

## A IMPORTÂNCIA DA ANÁLISE ESPACIAL NA RECONSTITUIÇÃO DA OFERTA EDUCATIVA EM PORTUGAL

**Jorge Rocha**

CEG – IGOT - UL  
jorge.rocha@campus.ul.pt

**Cristina Henriques**

CEG – IGOT - UL  
cristina.s.henriques@gmail.com

### Resumo

Este trabalho encontra-se inserido no projecto *Atlas-Repertório dos Municípios na Educação e na Cultura em Portugal (1820-1986)*. Como trabalho inicial, assentou na criação de uma base de trabalho, i.e. levantamento da informação para o ano zero (1800). Assim foram geocodificadas, com base no repertório toponímico nacional, todas as escolas existentes no referido ano e foram vectorizadas, recorrendo a informação proveniente de cartografia histórica, as estradas principais existentes na mesma data. Esta informação foi alvo de análises de estatística espacial com base num Sistema de Informação Geográfica (SIG).

**Palavras-chave:** Cartografia histórica; SIG; Análise espacial; Oferta educativa.

### Abstract

This work is inserted in the project *Atlas-Index of Municipalities regarding education and culture in Portugal (1820-1986)*. As an initial work is centered on the creation of a working basis, i.e. information surveys for the year zero (1800). So were geocoded, based on the national toponymic repertoire, all existing schools in that year and were vectorized, using information from historical cartography, the main roads existing on the same date. This information was then target of spatial statistical analysis on the basis of a Geographic Information System (GIS).

**Keywords:** Historic cartography; GIS; Spatial analysis; Education offer.

## Introdução

A Cartografia compreende, tradicionalmente, o conjunto de métodos e técnicas que conduzem à representação da superfície terrestre ou de outro planeta, sob a forma de mapas, globos, blocos diagrama, etc. A elaboração é feita a partir de observações de campo, da exploração de documentação variada e ainda de dados fornecidos por aparelhos que registem fenómenos terrestres. As primeiras referências cartográficas surgiram aproximadamente em 6200 a.C. na área que actualmente corresponde à Turquia. Recentemente a cartografia passou a incluir também na sua definição os estudos teóricos e empíricos sobre a utilização desses documentos.

Durante os séculos XV e XVI, com as inovações técnicas e de instrumentos de navegação realizadas pelos portugueses, e mais tarde com o aparecimento da Projeção de Mercator (Crane, 2002), a cartografia sofreu um grande desenvolvimento. Posteriormente, viria a ser alvo de sucessivas inovações significativas, e.g. precisão cartesiana, atingindo o auge em meados do século XX (Mota, 1958; Cortesão, 1969).

Mais recentemente, os Sistemas de Informação Geográfica (SIG) representaram a grande mudança de paradigma na Geografia em geral e na cartografia em particular. Apesar da recolha de informação espacial sobre fenómenos, que ocorrem à superfície da Terra, ser há muito tempo uma parte importante das actividades das sociedades organizadas (Burrough, 1986), o exponencial crescimento da informática (tanto ao nível de software como de hardware), que não poderá de todo estar dissociado da emergência dos SIG, contribuiu sobremaneira para uma nova forma de “ver a cartografia”.

Efectivamente, os avanços tecnológicos nos métodos de armazenamento e impressão e, principalmente, o aparecimento e vulgarização dos computadores vieram alterar profundamente a forma como os dados geográficos são adquiridos, processados e representados, bem como o modo como os interpretamos e exploramos. A descentralização da informação geográfica com elevada exatidão, a sua disponibilização em sistemas abertos através da internet, conduziu à massificação da sua utilização, que se tornou interativa, por vezes anulando mesmo a necessidade de impressão. Esta nova realidade alterou profundamente a forma como actualmente são concebidas a representação e a exploração dos dados geográficos (Coppock & Rhind, 1991).

O principal objetivo deste trabalho é reconstituir a acção dos Municípios na



educação, em Portugal, no período histórico que engloba os dois grandes movimentos: i) escolarização – da obrigatoriedade da Instrução primária à universalização da educação secundária; e ii) aculturação escrita. Tomando como unidade de observação o município, será feita uma reconstituição (panorâmica) das grandes manchas geográficas, históricas e culturais da alfabetização e da escolarização. Posteriormente, e como segundo objectivo, pretende-se levantar a rede viária e ferroviária nacional de modo a aferir da possível (ou não) relação entre a sua evolução e as mudanças observadas na rede escolar.

Para melhor responder ao nosso objetivo, adotou-se a seguinte estrutura: i) levantamento oferta educativa e escolar, pública e privada (discriminando a participação municipal e local); ii) análise e recolha dos limites administrativos; iii) análise e recolha dos grandes eixos viários e ferroviários; e iv) análise espacial dos resultados preliminares.

### **Base de Dados da Oferta Escolar**

Um dos instrumentos fundamentais do projeto é a Base de Dados, com duas unidades de observação: concelhias e escolares. A planificação dos campos e itens a incluir, a definição da estrutura e linguagem, foram objeto de trabalho por parte da equipa. Entre março e setembro de 2012, decorreu uma busca sistemática de catálogos nos arquivos centrais e municipais de que foi feito balanço na reunião de junho.

A investigação exploratória no Arquivo Nacional da Torre do Tombo aos Fundos do Ministério do Reino e de outras Secretarias de Estado revelou que há muita documentação não assinalada. Paralelamente, foram feitas investigações no Arquivo Histórico do Ministério da Educação e no do Tribunal de Contas, nos Arquivos Municipais de Lisboa e Coimbra, na Biblioteca Nacional de Portugal e na Biblioteca da Ajuda. Desde 2013 que a base de dados alfanuméricos tem estado a ser alimentada de forma regular.

### **Geocodificação e Georreferenciação**

Por geocodificação (*geocoding*) entende-se o processo de transformação de uma descrição numa localização, através de um sistema de coordenadas ou nome de uma determinada localidade. Através da geocodificação é possível, por exemplo, criar

uma tabela, em que constem inúmeras designações de locais, sejam aldeias, vilas, cidades, ou simples lugares, que posteriormente são transformados em pontos (x,y). O resultado destas localizações são localizações geográficas, onde constam determinados atributos que podem ser utilizados na cartografia ou em análises espaciais.

Ao aplicar a geocodificação, está-se automaticamente, de forma simples e dinâmica, a espacializar dados que inicialmente eram meros caracteres. Assim, passa a ser possível descortinar padrões de análise na informação resultante deste processo. Dentro deste método podemos observar diferentes tipos de geocodificação (imagem 1). Estes diferentes tipos de geocodificação irão possuir diferentes níveis de exactidão e/ou precisão. Assim temos a geocodificação directa, que se divide em precisa (G1) e aproximada (G2) e a geocodificação indirecta, i.e. vaga (G3).

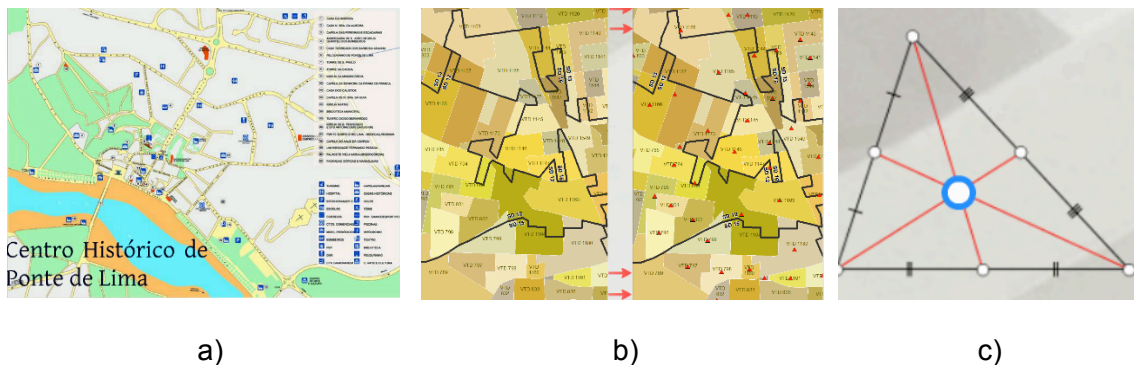


Imagem 1 – Tipos de geocodificação: directa precisa (a) e aproximada (b) e indirecta vaga (c).

A geocodificação directa precisa não requer um tratamento prévio dos dados para obtenção do ponto de localização, a directa aproximada corresponde à determinação do centro geométrico (centróide) de um elemento de referência para obter o ponto de localização, e.g. referência à localidade = centróide do núcleo histórico actual, e a indirecta vaga é utilizada quando não existem referências de localização identificáveis nas cartas actuais, e.g. acima do rio e antes da estrada.

A geocodificação está estreitamente ligada ao nosso dia-a-dia. Esta localização é incorporada de diferentes formas. No presente caso recorreu-se ao relatório toponímico nacional, do Instituto Geográfico do Exército, a uma escala 1:25 000. Esta base é a mais completa a nível nacional e conta com 167 423 registos, que incluem localidades, sedes de freguesia, rios, etc. (imagem 2).

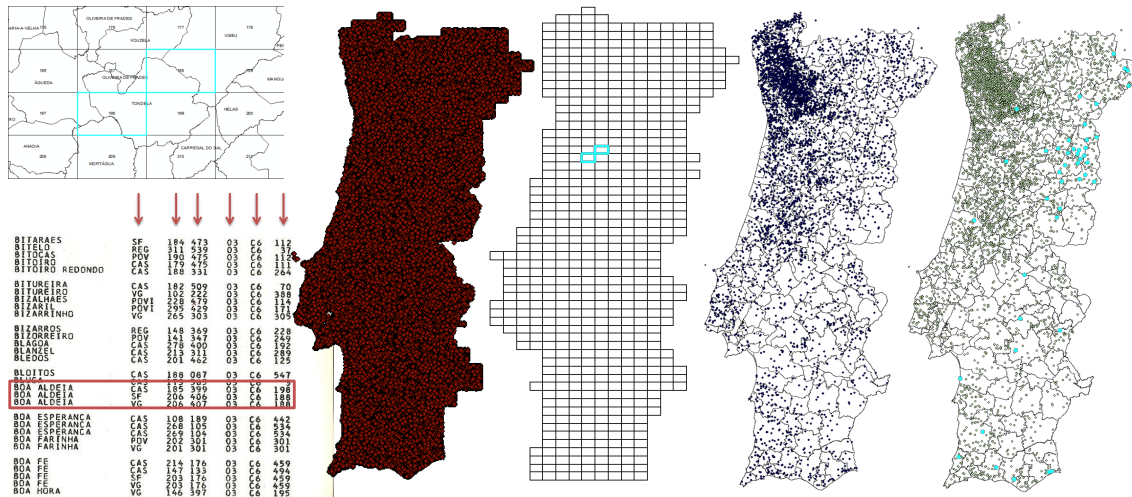


Imagem 2 – Relatório toponímico nacional.

### Análise do Vizinho Mais Próximo

O índice do vizinho mais próximo (*Nearest Neighbor Index*) foi desenvolvido por dois botânicos nos anos 1950, com o intuito de quantificar padrões em distribuições espaciais, ou seja, a partir do valor do índice é possível identificar se as distribuições espaciais são aleatórias, concentradas ou dispersas. Com os valores deste índice também é possível comparar duas ou mais distribuições distintas de forma a verificar qual é a distribuição mais ou menos dispersa.

Este índice consiste na medição de qual o tamanho da similaridade entre a distância média observada e a distância expectável numa distribuição aleatória. Assim sendo, o índice consiste na diferença entre a distância média observável e a expectável, ou então a relação entre a distância média observada e a distância média expectável.

Para saber a distância média observável, o processo descobre a distância entre cada elemento e o seu vizinho mais próximo, para depois calcular o valor médio destas distâncias. Sendo expresso da seguinte forma:

$$\bar{d}_o = \frac{\sum_i c_i}{n}$$

Partindo do princípio de que a distância média expectável tem por base uma distribuição aleatória com o mesmo número de elementos cobrindo a mesma área total, o cálculo é expresso da seguinte forma:

$$\bar{d}_e = \frac{0,5}{\sqrt{n/A}}$$

Uma vez conhecidos os valores das distâncias médias, estas são subtraídas e obtém-se o índice a partir da diferença:

$$d = \bar{d}_o - \bar{d}_e$$

Depois de calculado, o valor obtido pode variar em três possíveis resultados: i) Quando ambas as distâncias médias apresentam o mesmo valor, o valor da diferença é 0 e a distribuição observada é aleatória; ii) Quando a distância média expectável é superior à observável, o valor da diferença é inferior a 0 e a distribuição observada é concentrada; e iii) Quando a distância média expectável é inferior à observável, o valor da diferença é superior a 0 e a distribuição observada é dispersa.

A partir deste cálculo, existem também três resultados possíveis: i) Quando ambas as distâncias médias apresentam o mesmo valor, o valor da relação é 1 e a distribuição observada é aleatória; ii) Quando a distância média expectável é superior à observável, o valor da relação é inferior a 1 e a distribuição observada é concentrada. Quando o valor obtido mais se aproxima do 0, mais concentrada é a distribuição; e iii) Quando a distância média expectável é inferior à observável, o valor da relação é superior a 1 e a distribuição observada é dispersa (imagem 3).

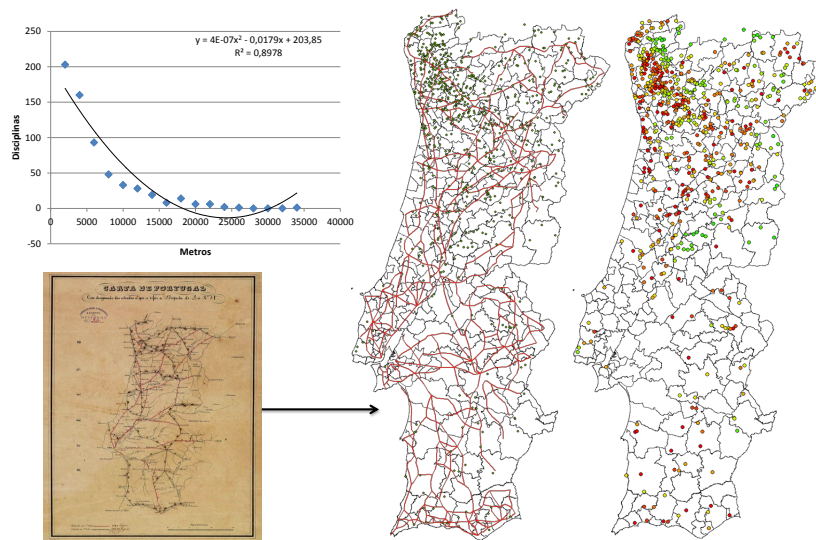


Imagem 3 – Distância ao vizinho mais próximo (estradas/escolas).



## Densidade de Kernel

Os métodos de interpolação espacial recorrem a técnicas de análise que possibilitam gerar superfícies contínuas de informação a partir de um determinado conjunto de dados amostrais. Na base destes métodos está a primeira lei de Tobler (1970), a qual refere que tudo está relacionado com tudo, mas as coisas mais próximas estão mais relacionadas do que as coisas distantes. Este princípio é também um dos alicerces da estimativa da densidade de Kernel. É uma técnica de interpolação e de análise de padrões espaciais de pontos que permite identificar, a partir de um conjunto de pontos conhecidos, a intensidade com que uma determinada variável se manifesta no espaço, revelando ser um método apropriado para aplicar neste estudo.

Consideremos então uma determinada área, onde ocorrem diversos eventos  $(s_1, s_2, \dots, s_n)$ . A intensidade  $(I)$  de uma variável numa localização  $(s)$ , pode ser definida como:

$$I(s) = \sum_{i=1}^n \frac{1}{r^2} k\left(\frac{s-s_i}{r}\right),$$

onde,  $k$  representa uma função de ponderação e  $r$  corresponde ao raio da área de influência de uma localização  $s$ .

A partir da lei de Tobler e da densidade de Kernel é possível determinar superfícies de atractividade. Assim, os eventos  $(s_i)$  contidos na área de influência definida para  $(s)$ , contribuem para o valor estimado de intensidade desta mesma localização de acordo com a proximidade dos eventos  $(s_i)$  a  $(s)$ . Com base nestes princípios e para obter uma superfície contínua com a variação de valores estimados de intensidade, podem ser utilizadas diversas funções. As diferentes opções não apresentam variações significativas no resultado da estimativa. Deste modo optou-se por utilizar a função quadrática de Epanechnikov. Esta função é uma aproximação da distribuição normal definida como:

$$\begin{cases} 3/4(1-t^2) & \text{para } t = d/h \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

com  $d$  a representar a distância entre a célula e o ponto e  $h$  o raio da área de influência.

O cálculo de densidades (imagem 4) é ideal para formular explicações e ilustrar conclusões, sendo um método estatístico de fácil compreensão para não matemáticos. Em combinação com os SIG torna-se possível não só a visualização da concentração de processos como também a descrição de alterações de processos a nível local.

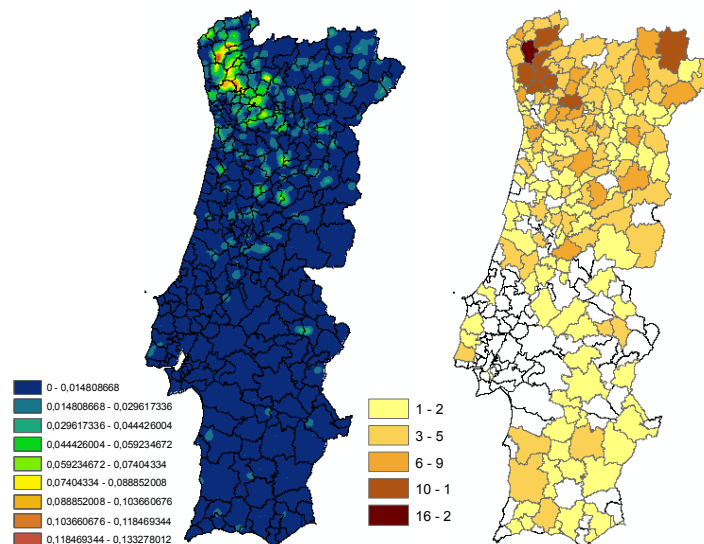


Imagem 4 – Densidade de Kernel versus mapa coropleto tradicional.

### Análise Hot Spot

A análise *Hot Spot* funciona num contexto de vizinhança e tem por base o cálculo estatístico *Getis-Ord Gi*, que apresenta quais é que são os valores altos (*hot spot*) ou baixos (*cold spot*) de aglomeração dos quais resultam os valores de *z-score* e de *p-value*.

Um ponto com um valor alto não significa que seja um *hot spot* estatisticamente significativo. Para ser um *hot spot* estatisticamente significativo, um ponto tem que ter um valor alto e ser “rodeado” por outros pontos com valores também altos. A soma local para um ponto e os seus vizinhos é comparada proporcionalmente à soma de todos os pontos. Quando a soma local é muito diferente da soma local esperada, e essa diferença é demasiado grande, resulta num valor *z* estatisticamente significativo. Os cálculos são expressos da seguinte forma:



$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} \chi_j - \bar{\chi} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{i,j}^2 - \left(\sum_{j=1}^n w_{i,j}\right)^2}{n-1}}}, \quad S = \sqrt{\frac{\sum_{j=1}^n \chi_j^2}{n} - (\bar{\chi})^2}, \quad \bar{\chi} = \frac{\sum_{j=1}^n \chi_j}{n}$$

, onde  $\chi_j$  é o valor atributo do elemento  $j$ ,  $w_{i,j}$  é o peso espacial entre os elementos  $i$  e  $j$ ,  $n$  é o número total de elementos.

Depois de calculados, os resultados podem variar entre valores positivos e negativos. Quando os valores de z-score são positivos, e quanto maior for o valor, maior é a aglomeração de valores de *hot spot*. Quando os valores de z-score são negativos, e quanto menor for o valor, maior é a aglomeração de valores de *cold spot* (imagem 5).

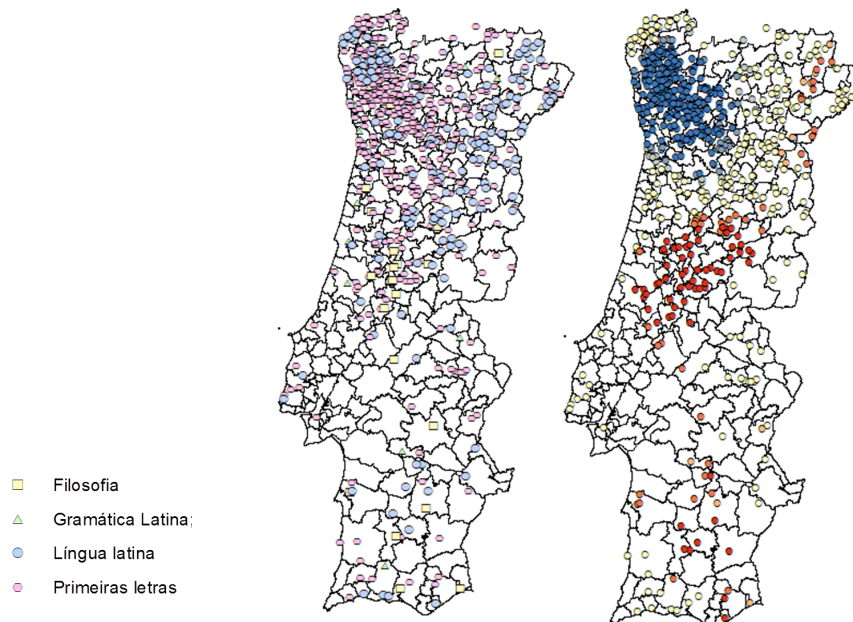


Imagem 5 – Análise de pontos quentes.

### Considerações Finais

A compreensão da distribuição espacial de dados constitui actualmente um grande desafio em várias áreas do conhecimento, sendo que a história não foge à regra. Os estudos que levam a esta compreensão são cada vez mais comuns face à disponibilidade de SIG, de baixo custo, e com ambientes de trabalho mais “amigáveis”. Estes sistemas permitem a fácil visualização e análise espacial dos dados. Para tal,

basta possuir uma base de dados apropriada. No caso da análise de dados pontuais, o principal interesse de estudo é a própria localização dos elementos.

O primeiro interesse ao analisar padrões de distribuição de pontos é determinar se os eventos observados exibem algum padrão sistemático, em oposição a uma distribuição aleatória. Procuramos identificar a existência de padrão de agrupamentos espaciais (*clusters*). Numa visão estatística, os processos pontuais são definidos como um conjunto de pontos irregularmente distribuídos num terreno, cuja localização foi gerada por um mecanismo estocástico.

Neste trabalho foram apresentados métodos de análise espacial aplicados à reconstituição histórica da oferta educativa em Portugal. Os métodos demonstraram a sua utilidade e eficácia, sendo no entanto bastante dependentes da qualidade da base de dados que os suporta. Desta forma, as próximas fases do trabalho passam pela aquisição de novos dados e melhoramento dos existentes, por exemplo através do recurso a mais fontes históricas.

### **Referências Bibliográficas**

- Burrough, P. (1986). *Principles of geographical information systems for land resources assessment*. Oxford: Clarendon Press.
- Coppock, J. & Rhind, D. (1991). The history of GIS. In D. J. Maguire, M. F. Goodchild & D. W. Rhind, *Geographical information systems: principles and applications* (Vol. 1, pp. 21-43). London: Longman.
- Cortesão, A. (1969). *História da cartografia portuguesa* (Vol. 1). Lisboa: Junta de Investigação do Ultramar.
- Crane, N. (2002). *Mercator: the man who mapped the planet*. New York: Henry Holt.
- Mota, A. (1958). L'art de naviguer en Méditerranée du XIV au XVII siècle et la création de la navigation astronomique dans les océans. In *Le navire et l'économie maritime du Moyen Âge au XVIII siècle principalement en Méditerranée* (pp. 127-154). Paris: Deuxième Colloque Internationale d' Histoire Maritime.