

ANÁLISE ESPACIAL EXPLORATÓRIA COM O EMPREGO DO ÍNDICE DE MORAN

Antonio José Rocha Luzardo*
Universidade Federal Fluminense**

Rafael March Castañeda Filho***
Universidade Federal do Rio de Janeiro****

Igor Brum Rubim*****
Universidade Federal do Rio de Janeiro*****

Resumo: Este trabalho aborda a análise exploratória de dados geoespaciais associados a feições de área, com o emprego de uma técnica de análise espacial baseada no Índice de Moran, nas modalidades global e local, objetivando compreender os padrões de associação espacial (*clusters*) existentes na distribuição dos dados da região de estudo, bem como identificar valores extremos desse conjunto de dados (*outliers*).

No desenvolvimento das atividades utilizou-se o aplicativo TerraView, um software desenvolvido e disponibilizado, gratuitamente, pelo Instituto Nacional de Pesquisas Espaciais (INPE), que possui ferramentas para o cálculo do Índice de Moran Global (I) e do Índice de Moran Local (I_l), e que permite a ponderação da matriz de dados pelo inverso da distância entre os polígonos das áreas e pela distância inversa ao quadrado.

A variável escolhida para este estudo foi o Índice de Desenvolvimento Humano Municipal (IDH-M), que representa uma evolução do Índice de Desenvolvimento Humano (IDH), proposto pelo Programa das Nações Unidas para o Desenvolvimento (PNUD), que, embora também considere saúde, educação e renda, na sua composição, trabalha com municípios, os quais são unidades político-administrativas menores.

Palavras-chave: IDH. Índice de Moran Global. Índice de Moran Local.

EXPLORATORY SPATIAL ANALYSIS USING MORAN INDEX

Abstract: This paper addresses the exploratory analysis of geospatial data associated with features of the area under study, with the use of a technique of spatial analysis based on Moran's Index, in both the global and local modes, in order to understand existing patterns of spatial association (*clusters*) in the distribution of data of the study area, and identify extreme values of this data set (*outliers*).

In the development of activities, we used the TerraView application, a software developed and provided free of charge by the National Institute for Space Research (INPE), which includes tools for calculating the Global Moran Index (I) and the Local Moran Index (I_l), and which allows the weighting of the data matrix by the inverse of the distance between the polygons of the areas and the inverse distance squared.

The variable chosen for this study was the Human Development Index for Counties (HDI-C), an upgraded version of the Human Development Index (HDI), utilized by the United Nations Development Program (UNDP). This novel version, which also measures average achievements in the same three dimensions of human development – health, education and income – works for counties, which are smaller political-administrative units.

Keywords: HDI. Global Moran Index. Local Moran Index.

ANÁLISIS ESPACIAL EXPLORATORIA UTILIZANDO EL ÍNDICE MORAN

Resumen: Este trabajo realiza un análisis exploratorio de atributos geoespaciales asociados a objetos de geometría areal, con el empleo de una técnica de análisis espacial basada en el Índice de Moran, en las modalidades global y local, con el objetivo de comprender los patrones de asociación espacial (*clusters*) existentes en la distribución de los datos de la región de estudio, así como identificar los valores extremos de ese conjunto de datos (*outliers*).

En el desarrollo de las actividades se utilizó la aplicación TerraView, un software desarrollado y disponible gratuitamente por el Instituto Nacional de Investigaciones Espaciales (INPE), que posee herramientas para el cálculo del Índice de Moran Global (I) y del Índice de Moran (I_l), y que permite la ponderación de la matriz de datos por el inverso de la distancia entre los polígonos de las áreas y por la distancia inversa al cuadrado.

La variable elegida para este estudio fue el Índice de Desarrollo Humano Municipal (IDH-M), que representa una evolución del Índice de Desarrollo Humano (IDH), propuesto por el Programa de las Naciones Unidas para el Desarrollo (PNUD), que toma en consideración indicadores de salud, educación y renta en su composición, además de trabajar con municipios, los cuales son unidades político-administrativas más pequeñas.

Palabras clave: IDH. Índice de Moran Global. Índice de Moran Local.

*Doutorando em Geografia (UFF), professor do Ensino Básico, Técnico e Tecnológico do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET/RJ). E-mail: antonioluzardo@gmail.com

**Instituto de Geociências – Campus Praia Vermelha – Universidade Federal Fluminense – Rua General Milton Gonçalves, s/n – Boa Viagem – Niterói, RJ – CEP 24210-346. Tel.: 21-2629-5953

***Doutorando em Geografia (UFRJ), pesquisador em Informações Geográficas e Estatísticas do Instituto Brasileiro de Geografia e Estatística (IBGE). E-mail: rafaelmarch43@gmail.com

****GeoCart – Laboratório de Cartografia – Departamento de Geografia – Universidade Federal do Rio de Janeiro. Avenida Athos da Silveira Ramos 274 – Cidade Universitária – Rio de Janeiro, RJ. CCMN – Sala H1 – 017 – CEP 21941-916. Tel.: 21-2590-1880.

*****Mestrando em Geografia (UFRJ), analista em Reforma e Desenvolvimento Agrário do Instituto Nacional de Colonização e Reforma Agrária (INCRA). E-mail: igorbr1@gmail.com

*****Laboratório Espaço de Sensoriamento Remoto e Estudos Ambientais – Departamento de Geografia – Universidade Federal do Rio de Janeiro. Avenida Athos da Silveira Ramos 274 – Cidade Universitária – Rio de Janeiro, RJ – CCMN – Bloco I – Sala 012 – CEP 21949-900. Telefone: 21-2598-9535.

Introdução

Este trabalho aborda a análise exploratória de dados geoespaciais associados a feições de área, com o emprego de uma técnica de análise espacial baseada no Índice de Moran, o qual foi proposto como uma versão da medida estatística não espacial de correlação, adaptada para o contexto espacial. Há duas expressões estatísticas distintas para este índice: a mais antiga, de 1950, refere-se ao índice global; a mais recente, de 1995, proposta por Luc Anselin, refere-se ao índice local.

O Índice de Moran Global (I) avalia a relação de interdependência espacial entre todos os polígonos da área de estudo e a expressa por meio de um valor único para toda a região (Moran, 1950, citado por O'Sullivan; Unwin, 2010). O Índice de Moran Local (I_i) identifica a relação existente entre um determinado polígono e a sua vizinhança, a partir de uma distância predefinida, por intermédio da covariância existente entre eles, permitindo o exame da homogeneidade/diversidade dos dados (Anselin, 1995, citado por O'Sullivan; Unwin, 2010).

Fundamentação teórica

Análise espacial exploratória de dados de área

Em várias situações práticas, a localização geográfica pontual de eventos ou fenômenos de interesse não está disponível. Muitas vezes, para onde uma variável tenha sido observada ou medida, em vez de por pontos, os dados são disponibilizados de maneira agregada por unidade de área, tais como municípios, bairros, setores censitários.

Geralmente, o valor atribuído a essa variável refere-se a uma contagem (enumeração), a uma taxa, ou a um índice. De acordo com Xavier da Silva (2001), os índices são exemplos típicos de dados que utilizam a escala de intervalo; enquanto as taxas e contagens são exemplos de dados que empregam a escala de razão ou racional.

Uma forma usual de visualização de dados agregados por áreas é o uso de mapas coloridos segundo o padrão espacial do fenômeno; são os chamados mapas temáticos, nos quais as cores indicam diferentes classes ou faixas de valores da variável mapeada. A maioria dos usuários limita-se a essas operações de visualização, tirando conclusões preliminares, de modo intuitivo. Contudo, é possível ir muito além dessa análise superficial. Quando se visualiza um padrão espacial, é muito útil traduzi-lo em questões objetivas, como as apresentadas a seguir:

- O padrão observado é aleatório ou apresenta uma agregação definida?
- Essa distribuição pode ser associada a causas mensuráveis?

- Os valores observados são suficientes para analisar o fenômeno espacial a estudar?

- Existem agrupamentos de áreas com padrões diferenciados na região de estudo?

Tais questões são objeto das técnicas de análise espacial de dados agregados por área. O modelo de distribuição mais utilizado para dados de área é o de variação espacial discreta, no qual se considera a existência de um processo estocástico $[Z_i, i = 1, 2, \dots, n]$, onde $[Z_i]$ é a realização do processo espacial na área $[i]$, ou seja, o valor assumido pela variável em estudo na área $[A_i]$, e $[n]$ é o número total de áreas $[A_i]$.

As referidas técnicas são essenciais para o desenvolvimento das etapas de modelagem estatística espacial. Na análise espacial devem-se investigar os *outliers*, não só no conjunto dos dados, mas também em relação aos vizinhos; bem como a não estacionariedade (variabilidade dos dados) do processo espacial na região de estudo, nos seus vários aspectos: variação na média, variância e covariância espacial.

As técnicas de análise exploratória visam a identificar a estrutura de correlação espacial, ou seja, a medida estatística, associada à covariância espacial, que melhor descreva os dados, a fim de estimar a magnitude da autocorrelação espacial entre as áreas. Para tanto, as ferramentas disponíveis são o Índice de Moran, o Índice de Geary e o Variograma (Câmara et al., 2002).

Matriz de proximidade espacial

Em Estatística, a correlação é uma medida que indica quanto uma determinada variável varia em função de outra, podendo ser de três tipos: direta (ou positiva), quando as duas variáveis aumentam ou diminuem ao mesmo tempo; inversa (ou negativa), quando elas variam em sentido contrário; ou nula. Porém, não se trata aqui de analisar apenas como duas variáveis se correlacionam numericamente, mas sim de investigar o quanto o valor observado de uma única variável de área, numa região do espaço geográfico, é dependente dos valores dessa mesma variável nas localizações vizinhas.

A questão de fundo, como bem enfatizam O'Sullivan e Unwin (2010), é: Como incorporar a noção de proximidade espacial numa medida de autocorrelação? Para tanto, é necessário capturar a relação espacial entre todos os pares de localizações (áreas), e isso se faz mediante uma matriz, que os autores citados denominam matriz de estrutura espacial ou matriz de pesos espacial, denotada por $[W]$ – referente a *weight*, que significa peso, em inglês –, em que cada valor $[w_{ij}]$ depende da relação espacial entre as localizações $[i]$ e $[j]$ e de como se escolhe representar esse relacionamento.

Para atribuir valores a cada elemento da matriz, podem-se usar diferentes critérios, dentre os quais os

mais comuns são:

- Critério da distância: $[w_{ij} = 1]$, se o centroide de $[A_i]$ está dentro de uma determinada distância (estabelecida pelo analista) do centroide de $[A_j]$; caso contrário, $[w_{ij} = 0]$;
- Critério da contiguidade ou adjacência: $[w_{ij} = 1]$, se $[A_i]$ compartilha um lado comum com $[A_j]$; caso contrário, $[w_{ij} = 0]$;
- Critério da vizinhança: $[w_{ij} = 1]$, se $[A_j]$ é um dos $[p]$ vizinhos mais próximos de $[A_i]$ (o valor $[p]$ é estabelecido pelo analista); caso contrário, $[w_{ij} = 0]$.

A matriz $[W]$ considera a proximidade ou a vizinhança de primeira ordem, mas pode-se trabalhar com o conceito de matriz $[W]$ de segunda ordem, representada por $[W^{(2)}]$. No critério de contiguidade, os polígonos vizinhos de segunda ordem são aqueles adicionados aos de primeira ordem quando se eliminam as fronteiras entre os polígonos vizinhos de primeira ordem. Esse conceito pode ser estendido à k -ésima ordem, com a matriz resultante sendo representada por $[W^{(k)}]$. Vale ressaltar que, no caso da matriz de primeira ordem, a notação $[W^{(1)}]$ é simplificada para $[W]$.

Como se pode observar, a adjacência permanece uma grandeza binária, já que $[w_{ij} = 0]$ indica polígonos não conectados e $[w_{ij} = 1]$, polígonos conectados. No entanto, conforme ressaltam O'Sullivan e Unwin, anteriormente citados, algumas relações de proximidade podem ser consideradas mais fortes que outras, mesmo que, inicialmente, tenham o mesmo valor 1. Nesse caso, os valores $[w_{ij}]$ poderão variar de 0 (interação fraca) a 1 (interação forte). Isso se consegue pela ponderação da interação entre dois polígonos (aplicação de pesos), mediante a seguinte fórmula geral:

$$w_{ij} \propto \frac{l_{ij}}{d_{ij}^m l_i}$$

Onde:

- $[d_{ij}]$ é a distância entre os centroides dos polígonos $[i]$ e $[j]$;
- $[m]$ é um inteiro (fator de potência);
- $[l_{ij}]$ é o comprimento da fronteira comum entre os polígonos $[i]$ e $[j]$;
- $[l_i]$ é o perímetro do polígono $[i]$.

Com essa abordagem, torna-se necessário aplicar um fator de escala nos pesos, de modo a assegurar que todos pertençam ao intervalo entre zero e um, por intermédio de uma normalização, onde o somatório dos valores de cada linha da matriz é igual a um, ou seja, 100%. Os elementos da diagonal principal da matriz $[W]$ são igualados a zero. A simetria da matriz de pesos é, geralmente, requerida de tal forma que $[w_{ij} = w_{ji}]$ em todos os casos. Porém, alguns métodos de construção da

matriz $[W]$ não asseguram simetria. Para resolver esse tipo de situação é válido forçar a simetria, impondo:

$$W_{\text{final}} = (W + W^T) / 2$$

- $[W^T]$ representa a matriz transposta de $[W]$.

No estudo de processos sociais utilizam-se, frequentemente, setores censitários ou outras unidades administrativas para representar a agregação dos dados. Nesses casos, é recomendável trabalhar com critérios simples para o estabelecimento da matriz $[W]$, tais como o de adjacência, pelo menos na fase exploratória da análise.

Segundo Câmara et al. (2002), para estimar a variabilidade espacial de dados de área, uma ferramenta básica é a matriz de proximidade espacial, também referenciada como matriz de vizinhança $[W]$. Dado um conjunto de $[n]$ áreas, $[A_1, \dots, A_n]$, constrói-se a matriz $[W^{(1)}]$, $n \times n$, onde cada um dos elementos $[w_{ij}]$ representa uma medida de proximidade entre $[A_i]$ e $[A_j]$, que pode ser calculada por um dos seguintes critérios:

- $[w_{ij} = 1]$, se o centroide de $[A_i]$ está a uma determinada distância de $[A_j]$; caso contrário, $[w_{ij} = 0]$;
- $[w_{ij} = 1]$, se $[A_i]$ compartilha um lado comum com $[A_j]$; caso contrário, $[w_{ij} = 0]$;
- $[w_{ij} = l_{ij}/l_i]$, onde $[l_{ij}]$ é o comprimento da fronteira entre $[A_i]$ e $[A_j]$ e $[l_i]$ é o perímetro de $[A_i]$.

Nota-se que os dois primeiros critérios de proximidade são idênticos para os autores estudados (anteriormente citados) e produzem resultados binários; a diferença reside no terceiro critério. No primeiro caso, os $[p]$ vizinhos mais próximos não são incluídos, mas há um caminho alternativo que produz resultados não binários, coincidentes com o segundo caso, exceto o fator inverso da distância $[d_{ij}]$ entre os centroides dos polígonos $[i]$ e $[j]$ elevado a uma potência inteira $[m]$.

Índice de Moran Global (I)

Uma vez determinada a estrutura espacial de análise, expressa pela matriz $[W]$, qualquer medida particular de autocorrelação pode ser concebida mediante a definição de um modo de se mensurar a diferença entre valores do atributo associado às localizações (áreas). A medida mais usada para este fim é o Índice de Moran Global (I), que é a versão espacial do coeficiente de correlação linear de Pearson (Costa Neto, 2002) e que representa o coeficiente de correlação para o relacionamento entre os valores de uma variável espacial (atributo) e o valor médio desta variável.

O Índice de Moran Global (I) é aplicado, usualmente,

em unidades de área às quais estejam associados atributos do tipo racional ou de intervalo (Longley et al., 2005) e expressa a autocorrelação, considerando apenas o primeiro vizinho, ou seja, emprega a matriz de vizinhança de primeira ordem, independentemente do critério de composição da matriz [W] escolhido. Usa-se esta expressão (Câmara et al., 2002):

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (Z_i - \mu_z)(Z_j - \mu_z)}{\sum_{i=1}^n (Z_i - \mu_z)^2}$$

Onde:

- [n] é o número de áreas;
- [Z_i] é o valor do atributo considerado na área [i];
- [μ_z] é o valor médio do atributo na região de estudo;
- [w_{ij}] é o elemento [ij] da matriz de vizinhança normalizada.

Observa-se na equação do Índice de Moran Global (I) a seguinte função operacional dos seus componentes: o numerador da fração é um termo de covariância; os subscritos [i] e [j] referem-se a diferentes unidades de área ou polígonos na região em estudo; [Z_i] é o valor assumido pelo atributo na área [i].

O cálculo do produto das diferenças entre o valor de [Z_i] em duas áreas, e a média global [μ_z] determina o quanto essas diferenças covariam. Se ambos os valores [Z_i] e [Z_j] estão do mesmo lado da média (acima ou abaixo), o produto é positivo; se um dos valores é superior à média e o outro é inferior, o produto é negativo; e o valor absoluto total depende de quão próximos da média global [μ_z] estão os valores de [Z].

Os termos de covariância são multiplicados por [w_{ij}], o que faz com que as parcelas de covariância sejam ponderadas de acordo com a intensidade de sua interação espacial. Considerando o critério da contiguidade (em modo binário), para cada valor fixado de [i], os elementos [w_{ij}] e [j = 1...n] serão iguais a um para os vizinhos adjacentes à área [i] e iguais a zero para os vizinhos não adjacentes. Desse modo, o fator [w_{ij}] faz com que o termo de covariância seja incluído no cálculo de [I] somente para os pares de localizações (áreas) vizinhas, segundo o critério de proximidade espacial considerado no cálculo da matriz [W].

O denominador da equação de [I] representa divisão pela variância (a menos do fator [1/n], que não aparece explicitado, devido à normalização prévia da matriz [W]). Assim, assegura-se que o valor resultante de [I] não seja alto simplesmente porque os valores e a variabilidade em [Z] são altos. Em essência, a divisão pela variância é uma forma de normalização da covariância em relação à variação dos valores do atributo [Z].

O resultado da equação de [I] é análogo ao que se obtém para um coeficiente de correlação não espacial.

Se houver correlação positiva dos dados, então a maioria dos polígonos vizinhos terá valores do mesmo lado da média e o índice será positivo, ou seja, [I > 0], indicando correlação espacial direta. Se os dados se correlacionam negativamente, então a maioria dos polígonos vizinhos terá valores de atributos em lados opostos da média e o índice será negativo, ou seja, [I < 0], caso este seja de correlação espacial inversa. [I = 0] indica ausência de correlação espacial.

Análise diagnóstica da estatística global de Moran

Na teoria estatística de regressão a equação do Índice de Moran Global (I) pode ser reescrita na seguinte forma matricial:

$$I = \frac{Z^T W Z}{Z^T Z}$$

Onde: [Z] é o vetor coluna de componentes [Z_i - μ_z], com [i = 1...n].

A expressão matricial do Índice de Moran Global (I) abre caminho para o emprego de estatísticas convencionais de diagnóstico de regressão linear, com vistas a associar os *p-valores* (níveis de significância) aos resultados do índice [I]. Porém, pelo fato de a estrutura espacial do mapa também ser um parâmetro na análise (a matriz [W]), uma abordagem mais usual de cálculo de *p-valores* associados ao [I] baseia-se no método de Monte Carlo (O'Sullivan; Unwin, 2010).

Considerando que o analista não sabe, *a priori*, como a variável se comporta, ele, por meio da matriz [W], faz uma hipótese implícita sobre a estrutura espacial do mapa. Daí a necessidade de ele simular uma distribuição empírica para a realização do teste de significância do valor observado de [I].

O Índice de Moran Global (I) pode ser submetido a um teste estatístico cuja hipótese nula representa a de independência espacial, condição esta dada por [I = 0]. Portanto, a hipótese nula somente poderá ser rejeitada caso [I] resulte estatisticamente diferente de zero, para um nível de significância preestabelecido. Normalmente, trabalha-se com [*p-valor* = 0,05] ou [*p-valor* < 0,05], ou seja, com uma margem de confiança igual ou superior a 95%.

Para estimar a significância do índice [I], sem pressupostos em relação à distribuição, pode-se realizar um teste de pseudossignificância, que consiste na geração de um determinado número de permutações, distintas entre si, dos valores de atributos associados às áreas. O TerraView disponibiliza duas opções para o cálculo do Índice de Moran Global (I): 99 e 999 permutações; a segunda é, geralmente, a mais utilizada.

Cada permutação produz um novo arranjo espacial,

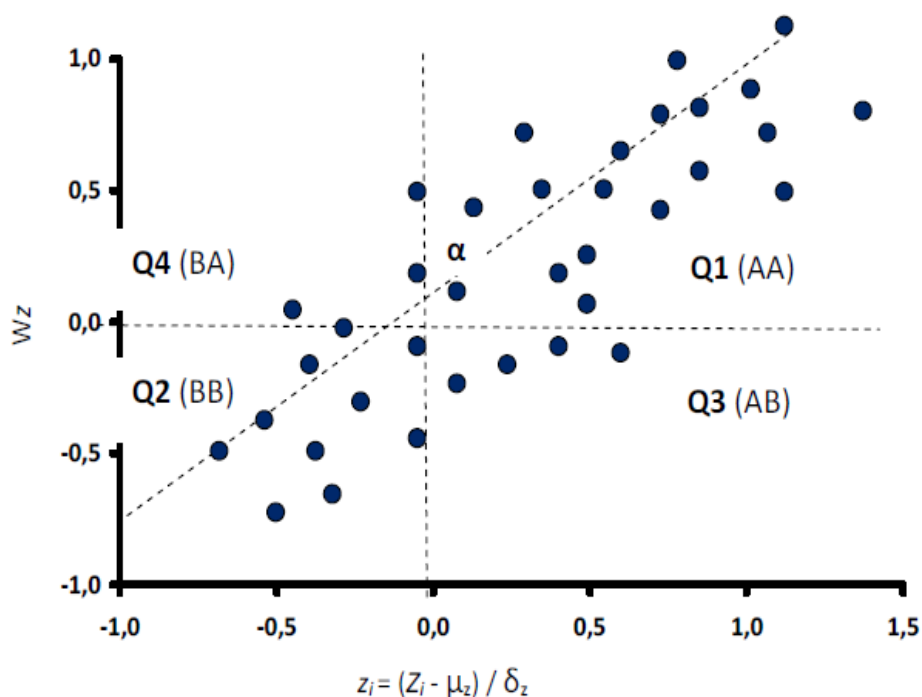
onde os valores são redistribuídos entre as áreas e um valor de $[I]$ é calculado. Como apenas um dos arranjos corresponde à situação observada, pode-se construir uma distribuição empírica de $[I]$. Se o valor observado de $[I]$ para a região de interesse corresponder a um extremo da distribuição simulada, então se trata de um valor com significância estatística, pois isso indica que o resultado obtido é atípico em relação à distribuição simulada.

Diagrama de espalhamento do Índice de Moran Global

O objetivo do diagrama de Moran Global – também referenciado como Moran Scatterplot – é proporcionar os elementos necessários para a comparação dos valores normalizados do atributo, em cada área, com a média dos seus vizinhos, construindo um gráfico bidimensional de $[z]$ (valores normalizados) por $[W_z]$ (média dos vizinhos), dividido em quatro quadrantes, onde cada ponto representa uma área (polígono).

A Figura 1 ilustra o Moran Scatterplot, o qual exhibe o relacionamento entre o valor normalizado do atributo em cada área $[i]$ (eixo horizontal) e o valor médio normalizado do atributo nas localizações vizinhas da área $[i]$. É importante ressaltar que esse diagrama é construído com base nos valores normalizados da variável, ou seja, os valores originais subtraídos do valor médio global, divididos pelo desvio-padrão.

Figura 1 – Moran Scatterplot



Adaptado de Câmara *et al.* (2002), p. 20

Onde:

- O quadrante superior direito (Q1) contém os casos onde o valor do atributo de cada polígono e o valor médio do atributo nos polígonos vizinhos são maiores que a média global (ambos representados por alto): alto-alto (AA);

- O quadrante inferior esquerdo (Q2) contém os casos onde o valor do atributo de cada polígono e o valor médio do atributo nos polígonos vizinhos são menores que a média global (ambos representados por baixo): baixo-baixo (BB);

- O quadrante inferior direito (Q3) contém os casos onde o valor do atributo de cada polígono está acima da média global (representado por alto), enquanto o valor médio do atributo em polígonos vizinhos está abaixo daquela média (representado por baixo): alto-baixo (AB);

- O quadrante superior esquerdo (Q4) contém os casos onde o valor do atributo de cada polígono está abaixo da média global (representado por baixo), enquanto o valor médio do atributo em polígonos vizinhos está acima daquela média (representado por alto): baixo-alto (BA).

A maior concentração de pontos nos quadrantes Q1 e Q2 é uma indicação de autocorrelação espacial positiva [$I > 0$]. Se os quadrantes Q3 e Q4 tivessem maior quantidade de pontos, a indicação seria de autocorrelação espacial negativa [$I < 0$]. Caso os pontos estivessem igualmente distribuídos pelos quatro quadrantes, haveria uma indicação de ausência de autocorrelação [$I \approx 0$].

Com base na teoria estatística de regressão linear, é correto afirmar que o índice [I] equivale ao coeficiente de regressão linear da reta de regressão [a] de [W_z] em [z]; e que o Índice de Moran Global (I) é o coeficiente de correlação para o relacionamento entre os valores de atributos e os valores médios de atributo dos vizinhos.

O diagrama de espalhamento de Moran também pode ser apresentado na forma de um mapa temático bidimensional, no qual cada polígono está associado ao seu respectivo quadrante no diagrama de espalhamento. Esse mapa temático contendo quatro classes (AA, BB, AB e BA) pode ser chamado Mapa de Espalhamento do Índice de Moran Global (I). É recomendado que os quadrantes Q1/Q2 e Q3/Q4 sejam representados em pares de cores diferentes, usando-se o efeito *dégradé* em cada par.

Índice de Moran Local (I_i)

Os indicadores globais de autocorrelação espacial fornecem um valor único como medida de associação espacial para todo o conjunto de dados, o que é útil na caracterização da região de estudo como um todo. Porém, quando se lida com um grande número de áreas em uma região, é muito provável que ocorram diferentes regimes de associação espacial e que apareçam máximos locais de autocorrelação onde a dependência espacial é ainda

mais pronunciada.

Assim, muitas vezes, é desejável examinar padrões mais detalhadamente. Para fazê-lo é preciso utilizar indicadores de associação espacial que possam ser relacionados às diferentes localizações de uma variável distribuída espacialmente. Os indicadores locais produzem um valor específico para cada área, permitindo, desse modo, a identificação de agrupamentos.

O Índice de Moran Local (I_i) foi concebido como uma ferramenta estatística para testar a autocorrelação local e para detectar objetos espaciais com influência no Índice de Moran Global (I); ele pode ser expresso para cada área [i] a partir dos valores normalizados [z_j] do atributo mediante a seguinte equação:

$$I_i = \frac{z_i \sum_{j=1}^n w_{ij} z_j}{\sum_{j=1}^n z_j^2}$$

Onde: [z_i] = [$(Z_i - \mu_z) / \delta_z$] é a variável normalizada.

Da equação de [I_i] percebe-se que resultados positivos são obtidos onde ocorrem concentrações de valores baixos ou de valores altos do atributo, enquanto resultados negativos decorrem da proximidade entre valores baixos e altos na mesma área. Assim, o Índice de Moran Local (I_i) dá uma indicação da homogeneidade e da diversidade dos dados.

Diagrama de espalhamento do Índice de Moran Local

Ao se trabalhar com o Índice de Moran Local (I_i) o diagrama de espalhamento torna-se uma ferramenta analítica por si mesma, pois cada ponto representativo de uma área (polígono) assume uma significação própria, ao contrário do que se verifica no diagrama de espalhamento de Moran Global, onde os pontos, individualmente, apenas contribuem como parcelas no cômputo do Índice de Moran Global (I).

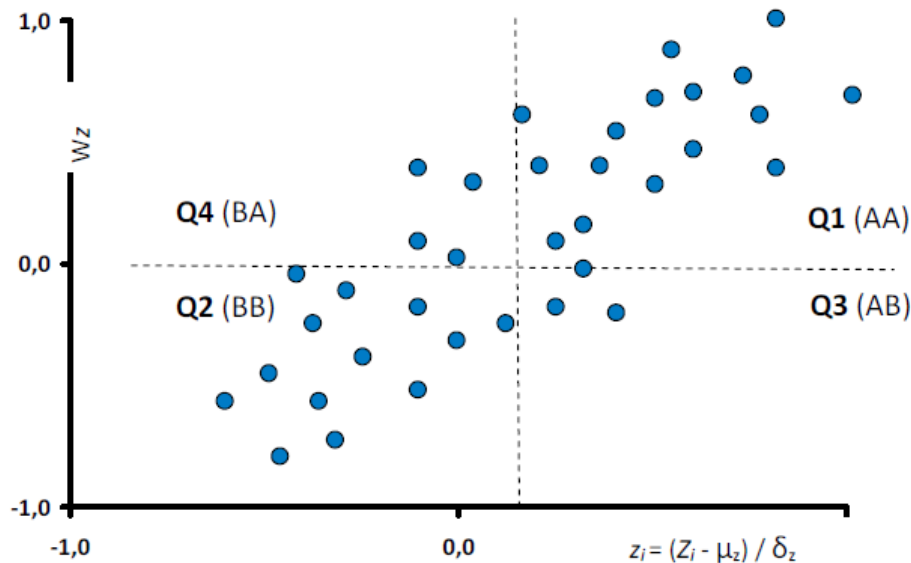
A Figura 2 ilustra o diagrama de espalhamento do Índice de Moran Local (I_i), onde os quadrantes são definidos pelo valor médio global do atributo normalizado: no eixo horizontal, a reta [$z_i = 0$]; no eixo vertical, a reta [$\sum w_{ij} z_j = 0$], indicativa do valor médio global do atributo normalizado nos polígonos vizinhos de cada polígono [i]. Os quatro quadrantes correspondem, cada um, às diferentes combinações possíveis do valor do atributo na área [i] com o valor médio do atributo em seus vizinhos.

Onde:

- AA (alto-alto) e BB (baixo-baixo): representam áreas que contribuem para autocorrelação positiva; correspondem aos quadrantes Q1 e Q2 e indicam clusters (aglomerações) de valores altos e baixos, respectivamente;

- AB (alto-baixo) e BA (baixo-alto): representam áreas que contribuem para autocorrelação negativa; correspondem aos quadrantes Q3 e Q4 e indicam áreas de transição, com os valores altos cercados por valores baixos do atributo, e vice-versa.

Figura 2 – Diagrama de espalhamento de Moran Local

Adaptado de Câmara *et al.* (2002), p. 20

Significância estatística do Índice de Moran Local

O Índice de Moran Local (I_i) explora o conceito de bolsões locais de correlação significativa ao permitir a identificação de agrupamentos de objetos (áreas) com valores de atributos semelhantes ou objetos anômalos (Oliveira *et al.*, 2013). Mas como fazer para determinar a significância estatística do resultado obtido para cada valor de $[I_i]$? Segundo os autores anteriormente citados, o procedimento é similar ao adotado para o Índice de Moran Global (I), utilizando o método de simulação de Monte Carlo.

Para cada área, uma vez calculado e fixado o respectivo Índice de Moran Local (I_i), permuta-se, aleatoriamente, o valor de atributo das demais áreas (tantas permutações quantas forem estipuladas pelo analista), e, para cada permutação, calcula-se o valor correspondente da estatística local até que se obtenha uma pseudodistribuição, para a qual seja possível calcular os parâmetros de significância de $[I_i]$ (pelo fato de se manter fixo o valor de $[I_i]$ em cada rodada de permutações).

Esse método é denominado permutação condicional.

Valores observados de $[I_i]$ que sejam muito baixos ou muito altos, relativamente a cada distribuição simulada resultante de cada rodada de permutações, são considerados de interesse. Um *p-valor*, ou seja, um valor de pseudosignificância, pode ser determinado a partir da posição do valor da estatística local $[I_i]$, relativamente ao conjunto de valores obtidos para todas as permutações realizadas, fixando-se $[I_i]$. Por exemplo, se a estatística $[I_i]$ é a mais alta dentre todos os valores resultantes de 999 permutações, então estima-se que seja uma ocorrência em mil: [*p-valor* \approx 0,001].

Uma vez determinada a significância estatística de $[I_i]$, é recomendável gerar um mapa temático, indicando as regiões com correlação local significativamente diferente do resto dos dados. Essas áreas podem ser expressas como bolsões de estacionariedade, com dinâmica espacial própria e, por essa razão, merecem estudo mais detalhado.

Análise da estrutura espacial no TerraView – Função LISA

O TerraView calcula o índice $[I_i]$ através de sua função LISA – sigla que significa Local indicators of spatial association, em inglês –, cujo resultado é uma tabela com sete colunas, a partir da qual é possível gerar mapas temáticos para análise exploratória baseada no Índice de Moran Local (I_i). A seguir está o significado de cada coluna:

- [Z]: vetor dos desvios dos valores de atributo normalizados; os desvios são em relação à média global; cada elemento de [Z] corresponde ao valor do atributo normalizado numa área $[i]$;

- [Wz]: vetor da média ponderada local, ou seja, dos vizinhos da área $[i]$; cada elemento de [Wz] corresponde à média ponderada, segundo a matriz de pesos espacial, dos valores de atributo normalizados dos polígonos vizinhos da área $[i]$;

- [MoranIndex]: Índice de Moran Local (I_i); é o resultado obtido para o índice $[I_i]$, um para cada polígono;

- [LISASig]: valor da estatística p (p -valor); corresponde ao nível de significância do índice $[I_i]$, sendo um para cada polígono; quando $[LISASig > 0,05]$, a indicação é de ausência de correlação espacial localizada;

- [BoxMap]: este campo assume um dos seguintes valores numéricos, de um a quatro: 1 [Q1:AA], 2 [Q2:BB], 3 [Q3:AB], 4 [Q4:BA], que representam a relação entre os valores de [Z] e [Wz] (vizinhos) em um diagrama de dispersão dividido em quadrantes [Q]. O BoxMap permite a geração de mapas temáticos de valores únicos. Vale ressaltar que os conceitos de alto e de baixo podem estar associados às melhores ou às piores condições, dependendo da natureza da variável analisada.

- [LISAMap]: este campo assume um dos seguintes valores numéricos, de zero a três, que são gerados após a marcação da caixa [Avaliar significância] e da escolha da quantidade de permutações (99 ou 999) para a realização dos testes de pseudossignificância, no TerraView. Os significados de cada valor do LISAMap são os seguintes: 0 [não significante]; 1 [p -valor = 0,05] ou grau de confiança igual a 95%; 2 [p -valor = 0,01] ou grau de confiança igual a 99%; 3 [p -valor = 0,001] ou grau de confiança igual a 99,9%;

- [MoranMap]: este campo assume um dos seguintes valores numéricos, de zero a quatro, que são gerados após a marcação da caixa [Avaliar significância], no TerraView. Este campo apresenta somente as sub-regiões cujos valores foram considerados significantes (com intervalo de confiança superior a 95%), diferentemente do BoxMap, que não considera significância e apenas indica o quadrante de cada polígono. O MoranMap funciona como uma combinação entre os resultados do BoxMap e do LisaMap. Os significados de cada valor do MoranMap são os seguintes: 0 [não significante]; 1 [Q1:AA]; 2

[Q2:BB]; 3 [Q3:AB]; 4 [Q4:BA].

Considerando os agrupamentos, cada polígono é classificado conforme sua posição em relação aos quadrantes, recebendo uma cor correspondente no mapa temático a ser gerado. Segundo Oliveira et al. (2013), o MoranMap permite mapear áreas de agrupamentos (AA e BB) e de transição (AB e BA), atípicos ou não:

- [AA]: demonstra o atributo acima da média para a unidade e seus vizinhos, indicando a existência de *clusters* de valores altos da variável analisada;

- [BB]: representa os atributos abaixo da média para a unidade e seus vizinhos, indicando a existência de *clusters* de valores baixos da variável analisada;

- [AB]: destaca o atributo acima da média para a unidade e abaixo da média para os seus vizinhos, indicando a existência de áreas de transição;

- [BA]: indica o atributo abaixo da média para a unidade e acima da média para os seus vizinhos, indicando a existência de áreas de transição.

Planejamento da investigação

Área de estudo

A área de estudo selecionada para o presente trabalho é o estado do Rio de Janeiro, unidade da Federação com 92 municípios e 16.001.103 habitantes, dos quais 80%, ou seja, cerca de 13 milhões, vivem na capital, de mesmo nome, e na sua região de influência. Essa população conduz o estado à terceira posição no País, em termos de volume populacional, ficando apenas atrás dos estados de São Paulo e Minas Gerais, primeiro e segundo colocados, respectivamente (IBGE, 2010a).

Ferramenta empregada

A principal ferramenta utilizada na realização deste trabalho é o aplicativo TerraView, um *software* que utiliza a biblioteca TerraLib e emprega o C++ como linguagem de código-fonte, o qual é desenvolvido e disponibilizado, gratuitamente, pelo Instituto Nacional de Pesquisas Espaciais (INPE). Trata-se de um visualizador de bases cartográficas para aplicações de Sistemas de Informações Geográficas (SIG), que possui uma interface amigável, bem como a capacidade de manipular dados vetoriais (pontos, linhas e polígonos) e dados matriciais (grades e imagens).

O TerraView é um aplicativo ideal para aplicações acadêmicas, pois disponibiliza vários recursos geoestatísticos para análises espaciais, o que facilita o estudo de conceitos mais técnicos, bem como disponibiliza ferramentas para o cálculo do Índice de Moran Global (I) e do Índice de Moran Local (I_i), permitindo que a matriz [W] seja ponderada pelo inverso

da distância entre os polígonos [$m = 1$] e pela distância inversa ao quadrado [$m = 2$] (INPE, 2014).

Temática em análise

A variável escolhida como ponto de partida para a presente investigação é o Índice de Desenvolvimento Humano (IDH), que foi proposto em 1947, dentro do Programa das Nações Unidas para o Desenvolvimento (PNUD), com o propósito de medir o desenvolvimento humano dos países no pós-Guerra, e considera três dimensões envolvidas na identificação do bem-estar do ser humano: vida longa e saudável (saúde), acesso ao conhecimento (educação) e padrão de vida digno (renda) (PNUD, 2014).

O PNUD estabelece três níveis de qualidade para o desenvolvimento humano: IDH baixo, para valores entre 0 e 0,5; IDH médio, para valores entre 0,5 e 0,8; IDH alto, para valores entre 0,8 e 1,0. A fórmula geral, que permite a construção e a combinação dos índices envolvidos, é a seguinte:

$$\text{Índice} = (\text{valor observado} - \text{valor mínimo}) / (\text{valor máximo} - \text{valor mínimo})$$

A Fundação João Pinheiro (FJP) associou-se ao Instituto de Pesquisa Econômica Aplicada (IPEA), em 1996, para realizar uma adaptação na metodologia do IDH, desagregando-o segundo unidades político-administrativas menores. O novo índice criado, Índice de Desenvolvimento Humano Municipal (IDH-M), resulta da composição de três dimensões: longevidade, educação e renda (Santos, 2005).

A longevidade é medida pela esperança de vida ao nascer. A educação é obtida pela combinação ponderada da taxa de analfabetismo da população adulta de 15 anos ou mais, com peso igual a 2/3, e do número médio de anos de estudos da população de 25 anos ou mais, com peso igual a 1/3. A renda é representada pela renda familiar per capita média ajustada (RFPC), expressa em salários mínimos, tendo por base o valor de setembro de 1991. Os valores calculados situam-se entre zero e um, onde o primeiro significa a pior situação e o segundo representa a melhor.

Segundo Santos, anteriormente citado, indicadores podem ser definidos como medidas que sintetizam e quantificam um subconjunto limitado de um universo de variáveis observadas, regularmente, por um longo período (como, por exemplo, dados censitários); ou de um conjunto de informações extraídas de fontes secundárias, desse modo auxiliando na apreensão de fenômenos complexos por públicos diversos. Para essa investigação, cujos fenômenos de interesse são de natureza socioeconômica, as dimensões escolhidas para serem exploradas são as do IDH-M.

Etapas metodológicas

Os valores do IDH-M relativos a cada município fluminense, referentes aos censos demográficos realizados pelo Instituto Brasileiro de Geografia e Estatística (IBGE), são obtidos para três momentos distintos: 1991, 2000 e 2010. Portanto, a base de dados estabelecida consiste de 92 feições de área (polígonos), descritoras dos municípios do estado do Rio de Janeiro, a cada uma das quais estão associados três valores de IDH-M, relativos às três épocas anteriormente indicadas.

Com a inclusão da dimensão temporal na investigação, objetiva-se ampliar o potencial da análise espacial exploratória na área de estudo, pois as conclusões referentes a uma determinada época poderão ser confrontadas com as de outro momento, proporcionando uma base de comparação relevante na análise dos dados.

Outro método empregado para ampliar o potencial da investigação consiste na aplicação de diferentes critérios de composição e ponderação da matriz de vizinhança [W], como, por exemplo, a função inversa da distância. Neste trabalho são empregados dois critérios, dentre os mais utilizados nos ensaios com o Índice de Moran Global (I): contiguidade (ou adjacência) e distância.

Em linhas gerais, as etapas metodológicas previstas nos ensaios realizados com o Índice de Moran são as seguintes:

- 1) Mapeamento do IDH-M para os municípios, em cada ano;
- 2) Geração da matriz de vizinhança [W] para os municípios, em cada ano, segundo o critério da contiguidade (1ª ordem), inicialmente sem ponderação. Numa segunda fase, aplicam-se pesos baseados na função inverso da distância;
- 3) Cálculo do Índice de Moran Global (I) para o estado do Rio de Janeiro, em cada ano, e o respectivo *p-valor* (que representa o nível de significância do índice);
- 4) Cálculo do Índice de Moran Local (I_i) e dos indicadores associados para os municípios, em cada ano;
- 5) Geração do Mapa de Moran, baseado no indicador MoranMap, para os municípios, em cada ano;
- 6) Análise dos resultados obtidos nas etapas 3 a 5, em cada ano, tendo como referência os mapas do IDH-M gerados na etapa 1. Comparação dos resultados obtidos, com e sem aplicação de pesos na matriz de vizinhança [W];
- 7) Repetição das etapas 2 a 6, usando o critério da distância para cálculo da matriz de vizinhança [W] e aplicando pesos baseados na função inverso da distância;
- 8) Análise comparativa dos resultados obtidos, segundo os dois critérios de cálculo da matriz de vizinhança [W], contiguidade e distância, ambos com aplicação de pesos baseados no inverso da distância.

Coleta e tratamento dos dados

Fontes de dados

Os dados empregados na realização do presente trabalho foram obtidos na internet. Os *sites* do IBGE e do PNUD possuem dados digitalizados nos formatos necessários para as análises espaciais realizadas neste trabalho. No IBGE a base cartográfica municipal do estado do Rio de Janeiro está disponível, bem como no *site* do PNUD há dados sobre o IDH-M.

Os limites municipais obtidos no site do IBGE estão em formato *shapefile*, no sistema geodésico SIRGAS 2000, em coordenadas geográficas. Esses dados geográficos digitais podem ser utilizados do modo como estão disponibilizados, sem necessidade de qualquer tratamento cartográfico prévio (IBGE, 2010b).

Os dados de IDH-M para os três anos (Figuras 3, 4 e 5) foram obtidos através de ferramentas de consulta no próprio *site* do PNUD, onde também é possível baixar as informações, em formato CSV, que significa valores separados por vírgula (PNUD, 2014).

Procedimentos preparatórios

Uniram-se os dados do arquivo vetorial contendo a malha municipal do estado do Rio de Janeiro e os dados do PNUD com o emprego do aplicativo ArcGIS, gerando, então, a base de dados completa para as análises espaciais. A Tabela 1 apresenta a distribuição dos municípios do estado do Rio de Janeiro por classe de IDH-M:

Tabela 1 – Distribuição dos municípios do estado do Rio de Janeiro por classe de IDH-M.

Classe	Quantidade de municípios			Percentual de municípios		
	1991	2000	2010	1991	2000	2010
Muito baixo (0,000 a 0,499)	59	0	0	64,13	0,00	0,00
Baixo (0,500 a 0,599)	31	42	0	33,70	45,65	0,00
Médio (0,600 a 0,699)	2	48	34	2,17	52,17	36,96
Alto (0,700 a 0,799)	0	2	57	0,00	2,17	61,96
Muito alto (0,800 a 1,000)	0	0	1	0,00	0,00	1,09
Total	92	92	92	100,00	100,00	100,00

IDH-M: Índice de Desenvolvimento Humano Municipal.

Ensaio e resultados

Com o arquivo da base de dados preparado, utilizou-se o programa TerraView para realizar os ensaios planejados. A construção da matriz de vizinhança [W], utilizando a ferramenta [Criar matriz de proximidade / menu Análise], foi realizada, para cada ano, com base em diferentes critérios de proximidade espacial e de ponderação, doravante denominados [modelos]. Foram considerados na composição da matriz de vizinhança [W] os seguintes [modelos]:

- [1º modelo]: critério da contiguidade (adjacência), sem aplicação de pesos;
- [2º modelo]: critério da contiguidade, com pesos proporcionais ao inverso da distância entre áreas vizinhas;

• [3º modelo]: critério da distância ($d = 67\text{km}$), com pesos proporcionais ao inverso da distância entre áreas vizinhas.

O [3º modelo] requer uma distância-base para avaliação de vizinhança entre áreas, a qual foi obtida a partir de uma avaliação visual do grau de interação espacial entre municípios, nos mapas, considerando a escala gráfica. Foi considerada a incidência de um fluxo elevado de serviços e pessoas entre os municípios do Rio de Janeiro e de Niterói, cujas sedes estão distantes entre si cerca de 40km. Desse modo, fixou-se uma distância linear de 0,6º de arco, o que corresponde a 67km, aproximadamente.

Uma vez gerada a matriz de vizinhança [W], o Índice de Moran Global (I) e o Índice de Moran Local (Ii) foram

calculados para cada ano. Para cômputo do Índice de Moran Global (I) empregou-se a ferramenta [Estatística espacial]. As Tabelas 2 e 3 apresentam as configurações

utilizadas nas análises e os respectivos valores do Índice de Moran Global (I), para cada ano, bem como o p-valor correspondente.

Tabela 2 – Modelos testados e os respectivos valores do Índice de Moran Global (I)

Critério de proximidade	Função de ponderação	Índice de Moran Global (I)		
		1991	2000	2010
Contiguidade	sem peso	0,133699	0,0990523	0,170966
Contiguidade	inverso da distância	0,140457	0,11562	0,181096
Distância (67km)	inverso da distância	0,118997	0,069896	0,144397

IDH-M: Índice de Desenvolvimento Humano Municipal.

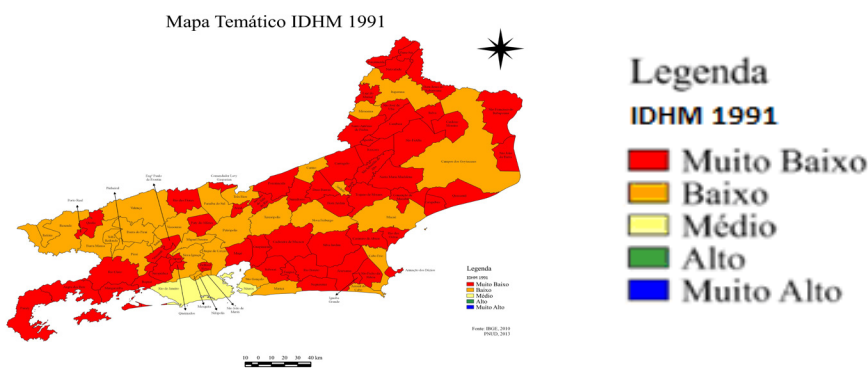
Tabela 3 – Resultados de p-valor para o Índice de Moran Global (I).

Critério de proximidade	Função de ponderação	p-valor		
		1991	2000	2010
Contiguidade	sem peso	0,04	0,07	0,03
Contiguidade	inverso da distância	0,05	0,07	0,02
Distância (67km)	inverso da distância	0,01	0,013	0,01

Obs.: p-valor obtido com base em 99 permutações.

As Figuras 3, 4 e 5 apresentam os mapas temáticos do IDH-M no estado do Rio de Janeiro, nos anos de 1991, 2000 e 2010, respectivamente.

Figura 3



Mapa Temático IDHM 2000

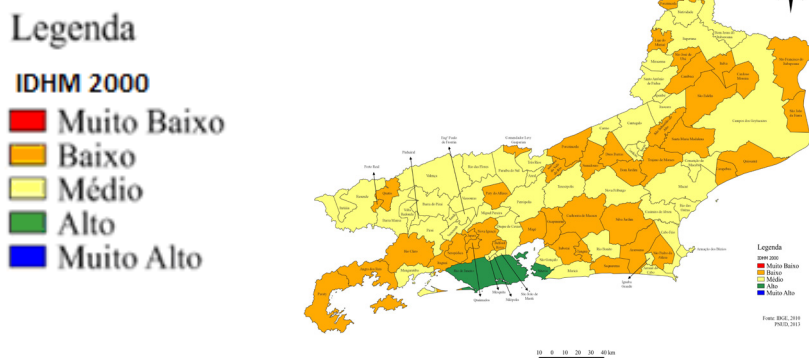
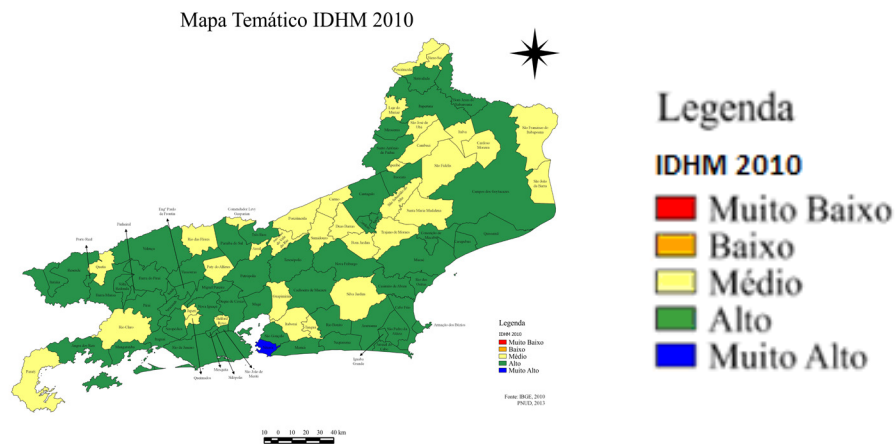


Figura 4

Figura 5



O Índice de Moran Global (I) permite detectar a incidência de autocorrelação espacial significativa na variável analisada (IDH-M) na área de estudo. No entanto, dado o número relativamente elevado de municípios envolvidos nesta investigação, o cálculo de um único índice global não é suficiente para identificar aglomerados de municípios com valores significativamente altos ou

baixos da variável, ou, então, discrepantes entre si. Para esse fim utiliza-se o Índice de Moran Local (Ii).

A Tabela 4 e as Figuras 6, 7 e 8 representam a distribuição dos municípios do estado do Rio de Janeiro pelo MoranMap, no [1º modelo]: critério da contiguidade (adjacência), sem aplicação de pesos.

Tabela 4 – Distribuição dos municípios do estado do Rio de Janeiro pelo MoranMap.
(Critério de contiguidade, sem peso)

Classe	Quantidade de municípios			Percentual de municípios		
	1991	2000	2010	1991	2000	2010
NS	77	86	84	83,70	93,48	91,30
AA	5	2	1	5,43	2,17	1,09
BB	4	0	4	4,35	0,00	4,35
AB	6	4	3	6,52	4,35	3,26
BA	0	0	0	0,00	0,00	0,00
Total	92	92	92	100,00	100,00	100,00

NS: não significativa; AA: alto-alto; BB: baixo-alto; AB: alto-baixo; BA: baixo-alto.

Figura 6

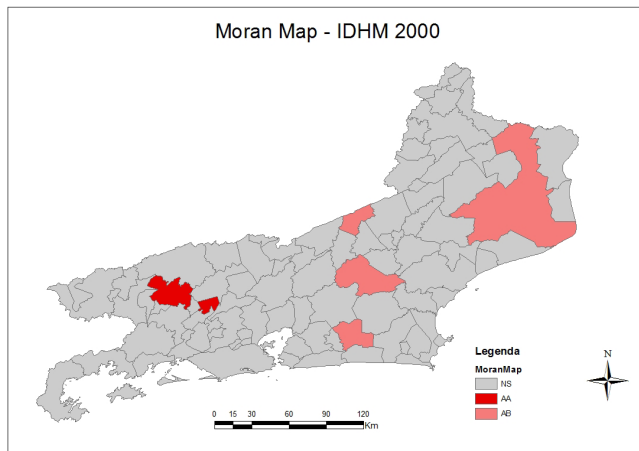
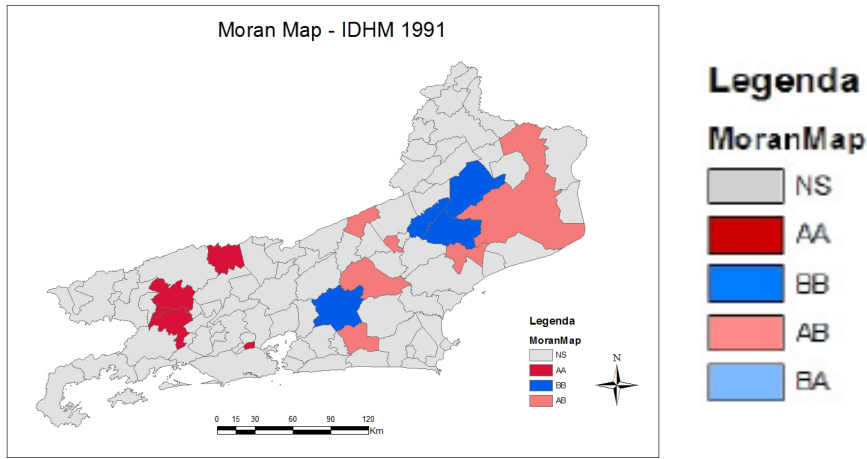
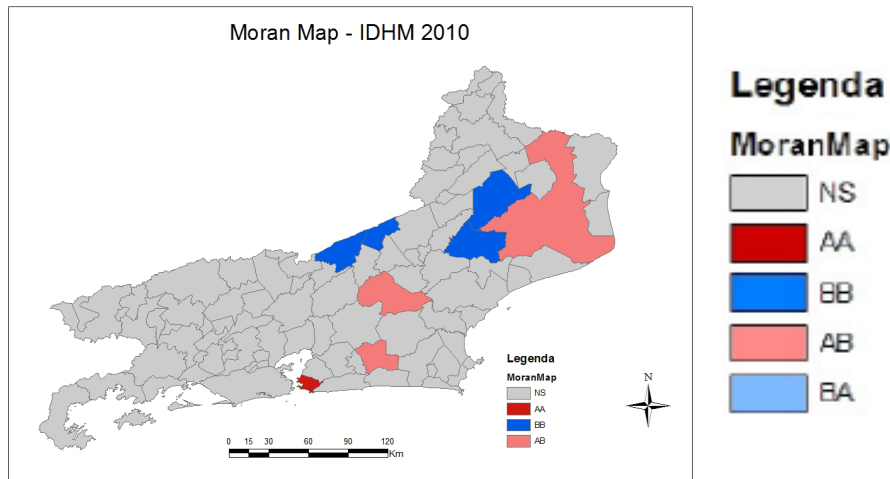


Figura 7

Figura 8



A Tabela 5 e as Figuras 9, 10 e 11 representam a distribuição dos municípios do estado do Rio de Janeiro pelo MoranMap, no [2º modelo]: critério da contiguidade,

com pesos proporcionais ao inverso da distância entre áreas vizinhas.

Tabela 5 – Distribuição dos municípios do estado do Rio de Janeiro pelo MoranMap.
(Critério de contiguidade, com pesos proporcionais ao inverso da distância)

Classe	Quantidade de municípios			Percentual de municípios		
	1991	2000	2010	1991	2000	2010
NS	76	80	78	82,61	86,96	84,78
AA	7	4	3	7,61	4,35	3,26
BB	3	3	4	3,26	3,26	4,35
AB	5	5	7	5,43	5,43	7,61
BA	1	0	0	1,09	0,00	0,00
Total	92	92	92	100,00	100,00	100,00

NS: não significativa; AA: alto-alto; BB: baixo-alto; AB: alto-baixo; BA: baixo-alto.

Figura 9

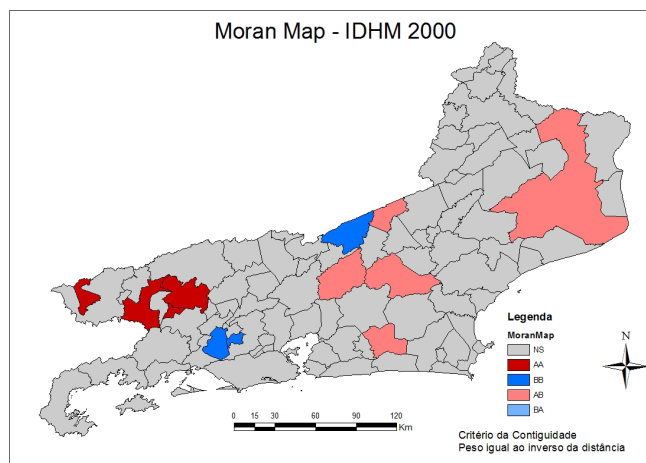
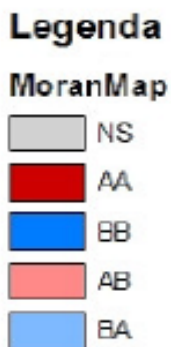
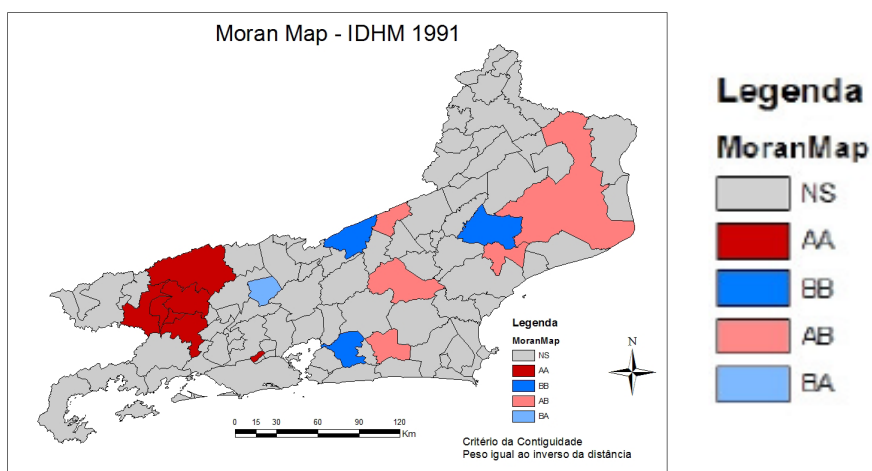
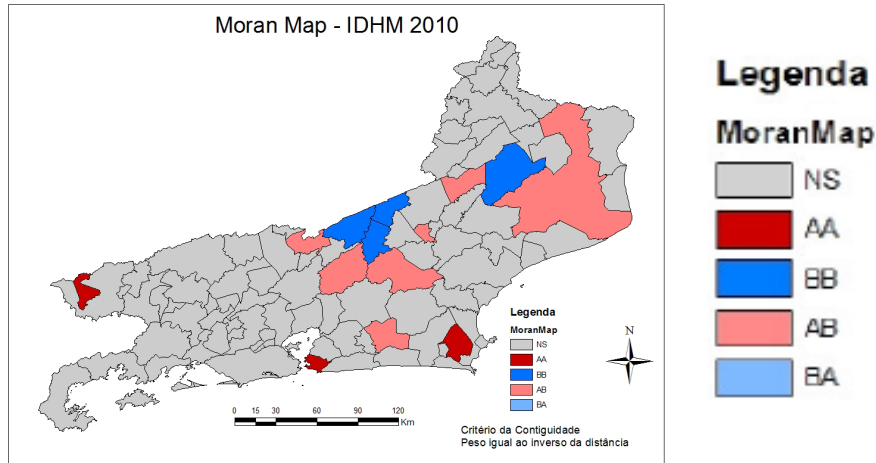


Figura 10

Figura 11



A Tabela 6 e as Figuras 12, 13 e 14 representam a distribuição dos municípios do estado do Rio de Janeiro pelo MoranMap, no [3º modelo]: critério da distância,

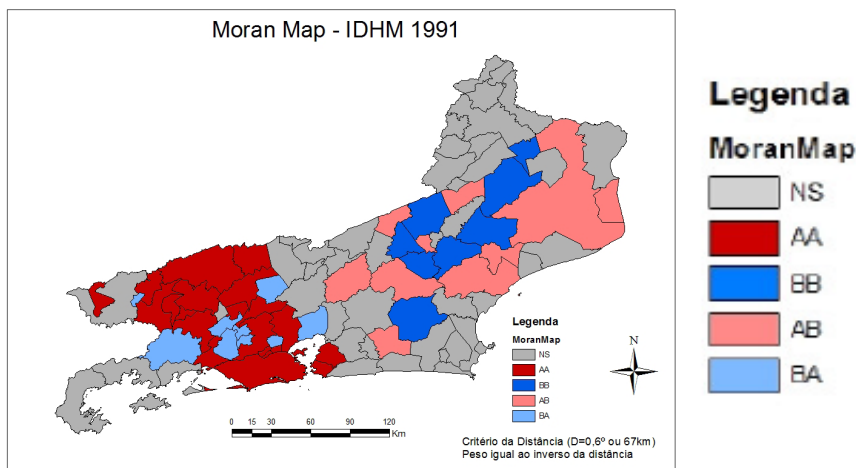
com pesos proporcionais ao inverso da distância entre áreas vizinhas.

Tabela 6 – Distribuição dos municípios do estado do Rio de Janeiro pelo MoranMap. (Critério da distância, com pesos proporcionais ao inverso da distância)

Classe	Quantidade de municípios			Percentual de municípios		
	1991	2000	2010	1991	2000	2010
NS	44	69	48	47,83	75,00	52,17
AA	21	7	16	22,83	7,61	17,39
BB	8	5	14	8,69	5,43	15,22
AB	10	6	10	10,87	6,53	10,87
BA	9	5	4	9,78	5,43	4,35
Total	92	92	92	100,00	100,00	100,00

NS: não significante; AA: alto-alto; BB: baixo-alto; AB: alto-baixo; BA: baixo-alto.

Figura 12



Legenda
MoranMap
 NS
 AA
 BB
 AB
 BA

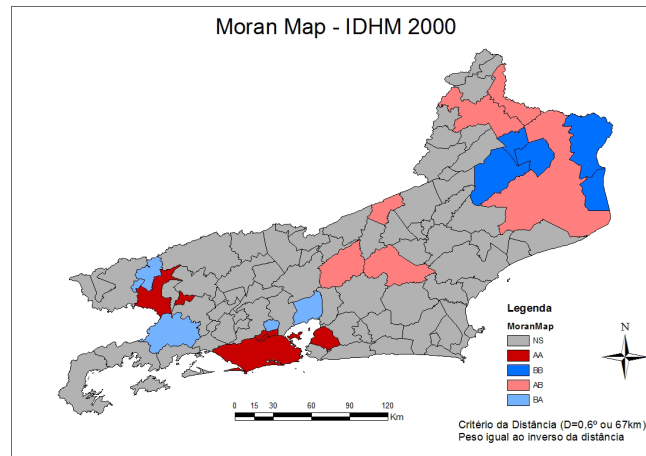
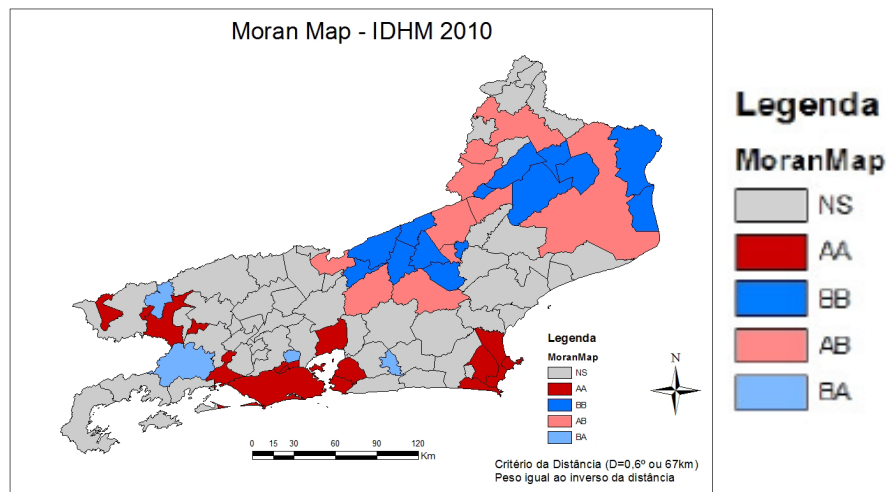


Figura 13

Figura 14



Análise dos resultados

Os mapas temáticos de IDH-M fornecem várias informações. Em 1991, havia dois municípios com IDH-M médio (Rio de Janeiro e Niterói), 31 municípios com IDH-M baixo e 59 municípios com IDH-M muito baixo. Em 2000, havia dois municípios com IDH-M alto (Rio de Janeiro e Niterói), 48 municípios com IDH-M médio e 42 municípios com IDH-M baixo. Rio de Janeiro e Niterói passam do IDH-M médio para o IDH-M alto. Em 2010, havia um município com IDH-M muito alto (Niterói), 57 municípios com IDH-M alto e 34 municípios com IDH-M médio. Niterói passou do IDH-M alto para IDH-M muito alto; Rio de Janeiro manteve-se no IDH-M alto.

Das três épocas para as quais o IDH-M foi mapeado, o ano 2000 apresenta a distribuição mais homogênea, onde 90 dos 92 municípios fluminenses são classificados nas faixas vizinhas centrais de baixo e médio, em quantidades bem distribuídas: 42 e 48, respectivamente. Em face de tal equilíbrio, o *p-valor* do Índice de Moran Global (I) apresenta todos os valores superiores a 0,05.

Essa constatação, porém, não impede que o ano 2000 seja incluído nos ensaios seguintes, realizados com base no Índice de Moran Local (I_i), já que este permite identificar sub-regiões com diferentes regimes de correlação espacial, sendo o seu emprego tão mais

necessário quanto maior o número de áreas envolvidas.

No [1º modelo] do Índice de Moran Local (I_i), para a composição da matriz de vizinhança [W], que emprega o critério da contiguidade (adjacência), sem aplicação de pesos, não foram encontradas áreas do tipo BA. No ano de 2010, somente as classes AA e AB foram representadas, com dois e quatro municípios, respectivamente, e a pequena quantidade de *clusters* e áreas de transição é coerente com o índice estatisticamente igual a zero: [*p-valor* < 0,05].

Os *clusters* de correlação positiva diminuíram: a classe AA começou com cinco municípios, em 1991, e terminou com apenas um (Niterói) em 2010; a classe BB se manteve com quatro municípios (dois passaram a NS e foram substituídos por dois outros até então não detectados). As áreas de transição AB também diminuíram, passando de seis para três municípios. Não houve ocorrência na classe BA. O destaque ficou com a emergência, aparentemente isolada, de Niterói na classe AA, em 2010.

No cômputo geral, trata-se de resultados pouco expressivos para a dinâmica constatada na variação do IDH-M, no estado do Rio de Janeiro, no período 1991 a 2010, o que conduz ao questionamento de que até que ponto essa baixa expressividade se correlaciona com o [1º modelo]. Será que a inclusão de pesos no

modelo poderia aumentar o potencial de explicação do comportamento espacial do IDH-M no estado do Rio de Janeiro? A resposta é sim. Uma constatação prévia pode ser feita, visualmente, nos MoranMaps dos três anos.

O [2º modelo] do Índice de Moran Local (I) para a composição da matriz de vizinhança [W], que emprega o critério da contiguidade, com pesos proporcionais ao inverso da distância entre áreas vizinhas, permite, por intermédio da ponderação, diferenciar a intensidade da interação espacial entre vizinhos, com base na função inversa da distância de cada área para seus vizinhos (áreas adjacentes).

No ano de 1991, observa-se, pela primeira vez, uma área de transição do tipo BA, onde um município, com um IDH-M muito baixo, é totalmente cercado por municípios com IDH-M baixo, o que significa uma superioridade destes em relação àquele. Esse bolsão de heterogeneidade não foi detectado pelo [1º modelo], no mesmo ano de 1991. Observou-se, também, que o [2º modelo] ampliou em 40% o número de municípios do tipo AA, agregando-os ao grupo de mesma classe identificados no [1º modelo].

O ano 2000 é o que apresenta menor variabilidade nos valores do IDH-M. Na classe AB, houve uma inclusão em relação ao [1º modelo], enquanto os demais municípios permaneceram inalterados. Na classe AA, foi confirmada a tendência de agregação de municípios já observada no [1º modelo]. Também foram significativos os resultados para áreas do tipo BB, onde foram detectadas três novas áreas, contra nenhuma detecção nessa classe pelo [1º modelo].

Nota-se que os *clusters* de correlação positiva diminuíram: a classe AA começou com sete municípios, em 1991, e terminou com três, confirmando a emergência de Niterói, mas ainda sem a presença do Rio de Janeiro; a classe BB começou com três municípios e terminou com quatro, em 2010. As áreas de transição AB aumentaram de cinco para sete municípios. A análise de resultados demonstra que o [2º modelo] mostra-se mais adequado para detectar os bolsões de heterogeneidade nos dados de IDH-M, produzindo resultados mais completos.

Mas como explicar a não detecção do município do Rio de Janeiro, de alto potencial turístico e boa qualidade de vida? Não seria razoável esperar que ele fosse classificado como AA ou AB? Será que a modificação do critério de composição da matriz de estrutura espacial pode aumentar o potencial de explicação do comportamento espacial do IDH-M no estado do Rio de Janeiro? As respostas a essas perguntas também são sim. Uma constatação prévia pode ser feita, visualmente, nos MoranMaps, nos três anos.

No [3º modelo] do Índice de Moran Local (I) para a composição da matriz de vizinhança [W], que emprega o critério da distância ($d = 67\text{km}$), com pesos proporcionais ao inverso da distância entre áreas vizinhas, ocorre

apenas uma alteração na composição da matriz de estrutura espacial, em relação ao [2º modelo], mantendo-se a mesma função de ponderação.

No ano 1991, observa-se uma quantidade significativa de municípios do tipo BA: são nove ocorrências nessa classe, contra apenas uma no [2º modelo] e nenhuma no [1º modelo]; a maioria dos quais formando áreas de transição para o *cluster* de municípios do tipo AA. A classe AB apresenta dez municípios, todos identificados pelo [2º modelo]. A classe AA apresenta 21 ocorrências, em 1991, e na classe BB aparecem oito municípios, contra apenas três detectados pelo [2º modelo].

O número de ocorrências em classes diferentes da NS, no [2º modelo], cai, consideravelmente, de 1991 para 2000, o que demonstra coerência com a constatação, anteriormente citada, de pouca variabilidade nos valores de IDH-M no ano 2000. Mesmo assim, pode-se constatar que o número total de 23 de ocorrências nas classes AA, BB, AB e BA é significativamente maior que o número 12, correspondente ao [2º modelo]. Vale destacar que a classe BA, no [3º modelo], apresenta cinco ocorrências contra nenhuma no [2º modelo].

A diferença entre o [2º modelo] e o [3º modelo] torna-se ainda mais acentuada na comparação de resultados para o ano 2010: são 44 ocorrências em classes diferentes da NS no [3º modelo], contra apenas 14, na mesma classe, no [2º modelo]. O balanço pende para as classes de autocorrelação positiva (AA e BB), conclusão coerente com o Índice de Moran Global (I) calculado para o [3º modelo], ano 2010: 0,144397 para um [$p\text{-valor} = 0,01$].

O valor positivo relativamente baixo do Índice de Moran Global (I), numa situação em que as classes AA e BB são em número bem maior que as de tipo AB e BA, indica que as parcelas negativas de desvios em relação à média do IDH-M (classes AB e BA), nos municípios fluminenses, pesam relativamente mais que as parcelas positivas (classes AA e BB). Em resumo: os contrastes entre os valores extremos de IDH-M (*outliers*) nas áreas de transição superam, em valor absoluto, as homogeneidades de concentração de valores de IDH-M acima ou abaixo da média (*clusters*).

Conclusão

As análises espaciais exploratórias realizadas neste trabalho foram centradas no uso dos Índices de Moran, em especial o Índice de Moran Local (I). Elas demonstram o potencial dessa ferramenta no estudo da distribuição espacial de uma variável associada a feições de área numa dada região de estudo.

Os ensaios efetuados mostraram a importância da definição do modelo de estrutura espacial mais adequado para descrever o comportamento da variável;

no caso, o IDH-M dos municípios do estado do Rio de Janeiro. Divididos em duas partes, os ensaios mostraram, primeiramente, a importância da aplicação de pesos que levassem em conta a distância entre áreas vizinhas, os quais permitiram diferenciar a intensidade da interação espacial entre vizinhos.

Em primeiro lugar, a função de ponderação escolhida assume que os pesos são proporcionais ao inverso da distância entre áreas vizinhas, que é o caso mais simples. De acordo com o critério da contiguidade, empregado no [1º modelo] e no [2º modelo], áreas vizinhas são aquelas que compartilham uma fronteira comum.

Uma vez fixado o critério da contiguidade na composição da matriz de vizinhança, os ensaios mostraram que a aplicação e a não aplicação de pesos pode fazer uma diferença significativa na detecção de bolsões de homogeneidade (*clusters*) e heterogeneidade (áreas de transição) nos dados, permitindo uma melhor compreensão do comportamento da variável na região de estudo.

Em segundo lugar, ficou evidente que o critério utilizado na construção da matriz de vizinhança ou proximidade espacial [W] promove uma diferença ainda mais significativa na capacidade de o modelo detectar sub-regiões com diferentes regimes de correlação espacial. Esse resultado ficou evidente na comparação entre o [2º modelo] e o [3º modelo], sendo este último baseado no critério da distância.

A distância-base definida, de 67km, é considerada satisfatória para expressar o conceito de proximidade entre os municípios fluminenses, que na sua maioria não são extensos e são conectados por uma malha rodoviária densa. Ao impor este critério, o [3º modelo] permite que municípios como Rio de Janeiro e Niterói, que, separados pelo extenso corpo d'água da Baía de Guanabara, mas ligados por vias terrestres e marítimas de fluxo intenso, se tornem vizinhos, virtualmente, com alto grau de interação espacial.

A distância entre o Rio de Janeiro e Niterói, de cerca de 40km, é relativamente pequena, se comparada à distância média entre municípios vizinhos no estado do Rio de Janeiro, o que faz aumentar o peso da ligação entre esses dois municípios que, no critério da contiguidade do [2º modelo], permaneceram desconectados.

A premissa de fundo no [3º modelo] é que a matriz de estrutura espacial mais adequada para o estado do Rio de Janeiro tem por base o critério da distância, em vez do critério da contiguidade, considerando que, num certo sentido, a definição de [W] representa uma hipótese sobre o fenômeno em estudo. Idealmente, a estrutura espacial representada na matriz [W] deve corresponder a algum aspecto do problema, que é significativo à luz dos processos estudados.

O caso da interação entre os municípios do Rio de Janeiro e de Niterói é notório, mas, certamente, não é

único. Além disso, é razoável supor que as conexões terrestres entre municípios fluminenses sejam um fator até mais importante no estabelecimento de interações espaciais entre eles do que a existência de fronteiras intermunicipais comuns. Paralelamente, pode-se argumentar que o grau de interação espacial entre os municípios é diretamente proporcional ao potencial de influência mútua, capaz de levar a uma homogeneização dos valores de IDH-M.

As análises exploratórias não esgotam as possibilidades de estudo do comportamento espacial do IDH-M no estado do Rio de Janeiro. Desdobramentos futuros desta pesquisa apontam para a necessidade de aprofundar o entendimento dos fatores socioeconômicos e ambientais capazes de influenciar o IDH-M e de como esses fatores variam nas diferentes sub-regiões do estado.

Nesse sentido, seria interessante investigar o comportamento espacial das variáveis componentes do IDH-M, bem como pesquisar, em trabalhos publicados, sobre que valores poderiam ser considerados na parametrização da distância de influência entre os municípios fluminenses ao se aplicar o critério da distância na composição da matriz de proximidade espacial. Essas futuras investigações poderiam ser subsidiadas por trabalhos socioeconômicos que buscassem explicar a dinâmica entre os municípios fluminenses.

Referências

- ANSELIN, L. (1995) Local indicators of spatial association – LISA. *Geographical analysis*, 27(2), p. 93-115. Disponível em: <http://dces.wisc.edu/wp-content/uploads/sites/30/2013/08/W4_Anselin1995.pdf>. Acesso em: 4 jul. 2014.
- CÂMARA, G.; CARVALHO, M. S.; CRUZ, O. G.; CORREA, V. (2002) Análise espacial de áreas. In: FUKS, S.D.; CARVALHO, M.S.; CÂMARA, G. (Orgs.) *Análise espacial de dados geográficos*. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/analise/cap5-areas.pdf>>. Acesso em: 4 jul. 2014.
- COSTA NETO, P. L. de O. (2002) *Estatística*. São Paulo: Edgard Blücher. 266 p.
- IBGE. Instituto Brasileiro de Geografia e Estatística. (2010a) *Censo 2010*. Downloads (pasta malhas digitais/censo_2010). [homepage na internet]. Disponível em: <http://downloads.ibge.gov.br/downloads_geociencias.htm>. Acesso em: 4 jul. 2014.
- _____. (2010b) *Dados vetoriais*. Downloads (pasta “malhas digitais/municipio_2010”). [homepage na internet]. Disponível em: <http://downloads.ibge.gov.br/downloads_geociencias.htm>. Acesso em: 4 jul. 2014.
- INPE. Instituto Nacional de Pesquisas Espaciais. (2014) *TerraView* [homepage na internet]. Disponível em: <<http://www.dpi.inpe.br/terraview/index.php>>. Acesso em: 4 jul. 2014.
- LONGLEY, P. A.; GOODCHILD, M. F.; MAGUIRE, D. J.; RHIND, D. W. (2005) *Geographic information systems and science*. Chichester: John Wiley and Sons. 517 p.
- MORAN, P. A. P. (1950) *Notes on continuous stochastic phenomena*. Disponível em: <[http://dds.cepal.org/infancia/guia-para-estimar-la-pobreza-infantil/bibliografia/capitulo-IV/Moran%20Patrick%20A%20P%20\(1950\)%20Notes%20on%20continuous%20stochastic%20phenomena.pdf](http://dds.cepal.org/infancia/guia-para-estimar-la-pobreza-infantil/bibliografia/capitulo-IV/Moran%20Patrick%20A%20P%20(1950)%20Notes%20on%20continuous%20stochastic%20phenomena.pdf)>. Acesso em: 4 jul. 2014.
- O’SULLIVAN, D.; UNWIN, D. J. (2010) *Geographic information analysis*. New Jersey: John Wiley and Sons. 405 p.
- OLIVEIRA, A. P. de; MACHADO, V. T.; MENDONÇA, T. S.; SOUZA, E. M. F. R. de; CARVALHO, V. M. S. G. de; CRUZ, C. B. M. (2013) Autocorrelação espacial no estudo das diferenças sociais no Rio de Janeiro: antigas zonas, novas fronteiras. *Anais do XVI Simpósio Brasileiro de Sensoriamento Remoto – SBSR*. Disponível em: <<http://www.dsr.inpe.br/sbsr2013/files/p0513.pdf>>. Acesso em: 4 jul. 2014.
- PNUD. Programa das Nações Unidas para o Desenvolvimento. (2014) *Desenvolvimento humano e IDH*. [homepage na internet]. Disponível em: <<http://www.pnud.org.br/IDH/DH.aspx?indiceAccordion=0>>. Acesso em: 4 jul. 2014.
- SANTOS, C. A dos (2005) *Integração de dados socioambientais por indicadores associados a limites políticos municipais*. Tese (Doutorado em Geografia) – Universidade Federal do Rio de Janeiro. Disponível em: <<http://ppegeo-local.igc.usp.br/pdf/anigeo/v28n2/v28n2a20.pdf>>. Acesso em: 4 jul. 2014.
- XAVIER DA SILVA, J. (2001) *Geoprocessamento para análise ambiental*. Rio de Janeiro: Ed. do Autor. 227 p.