e a existência de ferramentas de geoprocessamento nos software, que, mesmo sem que apresentem todo o modelo propriamente dito, permitem construir os passos necessários para o processo.

O modelo pode ser traduzido a uma álgebra de mapas, o que favorece que o problema seja subdividido em seus principais componentes: as variáveis que interferem no seu comportamento, a maneira como eles são integrados e a combinação deles, a fim de se construir o sistema.

A ideia de multicritério não é nova. Ela começou com a abordagem sistêmica, a partir dos anos 1950, com as investigações do biólogo Ludwig von Bertalanffy (1975), que propuseram a estudos interdisciplinares segundo a teoria geral dos sistemas. A realidade é vista como a justaposição e interposição de variáveis compondo um sistema em que as variáveis são interdependentes e qualquer alteração resulta em alterações de todo o conjunto.

A ideia da Análise como um processo é baseada na promoção da representação abstrata do fenômeno no modo a favorecer sua explicação por modelos que permitem que suas descrições. Na década de 1960, Chorley e Hagget (1967) desenvolveram estudos sobre sua aplicação à geografia, contribuindo para a compreensão dos modelos de análise espacial.

É necessário entender que a abordagem sistêmica não é apenas a decomposição em partes e reconhecimento do fenômeno através de seus componentes, mas é, acima de tudo, a compreensão de que essa decomposição só faz

sentido se os elementos são analisados em relação ao contexto em que eles estão inseridos. Os elementos têm características que os definem (intensidade, massa, idade, localização, entre outras) e os valores dessas características significam os estados dos elementos em um determinado local em um determinado período de tempo.

O modelo é uma simplificação da realidade, de forma que ela é representada por sua essência, eliminando o que não é significativo para a caracterização do fenômeno e de acordo com os objetivos da análise. Segundo Chorley e Hagget (1967), os modelos são a apresentação formal de uma teoria, que utilizam ferramentas da lógica matemática e da teoria dos conjuntos. O modelo pode ser uma teoria, uma lei, uma hipótese, uma ideia estruturada, um relacionamento, uma função, uma equação, a síntese de dados ou a descrição do mundo real. Como simplificações da realidade, os modelos têm como ponto importante a seleção dos aspectos mais relevantes. O sistema é estudado com a determinada finalidade, e tudo o que não afeta o objetivo definido deve ser eliminado.

Huggett (1980) explica o processo de estruturação do modelo para o desenvolvimento da análise sistêmica: "By discriminating system parts of sub-systems at different levels of resolution, a complex system is simplified in a logical and realistic way which avoids the bewildering bulk of information at microscopic levels."

Entre os modelos simplificados, compostos por processos iniciais, e com base na lógica de se decompor para compor, encontra-se o trabalho de Ian MacHarg no "Design com a natureza" (1969), a primeira importante aplicação na área da análise espacial com aplicação desses princípios. O autor caracteriza o território por uma série de mapas temáticos e propõe a identificação de áreas adequadas para diferentes usos, incluindo manutenção e instalação de atividades antrópicas. O resultado é uma combinação de mapas temáticos, naquele momento promovido por simples sobreposição de mapas, o que permitia a identificação dos lugares notáveis, resultante da síntese de variáveis que melhor atendessem aos objetivos da análise. Era um processo analítico por decomposição e síntese através da integração.

Uma vez identificados os principais componentes do sistema, eles podem ser isoladas em planos de informação ou camadas. Com este conjunto de dados, diversos profissionais, com os seus conhecimentos e os seus objetivos específicos, podem propor diferentes processos de integração de variáveis, resultando em interpretação mais dinâmica da realidade. As descrições tornam-se mais ricas, porque eles refletem "o que", "onde" e "como".

São etapas do processo metodológico em Multicritérios, segundo Moura (2003):

- *Definição de objetivos e aplicações no uso do sistema.*
- *Organização de dados alfanuméricos e cartográficos.*
- *Tratamento dos dados para a composição de representações matriciais (rasters) para análise e representação da distribuição potencial do fenômeno.*
- *Definição de pesos das variáveis e de valores para seus componentes de legenda.*
- *Integração das variáveis na Análise de Multicritérios.*
- *Comparação com a realidade existente.*
- *Calibração do Sistema e retorno para a fase da análise.*
- *A construção de cenários (if/then).*
- *Elaboração de propostas de intervenção, gestão e restrições.*

Segundo Anselin (1999) há alguns passos destinados à compreensão do problema, por análise exploratória, que têm como objetivo a melhor compreensão sobre o objeto investigado; e há etapas baseadas na estruturação do sistema e na construção de modelos que representam a realidade. Devido a estes procedimentos a Análise Multicritérios se destina, em essência, como apoio à tomada de decisões.

Após compor as camadas das variáveis, a álgebra de mapas é estruturada por média ponderada (1):

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

Onde:

A_{ij} - A posição na análise da matriz (linha/coluna), ou de um pixel no mapa

n - Número de mapas ou camadas de variáveis combinadas.

P_k - Pontos percentuais ou peso atribuído ao mapa ou camada da variável k no conjunto de camadas.

N_k - Valores que significam a importância do grau de pertinência de cada componente da legenda da variável k segundo o objetivo da análise

A escolha de pesos devem ser muito bem documentada e justificada, deve ser aberta a revisões, e, na maioria dos casos, é onde as calibrações do sistema acontecem. Os diferentes modos de escolha de pesos podem ser resumidos em dois grupos de procedimentos: a avaliação orientada pelos dados e a avaliação orientada por conhecimentos especializados, que Bonham-Carter (1994) denomina "data driven evaluation" ou "knowledge driven evaluation".

Entre os processos de definir os pesos por "*data driven*" podem ser citados os métodos que se baseiam em mineração de dados, que visam identificar as tendências da hierarquia das variáveis, de acordo com o que acontece na realidade, medido por amostras escolhidas no território investigado (Castro, 2010).

Entre os processos para se definir os pesos por "*knowledge driven*", por consulta a especialistas, o objetivo é receber *feedback* de pessoas que entendem o fenômeno e a realidade investigada, e emitem suas opiniões de acordo com probabilidade e o estado-da-arte da variável. Existem diferentes métodos para se realizar essa consulta, sendo os mais populares o Thomas Saaty e o método Delphi.

O método de Saaty (1980), denominado "Análise Hierárquica de Pesos", foi desenvolvida em 1978 na Universidade da Pensilvânia. O processo visa a definição do peso de cada plano de informação segundo a sua contribuição relativa para o conjunto de planos. A meta é a decompor-se a complexidade da situação, a fim de tomar as decisões sobre o que é mais importante. O método propõe a comparação de variáveis par-a-par segundo o critério de importância relativa entre elas.

O método Delphi foi inicialmente proposto na década de 1950 pela indústria militar norte-americana, o Research and Development (RAND), com o objetivo de dividir responsabilidades e obter a melhor solução que tenha sido consenso entre as pessoas envolvidas. O nome vem do Oráculo de Delfos, porque o objetivo é apoiar as decisões. A justificativa é que as decisões dos grupos estão mais próximas da realidade, sendo mais precisas que simples estudos individuais.

Segundo Dalkey e Helmer (1963) o método é composto por questionários aplicados a especialistas em rodadas, e essas rodadas são intercalados com *feedbacks*, que visam a convergência das opiniões expressas pelos participantes. Outros autores, como Linstone e Turoff (2002), argumentam que o número de rodadas deve ser tanto quanto sejam necessárias para se atingir a convergência de opiniões. Contra esta opção defendemos que o consenso não existe, e o que pode ser obtido é a maximização de consenso e, por isso, com base em um número de expressivos estudos de caso por nós desenvolvidos, fizemos a escolha pelo emprego de três rodadas.

Há etapas muito importantes para que o modelo represente a síntese da maximização de consenso ou a síntese do comportamento dos dados. Essas etapas estão relacionadas à clara definição dos objetivos de análise, a preparação dos dados, a escolha entre o a avaliação guiada pelo conhecimento (*knowledge driven*) ou a avaliação guiada pelos dados (*data driven*) na definição dos pesos a cada variável (o que significa a importância de cada uma das variáveis para o objetivo de investigação). Mais recentemente, os estudos de Análise Multicritério receberam outra contribuição: a possibilidade de controlar a incerteza nos estudos, usando Análise de Sensibilidade.

Este trabalho apresenta as etapas de preparação dos dados, a integração dos mesmos por Análise Multicritério em *raster* ou em representação vetorial e a aplicação da simulação de Monte Carlo proposta por Jankowski e Ligmann-Zielinska (2010, 2012, 2014) para desenvolver abordagem SASE (Sensitivity Analysis to Suitability Evaluation). A contribuição é a discussão sobre a preparação de dados e sobre a definição de intervalos de simulação de Monte Carlo.

3. PREPARAÇÃO DOS DADOS PARA AS ANÁLISES

Segundo os objetivos da análise, as variáveis são definidos e organizadas em camadas. Essas camadas representam a "superfície potencial de distribuição do fenômeno ou ocorrência espacial", o que significa que cada ponto ou célula que representa o território recebe um valor ou índice segundo a variável.

Todos os dados ou camadas devem ser standardizados. Isso significa que serem organizados na mesma escala de representação espacial (mesma resolução espacial ou tamanho da célula, o que estabelece a unidade territorial de integração) e na mesma escala de valores (definido, por exemplo, entre 0 e 1, com os valores na mesma razão escala numérica, o que significa que todos comecem no mesmo valor mínimo e terminam no mesmo valor máximo).

3.1 Tipos de dados iniciais

Dados pode ser dividida em: seletivo ou nominais, e em qualitativos ou quantitativos.

3.1.1. Dados Seletivos ou Nominais

São dados apresentados de uma forma que define diferentes tipologias. Exemplo: os tipos geológicos, e assim por diante. Há duas maneiras de se trabalhar com esses dados: separar os componentes de legenda (cada tipo identificado) em um mapa e compor este mapa na forma de superfície potencial de distribuição do dado; ou apresentar todos os componentes de legenda no mesmo mapa e classificá-los com valores que representem a relevância do componente em relação ao objetivo da investigação.

3.1.1.a. Apresentação do dado por camada, selecionando os componentes de legenda e preparando uma camada por componente, na forma de superfície potencial de distribuição do componente.

Exemplos são situações como o mapa de uso e cobertura do solo, que pode ser decomposto em mapas de distribuição de solo exposto, cobertura vegetal, mancha urbana, água ou sombra, entre outros os componentes.

O primeiro passo consiste em separar cada um desses componentes de legenda e prepará-los em mapa de representação binária apenas com presença ou ausência do tema (Figura 1 e Figura 2). A etapa seguinte consiste em

transformar a representação binária em superfície potencial de distribuição da ocorrência, considerando o efeito de borda entre as ocorrências, o que significa: nas bordas da área verde é natural que sua representação sofra redução gradual por influência da vizinhança. Se a vizinhança de um pixel ou célula é predominantemente a informação “ausência” do elemento, a fronteira será mais abrupta, mas se a vizinhança é composta também por outras células ou *pixels* com “presença” do elemento, a fronteira será mais suavizada ou difusa em função da composição da vizinhança.

É necessário que se faça a conversão da matriz em pontos segundo cada posição da célula ou para cada equidistância de análise (unidade territorial de análise, por exemplo, 10 metros), de modo que o ponto receba o atributo de valor (no caso ausência ou presença da ocorrência). Se o mapa está em polígonos, é feita a conversão para *raster* inicialmente, e depois de *raster* para pontos, usando no raster a dimensão de célula escolhida. Se o mapa já está em *raster*, basta converter para pontos (Figura 3).

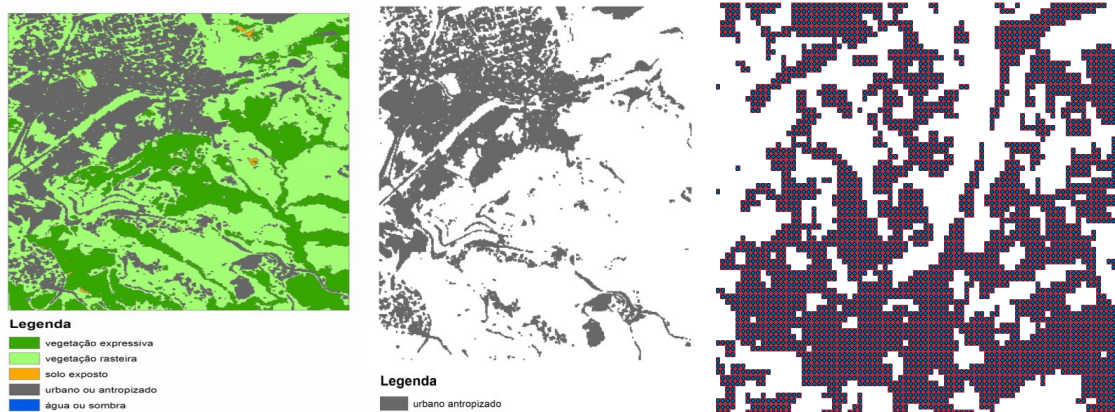


Fig. 1 – Uso do Solo Fig. 2 - Uso do Solo apenas um tema (urbano) Fig. 3 - Raster em Pontos

Em seguida, aplica-se a densidade do Kernel ponderada pelos valores dos pontos, para se compor a "superfície potencial de distribuição do fenômeno ou ocorrência". O procedimento se justifica, pois considera a vizinhança do ponto e a concentração desses valores, e o resultado é a distribuição gradual de valores, e não apenas o “sim” ou “não” da ocorrência (Figura 4). Destaca-se, contudo, que nas partes internas da ocorrência, onde não há efeito de borda, o valor não se altera. As mudanças acontecem somente nas bordas.

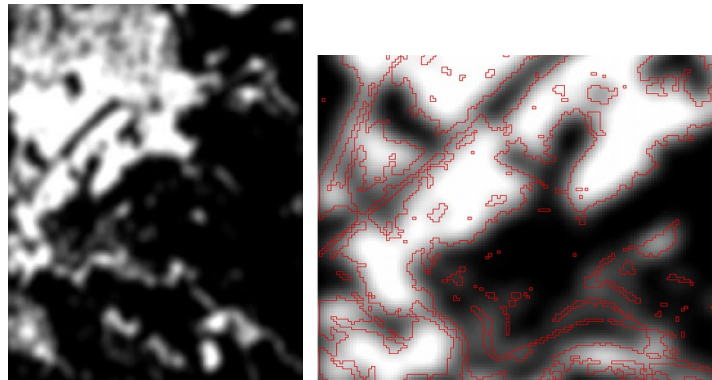


Fig. 4 - Kernel, apresentado como simbolização de cores graduais, e distribuição potencial do fenômeno, que suaviza as fronteiras e considera a influência da vizinhança.

3.1.1.b. Apresentação do dado com diferentes componentes no mesmo mapa, mas já classificadas de acordo com sua relevância para o objetivo de análise.

Observe que no mapa de uso e cobertura do solo há diferentes componentes de legenda. Para esse processo os componentes foram reclassificados, recebendo valores numéricos de acordo com sua relevância para o motivo de investigação. Como exemplo, vamos propor a análise visando identificação de áreas adequadas para o adensamento e crescimento urbano, e os valores numéricos representam o quanto cada componente de legenda pode receber esse novo uso (Figura 5).

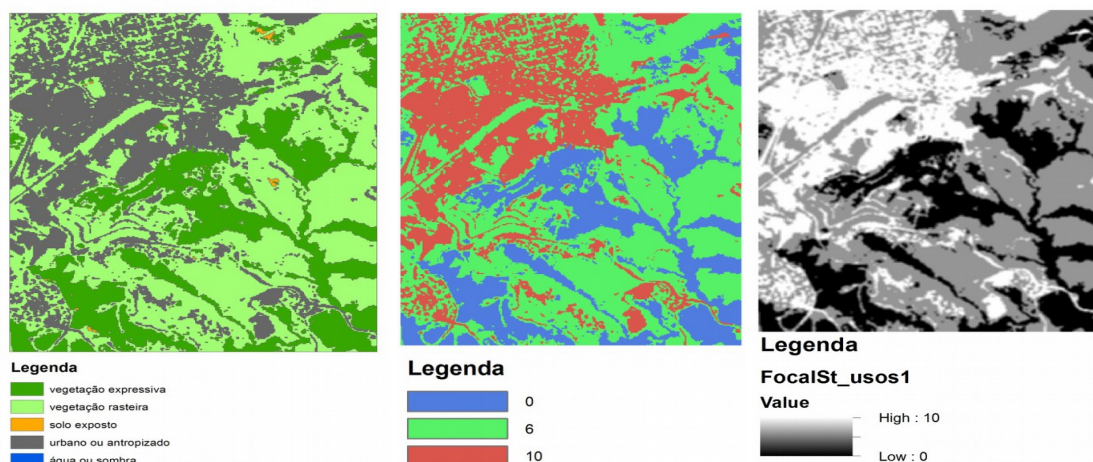


Fig. 5 - (a) uso do solo com os componentes da legenda; (b) reclassificação segundo a relevância para a investigação. (c) dados os valores numéricos aos componentes, emprego de estatística focal.

Contudo, sabe-se que a distribuição de usos no solo ou de fenômenos de natureza semelhante (categóricos, também chamados de seletivos ou nominais) não acontece de modo a termos clara definição de limites, mas acontecem efeitos de borda entre as ocorrências, de modo que é necessário aplicar o filtro que recalcula os valores de acordo com os vizinhos mais próximos combinados (Figura 5-c Estatística Focal). Seguindo a mesma lógica usada nos processos com uma variável isolada, onde o valor entre vizinhos é igual não há mudança no resultado, mas onde há mudanças de ocorrências ou de valores haverá interferência do filtro de vizinhança e o efeito de borda acontecerá.

A fim de obter esse novo efeito de borda no mapa, é possível seguir o mesmo procedimento indicado no item anterior (usando a densidade de pontos com valores que representam a distribuição espacial das ocorrências, o que inicialmente era representado em *raster*), ou pode ser aplicado o filtro de vizinhança, o *Focal Statistics*, que considera a média os valores dos vizinhos segundo a distância definida pelo pesquisador, o que é representado pelo número de células ao redor do pixel, a partir do mínimo de 3x3 células.

3.1.2. Dados Quantitativos ou Qualitativos

Os dados qualitativos os componentes de legenda ordenados segundo os objetivos de análise. Exemplo: estado de conservação da edificação (excelente, bom, médio e ruim, terrível). Trabalhar com variáveis qualitativas requer a ampla documentação para justificar a taxonomia utilizada (classificação), pois é necessário decodificar a subjetividade por trás da decisão. Quando os dados são quantitativos eles são necessariamente ordenados, mas nem todos os dados ordenados são quantitativos, pois eles podem ser de natureza qualitativa. Contudo, para procedimentos de análise espacial dados qualitativos precisam receber valores numéricos, e os números atribuídos devem ser representantes de uma ordem escolhida.

3.1.2.a. Dados qualitativos ou quantitativos que apresentam efeito de borda

Algumas variáveis, medidas por qualquer unidade territorial se comportam de forma que a ocorrência territorial interfere na sua vizinhança e pode ter efeito de mudar os valores nas fronteiras.

Por exemplo: no mapa de distribuição de valores dos terrenos urbanos, se há na vizinhança de áreas de menores valores áreas de ocorrências de maiores valores (no caso da favela localizada no bairro alto valor), no contato das fronteiras há uma natural diminuição de valor da zona mais valorizada (que sofre o efeito vizinhança) aumento natural dos valores dos menos valorizados.

O modelo de Estatística Focal considera essas interferências e gera as reduções ou ampliações de valores nas células de bordas de contato. Lembre-se que os dados qualitativos devem ser previamente convertidos em quantitativos, recebendo os valores numéricos de acordo com a pertinência para o objetivo de análise.

3.1.2.b. Dados qualitativos ou quantitativos que não apresentam efeito de borda

Há situações, tanto em dados qualitativos como quantitativos, em que as integrações entre fronteiras não é viável, porque não é possível arbitrar que a proximidade territorial dos valores diferentes é um fator de interferência nos resultados de ambos os lados.

É o caso, por exemplo, no contexto socioeconômico, de dados representados em distribuição que signifique recorte territorial de armazenamento da média de amostras coletadas, como acontece com os setores censitários. Os mapas que simbolizam a distribuição do nível de escolaridade, do nível de renda ou outros simbolizam o valor médio ou o valor agrupado por unidade amostral. Ao se converter a superfície da unidade amostral (por exemplo o polígono do setor censitário) para *raster*, em toda a superfície será aplicado o mesmo valor, mesmo que isto não aconteça na realidade Além disto, não é possível aplicar efeitos de vizinhança pois não se pode dizer que entre dois territórios, um

com o valor " 4 " e o outro com o valor " 2 ", pode existir o território com valor "3 ". Neste caso, cabe apenas a normalizar os valores com vistas a permitir comparações e combinações, mas não devem ser aplicados filtros de borda ou vizinhança.

3.1.2.c. Dados qualitativos ou quantitativos que os valores não cobrem toda o território, mas foram coletados em amostras pontuais.

Muitas variáveis territoriais não são coletados exaustivamente sobre todo o território de análise, mas em pontos de coleta ou amostrais. É o caso, por exemplo, de dados de precipitação, dos animais ou espécies de plantas registrados em eventos de ocorrência.

Em situações como esta, para a criação da “superfície potencial de distribuição do fenômeno ou ocorrência” é indicado a utilização do interpolador IDW (inverso da distância ponderada) que é o modelo gravitacional que considera a distância entre cada posição do em território e o ponto onde foram coletados dados. Isso significa calcular o inverso da distância entre cada célula (na matriz ou no raster onde estão representadas) e o ponto de amostragem, considerando o valor registrado no ponto amostral. É natural que os valores dos pontos amostrais mais próximos da célula tenham mais influência sobre o seu valor.

3.2. Normalização para integração e comparação de variáveis

A integração ou a comparação das variáveis requer que elas sejam apresentadas de modo normalizado, o que significa que apresentem o mesmo valor mínimo e o mesmo valor máximo, ou seja, que estejam dentro de uma mesma escala, ainda que dentro dos limites haja variações diferentes de valores. Observando os gráficos (Figura 6) é possível avaliar a importância de se promover a normalização de valores para a comparação do comportamento das variáveis. Como comparar as variáveis na figura à esquerda? A adequada comparação só é possível quando há normalização da distribuição, de acordo com o apresentado no gráfico da direita.

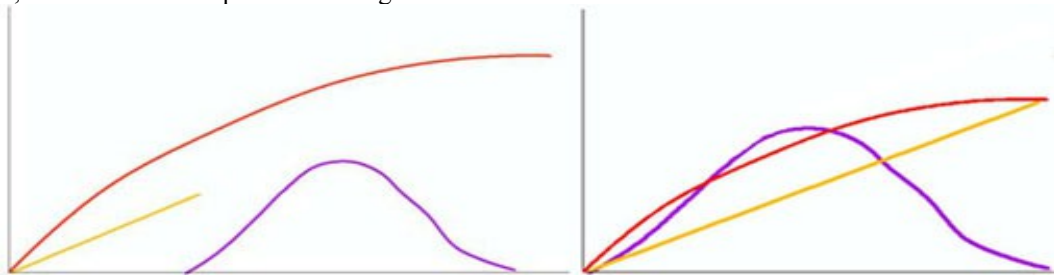


Fig. 6 - Distribuição das variáveis. Comparação e integração possível por normalização da distribuição.

Isto significa que a distribuição de valores entre a mesma origem e o mesmo fim, como o exemplo da distribuição de 0 a 1. Observando o exemplo do mapa de Acessibilidade e Capilaridade, que apresenta distribuição dos valores de 0 a 0,048494 (Figura 7-a), a normalização é feita na forma de aplicar o fator de escala da seguinte maneira: para cada pixel matriz usar a álgebra para identificar a sua nova posição proporcional à escala entre "zero" e "um" (Figura 7-b).

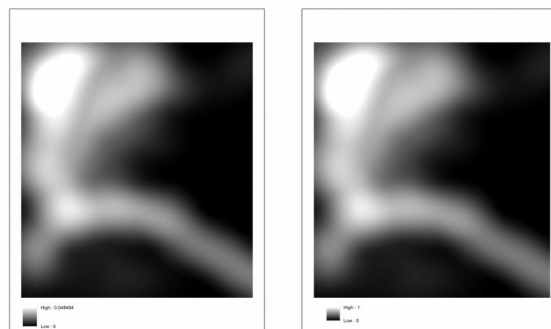


Fig. 7 - a) mapa inicial (valores de 0 a 0,048494) e (b) o mapa normalizado (valores de 0 a 1)

A álgebra (2) é estruturada como uma regra de proporção, para o caso da escala de 0 a 1:

$$\frac{(\text{valor antigo} - \text{valor mínimo})}{(\text{valor máximo} - \text{valor mínimo})} = \frac{(\text{novo valor})}{(1-0)} \quad (2)$$

Logo: "novo valor = [(valor antigo - valor mínimo antigo) * (maior valor nova escala – menor valor nova escala)] / (valor máximo antigo – valor mínimo antigo) ".

Observe que quando a variável é entendida como benefício, o valor mais alto é o mais positivo. Por outro lado, quando a variável é entendida como custo ou impedância, o valor mais baixo é o mais positivo, o que significa que a escala de valores deve ser invertida. Assim, nos casos de “custo ou impedância” após a normalização para a escala de 0 a 1 conforme procedimento explicado, a camada deve ter seus valores invertidos. Inverter uma escala de valores no caso de 0 a 1 é realizado pelo procedimento “1-valor existente”.

4. AVALIAÇÃO MULTICRITÉRIOS - PROCEDIMENTOS EM RASTER OU EM VETOR

A avaliação multicritérios pode ser feita pelo uso de uma coleção de matrizes ou mapas *rasters*, considerando os pesos para cada camada ou matriz e observando os valores contidos em cada célula da mesma. Pode-se, também, realizar a análise multicritério por coleção de arquivos vetoriais especificando os pesos para cada camada vetorial e reconhecendo as notas de cada componentes de legenda contidas nas tabelas associadas às primitivas gráficas.

Importante passo para a integração de dados em análise multicritérios é a indicação do peso escolhido para cada variável em função dos objetivos definidos para a análise. Há muitas maneiras de definir os pesos, que se distinguem em dois processos: o guiado pelo conhecimento especialista e o guiado pela observação dos dados. Os estudos sobre a definição de pesos guiados pelos dados, conhecido como “data driven evaluation” seguem a lógica da mineração de dados (Moura et al, 2010; Castro, 2010) ou da assinatura (Borges, 2009); ao passo que estudos guiados pelos especialistas, conhecidos como “knowledge driven evaluation” apresentam como os procedimentos mais conhecidos o método de Thomas Saaty (Moura, 2007) e o método Delphi (Moura et al, 2010; Silveira, 2013).

4.1. Análise Multicritérios por camadas *raster* - utilização do *Raster Calculator*

Em qualquer ferramenta de álgebra de mapas existentes nos softwares, o processo é realizado pela aplicação da média ponderada segundo os procedimentos básicos:

- As camadas de informações devem ser representada no grid ou formato tiff, com controle dimensão da matriz (extensão), o que significa controle do número de linhas e colunas e da dimensão do pixel (célula).
- As camadas devem representar, por meio de seus valores em *raster*, o grau de pertinência de cada componente de legenda, de acordo com a sua adequabilidade para o objetivo de investigação.
- Os pesos devem ser definidos, por algum processo.
- O processo de integração das camadas é feito por álgebra de mapa, exemplo: 'variável 1' * peso da variável 1 + " variável 2" * peso da variável 2 + " variável 3" * peso da variável 3 + " variável 4" * peso da variável 4., etc.....

4.2. Análise Multicritérios por camadas vetoriais - usando o “*Weighted Sum for Feature Class*”

A integração ou álgebra de mapas também pode ser realizada em mapas construídos através de feições vetoriais, mas a condição é que as feições tenham a a mesma unidade territorial de integração, que pode ser o mesmo polígono em todos as camadas de variáveis (como o exemplo de setores censitários, polígonos de municípios em análise regional, ou outras unidades territoriais utilizadas para representação da informação). No caso de coleção de dados que apresentam diferentes formas de espacialização da informação, a forma de se garantir que todos tenham e mesma unidade de integração é a conversão de dados para uma grade regular, conformada por células ou pontos, e nesses elementos associar as informações em tabelas alfanuméricas. Uma vez organizados os dados em tabelas, a álgebra de mapas na análise multicritérios é elaborada através das colunas, por exemplo: "O valor da coluna 1 * peso definido é a variável 1 + valor da coluna 2 * peso definido é a variável 2 + valor da coluna 3 * peso definido é a variável 3, etc...".

Os professores Jankowski e Ligmann-Zielinska (2010, 2012, 2014) desenvolveram uma ferramenta que promove a combinação multicritério em camadas vetoriais. O peso de cada camada é dado na interface do aplicativo, mas o conjunto de valores sobre cada variável deve estar registrado nas colunas das tabelas, associadas ao mapa segundo a expressão vetorial em primitivas gráficas (conjunto de pontos, conjunto de polígonos, etc). Apresentamos uma forma de aplicar a ferramenta, partindo da conversão de matrizes de camadas *rasters* em vetores:

4.2. a. Preparo dos dados para análise

Em primeiro lugar, é preciso que todas as variáveis que serão combinadas tenham suas representações territoriais em vetores, no caso em pontos, pois contemplam qualquer tipo de especialização. O primeiro passo é gerar uma camada de pontos que representem as células das matrizes. Isto pode ser feito selecionando qualquer uma das camadas e aplicando a ferramenta de conversão de *raster* para pontos. Tendo a primeira camada de pontos ela é usada para extrair os valores das outras camadas, de modo que no final nessa camada estarão disponíveis colunas com informações de cada camada.

Uma informação importante: o método Monte Carlo na ferramenta disponibilizada pelo mencionado professor trabalha apenas com valores de 0 a 1 nas colunas a serem avaliadas. Portanto, é necessário observar se aconteceram arredondamento de casas decimais, pois os valores que excederem devem se limitar a 1, o que pode ser ajustado solicitando que tudo o que for maior que “1” seja substituído por “1”.

4. 2.b. Multicritérios em arquivos vetoriais - “*Weighted Sum for Feature Class*”

Após organizar todas as informações das variáveis em colunas de uma tabela associada aos pontos, com valores apresentando de 0 a 1 significando a “superfície potencial de distribuição dos fenômenos ou ocorrência”, valores normalizados, emprega-se a ferramenta. Neste ponto é indicado para usar a ferramenta “*Weighted Sum for Feature Class*” desenvolvido pelo Prof. Jankowski para ser utilizada no ArcGIS.

Na ferramenta devem ser especificadas as colunas com os atributos que serão combinadas, e indicar os pesos do item “*Weights String*”, colocando os valores de acordo com cada variável. O resultado é apresentado na coluna “score”, e a ordenação dos elementos segundo os valores é apresentada na coluna “rank”. O resultado pode ser apresentado em mapa temático de pontos, ou o mapa vetorial pode ser convertido em *raster*.

5. AVALIAÇÃO DA INCERTEZA E EMPREGO DA ANÁLISE DE SENSIBILIDADE - CONFIABILIDADE NOS RESULTADOS DA MULTICRITÉRIO MEDIANTE PROCESSO DE MONTE CARLO

Uma vez feita a análise de multicritérios, é importante avaliar a incerteza existente na resposta, para que ela funcione como apoio à tomada de decisão. O processo que ajuda nessa avaliação é o Monte Carlo, que consiste em simular diferentes possíveis pesos para as variáveis, dentro de limites de máximo e mínimo em uma faixa de valores. Utilizamos a ferramenta “*Monte Carlo Weighted Sum*”, cortesia do Professor Jankowski (op. cit.) no seu pacote “*Multicriteria Evaluation for Discrete Set of Options*”, no item “*Uncertainty and Sensitivity Analysis*”.

O professor indica que o peso atribuído a cada variável deve ser informado e, além disso, os valores que compõem o “*Minimum Weighted String*” (relação de valores mínimos para a simulação, em lista de acordo com as variáveis) e o “*Maximum Weighted String*” (relação de valores máximos). Exemplo: tendo definido o peso de 0,25 (25%) para uma variável, optou-se por simular possíveis pesos dentro do limite (*range*) de 0,10 a 0,30.

É necessário definir o processo metodológico para escolher esta janela de trabalho para simulação de pesos. Os mesmos processos usados para escolher os pesos podem ser utilizados aqui. Como contribuição do presente trabalho, sugerimos:

- a) que os valores mínimo e máximo sejam obtidos a partir da avaliação do conjunto de pesos indicados por especialistas.
- b) que seja utilizado o desvio padrão dos pesos sugeridos pelos especialistas.
- c) no caso de se optar por uma forma mais robusta de indicação de limites de pesos, empregar a lógica da função de densidade de probabilidade

Exemplo – Investigação sobre lugares atrativos para a concentração urbana. Quatro variáveis foram definidas e três especialistas consultados, que indicaram os seguintes valores de pesos para as variáveis:

Presença de Cobertura Vegetal Expressiva - 30 %, 25 %, 20% - média 25%

Densidade Residencial - 23 %, 20 %, 32% - média 25%

Densidade Comercial - 20 %, 40 %, 15% - média 25%

Acessibilidade e Capilaridade - 25 %, 25 %, 25% - média 25%

Para usar o desvio padrão como referência - calcular a variância (3) e o desvio padrão. A variância é calculada subtraindo-se o valor observado do valor médio. Esta diferença significa o quanto um valor observado fica muito longe do valor médio.

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \mu)^2$$

Variância

(3)

$$\text{Vegetação} = \frac{(30-25)^2 + (25-25)^2 + (20-25)^2}{3-1} = \frac{25+0+25}{3-1} = 25$$

$$\text{Densidade Residencial} = \frac{(23-25)^2 + (20-25)^2 + (32-25)^2}{3-1} = \frac{4+25+49}{3-1} = 39$$

$$\text{Densidade Comercial} = \frac{(20-25)^2 + (40-25)^2 + (15-25)^2}{3-1} = \frac{25+225+100}{3-1} = 175$$

$$\text{Acessibilidade e capilaridade} = \frac{(25-25)^2 + (25-25)^2 + (25-25)^2}{3-1} = \frac{0+0+0}{3-1} = 0$$

O desvio padrão é calculado extraindo a raiz quadrada da variância. A variância determina o desvio em relação à média aritmética e desvio padrão analisar os valores regularidade. Logo: Vegetação = $\sqrt{25} = 5$; Densidade Residencial = $\sqrt{39} = 6,24$; Densidade Comercial = $\sqrt{175} = 13,23$; Acessibilidade e Capilaridade = $\sqrt{0} = 0$

Isso significa que não há irregularidade relativa ao peso dado por três especialistas para a acessibilidade e capilaridade, há baixa irregularidade relativa ao peso dado à vegetação, há média irregularidade no peso dado à densidade residencial, e há alta irregularidade no peso dado pelos especialistas para a densidade comercial. Um processo operacional seria aceitar a variação do peso da vegetação entre 20 e 30% (média de 25 %), da densidade residencial entre 18,76% e 31,24% (média de 25 %), da densidade comercial entre 11,77 %; e 28,23% (média de 25 %) e da acessibilidade e capilaridade de 25 a 25% (média de 25 %).

Outro processo de definição de faixas de peso que estamos propondo é a aplicação da função de densidade de probabilidade, reconhecendo que as variáveis tendem a se distribuírem segundo valores que fazem a curva normal ou curva do sino, como a representada a seguir. Estatisticamente, se optarmos por escolhermos faixas a determinadas distâncias em relação à média, segundo a tabela abaixo, isto significa definir o nível de confiança em faixas de 68,2 %; 95,4% e 99,6% (Figura 8).

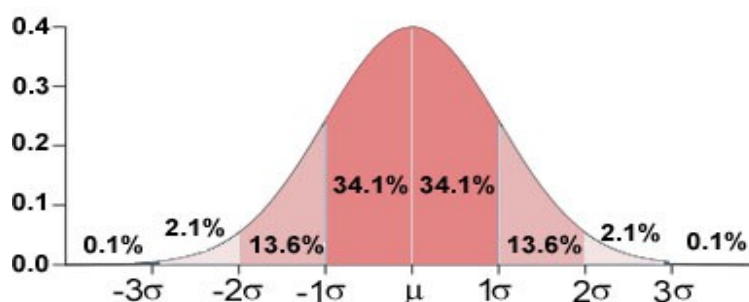


Fig. 8 - Diagrama de desvio padrão em função de densidade de probabilidade

$T = T' \text{ mais ou menos } k * St.$

Sendo T - a nova posição, T' , a posição média, K pode variar entre 1, 2 e 3, e sendo o desvio padrão

(na figura 8 representada por 1σ).

$P(T' - ST_{n-1}, T' + ST_{n-1}) = 68,2\%$ (porque é 34,1% para cada lado)

$P(T' - 2 ST_{n-1}, T' + 2 ST_{n-1}) = 95,4\%$ (porque é 34,1+13,6% para cada lado)

$P(T' - 3 ST_{n-1}, T' + 3 ST_{n-1}) = 99,6\%$ (porque é 34,1+13,6+2,1% para cada lado)

A primeira faixa, mais próximo à média, compreende 68,2 % do conjunto de possíveis valores, a segunda faixa compreende 95,4% e na última e terceira faixa 99,6 %. Assim, pode-se escolher a referência desses limites para se definir o máximo e o mínimo no intervalo de valores dentro do qual o processo de Monte Carlos vai simular valores aleatoriamente. Observa-se que a escolha de faixas mais estreitas devem ser usadas em situações de menos dúvidas sobre a variável (pois as escolhas aleatórias de pesos serão feitas dentro de menor espectro de possibilidades), ao passo que faixas mais largas exigirão muitas integrações pra uma boa análise, pois aumentam as possibilidades de escolhas de valores. Não se aconselha o uso da faixa 3, pois ela tem é tão larga que seria praticamente simular em todo o universo possível. Exemplo - Motivo de investigação - localização de áreas adequadas para a densificação urbana. Escolha de 4 variáveis com 25% de peso para cada, com a seguinte relação entre a máximo e mínimo para realização da simulação de Monte Carlo (Tabela 1):

TABELA 1 - FUNÇÃO DE DENSIDADE DE PROBABILIDADE USAR EXEMPLO

Variável	Relação custo/benefício	Peso	Mínimo/Máximo	Explicação
Vegetação	Custo	25	Desvio Padrão = 5 Faixa 1 = 5 30 Mín 20 Max	Não há muita dúvida sobre a distribuição, importância e caracterização da vegetação - faixa 1 da curva normal
Concentração Residencial	Benefício	25	Desvio padrão = 6,25 Faixa 2 = 12,5 Mín 12,5 Máx 37,5	Existem algumas dúvidas sobre a importância da variável como elemento de atratividade - faixa 2 da curva normal
Concentração Comercial	Benefício	25	Desvio Padrão = 13,23 Faixa 3 = 39,69 Mín 0 Max 64,69	Há dúvidas quanto à distribuição das variáveis e o seu papel – faixa 3 da curva normal
Acessibilidade e Capilaridade	Benefício	25	Desvio Padrão = 0 Faixa 1 = 0 O valor 25 % não se altera Mín 25 Max 25	Não há dúvidas quanto à distribuição das variáveis e quanto ao significado da variável - faixa 1 da curva normal

Lembrando sobre o exemplo: Vegetação = $\sqrt{25} = 5$, Densidade Residencial = $\sqrt{39} = 6,24$, Densidade Comercial = $\sqrt{116,6} = 10,79$, Acessibilidade e Capilaridade = $\sqrt{0} = 0$

Estamos propondo a adoção de um ou mais dos procedimentos para se decidir em qual faixa colocar a variável:

a) Observe se há muita diferença entre o menor e o maior valor sugerido entre os especialistas. Opte pela faixa 1 (uma vez o desvio-padrão para cada um dos lados do valor médio) quando a variação de opiniões é menor, e por faixas mais amplo (faixa 2 – duas vezes o desvio-padrão para cada um dos lados do valor médio), mas evite a faixa 3, pois significa simular em faixa que engloba praticamente todos os dados.

b) Ao aplicar o Delphi ou o AHP, no processo de consulta ao especialista sobre o grau de importância da variável, acrescente a pergunta sobre o grau de certeza/segurança com que ele opinou sobre variável (por exemplo - Você tem dúvidas sobre isso? Escolha o seu nível de dúvida de 0 a 10). Identifique assim as variáveis onde os próprios peritos declararam que tinham dúvidas ou incertezas na atribuição de pesos, e escolha para essas variáveis amplitudes maiores (faixa 2 – 2 vezes o desvio padrão para cada lado da média) para a simulação de Monte Carlo.

c) Analisar a qualidade do mapa que representa a variável. Se o mapa por algum motivo resultar em redução de confiança em seus dados, colocar a variável na faixa de maior amplitude de simulação (faixa 2). Essas faltas de confiança podem ser problemas relacionados com a escala cartográfica (resolução espacial), modo de representação (se o mapa foi produzido por interpolações ou generalizações), ou qualquer outra condição que fazem o usuário pensar que poderia não estar tão seguro sobre a qualidade da informação contida no mapa.

5.1. Simulação de Monte Carlo com a ferramenta "Monte Carlo Weighted Sum"

A ferramenta requer que sejam especificadas as variáveis envolvidas, assim como a relação de valores mínimos e de valores máximos (Figura 9).

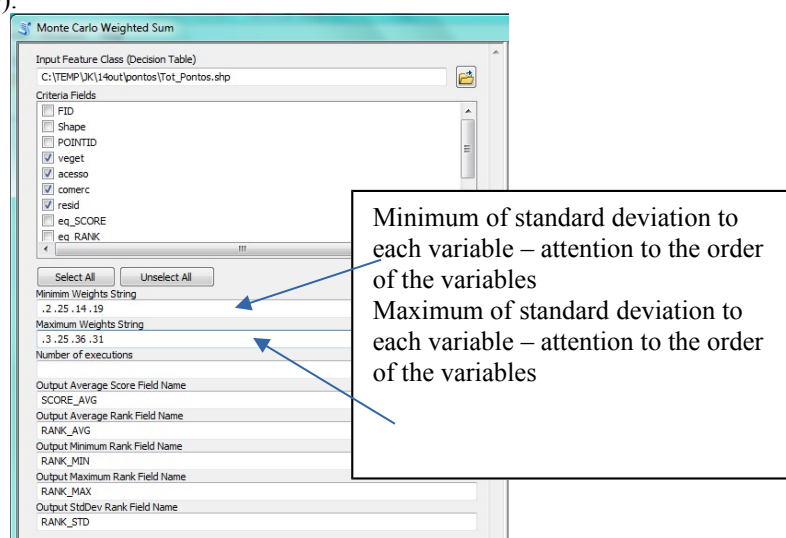


Fig. 9 - A utilização de "Monte Carlo Weighted Sum", ferramenta desenvolvida pela Prof. Jankowsky.

Jankowski e Ligmann-Zielinska (2010, 2012, 2014) indicam que os resultados podem ser:

1. Alta posição no rank da média e baixa posição no rank do desvio padrão - significa a posição vencedora e com alta confiabilidade.
2. Baixa posição no rank da média e baixa posição no rank do desvio padrão - representam a posição perdedora e com alta confiabilidade, logo a opção pode ser descartada.
3. Alta posição no rank da média e alta posição no rank do desvio padrão - significa a posição vencedora mas com alta incerteza, exigindo estudos mais detalhados em função da instabilidade.
4. Baixa posição no rank da média e alta posição no rank do desvio padrão - significa posição perdedora que pode ser eliminada, mas com menos confiança sobre a decisão do que o caso 2.

O referido Professor indica a comparação da média da classificação, que é realmente muito semelhante à integração multicritério, mas colocado na classificação. Assim, não importa se o rank AVG (ranking da média, ou ranking do valor de multicritérios) ou o multicritério é usado sobre a comparação com o desvio padrão (Figura 10). É importante lembrar que, no caso de escolher o RANK, os menores valores numéricos significam os primeiros do ranking, logo são os de maior valor de resultado na análise de multicritérios (melhor resultado para a investigação realizada). Os resultados podem ser vistos no mapa temático de pontos, ou eles podem ser convertidos em *raster*.

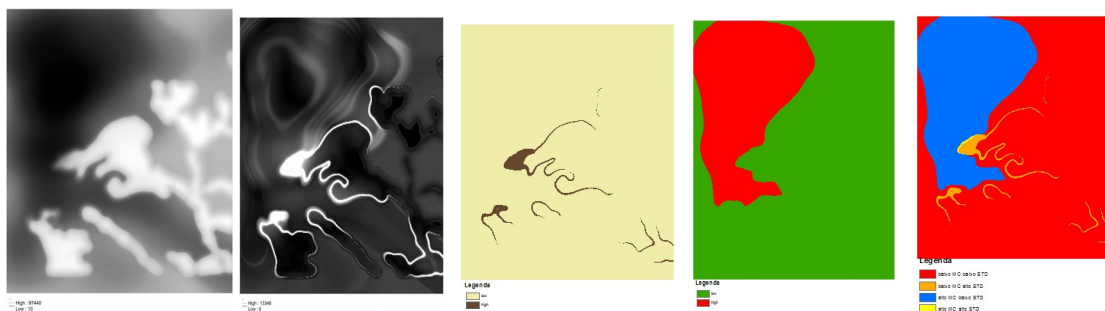


Fig. 10 - Monte Carlo. (a) Rank da Média (AVG), (b) Rank do Desvio Padrão (STD) (c) Rank STD com apenas alto/baixo, (d) Resultado Multicritério apenas a alta/baixa, (e) comparação STD & MC.

Interpretação:

- Branco em MC+ preto em STD (Baixo - Baixo) = baixo nível de média e baixo desvio padrão significa não interesse e com confiabilidade aos resultados. Lugar que pode ser descartado.
- Preto MC + preto em STD (Alto - Baixo) = alto nível de média e baixo desvio padrão – lugar de interesse e sem dúvidas no resultado.
- Branco em MC + branco em STD (Baixo - Alto) = baixo nível de média e alto desvio padrão – não é um lugar de interesse, mas há dúvidas, então pode ser melhor investigado.
- Preto em MC + Branco em STD (Alto - Alto) = alto nível de média e alto desvio padrão, então é alto interesse mas com muitas dúvidas, então cabe melhor investigação.

A fim de tornar a visualização mais fácil, foi decidido representar o dado simbolizado em cores de duas faixas de legenda (estudos mais detalhados pode ser feito através da comparação de média-baixa). O processo consiste em classificar a imagem *raster* segundo duas bandas (no caso da combinação simples, figura 11) ou em 3 ou mais bandas (no caso de combinação complexa, figura 12) através do processo de simbolização por quebra natural (*natural breaks*). Um exemplo de combinação simples é apresentado na figura 10 (c), (d) e (e).

5.2 . Cotejo entre os resultados de Análise Multicritérios (ou AVG) e o nível de Incerteza pelo desvio padrão (STD) para avaliação da confiabilidade os resultados

A comparação pode ser feita de forma simplificada (Tabela 2), considerando-se apenas os valores máximos e mínimos, ou mesmo, de forma mais complexa (Tabela 3), considerando baixo, médio e alto valor ou fatiamentos ainda mais detalhados.

TABELA 2 - COTEJO SIMPLIFICADO RANK STD & RANK MÚLTICRITÉRIOS

Multicritério Resultados	Rank STD		
	Valores para combinação	0	1
	0	0	1
	2	2	3

Matriz de combinação é a identificação das relações: 0 (alto STD, alto MC), 1 (baixo STD, alto MC), 2 (alto STD, baixo MC), 3 (baixo STD, baixo MC).

TABELA 3 - COTEJO MAIS COMPLEXO RANK STD & RANK MÚLTICRITÉRIOS

Multicritério Resultados	Rank STD			
	Valores para combinação	0	1	2
	0	0	1	2
	3	3	4	5
	6	6	7	8

Matriz de combinação é identificação das relações: 0 (alto STD, alto MC), 1 (médio STD, alto MC), 2 (baixo STD, alto MC), 3 (baixo STD, médio MC), 4 (médio STD, médio MC), 5 (baixo STD, médio MC), 6 (alto STD, baixo MC), 7 (médio STD, baixo MC), 8 (baixo STD, baixo MC).

5.3. “Variance-Based Global Sensitivity Analysis”

O objetivo é definir quantitativamente os pesos que têm mais influência no resultado da análise de multicritérios. Ela apresenta duas medidas de sensibilidade: a de primeira ordem (S) e o efeito total do índice de sensibilidade (ST). O processo define o “Average Shift in Ranks” (média de mudanças no ranking) e o “Robust Winner” (vencedor mais robusto), o que significa a classificação média e a importância ou qualidade do vencedor. O aplicativo

foi desenvolvido por Jankowski e Ligmann-Zielinska (2010, 2012, 2014), chamado "*Variance Decomposition*", e que se encontra no pacote de ferramentas "*Uncertainty and Sensitivity Analysis*".

Continuamos com o exemplo de combinação de vegetação, acessibilidade, concentração residencial e concentração comercial. A mesma relação de "pesos mínimos *string*" e dos "pesos máximos *string*" utilizado antes é informado (no exemplo, mínimos 2 .25 .14 .19) (no exemplo, máximos .3 .25 .36 .31). Note que o processo irá fazer mil interações e que a melhor opção do ID do Objeto é -1, o que significa o seguinte: um elemento que não existe, porque este valor do ID não existe. Isso significa que ele não definirá qualquer relação/resultado com a melhor opção *Object ID*, o que significa que a melhor opção não foi calculada, e todos os ASR (*Average Shift in Ranks*) serão considerados.

Testamos com 10.000, 1.000 e 500 amostras de seleção aleatória de pesos, mas devido ao número de informações envolvidas no cálculo o processo só funcionou com 100 interações. O exemplo foi feito com pontos de amostragem a cada 10 metros, em área de 3Km por 3Km, e para 4 variáveis. Seria necessária uma infraestrutura computacional de melhor desempenho para o elevado número de cálculos (4 critérios + 2 * 100 amostras * 90000 células equivale a 54 * 10⁶ computações).

Para analisar os resultados obtidos Jankowski explica:

"1. Cada *S* deve estar dentro da faixa de 0 a 1 para que a soma dos *S* ≤ 1 . Se são obtidos grandes valores negativos ou valores maiores do que "1", a amostra de Monte Carlo foi provavelmente demasiado pequena a abordagem é insuficiente (valores negativos próximos de zero estão OK).

2. A soma de *ST* deverá ser ≥ 1 (Se $\sum S = \sum ST$ então o modelo é perfeitamente linear)."

A interpretação dos dois valores é resumida no seguinte esquema apresentado por Jankowski e Ligmann-Zielinska (2010, 2012, 2014) acerca de *Si* e *Sti*. *Si* é o índice de sensibilidade de primeira ordem e o *Sti* o índice de interação de sensibilidade:

- a) *Si* relativamente alto - o fator que é individualmente influente sobre a variabilidade do modelo resultante.
- b) Soma de *Si* sobre todos os *inputs* - Percentual de variabilidade de resultado em função dos *inputs* independentemente; o restante (em 100%) é a fração de variação de resultado em função das interações entre os *inputs* (*input* - variáveis de entrada).
- c) *STi* - *Si* relativamente alta - Input "*i*" é altamente envolvido nas interações com outros fatores, todas os *inputs* com grande diferença são envolvidas nas interações entre si. Note que o input "*i*" pode ser individualmente insignificantes (baixo *Si*), mas ser influente quando envolvido em interações com outros fatores (alto *Sti*). Ele então influencia a variância do resultado mais através de interações do que individualmente.
- d) Valor relativamente baixo de *STi* - um fator insignificante.

No original:

- "a) Relatively high *Si* - a factor that is singly influential on the variability of model output;
- b) Sum of *Si* over all inputs - Percent of output variability due to the inputs taken independently; the remainder (to 100%) is the fraction of output variance due to the interactions among inputs.
- c) *STi* - *Si* relatively high - Input *i* is highly involved in interactions with other factors, all inputs with high difference are involved in interactions among each other. Note that input *i* can be singly insignificant (low *Si*), but influential when involved in interactions with other factors (high *Sti*). It would therefore influence the output variance more through interactions than individually.
- d) Relatively low value of *STi* - An insignificant factor."

6. ESTUDO DE CASO - REGIÃO DA PAMPULHA EM BELO HORIZONTE - DEFINIÇÃO DE ÁREAS QUE DEVEM SER PRESERVADAS DE ACORDO ÀS CONDIÇÕES AMBIENTAIS

A fim de apresentar um estudo de caso para ilustrar a proposta metodológica, foi escolhida a região administrativa da Pampulha, zona norte de Belo Horizonte, e devido à sua importância como patrimônio arquitetônico, cultural e ambiental da cidade. Pampulha foi o primeiro bairro modernista planejado no Brasil por Oscar Niemeyer, antes do planejamento de Brasília. Por esse valor o conjunto é candidato a patrimônio da humanidade pela Unesco. Além disso, devido ao expressivo crescimento urbano e das transformações da paisagem, atualmente a região da Pampulha apresenta muitos conflitos de interesses, em função da justaposição de valores ambientais, econômicos (eixo de crescimento da cidade), problemas sociais e presença de valores culturais e históricos.

Três variáveis foram selecionadas (Figura 11):

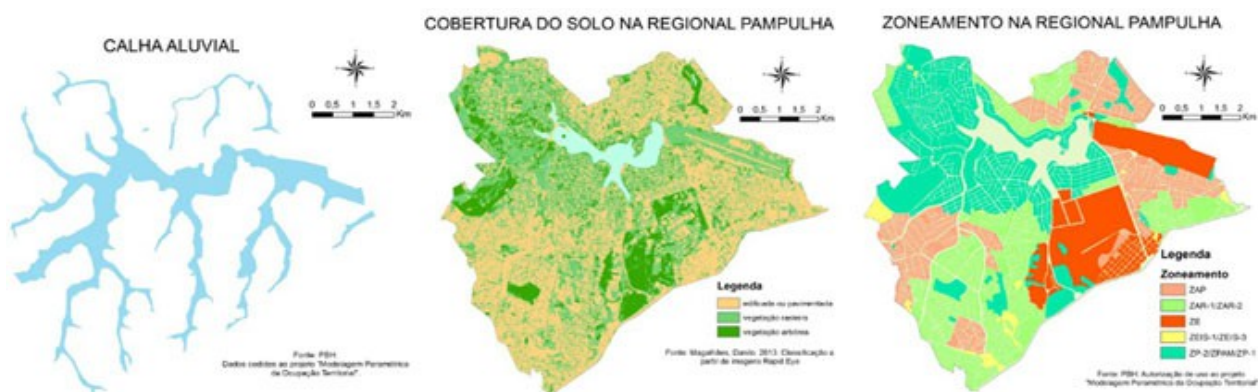


Fig. 11 - Variáveis: (a) Canal Aluvial (b) Cobertura do Solo (c) Zoneamento do Plano Diretor.

Os mapas das variáveis foram transformados em *raster*, e para serem apresentados como *superfície potencial de distribuição do fenômeno ou ocorrência*, eles foram tratados por interpoladores de análise de vizinhanças e de agrupamentos, conforme o discutido no tópico 3. No caso de canal aluvial, foi aplicado o *Distância Euclidiana* que define a área de influência desta ocorrência (Figura 12-a). No caso da cobertura do solo, foi realizada classificação em função dos interesses de preservação, e a camada *raster* foi convertida em pontos com controle da resolução espacial, em função dos objetivos de análise (Figura 12b). O zoneamento também foi convertido para valores em *raster* e classificado em função de sua relevância para os valores ambientais e, da mesma forma, depois convertido em pontos (Figura 12-c).

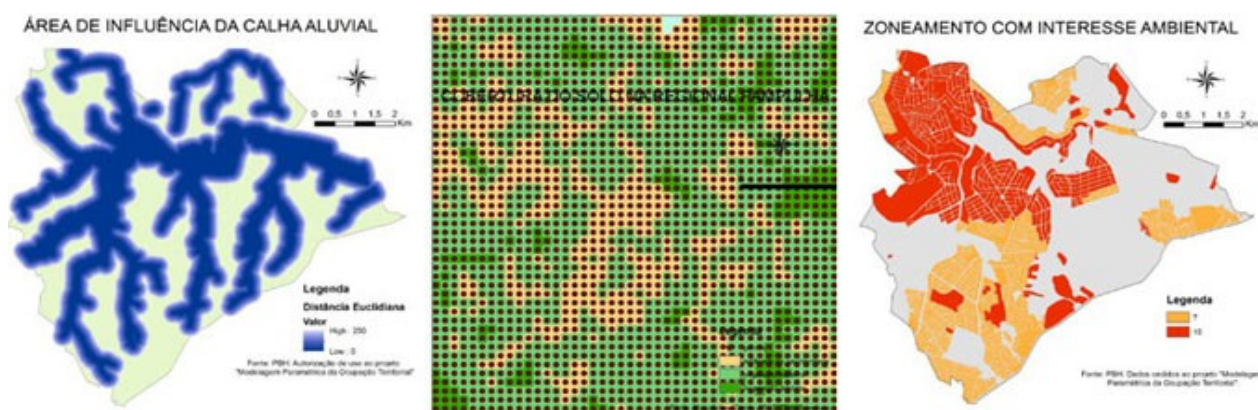


Fig. 12: (a) Canal Aluvial - Distância Euclidiana (b) Convertidos -pontos(c) Classificados *raster*

Os mapas convertidos para pontos foram submetidos ao interpolador de *Densidade de Kernel* considerando-se a distribuição de valores. Como os pontos correspondia a cada célula da matriz, eles compunham grade de distribuição espacial homogênea, então por esse motivo os valores registados nos pontos definiram a *superfície potencial de distribuição do fenômeno ou ocorrência*, como resultado da influência da vizinhança entre os pontos. Isso resultou na distribuição de interesses em proteção ambiental segundo a cobertura do solo (Figura 13-a), e de acordo com o zoneamento (Figura 13b).

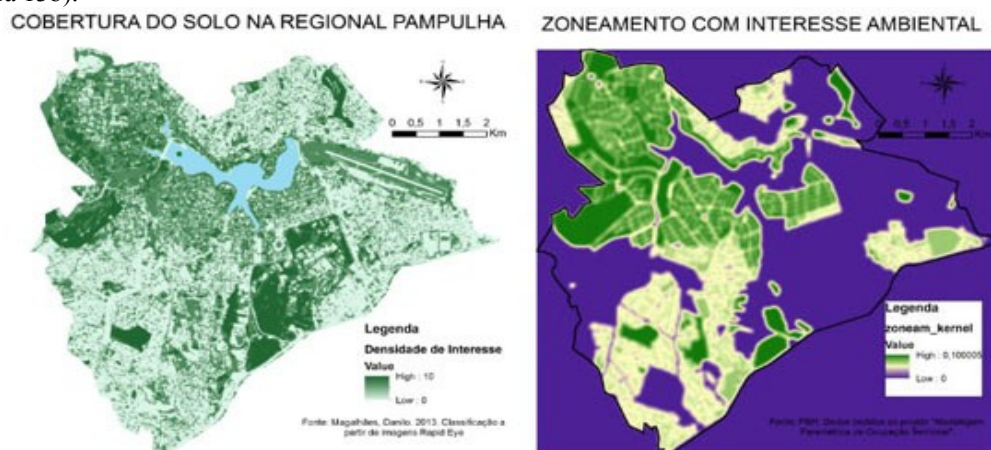


Fig. 13 (a) Interesse em proteção ambiental na cobertura do solo; (b) Interesse em proteção ambiental no zoneamento.

As imagens *raster* representando interesse na proteção ambiental segundo a cobertura do solo, segundo o zoneamento e segundo a distância à calha aluvial foram normalizadas para permitir comparações e combinações. Todos os 3 mapas foram convertidos em valores de 0 a 1, de acordo com o item 3.2. Nota-se que o mapa não se altera, mas só recebe os valores de 0 a 1 para permitir a comparação e a integração entre os níveis de *raster* (Figura 14).



Fig.14: Os mesmos mapas antes e depois da normalização.

Com os mapas normalizados eles foram combinados em análise multicritérios, usando os pesos: calha aluvial - 40 %, zoneamento - 30% e cobertura do solo - 30% (Figura 15-a). Em seguida, foi aplicada a simulação de Monte Carlo, usando a faixa de variação de acordo com o nível de suas dúvidas sobre a utilização da variável e sobre a qualidade cartográfica da variável, interpretação ajustada pela curva de função de densidade e probabilidade, como proposto no item 5 (Figura 15-b):

a) Calha aluvial - não há dúvidas sobre a importância e distribuição espacial, por isso a opção pela primeira faixa da função de densidade de probabilidade, o que significa uma vez o desvio padrão antes de depois do valor médio, resultando em variedade de simulação de Monte Carlo de 39% a 41 %, a média é de 40 % e o desvio padrão foi 1.

b) Cobertura do solo - há dúvidas sobre a qualidade da informação, uma vez que ela foi obtida por classificação da imagem RapidEye que provoca mistura entre solo exposto e ocupação urbana e, sobretudo porque mapeia a copa das árvores por cima, sem avaliação do que acontece no nível do solo propriamente dito. Por este motivo, optou-se pela segunda faixa da função de densidade de probabilidade e função densidade (duas vezes o desvio padrão da média), os valores vão de 20% a 40 %, em média de 30 %, desvio padrão de 14,142, por isso a variedade de simulação de Monte Carlo vai de 1,716 para 58,284. A fim de analisar o impacto de tão ampla faixa, foi simulada a faixa de uma vez o desvio padrão para cada lado do valor médio (a partir de 15,858 a 44,142) e o resultado foi praticamente o mesmo. Assim, os estudos finais que identificam a importância da variável não são induzidos pela dúvida sobre a qualidade do dado, mas sim pela sua importância real em comparação às outras variáveis.

c) Zoneamento – o nível de segurança sobre a informação da variável foi traduzida na escolha da primeira faixa da função de densidade de probabilidade (uma vez o desvio-padrão em relação à média), de modo que os valores vão de 25% a 35 %, em média de 30 %, desvio-padrão de 7,071, para o intervalo de simulação de Monte Carlo vai de 22,929 a 37.071 %.

O "Monte Carlo Weighted Sum" fornece os resultados a classificação da média (Classificação AVG), o que significa a ordem de valores obtidos pela análise de multicritérios, e o ranking do desvio-padrão (Rank STD). (Figura 15-b e Figura 15-c).

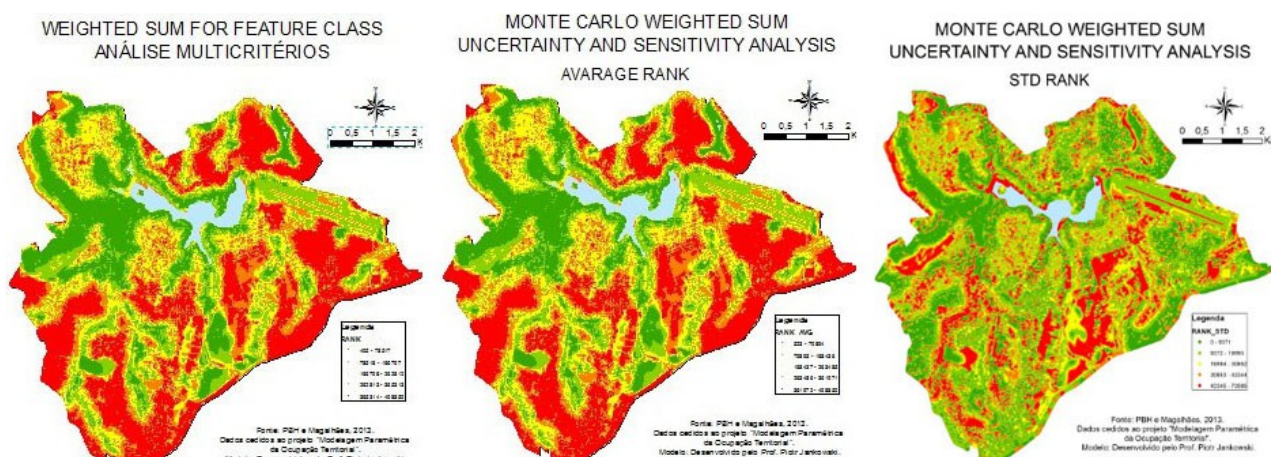


Fig. 15 - (a) resultado da multicritérios; (b) Ranking do AVG; (c) Ranking do STD

Verde - grande interesse na proteção ambiental, mas alta dúvida devido à incerteza. Vermelho - pouco interesse na proteção ambiental, e baixa dúvida devido à incerteza.

O ranking do desvio-padrão (Rank STD) foi comparado com o ranking de média (Rank AVG). Para tornar a comparação mais simples, os mapas foram classificados somente na representação das altas e baixas (Figura 16). Ao se analisar as combinações, segundo Jankowski e Ligmann-Zielinska (2010, 2012, 2014), as regras são as seguintes:

1. Alta posição no ranking da média e baixa posição no ranking do desvio padrão significa o vencedor na escolha do melhor local e alto nível de confiança na sua priorização.
2. Baixa posição no ranking da média e baixa posição no ranking do desvio padrão representam o perdedor na posição da escolha do melhor lugar e alto nível de confiança no resultado, logo a área pode ser ignorada.
3. Alta posição no ranking da média e alta posição no ranking do desvio padrão - caracteriza as opções que são potencialmente boas, mas carecem de um estudo mais aprofundado devido à incerteza.
4. Baixa posição no ranking da média e alta posição no ranking do desvio padrão - significa que é uma área fora de interesse e sem incerteza disto - entre as opções as áreas podem ser retiradas da análise, mas com menor confiança que em relação ao caso 2.

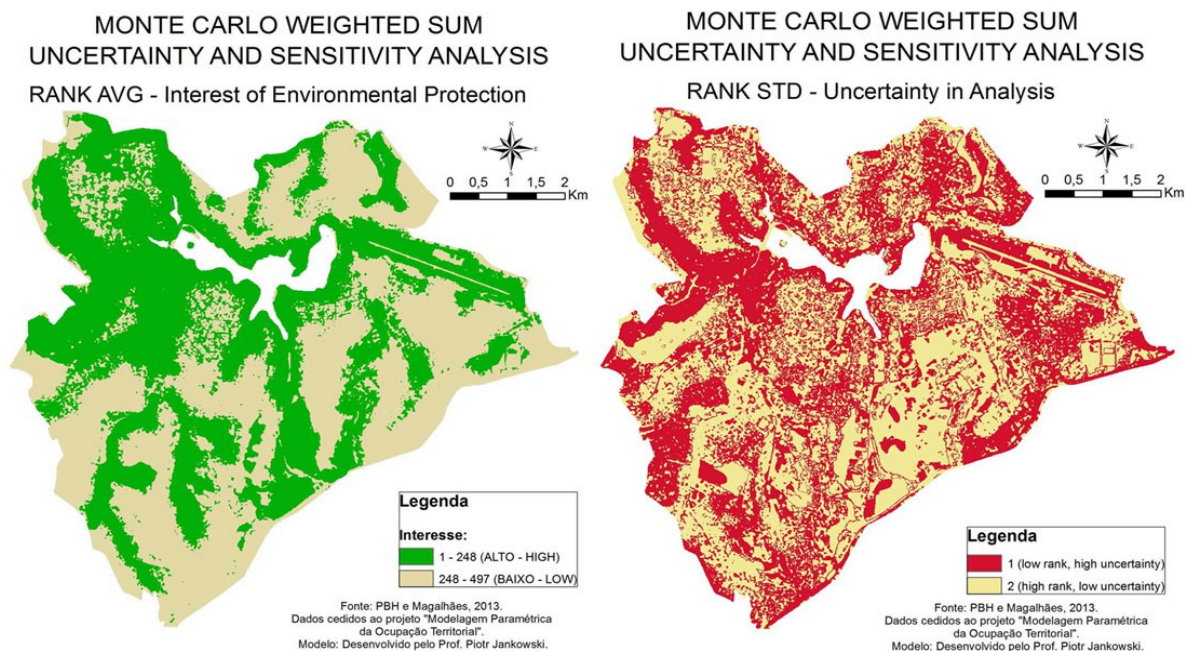


Fig. 16 - (a) Ranking AVG ou Média - alta/ baixa, (b) Ranking STD alta/ baixa.
Verde: grande interesse na proteção ambiental, Vermelho: Alta dúvida ou incerteza.

As duas representações de alto e baixo foram combinadas para promover a identificação dos quatro possibilidades (Figura 17).

MONTE CARLO WEIGHTED SUM UNCERTAINTY AND SENSITIVITY ANALYSIS

RANK AVG & RANK STD Enviroment Protection & Uncertainty in Analysis

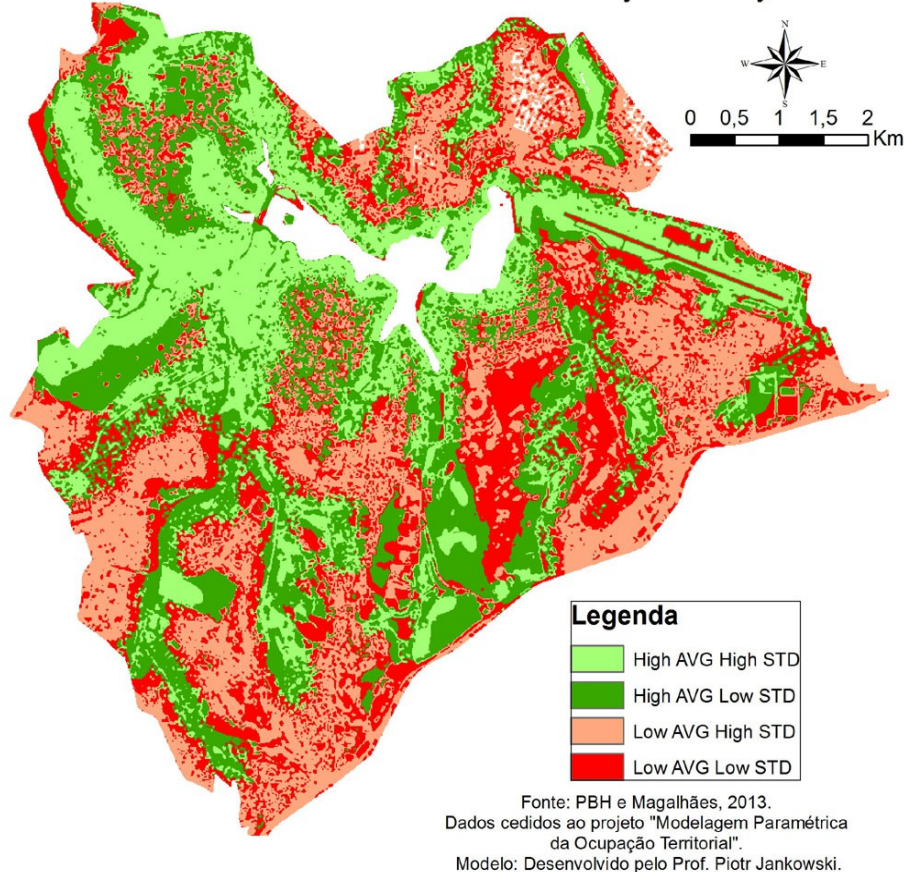


Fig.17 - Comparação ranking AVG e ranking STD, resultado da análise de sensibilidade

Para se definir quantitativamente quais variáveis tiveram maior influência sobre o resultado do ranking de multicritérios, a aplicação "Variance Decomposition" desenvolvido pelo Prof. Jankowski foi utilizada. Ela apresenta duas medidas de sensibilidade: a primeira ordem (S) e o efeito total do índice de sensibilidade (ST). São definidos a média de mudanças de posições no ranking ("Average Shift in Ranks") e o vencedor mais robusto ("Robust Winner"), este último resultante do ranking da média e do seu nível de significância. Os resultados foram os seguintes:

Variância Decomposição: GSA: média de câmbio das posições

Fator	S	ST.
Usosolo	1.043	0.954
Calhaaluv	-0,045	0.003
O zoneamento	-0.037	IN
		0,13

Soma(soma) 0,961 1,087

Fator	%S	%S.
Usosolo	104,3	87,8
Calhaaluv	-4.5	0,2
O zoneamento	3.3 -3.7	12.0

NONL 3.9

Ao analisar os resultados, o primeiro passo é interpretar os valores S: para "calhaaluv" (calha aluvial) e para o "zoneamento" os valores são próximos a zero ou praticamente igual a zero. Isto significa que estes dois critérios não

têm influência (zero) sobre a incerteza (desvio padrão) da média da aptidão ou adequabilidade. A incerteza (variabilidade) da média de aptidão ou adequação é praticamente (100%) determinado apenas pelo "usosolo" (cobertura do solo).

Os pequenos valores negativos e o valor de S que está um pouco acima de 100% (104,3) são pequenos erros, devido a dados ou arredondamentos no processo de amostragem. Os valores de índice de S são 100 %, 0 %, 0 %.

Os valores do índice total de interação ST confirmam aquilo que os valores S demonstraram, ou seja, que quase todas as propostas da média de aptidão ou de adequação é devido à distribuição de valores de uso do solo; $ST = 0.954$ significa que a cobertura do solo é uma variável que interage bastante com os outros dois critérios. Além disso, não existe nenhuma interação entre "calhaaluv" (calha aluvial) critério e os outros dois critérios, e pouca interação entre o zoneamento e os outros dois.

NONL valor igual a 3,9, que é muito baixa, diz-nos que em geral há pouca interação entre todos os três critérios. Isso corrobora com o diagnóstico anterior, quando se observou que a maioria da incerteza na média da aptidão é devido à cobertura do solo ao longo da área de estudo. Isso é bastante interessante, e exige uma nova interpretação com a observação da Figura 19. No mapa de comparação dessa figura as condições Alto-Alto e Alto_baixo são territorialmente adjacentes. Porque é que é assim? Estaria relacionado a um padrão de categorias de uso do solo? Acreditamos que a resposta a esta pergunta é que a área reconhecida como de importância para a preservação ambiental tem uma expressiva área territorial com diferentes usos quanto à cobertura do solo, e embora seja região importante de se proteger, há dúvidas sobre o fato disto ainda ser possível, por esses diferentes usos.

Um dos possíveis resultados práticos do modelo de análise de sensibilidade global é identificar fatores (critérios) que não são importantes. Segundo esta, poder-se-ia afirmar que o peso de zoneamento e de calha aluvial não são importantes. De fato, seria interessante executar uma experimentação dando um grande peso ao uso da terra e baixo peso para os outros dois critérios, e comparar o resultado com o mapa inicial de aptidão ou potencial de proteção ambiental. Outra implicação prática global da análise de sensibilidade é a observação de que a camada de cobertura do solo deve ser examinada quanto à sua precisão, posto que é resultante de classificação de imagem de satélite por PDI. É possível (mas não é garantido) que, com maior precisão da cobertura do solo possa se reduzir a incerteza no mapa de aptidão, o que se manifestaria por valores menores no mapa de desvio padrão.

7. CONCLUSÕES:

A análise de sensibilidade complementa e responde a muitas questões que os pesquisadores têm de enfrentar ao apresentarem estudos utilizando análise multicritérios, que são as seguintes:

- Como você pode fazer estas declarações?
- Por que você escolheu esses pesos e como as alterações nos pesos interferem nos resultados?
- O trabalho apresenta critérios defensáveis e reproduzíveis? Isso significa que, se você apresentar este trabalho em público ou reunião técnica, você é capaz de defendê-la? Caso outros pesquisadores sigam os métodos que você escolheu eles chegam a resultados semelhantes?

Quem trabalha na área de geoprocessamento ou geodesign, certamente já teve que responder a essas perguntas, ou mesmo já fez estas perguntas a si mesmo. O estudo com a análise de sensibilidade confere maior robustez aos estudos de análise multicritérios, hoje amplamente difundida entre técnicos geodesigners.

Assim, a análise de sensibilidade tem como objetivo demonstrar os resultados para tornar claros os limites das declarações apresentadas. Com os resultados da análise de sensibilidade é possível para começar outra fase de análise espacial, que é a identificação de áreas que devem ser mais profundamente estudadas em virtude da falta de certeza sobre suas condições.

O método responde também por possíveis inseguranças na atribuição de pesos e notas na análise multicritérios, porque é possível trabalhar não só com um único valor de peso por variável, mas com a faixa onde existem possíveis valores.

A proposta de utilizar a função de densidade de probabilidade para definir as faixas de simulação de Monte Carlo foi testada e pode ser utilizado, a fim de se ter um parâmetro de como definir essa variação. Observamos também que o uso de faixas mais largas para a simulação (por exemplo duas vezes o desvio-padrão, na faixa 2, o que resulta em um total de 95,4% de possibilidades) podem ser evitados. Nós testamos as mudanças na cobertura do solo, simulando o intervalo usando tanto a faixa 1 como a faixa 2 e os resultados foram praticamente os mesmos, mas a vantagem da redução da faixa é liberar a condição computacional que, em lugar de simular muitos diferentes pesos em larga faixa, ganha condições de aumentar o número de interações. Quanto maior o número de interações, melhor a qualidade do resultado.

Também é importante entender que no SASE a resolução espacial (dimensão do pixel) pode ser reduzida, a fim de se promover menores matrizes e tornar possível a aplicação de maior número de simulações, lembrando que quanto maior o número de simulações, mais robusto o resultado. Os dados não precisam ter resolução muito detalhada, pois a sensibilidade tem como objetivo identificar regiões do território que precisam de mais estudos para definir as suas potencialidades, em zonas, e não identificar detalhes ou pontos. Da mesma forma, não cabe trabalhar com grandes

extensões territoriais, posto que quanto maior o território maior variabilidade de ocorrências acontecerá, e isto pode resultar em incertezas no resultado.

No estudo de caso da Pampulha esta é uma primeira análise para teste de análise multicritérios com estudos de sensibilidade, tendo sido escolhido o tema de proteção ambiental e realizadas análises para apenas três variáveis, mas os próximos passos seria simular a inserção de outras variáveis e troca de algumas utilizadas, para investigação de suas influências no conjunto.

AGRADECIMENTOS

Agradecemos à PBH pelo acesso à expressiva base de dados que tem tornado o nosso projeto possível.

Agradecemos à Fapemig pelo apoio financeiro para participação no evento.

Os autores da publicação agradecem, sobretudo, ao CNPq, pelo apoio através do projeto “Modelagem Paramétrica da Ocupação Territorial: proposição de novos recursos das geotecnologias para representar e planejar o território urbano”, Processo 405664/2013-3, Chamada MCTI/CNPq/MEC/CAPES N° 43/2013.

REFERÊNCIAS BIBLIOGRÁFICAS

ANSELIN, L. Interactive techniques and exploratory spatial data analysis. In.: Longley, P., Goodchild, M., Maguire, D. e Rhind, D. (eds.). **Geographical Information Systems: Principles, Techniques, Management and Applications**. New York, John Wiley & Sons. 1999.

BERTALANFFY, Ludwig Von. **Teoria Geral dos Sistemas**. São Paulo, Vozes. 1975.

BONHAM-CARTER, G. **Geographic Information Systems for Geoscientists; modelling with GIS**. Ottawa, Pergamon. 1994.

BORGES, Júnia de Castro. **Estudo de fragilidade e potencial de uso da paisagem e análise de capacidade de carga turística do Parque Nacional da Serra do Cipó**. Mestrado em Análise e Modelagem de Sistemas Ambientais, IGC-UFMG. Orientação Ana Clara Mourão Moura. 2009.

CASTRO, Dayan M. Procedimentos de data mining na definição de valores para as análises de multicritérios como apoio à tomada de decisões e análise espaciais urbanas. In.: XXIV Congresso Brasileiro de Cartografia. Aracaju. **Anais**, 2010.

CHORLEY, J., HAGGET, P. Models, Paradigms and the New Geography. In.: **Integrated Models in Geography**. London, Methuen. 1967.

DALKEY, N; HELMER, O. An experimental application of the Delphi method to the use of experts. **Management Science**, v. 9, n. 3. 1963.

HUGGETT, R. **Systems analysis in geography; contemporary problems in geography**. Oxford, Clarendon Press. 1980.

LIGMANN-ZIELINSKA, A., SUN, L. Applying Time Dependent Variance-Based Global Sensitivity Analysis to Represent the Dynamics of an Agent-Based Model of Land Use Change. **International Journal of Geographical Information Science** 24(12), 1829-1850. 2010.

LIGMANN-ZIELINSKA, A., JANKOWSKI, P. Impact of proximity-adjusted preferences on rank-order stability in geographical multicriteria decision analysis. **Journal of Geographical Systems**, 14(2), 167-187.2012.

LIGMANN-ZIELINSKA, A., JANKOWSKI, P. Spatially-explicit integrated uncertainty and sensitivity analysis of criteria weights in multicriteria land suitability evaluation. **Environmental Modelling & Software**, DOI: 0.1016/j.envsoft.2014.03.007. 2014.

LINSTONE, H., TUROFF, M. **The Delphi Method: Techniques and Applications**. Turoff and Linstone. 2002.

MCHARG, Ian L. **Design with nature**. Garden City - NY, Natural History Press. 1969.

MOURA, Ana Clara M. **Geoprocessamento na gestão e planejamento urbano**. Belo Horizonte, Ed. da autora. 2003.

MOURA, Ana Clara M. Reflexões metodológicas como subsídio para estudos ambientais baseados em Análise de Multicritérios. In.: XIII Simpósio Brasileiro de Sensoriamento Remoto, Florianópolis, Brasil, INPE. **Anais**, p. 2899-2906. 2007.

MOURA, Ana Clara M. et al. 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. In.: XXIV Congresso Brasileiro de Cartografia. Aracaju. **Anais**, 2010.

SAATY, T. L. **The Analytic Hierarchy Process**. N. York, McGraw-Hill. 1980.

SILVEIRA, Samuel João. **Loteamento sustentável quanto ao ciclo hidrológico com gestão através do CTM**. Universidade Federal de Santa Catarina. Tese de Doutorado. Orientação: Francisco Henrique de Oliveira. 2013.