

ABORDAGEM GEOESTATÍSTICA PARA O MAPEAMENTO DE POTENCIAIS CLUSTERS INDUSTRIAIS

Caio Peixoto Chain

Professor do Programa de Pós-graduação em Gestão e Estratégia e do Departamento de Administração Pública da Universidade Federal Rural do Rio de Janeiro (UFRuralRJ)

Luiz Gonzaga de Castro Junior

Professor do Programa de Pós-graduação em Administração e do Departamento de Gestão Agroindustrial da Universidade Federal de Lavras (UFLA)

RESUMO

O objetivo do presente artigo foi examinar a proximidade e a concentração de firmas no espaço contínuo e em direções específicas para o mapeamento de potenciais *clusters* industriais, por meio de uma abordagem geoestatística. Os métodos utilizados foram o semivariograma indicador, a modelagem da anisotropia e a krigagem. Foram aplicados inicialmente em dados simulados e posteriormente validados em dados empíricos, referentes à indústria do café em Minas Gerais. A abordagem desenvolvida foi suficiente para detectar os padrões de aleatoriedade e de dependência espacial, mensurar o alcance da proximidade entre firmas, identificar no mapa as regiões com elevado nível de concentração industrial e estimar um indicador de concentração no nível da firma. Os resultados também indicaram que os modelos anisotrópicos (direcionais) descreveram melhor a realidade dos fenômenos em comparação aos isotrópicos. Concluiu-se que a direção pode ser um determinante da teoria das aglomerações e que a abordagem geoestatística viabilizou o mapeamento de potenciais *clusters* industriais.

Palavras-chave: Teoria das aglomerações. Dependência espacial. Anisotropia.

ABSTRACT

The objective of this paper was to examine the proximity and concentration of firms in continuous space and in specific directions for the mapping of potential industrial clusters, using a geostatistical approach. The methods used were the semivariogram indicator, anisotropy modeling and kriging. They were initially applied in simulated data and later validated in empirical data, referring to the coffee industry in Minas Gerais. The approach developed was sufficient to detect patterns of randomness and spatial dependence, to measure the extent of proximity among firms, to identify regions with a high level of industrial concentration on the map, and to estimate a concentration indicator at the firm level. The results also indicated that the anisotropic (directional) models described better the reality of the phenomena in comparison to the isotropic ones. It was concluded that the direction can be a determinant of the agglomerations theory and that the geostatistical approach enabled the mapping of potential industrial clusters.

Keywords: Agglomeration Theory. Spatial dependence. Anisotropy.

Área temática: 04 - Geoprocessamento e análise espacial.

Classificação JEL: R12, C21.

1 INTRODUÇÃO

O conceito de *clusters* industriais pode ser subdividido em um sentido econômico e em um sentido geográfico (FESER; SWEENEY, 2000; FESER; ISSERMAN, 2009). Essa diferenciação objetivou demonstrar que as firmas com maior intensidade nas relações entre fornecedores e compradores estão propensas a se aglomerar no espaço. Segundo esses autores, o *cluster* econômico se refere aos vínculos ou interações entre firmas como resultado de relacionamentos econômicos e tecnológicos que incluem compra e venda de insumos, similaridades nas tecnologias de produção e/ou mercados, compartilhamento da força de trabalho e intercâmbio de informações em ambientes formais e informais. Por sua vez, o *cluster* geográfico refere-se às relações de dependência espacial entre empresas, isto é, *clusters* econômicos que se manifestam em aglomerações locais.

As economias das aglomerações se manifestam na concentração geográfica de atividades econômicas. Deve-se considerar também que indústrias geograficamente localizadas não garantem a existência de interação ou transbordamentos, porém um padrão de localização consistente entre várias regiões cria uma pressuposição forte de que esses fenômenos estão presentes (DELGADO; PORTER; STERN, 2016; PORTER, 2003). Assim, um conceito de potenciais *clusters* foi proposto por Carrol, Reid e Smith (2008) e definido como uma elevada concentração de firmas geograficamente próximas.

Um dos campos de conhecimento dentro da literatura das ciências regionais é direcionado para pesquisas em métodos e medidas para a identificação, classificação ou explicação de *clusters* de atividades econômicas. A análise do espaço influenciou a literatura deste campo e sustentou uma constante evolução das técnicas de análise regional, como os métodos baseados em Sistemas de Informações Geográficas (SIG) e as recentes contribuições da Econometria Espacial (CRUZ; TEIXEIRA, 2010).

O estado da arte da literatura sobre mensuração de aglomerações locais de firmas tem buscado contornar o Problema da Unidade de Área Modificável (MAUP) pela aplicação de funções de processos pontuais. Esta literatura segue a abordagem de métodos baseados em distâncias proposta no trabalho seminal de Duranton e Overman (2005). Uma revisão de literatura sobre esta abordagem *mainstream*, também conhecida por “Microeconometria Espacial”, e sua evolução pode ser encontrada em Marcon e Puech (2017).

O MAUP pode ser entendido como um fenômeno originado da arbitrariedade das fronteiras geográficas que nem sempre coincidem com as fronteiras econômicas. Como observado por Krugman (1991, p. 71) “there is no particular reason to think that national boundaries define a relevant region”. Esse problema pode se manifestar em um viés de escala (agregação) ou de zoneamento. No primeiro caso, podem ocorrer variações nos resultados das análises estatísticas dependendo do número de unidades de área utilizado, ou seja, quando as unidades de área são combinadas em unidades maiores ou menores. No segundo, o viés é derivado de diferentes divisões de uma mesma área, mantendo a área e o número de divisões constantes (CRESSIE, 1993; OPENSHAW, 1984; SCHOLL; BRENNER, 2016).

No entanto, dois tópicos relevantes vêm sendo negligenciados pela literatura. O primeiro é o viés direcional, ou anisotropia, que pode se manifestar em dados espacialmente dependentes (ARBIA, 2013; CHAIN et al., 2019), ou seja, quando a distribuição ou a variabilidade de um fenômeno espacial é mais intensa em uma determinada direção, em comparação com as demais direções. A outra crítica é que, embora a abordagem de processos pontuais seja estatisticamente robusta, ela não tem como objetivo zonestar em mapas a localização das aglomerações de firmas (GUILLAIN; LE GALLO, 2010; MARCON; PUECH, 2010).

Diante do exposto, questionou-se a possibilidade de incorporar o viés direcional e o mapeamento na análise da concentração de firmas no espaço sem descartar os fundamentos já

estabelecidos pela literatura *mainstream*, especificamente pelo trabalho de Duranton e Overman (2005, p. 1103): “(i) comparability across industries, (ii) control for the uneven distribution of overall manufacturing, (iii) control for industrial concentration, (iv) no aggregation bias, and (v) statistical significance”.

O objetivo do presente artigo foi examinar a proximidade e a concentração de firmas no espaço contínuo e em direções específicas, por meio de uma abordagem geoestatística. Especificamente, buscou-se identificar a estrutura de dependência espacial entre firmas, modelar o viés direcional, estimar a concentração setorial e mapear os potenciais *clusters* de atividades econômicas. Uma abordagem introdutória foi testada em dados simulados e, em seguida, validada com dados reais referentes à indústria do café no estado de Minas Gerais.

Entende-se ainda que o refinamento dos métodos de identificação de *clusters* geográficos se justifica visto que o padrão anisotrópico é mais uma regra do que uma exceção na observação de fenômenos da economia (ARBIA, 2013). Outro ponto observado por Scholl e Brenner (2016) é que a visualização dos mapas é uma ferramenta relevante para os *policy makers* envolvidos em políticas de *clusters* industriais.

2 REFERENCIAL TEÓRICO

Atualmente a geoestatística é aplicada em várias áreas do conhecimento. Seu objetivo final é quantificar e explorar o padrão espacial de uma variável regionalizada (LEUANGTHONG; KHAN; DEUTSCH, 2011). Uma variável regionalizada pode ser definida como uma função espacial numérica, variante no espaço, com uma aparente continuidade e sua variação não pode ser representada por uma função matemática determinística (MATHERON, 1963). A geoestatística pode ser subdivida, principalmente, nos conceitos de variograma e krigagem. O primeiro caso consiste na estimativa da continuidade (ou dependência espacial), enquanto o segundo está interessado na interpolação de valores em qualquer posição dentro de uma determinada região, geralmente mapas, de forma ótima (não tendenciosa e com mínima variância) baseada na estrutura definida pelo variograma.

A presente explanação da teoria geoestatística seguiu principalmente a proposição de Oliver e Webster (2014), no entanto, mais detalhes e as devidas demonstrações encontram-se formalizadas em Cressie (1993), Goovaerts (1999), Matheron (1963), entre outros.

O valor amostrado “ z ” em um local “ x ” (com coordenadas geográficas em duas dimensões x_1 e x_2) é considerado uma entre infinitas realizações de uma variável aleatória $Z(x)$ naquele respectivo local. Então, este valor amostrado (medido) “ z ” no local “ x ” pode ser escrito como $z(x)$. A variável $Z(x)$ pertence a um conjunto de variáveis aleatórias denominado processo (campo ou função) aleatório, também representado por $\{Z(x)\}$. Pelo fato de que os valores $\{z(x)\}$ em uma região tendem a apresentar alguma estrutura de dependência espacial, isto é, depender da sua vizinhança, a variável $z(x)$ é chamada de variável regionalizada.

As principais hipóteses sobre $Z(x)$ a serem assumidas estão relacionadas à estacionariedade do processo aleatório. A estacionariedade no espaço assume que o modo de variação é o mesmo de local para local. Um processo aleatório pode ser representado por:

$$Z(x) = \mu(x) + \varepsilon(x) \quad (01)$$

Em que $\mu(x)$ é a média do processo, se $\mu(x)$ é constante, então $\mu(x) = \mu$; $\varepsilon(x)$ é um valor aleatório com média zero e covariância $C(h)$. O termo $\varepsilon(x)$ pode ainda ser subdividido em $\varepsilon'(x)$ e $\varepsilon''(x)$ em que o primeiro é um termo estocástico espacialmente correlacionado e o segundo um ruído aleatório não correlacionado. Na hipótese de estacionariedade de 2ª ordem, a covariância depende somente de h e pode ser representada por:

$$\begin{aligned} C(h) &= E\{[Z(x) - \mu][Z(x+h) - \mu]\} \\ &= E[Z(x)Z(x+h) - \mu^2] \end{aligned} \quad (02)$$

Em que h é a separação entre os pontos em termos de distância e direção, $Z(x)$ e $Z(x+h)$ são os valores de Z nas localidades (x) e $(x+h)$ e E é a esperança. Se a média $\mu(x)$ não é constante então a modelagem se torna mais complexa. Pode-se assumir uma hipótese menos restritiva de estacionariedade, a hipótese intrínseca, na qual a esperança da diferença entre $Z(x)$ e $Z(x+h)$ é zero, e a covariância é substituída pela metade da variância das diferenças (semivariância), como demonstrada na Equação 03.

$$\begin{aligned} \gamma(h) &= \frac{1}{2} \text{var} [Z(x) - Z(x+h)] \\ &= \frac{1}{2} E\{[Z(x) - Z(x+h)]^2\} \end{aligned} \quad (03)$$

A semivariância também depende somente de h , enquanto $\gamma(h)$ pode ser chamada de semivariograma em função de h . O semivariograma passou a ser mais utilizado que a covariância por ser menos exigente estatisticamente. Na hipótese mais restritiva, de um processo estacionário de segunda ordem, as funções de covariância e de semivariância são equivalentes para caracterizar a dependência espacial, como demonstrado na Equação 04:

$$\gamma(h) = C(0) - C(h) \quad (04)$$

Em que $C(0)$ é a variância do processo aleatório. A hipótese de estacionariedade de 2ª ordem, e consequentemente a hipótese intrínseca, são suficientes para a interpolação ótima de uma superfície utilizando a Krigagem Simples e a Krigagem Ordinária.

Uma análise da bibliografia relevante acerca da geoestatística foi elaborada por Hengl, Minasny e Gould (2009) seguindo critérios bibliométricos para sistematização da literatura e indicou que o tema foi pouco explorado pela categoria *business & economics*. Nesse campo, foram identificadas 5 publicações pela palavra-chave *geostatistics* e 151 por *spatial statistics*, em um total de 1.047 publicações referentes a todas as áreas de conhecimento, como ciências da terra, engenharia, matemática, entre outras. Deve-se considerar que o segundo caso abrange basicamente dados de área e processos pontuais.

Na identificação de *clusters* de atividades econômicas foram encontradas duas aplicações da geoestatística, ainda como método secundário. Aguilar e Vlosky (2006) utilizaram o semivariograma isotrópico para analisar a dependência espacial entre firmas de produtos florestais em Louisiana, EUA, entretanto, a presença de *clusters* foi detectada por padrões pontuais. Por sua vez, Rybnikova e Portnov (2015) utilizaram a interpolação do Índice de Moran Local, por meio da krigagem, como tentativa de suavizar o MAUP e aumentar a precisão da delimitação da concentração setorial na Europa.

3 METODOLOGIA

3.1 Objeto de estudo - a indústria do café

A indústria de torrefação do café possui relevância socioeconômica e um elevado efeito multiplicador insumo-produto. Um novo emprego nesse setor tende a gerar 17 empregos diretos e indiretos e representa o quarto maior efeito multiplicador entre os 42 setores da economia brasileira (BRENE et al. 2014). Segundo Sakon et al. (2012), a indústria do café pode ser considerada setor-chave para o sistema produtivo brasileiro, principalmente por ser uma grande demandante de bens e serviços relativamente aos demais setores.

O Brasil apresenta também uma questão regional interessante referente ao agronegócio café. A cafeicultura em Minas Gerais produziu 56% dos grãos, enquanto seu vizinho São Paulo produziu 8%, segundo dados da Companhia Nacional de Abastecimento (CONAB, 2013). Em relação à industrialização, Minas Gerais processou 17% do café torrado e São Paulo 41%, Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE, 2013). Embora grande parte do café verde produzido em Minas Gerais seja direcionada para o mercado externo, nota-se uma desigualdade regional na agregação de valor ao produto final.

A política industrial brasileira buscou aperfeiçoar sistemas e mecanismos para promover a sustentabilidade econômica e social na agroindústria do café por meio da estratégia de *clusters*. A necessidade de identificação de potenciais *clusters* (ou Arranjos Produtivos Locais - APLs) foi expressa no “Plano Brasil Maior,” em que uma de suas metas foi “Definir as regiões e validar projetos de APL para a industrialização de café torrado e moído” (MDIC, 2013, p.159).

3.2 Banco de dados

Os padrões de localização utilizados nos cenários simulados se basearam nos critérios definidos por Scholl e Brenner (2016), para tornar a abordagem geoestatística replicável e comparável.

Todos os pontos foram gerados e referenciados em um quadro de 2x2 km, por meio de coordenadas “x” e “y”. As firmas simuladas como grupo controle (Categoria *benchmark*) foram geradas a partir de um Processo de Poisson Homogêneo (PPH) com intensidade de 500. As firmas na Categoria 1 também foram geradas com base no PPH de intensidade 75 para testar a capacidade do modelo em detectar o padrão de ausência de dependência espacial. Na Categoria 2 foram geradas firmas seguindo o PPH (intensidade 50) e adicionadas as firmas a partir de um processo Matérn, ou seja, dois *clusters* com raio de aproximadamente 0,1 quilômetro (km). Aos pontos do grupo controle foram atribuídos o valor zero e aos pontos das Categorias 1 e 2 foram atribuídos o valor de 1.

Em relação à abordagem empírica, o grupo controle foi gerado a partir dos endereços de firmas industriais coletadas no Cadastro Industrial de Minas Gerais (CIMG) disponibilizado pelo Centro Industrial Empresarial de Minas Gerais (CIEMG, 2014). O CIMG disponibilizou dados referentes a mais de 30 setores industriais que, embora demandem uma análise computacionalmente intensiva, não são censitários. As firmas de indústria de café tiveram seus endereços disponibilizados pela Relação Anual de Informações Sociais (RAIS) do MTE (2013), mediante solicitação, e abrangeram dados censitários referentes às empresas formais inseridas no grupo 108 (“Torrefação e moagem de café”) da Classificação Nacional de Atividades Econômicas (CNAE) do IBGE. O ano base dos dois bancos de dados foi 2013.

Para coletar as variáveis “latitude” e “longitude”, os endereços das firmas foram transformados em coordenadas geográficas pelo *Google Maps* e projetadas em *Albers Equal Conic* para o georreferenciamento em coordenadas métricas.

Os dados foram processados em ambiente R, com auxílio de diversos pacotes. Os pontos para a análise simulada foram obtidos a partir do spatstat (BADDELEY; TURNER, 2005). As análises geoestatísticas foram conduzidas com auxílio do gstat (PEBESMA, 2004), com exceção dos envelopes semivariográficos que foram obtidos pelo geoR (RIBEIRO Jr; DIGGLE, 2001).

3.3 Abordagem geoestatística

3.3.1 (Semi)variograma

O semivariograma experimental (empírico ou amostral) é aquele estimado com base no conjunto de dados $z(x_i)$ e na distância h , ou seja, baseado na distância entre pares de pontos. Esta propriedade atende ao pressuposto (iv) – de ausência do viés de agregação (DURANTON; OVERMAN, 2005). O estimador de semivariância ($\hat{\gamma}(h)$) pode ser calculado pelo método dos momentos (MATHERON, 1963), em outras palavras, uma forma de se estimar o semivariograma é pela média do quadrado da diferença de pares de observações que estão separados por uma mesma distância (Equação 05):

$$\hat{\gamma}(h) = \frac{1}{2m(h)} \sum_{\substack{\forall x_i, x_j \\ \text{tal que} \\ |x_i - x_j| = h}} \{z(x_i) - z(x_j)\}^2 \quad (05)$$

Em que $m(h)$ é o número de pontos amostrais a uma distância h . Com os incrementos em h é possível obter um conjunto de pontos que indicam a estrutura de dependência espacial dos dados e que podem ser analisados de forma gráfica (gráfico do semivariograma). Quanto menor a semivariância, mais similares são as informações amostradas e espera-se que os pontos se tornem mais dissimilares na medida em que a distância aumenta.

O semivariograma indicador é um caso particular por não trabalhar com variáveis contínuas, mas com atributos categóricos ou pontos de corte, que normalmente são transformados em uma variável indicadora do tipo 1 (pertence a determinada categoria) e 0 (caso contrário). Neste caso, $z(x_i) = i(u_\alpha; s_k)$ e $z(x_i + h) = i(u_\alpha + h; s_k)$, em que $i(u_\alpha; s_k)$ é a variável indicadora i na coordenada geográfica u_α referente ao atributo s da categoria k . Neste semivariograma os pares de pontos – separados por uma distância (h) – pertencentes a categorias opostas são os que mais contribuem para a estimativa da semivariância, quanto menor $\hat{\gamma}(h)$, mais conectada é a categoria tipo 1 (GOOVAERTS, 2009).

A propriedade indicadora atende aos pressupostos (ii) controlar a tendência das indústrias se aglomerarem no espaço e (iii) controlar a concentração industrial (DURANTON; OVERMAN, 2005).

O semivariograma experimental pode ainda ser comparado com um envelope de semivariogramas computados por permutações aleatórias e independentes (normalmente 999) dos dados entre as localizações observadas (DIGGLE; RIBEIRO Jr., 2007). Esse envelope simulado pelo método de Monte Carlo indica a hipótese de ausência de correlação espacial e um ponto fora dele confirma a significância estatística do semivariograma, ou seja, a presença de correlação espacial. Esta análise buscou atender ao pressuposto – (v) – significância estatística (DURANTON; OVERMAN, 2005).

Após a estimação do semivariograma experimental deve-se ajustar um modelo que descreva as principais características dos dados amostrados e ignore variações erráticas. Esta curva é denominada semivariograma teórico e é expressa por uma equação matemática. Alguns ajustes teóricos são Esférico, Exponencial, *Wave*, entre outros. Para selecionar o modelo mais bem ajustado foi utilizada a menor Soma dos Erros Quadrados Ponderados (SEQP) dos resíduos entre o semivariograma experimental e o teórico (CRESSIE, 1993).

Entre os exemplos mais populares, o modelo esférico isotrópico pode ser representado pela Equação 06 (OLIVER; WEBSTER, 2014):

$$\begin{aligned} \gamma(h) &= 0 \text{ para } |h| = 0 \\ &= c_0 + c \left\{ \frac{3|h|}{2r} - \frac{1}{2} \left(\frac{|h|}{r} \right)^3 \right\} \text{ para } 0 < |h| \leq r \\ &= c_0 + c \text{ para } |h| > r \end{aligned} \quad (06)$$

Os parâmetros c_0 , c e r são obtidos pela função do semivariograma. O efeito pepita (c_0) é a variação não explicada na distância zero e pode incluir também erros de mensuração, na teoria a semivariância em uma distância zero deveria ser zero; a contribuição – ou patamar parcial – (c) indica a variância espacialmente correlacionada; o patamar ($c_0 + c$) é uma estimativa da variância máxima das realizações do processo aleatório; por fim, o alcance (r) é o limite da correlação espacial, ou seja, uma distância a partir da qual os pares de realizações de um processo aleatório deixam de ser correlacionados (espacialmente dependentes).

3.3.2 Anisotropia

O viés direcional, ou anisotropia, pode ser definido como a variação no comportamento do semivariograma devido às mudanças na direção. Quando ignorada ou modelada indevidamente, a anisotropia pode induzir a resultados enganosos em relação ao fenômeno real (OLIVER; WEBSTER, 2014).

A anisotropia pode ser classificada em geométrica, zonal e mista (combinação dos dois casos). No primeiro caso, os semivariogramas direcionais apresentam alcances diferentes, enquanto no caso seguinte os patamares é que diferem. Quando esse comportamento é similar independente da direção, isto é, depende somente da distância entre pares de pontos, pode-se dizer que o fenômeno em estudo é isotrópico.

A anisotropia geométrica pode ser tratada ao se modelar o alcance como uma elipse definida pelo alcance máximo (maior direção) e mínimo (menor direção), por meio de uma transformação linear das coordenadas que torna a variação isotrópica. O novo parâmetro de alcance anisotrópico (r_ϕ) substitui o alcance no modelo teórico isotrópico (exemplo no modelo esférico da Equação 06), seguindo a Equação 07 (ERIKSSON; SISKKA, 2000):

$$r_\phi = \eta r / \sqrt{\eta^2 \cos^2(\phi - \theta) + \text{sen}^2(\phi - \theta)} \quad (07)$$

Em que r_ϕ é o alcance do semivariograma na direção ϕ que, por sua vez, é o ângulo de separação entre os pares de pontos nas coordenadas originais e transformadas; θ é a direção do semivariograma de maior dependência espacial ($r_{m\acute{a}x}$); e η é a razão entre o alcance na direção θ e o alcance na direção perpendicular ($r_{m\acute{i}n}$); se $\eta < 1$, $r_{m\acute{a}x} = r$ e $r_{m\acute{i}n} = \eta r$.

As formas de modelar os tipos de anisotropia em uma elipse e os respectivos cálculos foram apresentadas por Eriksson e Siska (2000). Segundo esses autores, a anisotropia zonal pode ser ajustada por meio da estrutura geométrica com uma elevada razão e um alcance tendendo ao infinito na direção do menor patamar. Para anisotropia mista, a modelagem ocorre por meio da adição do modelo geométrico ao zonal (semivariogramas aninhados).

3.3.3 Krigagem de variáveis categóricas

A previsão de variáveis categóricas (qualitativas) deve considerar a correlação espacial, a transição de probabilidades entre elas e gerar medidas de confiabilidade nas estimativas (GOOVAERTS, 1999). Para tal, é necessário transformar os dados categóricos em funções indicadoras que classificam dicotomicamente cada observação, sendo 1 (pertence a determinada categoria) e 0 (caso contrário) e assim viabilizar a krigagem indicadora. Este método indica a probabilidade de uma categoria ocorrer em uma localização específica não amostrada com base na função de probabilidade condicional (p) apresentada na Equação 08:

$$p(u; s_k | (n)) = \text{Prob}\{S(u) = s_k | (n)\} \quad k = 1, \dots, K \quad (08)$$

Em que u é a localização não amostrada; s_k é o atributo s da categoria k ; n é o número de observações disponíveis; $|(n)$ indica que a função está condicionada a informação da vizinhança local de s_k ; e $S(u)$ é um campo aleatório com realizações categóricas na localização u não amostradas. Para estimar os valores da função de probabilidade condicional pode-se utilizar a combinação linear da função indicadora dos vizinhos, por meio do estimador de krigagem ordinária (OK) descrito na Equação 09:

$$[p(u; s_k | (n))]_{OK} = \sum_{\alpha=1}^{n(u)} \lambda_{\alpha} i(u_{\alpha}; s_k) \quad (09)$$

Com λ_{α} sendo a solução do sistema de equações lineares que minimiza a variância do erro de previsão, denominado peso de krigagem. Estes pesos levam em consideração a configuração espacial dos dados, a proximidade entre os dados e as localizações não amostradas e o padrão modelado pelo semivariograma experimental (GOOVAERTS, 2009). A variável indicadora observada é dada por $i(u_{\alpha}; s_k)$, como no item 3.2.2. O resultado da krigagem pode ser visualizado no “mapa de krigagem” que é importante para delimitar as regiões (superfícies) com elevada concentração setorial.

A validação da krigagem foi utilizada para verificar a consistência das estimativas e, conseqüentemente, o melhor semivariograma ajustado. Neste processo, os pontos observados são excluídos e comparados com sua própria previsão (*leave one out*). Espera-se que o melhor modelo apresente a menor Soma do Quadrado dos Resíduos (SQR), o Resíduo Médio Padronizado (RMP) próximo de zero e a Raiz do Resíduo Médio Quadrado Padronizado (RRMQP) próxima de um (CRESSIE, 1993).

A probabilidade mapeada e prevista pela validação nada mais é do que um indicador de concentração no nível da firma, aqui denominado Índice Geoestatístico de Concentração (IGC). Em nível setorial o IGC pode ser obtido pela probabilidade média das firmas pertencentes a classe em análise. Esse índice varia entre 0, ausência de concentração, e 1, máxima concentração e buscou atender ao pressuposto (i) – comparabilidade entre indústrias (DURANTON; OVERMAN, 2005).

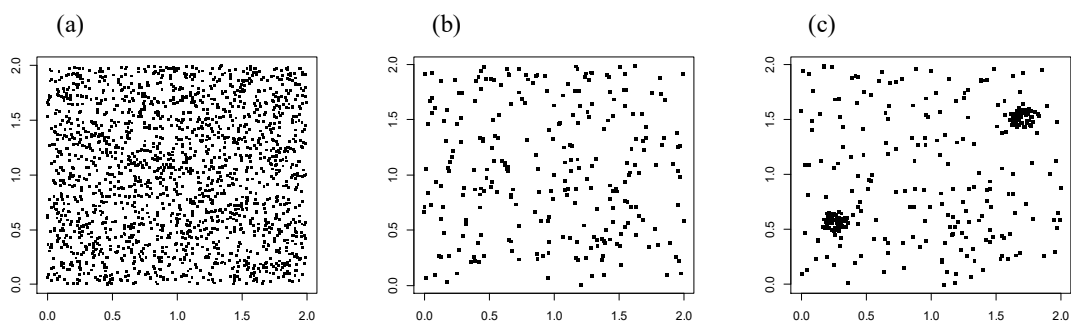
As firmas com elevado IGC indicam o centro dos *clusters* (*cluster core*). O ponto de corte “> 50%” de probabilidade foi selecionado para mapear os potenciais *clusters*, ou seja, maior a probabilidade de ocorrência da categoria em estudo devido ao perfil das firmas na vizinhança, dado o alcance do semivariograma, logo um alto nível de concentração industrial.

4 RESULTADOS E DISCUSSÃO

4.1 Análise simulada

Na Figura 1(a) estão localizadas as firmas simuladas que formam o grupo controle (Categoria *benchmark*), em um total de 1.904 pontos. A Figura 1(b) apresentou a distribuição espacial das firmas simuladas, na qual todas as 262 firmas seguem o PPH (Categoria 1). Na Figura 1(c) encontra-se a distribuição da Categoria 2, formada por um PPH com 206 firmas acrescidas de um Processo Matérn com dois *clusters* (canto inferior esquerdo e canto superior direito) com 49 firmas cada e distantes entre si em 1,8 km.

Figura 1- Distribuição espacial das firmas simuladas.

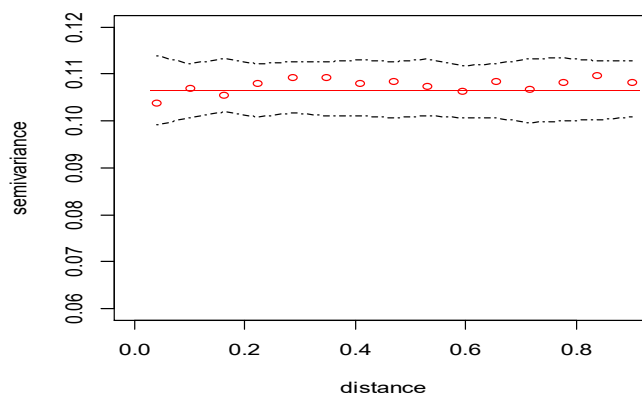


Fonte: Elaborado pelo autor com base em Scholl e Brenner (2016).

4.1.1 Cenário 1

O Cenário 1 foi formado pela sobreposição das firmas da Categoria 1 no *benchmark*. O gráfico do semivariograma isotrópico (omnidirecional) da variável indicadora está exibido na Figura 2.

Figura 2 - Semivariograma isotrópico, modelo ajustado e o envelope de confiança estatística, Cenário 01.



Fonte: Elaborado pelo autor.

O comportamento do semivariograma indicou corretamente a ausência de dependência espacial entre as firmas do grupo controle e da Categoria 1, visto que seus pontos se localizaram dentro do envelope de significância estatística. Pode-se afirmar que a Categoria 1 está aleatoriamente distribuída, independentemente da distância entre as firmas simuladas.

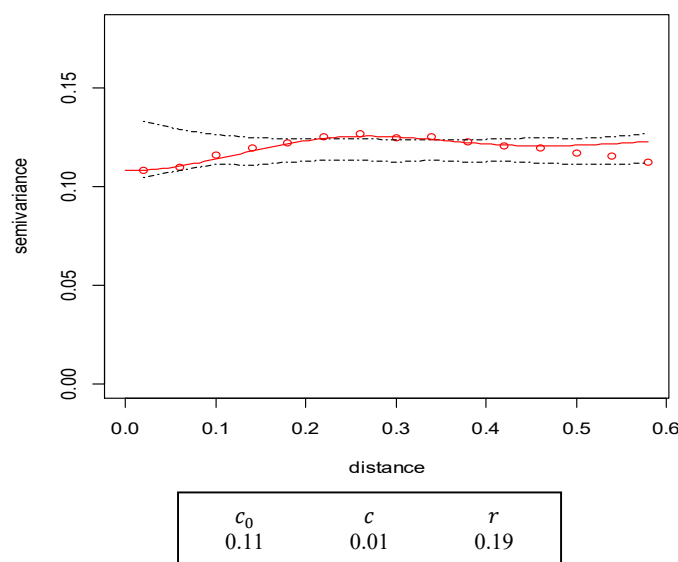
A semivariância estimada foi de aproximadamente 0,11 e o modelo teórico ajustável foi o “pepita puro” que não possui o parâmetro de alcance e , e, conseqüentemente, não possibilitou a identificação de *clusters* espaciais. Segundo Scholl e Brenner (2016) a função Kd também rejeitou o padrão de concentração/dispersão espacial de um conjunto semelhante de dados, visto que a mesma se localizou dentro das bandas do intervalo de confiança.

Os semivariogramas direcionais (0° , 45° , 90° e 135°) foram testados e não apresentaram diferenças no padrão de continuidade espacial em relação ao isotrópico. Esse conjunto de resultados, denominado efeito pepita puro, inviabilizou o mapa de krigagem.

4.1.2 Cenário 2

O Cenário 2 foi formado pela sobreposição da Categoria 2 ao *benchmark*. Dois modelos teóricos foram ajustados ao semivariograma experimental para fins de comparação: esférico e *wave*, porém a SEQP indicou um valor menor de erro para o modelo *wave* que foi escolhido para a análise do viés direcional. As medidas de erro foram 67,32 e 41,29, respectivamente. O ponto de corte entre pontos foi de 20% da distância máxima para o variograma omnidirecional e para os direcionais. A Figura 3 apresenta o semivariograma isotrópico da variável indicadora, o ajuste teórico *wave* e o envelope de significância.

Figura 3 - Semivariograma isotrópico, modelo ajustado e o envelope de confiança estatística, Cenário 02.



Fonte: Elaborado pelo autor.

A semivariância observada foi superior ao envelope e confirmou a presença de correlação espacial estatisticamente significativa. Os pontos de 6 a 9 do semivariograma isotrópico foram superiores ao envelope, nas distâncias entre 0,22 km e 0,34 km, respectivamente e após essa distância, o semivariograma retornou ao envelope. Esse resultado foi condizente com a função Kd aplicada à um padrão de pontos semelhantes por Scholl e Brenner (2016) em que a densidade excedeu a banda superior para curtas distâncias, devido ao elevado número de vizinhos dentro dos *clusters*, e retornou para o envelope de confiança estatística em seguida.

Os resultados da função Kd e semelhantes refletem se a contagem de vizinhos de cada ponto, normalizada pelo espaço ou por um *benchmark*, é maior ou menor do que o esperado ao acaso, nesse caso a dependência espacial reflete a noção de aglomeração ou dispersão (MARCON; PUECH, 2017). Assim, mesmo com os formatos e resultados semelhantes encontrados entre função Kd e o semivariograma indicador nos cenários simulados (1 e 2), o conceito de dependência espacial é diferente.

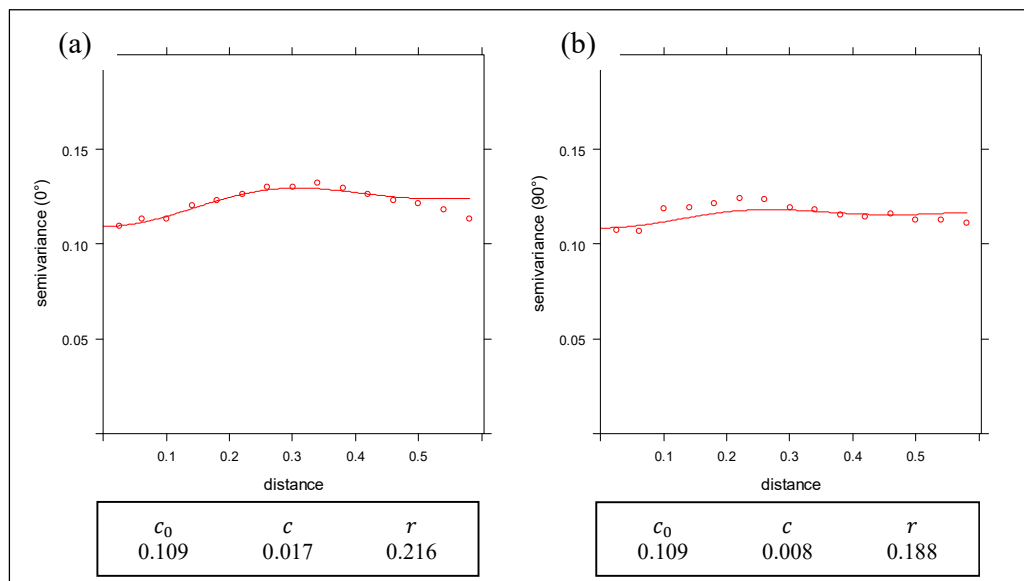
O ajuste teórico *wave* do semivariograma isotrópico apresentou um alcance de 0,19 km como era esperado, visto que os *clusters* foram simulados com raio aproximado de 0,1 km. No caso do semivariograma indicador de um fenômeno com a existência de *clusters* espaciais, o alcance corresponde ao tamanho médio destes *clusters* (GOOVAERTS, 2009), ou

seja, o diâmetro do *cluster*. No caso da função Kd, o ponto máximo da densidade pode ser interpretado como o próprio raio do *cluster* (MARCON; PUECH, 2010).

O efeito pepita de aproximadamente 0,11 indicou uma considerável heterogeneidade entre a Categoria 2 e o *benchmark* em curtas distâncias e foi aproximadamente o mesmo estimado no Cenário 1. O patamar parcial foi de 0,01.

O modelo *wave* foi ajustado nos semivariogramas da variável indicadora em 2 direções (0° e 90°) para análise da anisotropia. A tolerância utilizada foi de 45° de forma que todos os pontos fossem incluídos na análise. As direções 45° e 135° foram testadas, porém não apresentaram parâmetros consideravelmente diferentes do ajuste isotrópico. Os semivariogramas *wave* ajustados e os respectivos parâmetros se encontram na Figura 4.

Figura 4 - Semivariogramas direcionais (0° e 90°) e modelos ajustados, Cenário 02.

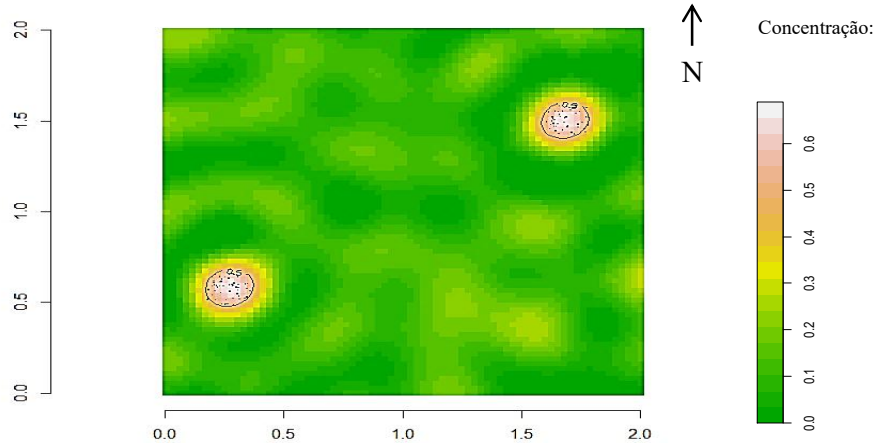


Fonte: Elaborado pelo autor.

Os parâmetros ajustados indicaram alcances distintos nas duas direções, com o maior alcance na direção de 0° (Norte) e o menor a 90° (Leste). Esse padrão espacial de anisotropia geométrica indica que o *cluster* espacial não é circular (GOOVAERTS, 2009). Nesse caso, os *clusters* possuem um formato elíptico com aproximadamente 0,22 km na direção 0° e 0,19 km na direção 90°. O menor patamar na direção de 90° semivariância indicou que os pares de pontos são mais homogêneos a Leste, ou seja, a Categoria 02 é mais conectada.

Para elaboração do mapa de krigagem foi utilizado o semivariograma anisotrópico, com os procedimentos de modelagem de anisotropia mista. A anisotropia zonal foi modelada com um alcance tendendo ao infinito na direção de 90° de forma a nivelar o patamar com a direção de 0°, enquanto a anisotropia geométrica considerou uma razão de 87% (0,188/0,216) entre o menor e o maior alcance. Em seguida, os dois ajustes foram aninhados. O mapa de krigagem (Figura 5) delimitou a superfície da concentração de firmas simuladas.

Figura 5 - Mapa da krigagem indicadora, Cenário 02.



Fonte: Elaborado pelo autor.

Os contornos delimitados na Figura 5 indicam uma probabilidade superior a 50% das firmas pertencerem ao processo Matérn da Categoria 2. Nessa faixa, 92% das firmas simuladas foram identificadas corretamente pelo processo de validação da krigagem e foram classificadas como *clusters cores*. Essas firmas apresentaram alto nível de concentração, porém as demais firmas no entorno e de mesma categoria também fazem parte do potencial *cluster* geográfico, mesmo com probabilidade inferior a 50%, pois estão dentro do alcance do semivariograma logo, podem ser consideradas geograficamente próximas.

Não foi identificado nenhum caso de “falso positivo” nos contornos acima de 50%, ou seja, firmas com perfil de potencial *cluster*, porém fora da região de elevada concentração.

O Índice Geoestatístico de Concentração (IGC) no Cenário 02 foi de 0,27, valor próximo a 0,29 encontrado pelo índice estimado por Scholl e Brenner (2016) para um padrão simulado equivalente de firmas. Esse valor indica o percentual aproximado de firmas localizadas dentro do *cluster*.

Se utilizado de forma desagregada, o IGC é útil por fornecer uma nova variável de concentração no nível da firma para modelos de regressão que buscam explicar, por exemplo, os determinantes das aglomerações ou do crescimento das firmas (ROSENTHAL; STRANGE, 2001; ALKAY; HEWINGS, 2013). Assim, a abordagem geoestatística pode ser utilizada não somente como uma pesquisa exploratória, mas também como explicativa.

O mapa da krigagem indicadora (Figura 05) apresentou uma interpretação simples em relação a abordagem para dados de área proposta por Guillain e Le Gallo (2010) e facilitou a conciliação entre análise estatística e econômica simultaneamente, visto que uma região é ou não altamente concentrada, dada uma probabilidade como ponto de corte. O mapa de Moran utilizado para a identificação de concentração de atividades econômicas apresenta resultados estatísticos conhecidos como “Alto-Alto”, “Alto-Baixo”, “Baixo-Alto” e “Baixo-Baixo”, no entanto, somente os *clusters* “Alto-Alto” e “Alto-Baixo” possuem interpretação econômica (GUILLAIN; LE GALLO, 2010).

A comparação entre a validação da krigagem utilizando o variograma isotrópico (Figura 3) e anisotrópico (Figura 4) foi apresentada na Tabela 1.

Tabela 1 - Medidas de erro, Cenário 02.

Modelo	SQR	RMP	RRMQP
Wave isotrópico	222	0,00	0,94
Wave anisotrópico	222	0,00	0,94

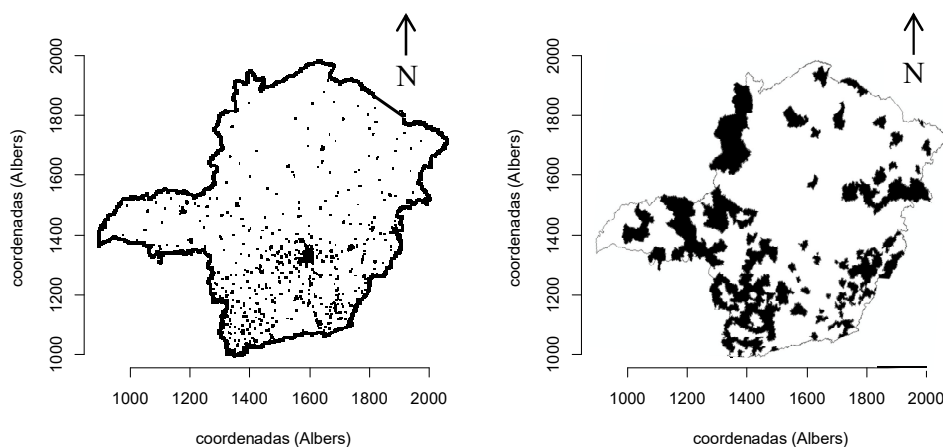
Fonte: Elaborado pelo autor.

Pode-se observar que as medidas de erro foram próximas entre o modelo isotrópico e anisotrópico, apenas diferenças de arredondamento. Os resultados indicaram boas previsões nos dois casos, com o RMP próximo de zero e o RRMQP próximo de 1, no entanto o modelo isotrópico identificou somente 85% das firmas aglomeradas (7% a menos que o anisotrópico).

4.2 Análise empírica da indústria do café em Minas Gerais

O mapa à esquerda da Figura 6 indica a localização de 8.533 firmas industriais de Minas Gerais que formaram o grupo controle da análise empírica e as 332 firmas da categoria de torrefadoras de café (mapa à direita da Figura 6) que foram agrupadas em municípios por motivo de confidencialidade dos dados. Visualmente, foi identificado um elevado número de torrefadoras na região Sul de Minas (canto inferior esquerdo), na Zona da Mata (canto inferior direito) e na região metropolitana (centro) que, em geral, aparentam seguir o mesmo padrão de localização do grupo controle.

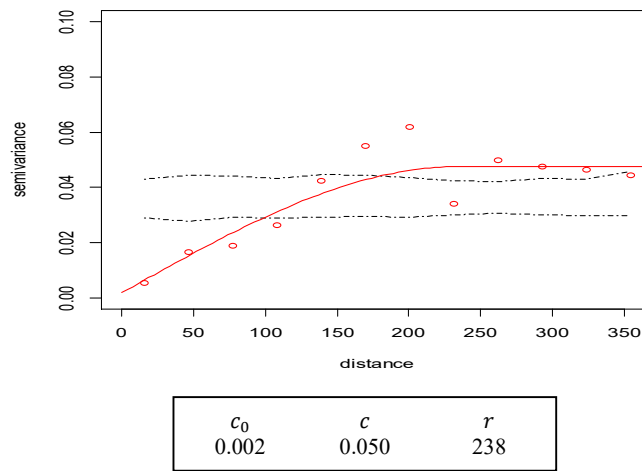
Figura 6 - Distribuição espacial das firmas *benchmark* e das torrefadoras de café.



Fonte: Elaborado pelo autor.

A análise empírica foi conduzida pela sobreposição das firmas de torrefação no *benchmark* formado pelos demais setores industriais da economia de Minas Gerais. Na Figura 7 foi representado o semivariograma isotrópico da variável indicadora, o ajuste esférico e o envelope simulado de significância estatística. O erro (SEQP) dos ajustes *wave* e esférico foram 0,07 e 0,05, respectivamente, logo o ajuste esférico foi utilizado para a análise direcional. O ponto de corte foi de 1/3 da distância máxima entre os pontos, tanto para os variogramas direcionais quanto para o isotrópico.

Figura 7 - Semivariograma isotrópico, modelo ajustado e envelope de confiança estatística, Cenário Empírico.

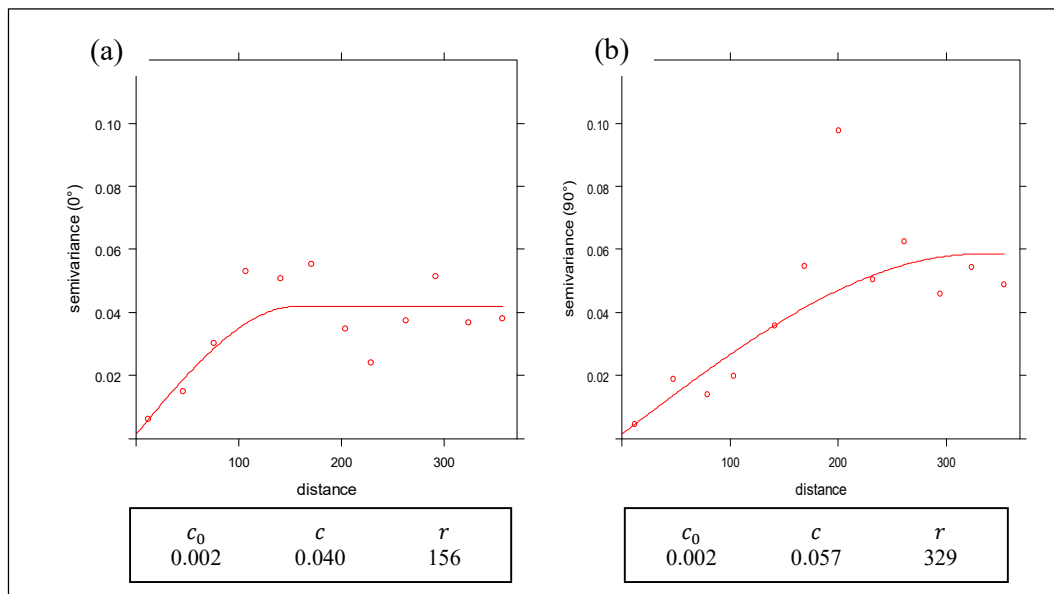


Fonte: Elaborado pelo autor.

Os pontos fora do envelope indicaram uma correlação espacial significativa. O alcance do semivariograma esférico foi de 238 km e confirmou a presença do padrão de *cluster* espacial. O efeito pepita próximo de zero (origem) indicou que o grupo controle também apresentou um padrão aglomerado, ou seja, em distâncias curtas, as firmas *benchmark* estão próximas de outras da mesma categoria, assim como as firmas da categoria torrefação. O patamar parcial foi de 0,05.

O modelo esférico foi ajustado nos semivariogramas de 0° e 90° para análise do viés direcional da variável indicadora (Figuras 8a e 8b, respectivamente) com tolerância de 45° para incluir todas as firmas nas estimativas. As direções 45° e 135° foram testadas, porém não apresentaram estruturas espaciais consideravelmente distintas da isotropia.

Figura 8 - Semivariogramas direcionais (0° e 90°) e modelos ajustados, Cenário Empírico.



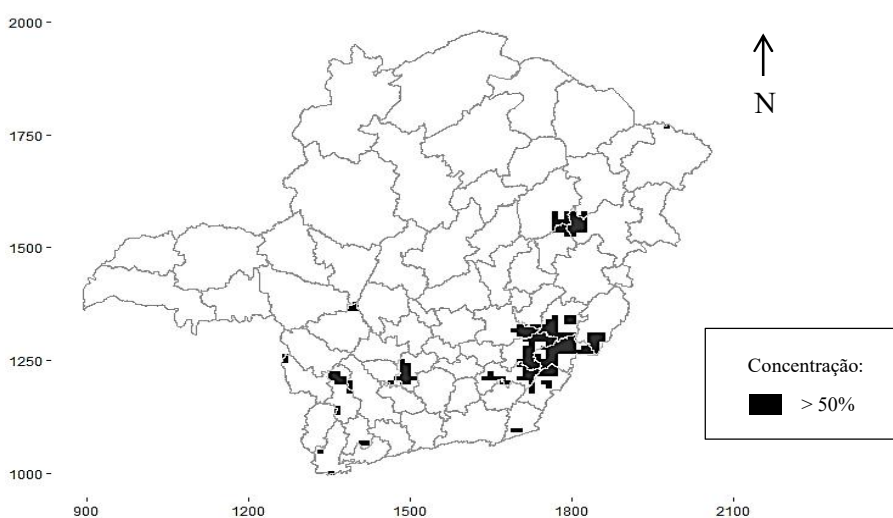
Fonte: Elaborado pelo autor.

A análise dos semivariogramas para indústria do café indicou uma anisotropia geométrica na direção de 90° com um alcance de 329 km contra 156 km na direção de 0°. Esse resultado indicou um *cluster* elíptico de maior alcance na direção Leste. A anisotropia zonal na direção de 0° também foi detectada, com um patamar de 0,04, enquanto na direção perpendicular foi de aproximadamente 0,06. Assim, na direção Norte as firmas são mais homogêneas, logo as torrefadoras são espacialmente mais conectadas. Na direção de 90° foi identificado um *outlier* espacial ao alcance de 200 km que indicou a elevada predominância da categoria de torrefadoras de café sobre o *benchmark*.

Para o mapa de krigagem indicadora foi utilizado o semivariograma anisotrópico. A anisotropia zonal foi modelada na direção de 0° de forma a nivelar o patamar com a direção de 90°, enquanto a razão de 47% (156/329) entre o menor e o maior alcance foi utilizada para modelagem da anisotropia geométrica. Os dois semivariogramas ajustados foram aninhados para modelagem da anisotropia mista.

Foi mapeada, na Figura 9, a superfície com probabilidades acima de 50% de firmas pertencerem à categoria de torrefação de café, por meio da krigagem indicadora e do semivariograma anisotrópico. O mapa de krigagem pode ser interpretado como as regiões com perfil de elevada concentração na indústria do café (*cluster core*).

Figura 9 - Mapa de krigagem Indicadora, Cenário Empírico.



Fonte: Elaborado pelo autor.

A maior concentração ocorreu nas microrregiões limítrofes Caratinga, Itabira, Manhuaçu, Ponte Nova e Viçosa. Outra região proeminente foi a da fronteira entre Capelinha e Teófilo Otoni. Em extensões menores, os demais *clusters core* se encontram no Sul de Minas, em especial Alfenas e Oliveira. Esses resultados foram condizentes com o mapeamento de *clusters* da indústria de café em Minas Gerais realizado por Chain et al. (2016), por meio do Índice de Moran Local, em nível municipal, que teve como insumo o Quociente Locacional do valor da produção. Segundo esses autores, as regiões de Caratinga, Manhuaçu, Ponte Nova, Viçosa, Capelinha e, por fim, Varginha (também Sul de Minas e limítrofe a Alfenas) indicaram a presença de concentração estatisticamente significativa

Os resultados semelhantes encontrados entre os mapas na abordagem geoestatística e a análise de dados de área no cenário empírico eram esperados, visto que a formulação da covariância e da semivariância na geoestatística são análogas aos Índices de Moran e de Geary, respectivamente (BIVAND, 1998). Em relação à interpretação dos resultados, essas

duas abordagens compartilham o conceito de correlação espacial, ou seja, valores semelhantes estão mais próximos entre si do que era de se esperar ao acaso.

Essas regiões são reconhecidas como grandes produtoras do café *in natura* e ao mesmo tempo, tem no setor de torrefação uma participação do número de empresas acima da média em relação aos demais setores da economia. O elevado número de unidades geográficas em Minas Gerais (853 municípios) tendeu a aproximar os resultados da análise de área (Índice de Moran) e da abordagem geoestatística. É relevante mencionar que a abordagem geoestatística indicou a concentração do número de firmas, enquanto o Índice de Moran indicou a concentração de produção regional (produção agregada de todas as firmas em cada município), e mesmo assim os resultados foram condizentes entre si.

O IGC para o cenário empírico foi de 0,19 e indicou uma concentração setorial fraca, mesma classificação encontrada por Chain et al. (2016) para a indústria do café em Minas Gerais pelo Índice de Moran Global. O IGC da indústria do café pode ser comparado com o índice da análise simulada que, por sua vez, se apresentou mais concentrada. No nível da firma, o maior IGC foi encontrado em uma torrefadora em Manhuaçu (0,93).

A Tabela 2 apresentou os resultados da validação da krigagem baseada no semivariograma isotrópico (Figura 7) e anisotrópico (Figura 8).

Tabela 2 - Medidas de erro, Cenário Empírico.

Modelo	SQR	RMP	RRMQP
Esférico isotrópico	301	0,00	3,56
Esférico anisotrópico	294	0,00	2,52

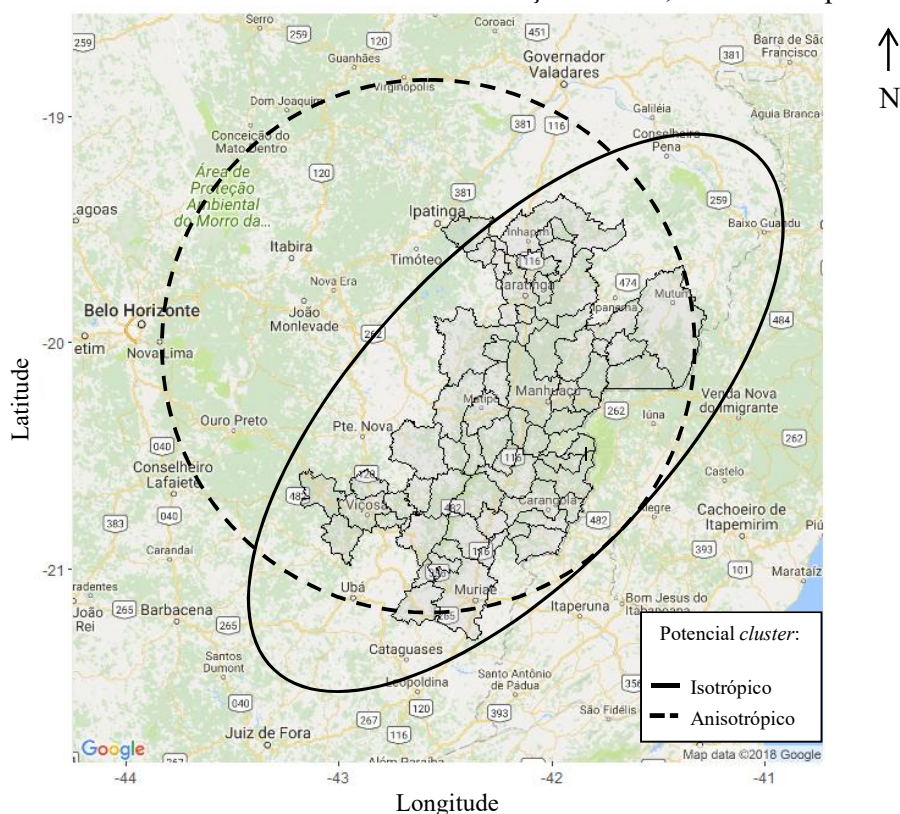
Fonte: Elaborado pelo autor.

Os critérios de erro apresentaram bons ajustes, no entanto a análise anisotrópica gerou estimativas mais robustas, com a SQR menor, a RRMQP mais próxima de 1 e aproximadamente 30% inferior que o RRMQP isotrópico.

A região das Matas de Minas, demarcada na Figura 10, foi selecionada para fins de visualização do viés direcional e da delimitação do potencial *cluster* industrial, pois apresentou o maior número de firmas com elevado IGC em Minas Gerais. A Figura 10 pode então ser interpretada como o mapa de krigagem indicadora no nível da firma, pois mostra a probabilidade de cada firma de pertencer ao *cluster*. Nesse mapa, está visível o direcionamento das probabilidades, interpretadas como o indicador de concentração (IGC).

Segundo Porter (2000), delimitar os limites do *cluster* envolve certo grau de criatividade, assim, para reduzir a subjetividade, os potenciais *clusters* geográficos foram delimitados com base no alcance dos semivariogramas isotrópico (círculo) e direcionais (elipse) para comparação. No caso da elipse, a inclinação a 90° tem uma tolerância de 45° utilizada no cálculo do semivariograma anisotrópico. O alcance do semivariograma definiu a distância que as firmas podem ser consideradas próximas.

Figura 10 – Potencial *cluster* industrial de torrefação de café, Cenário Empírico.



Fonte: Elaborado pelo autor.

A implicação econômica do mapeamento isotrópico em relação ao anisotrópico foi o agrupamento de regiões com perfis distintos. Enquanto o *cluster* anisotrópico alcançou as firmas na região das Matas de Minas, o *cluster* isotrópico alcançou, além da região das Matas de Minas, a região de João Monlevade até Belo Horizonte. O potencial *cluster* anisotrópico foi composto por 78 torrefadoras, sendo 27 o *cluster core* (IGC > 0,5) e 51 firmas periféricas. Todas as firmas com IGC acima de 50% na região abrangida pela Figura 10 se localizam nas Matas de Minas. A extensão do *cluster core* foi de aproximadamente 200 km, condizente com o *outlier* espacial no semivariograma anisotrópico da variável indicadora na Figura 8(b).

A região das Matas de Minas é regulamentada pelo Conselho das Entidades das Matas de Minas (CEMM), uma organização sem fins lucrativos de caráter representativo, científico, educacional e cultural, voltada ao desenvolvimento dos produtores de café em 63 municípios (demarcados na Figura 10) que compõem a região. Desde 2010, houve uma intensificação no processo de mobilização dos atores na busca pelo registro da Indicação de Procedência, uma estratégia de reconhecimento da singularidade do café da região, depositado em setembro de 2016 pelo CEMM no Instituto Nacional de Propriedade Intelectual (CARVALHO, 2016).

5 CONCLUSÕES

A abordagem geoestatística para identificação de potenciais *clusters* industriais foi condizente com a teoria das aglomerações ao mensurar a proximidade entre firmas, por meio do semivariograma indicador, e estimar a concentração industrial, por meio da krigagem. Além dos pressupostos consolidados pela literatura *mainstream*, foi agregada a análise do viés

direcional, a exibição dos resultados em mapas e um indicador de concentração para cada firma.

No caso da indústria do café em Minas Gerais, firmas podem ser consideradas próximas a uma distância de 238 km em termos isotrópicos ou 329 km a leste e 156 km a norte, em termos direcionais. O IGC indicou um nível de concentração setorial fraco, na magnitude de 0,19, com maior potencial de *cluster* localizado na região cafeeira das Matas de Minas. O maior IGC no nível da firma foi encontrado em Manhuaçu (0,93).

Como limitação do estudo coloca-se que não foram investigados os fatores que determinam a concentração da indústria do café em Minas Gerais. Alguns fatores conhecidos na literatura são as economias de urbanização (relacionadas com a infraestrutura da região) e as de localização (associadas à proximidade entre firmas). Assim, novos estudos podem ser conduzidos relacionando os índices de concentração em nível da firma (isotrópicos e anisotrópicos) em função de variáveis que determinam as aglomerações. Estudos futuros podem também analisar outros setores de atividades econômicas com diferentes padrões de localização das firmas.

Somente por meio de dados secundários e métodos quantitativos não é possível afirmar que o potencial *cluster* anisotrópico é viável e o isotrópico é inviável em termos de política industrial. Deve-se considerar que o modelo anisotrópico descreveu melhor a realidade dos dados com menor erro de predição (até 7% no cenário simulado e até 30% no cenário empírico) e agrupou firmas inseridas em uma região de perfil histórico, social e cultural semelhante que tende a facilitar a coordenação do *cluster*. Assim, a variável “direção” pode ser considerada um determinante da teoria das aglomerações nos casos analisados.

REFERÊNCIAS

- AGUILAR, F. X.; VLOSKY, R. P. Spatial analysis of forests products manufacturer clusters in Louisiana. **Wood and Fiber Science**, v. 38, n. 1, p. 121-131, 2007.
- ALKAY, E.; HEWINGS, G. J. D. The determinants of agglomeration for the manufacturing sector in the Istanbul metropolitan area. **The Annals of Regional Science**, v. 48, n. 1, p. 225-245, 2012.
- ARBIA, G.; BEE, M.; ESPA, G. Testing isotropy in spatial econometric models. **Spatial Economic Analysis**, v. 8, n. 3, p. 228-240, 2013.
- BADDELEY, A.; TURNER, R. Spatstat: an R package for analyzing spatial point patterns. **Journal of Statistical Software**, v. 12, n. 6, p. 1-42, 2005.
- BIVAND, R. S. **A review of spatial statistical techniques for location studies**. Bergen: Norwegian School of Economics and Business Administration. 1998
- BRENE, P. R. A. et al. Mudança estrutural e dualismo no Brasil: Uma análise insumo-produto para os anos 2000 e 2007. **Revista Capital Científico**, v. 12, n. 4, p. 29-46, 2014.
- CARROLL, M. C.; REID, N.; SMITH, B. W. Location quotients versus spatial autocorrelation in identifying potential cluster regions. **The Annals of Regional Science**, v. 42, n. 2, p. 449-463, 2008.
- CARVALHO, J. P. **Campo organizacional e adoção de Indicação Geográfica (IG): um estudo sobre a produção de cafés especiais no Brasil**. Tese de Doutorado. Universidade de São Paulo. 2016.
- CENTRO INDUSTRIAL EMPRESARIAL DE MINAS GERAIS. **Cadastro Industrial de Minas Gerais**. Belo Horizonte: CIEMG, 2014.
- CHAIN, C. P. et al. Concentração espacial na indústria do café em Minas Gerais. **Organizações Rurais & Agroindustriais**, v. 18, n. 2, p. 111-124, 2016.

- CHAIN, C. P. et al. Bibliometric Analysis of the Quantitative Methods Applied To the Measurement of Industrial Clusters. **Journal of Economic Surveys**, v. 33, n. 1, p. 60-84, 2019.
- COMPANHIA NACIONAL DE ABASTECIMENTO. **Safras: séries históricas**. Brasília: CONAB, 2013. Disponível em: <<http://conab.gov.br/site/>>. Acesso: 28 jun. 2016.
- CRESSIE, N. A. C. **Statistics for Spatial Data**. New York: Wiley & Sons, 1993.
- CRUZ, S. C. S.; TEIXEIRA, A. A. C. The evolution of the cluster literature: shedding light on the regional studies–regional science debate. **Regional Studies**, v. 44, n. 9, p. 1263-1288, 2010.
- DELGADO, M.; PORTER, M. E.; STERN, S. Defining clusters of related industries. **Journal of Economic Geography**, v. 16, n. 1, p. 1-38, 2016.
- DIGGLE, P. J.; RIBEIRO, P. J. **Model-based Geostatistics**. Springer, 2007.
- DURANTON, G.; OVERMAN, H. G. Testing for localization using micro-geographic data. **The Review of Economic Studies**, v. 72, n. 4, p. 1077-1106, 2005.
- ERIKSSON, M.; SISKKA, P. P. Understanding anisotropy computations. **Mathematical Geology**, v. 32, n. 6, p. 683-700, 2000.
- FESER, E.; ISSERMAN, A. The rural role in national value chains. **Regional Studies**, v. 43, n. 1, p. 89-109, 2009.
- FESER, E. J.; SWEENEY, S. H. A test for the coincident economic and spatial clustering of business enterprises. **Journal of Geographical Systems**, v. 2, n. 4, p. 349-373, 2000.
- GOOVAERTS, P. Geostatistics in soil science: state-of-the-art and perspectives. **Geoderma**, v. 89, n. 1, p. 1-45, 1999.
- GOOVAERTS, P. Combining area-based and individual-level data in the geostatistical mapping of late-stage cancer incidence. **Spatial and Spatio-temporal Epidemiology**, v. 1, n. 1, p. 61-71, 2009.
- GUILLAIN, R.; LE GALLO, J. Agglomeration and dispersion of economic activities in and around Paris: an exploratory spatial data analysis. **Environment and Planning B: Planning and Design**, v. 37, n. 6, p. 961-981, 2010.
- HENGL, T.; MINASNY, B.; GOULD, M. A geostatistical analysis of geostatistics. **Scientometrics**, v. 80, n. 2, p. 491-514, 2009.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Pesquisa Industrial Anual**: Produto. Brasília: IBGE, 2013.
- KRUGMAN, P. **Geography and Trade**. Cambridge, MA: MIT Press, 1991.
- LEUANGTHONG, O.; KHAN, K. D.; DEUTSCH, C. V. **Solved problems in geostatistics**. John Wiley & Sons, 2011.
- MARCON, E.; PUECH, F. Measures of the geographic concentration of industries: improving distance-based methods. **Journal of Economic Geography**, v. 10, n. 5, p. 745-762, 2010.
- MARCON, E.; PUECH, F. A typology of distance-based measures of spatial concentration. **Regional Science and Urban Economics**, v. 62, p. 56-67, 2017.
- MINISTÉRIO DO DESENVOLVIMENTO, INDÚSTRIA E COMÉRCIO EXTERIOR. **Plano Brasil Maior**: acompanhamento das medidas setoriais. Brasília: MDIC, 2013. Disponível em: <<http://www.brasilmaior.mdic.gov.br/images/data/201310/1b9c7442c23cf1e63033ac26f41ae903.pdf>>. Acesso em: 03 out. 16.

- MINISTÉRIO DO TRABALHO E EMPREGO. **Relação Anual de Informações Sociais**. Brasília: MTE, 2013. Disponível em: <<http://bi.mte.gov.br/bgcaged/rais.php>>. Acesso em: 23 nov. 2016.
- MATHERON, G. Principles of geostatistics. **Economic Geology**, v. 58, n. 8, p. 1246-1266, 1963.
- OLIVER, M. A.; WEBSTER, R. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. **Catena**, v. 113, p. 56-69, 2014.
- OPENSHAW S. **The Modifiable Areal Unit Problem (Concepts and techniques in modern geography)**. Norwich: Geo Books, 1984.
- PEBESMA, E. J. Multivariable geostatistics in S: the gstat package. **Computers & Geosciences**, v. 30, n. 7, p. 683-691, 2004.
- PORTER, M. E. *Clusters and competition*. in Gordon E. Clark, eds., **Oxford Handbook of Economic Geography**. Oxford: Oxford University Press, 2000, pp. 197-2
- RIBEIRO JR, P. J.; DIGGLE, P. J. geoR: a package for geostatistical analysis. **R news**, v. 1, n. 2, p. 14-18, 2001.
- ROSENTHAL, S. S.; STRANGE, W. C. The determinants of agglomeration. **Journal of Urban Economics**, v. 50, n. 2, p. 191-229, 2001.
- RYBNIKOVA, N. A.; PORTNOV, B. A. Using light-at-night (LAN) satellite data for identifying clusters of economic activities in Europe. **Letters in Spatial and Resource Sciences**, v. 8, n. 3, p. 307-334, 2015.
- SAKON, F. M. et al. A produção e industrialização do café no Brasil e a economia nacional: simulações sob um modelo inter-regional de insumo-produto. **Revista Econômica do Nordeste**, v. 43, n. 3, p. 443-462, 2012.
- SCHOLL, T.; BRENNER, T. Detecting spatial clustering using a firm-level cluster index. **Regional Studies**, v. 50, n. 6, p. 1054-1068, 2016.