

APLICAÇÃO DE CONCEITOS GEOESTATÍSTICOS PARA ANÁLISE DE GERAÇÃO DE VIAGENS URBANAS

Cira Souza Pitombo

Antonio Jorge de Sousa

CERENA, Departamento de Engenharia de Minas, Instituto Superior Técnico, Universidade Técnica de Lisboa.

RESUMO

Partindo do princípio básico da geoestatística de que objetos próximos espacialmente devem ter comportamentos similares, a principal proposta deste trabalho é utilizar a geoestatística como ferramenta auxiliar na análise de geração de viagens na Região Metropolitana de São Paulo. O atual trabalho tem como principal meta estimar taxas de produção e atração de viagens por modo de transporte e motivo de viagem. Para se atingir os objetivos do trabalho são sugeridos procedimentos geoestatísticos, através da utilização do *software geoMs*. Ao fim das análises, é possível estimar os valores das variáveis relativas a viagens em n pontos desconhecidos e, finalmente, através da validação cruzada mensurar a qualidade dos modelos obtidos. Vale ressaltar que o principal intuito não seria propor a substituição dos métodos tradicionais para geração de viagens e sim, apenas apresentar uma nova ferramenta capaz de prever com certa precisão taxas de viagens considerando a correlação espacial das variáveis.

ABSTRACT

Considering the basic principle of geostatistics that spatially near objects must have similar behaviors, the main proposal of this work is to use geostatistics as an auxiliary tool in the analysis of trip generation in São Paulo Metropolitan Area. The current work has as main goal to predict trip rates production and attraction for travel mode and trip purpose. To reach the objectives of the work, geostatistics procedures are suggested through the use of *geoMs*. Through the analyses, it is possible to forecast values of trip related variables in n unknown points and, finally, with crossed validation, to measure the quality of models. It is important to highlight that the main purpose is not to replace the traditional methods for trip generation, but only to present a new tool able to forecast adequately taxes of trip production and attraction considering the spatial correlation of the variables.

1. INTRODUÇÃO

Este trabalho envolve a aplicação de conceitos anteriormente utilizados em áreas como Engenharia de minas e petrolífera e que atualmente vem sendo aplicados com êxito a ciências sociais. Vale ressaltar que Planejamento de Transportes, mais especificamente estudo do comportamento relacionado a viagens e previsão de viagens urbanas, pode ser definido como parte de ciências sociais, já que considera uma infinidade de aspectos sociais de indivíduos e grupos humanos e fatores como características socioeconômicas e variáveis comportamentais.

O presente trabalho se propõe a aplicar a geoestatística ao Planejamento de Transportes. Com o uso de tal ferramenta é possível prever adequadamente tanto produção quanto atração de viagens por diferentes modos de transportes e motivos de viagens.

O estudo de caso é a Região Metropolitana de São Paulo (RMSP). Através de uma análise simplificada (geoestatística univariada) espera-se prever taxas de viagens levando-se em conta apenas a autocorrelação espacial de variáveis. Assim, um problema inicialmente multivariado não espacial (previsão de viagens através de variáveis independentes como densidade populacional, densidade de empregos e custos de transportes, por exemplo) pode ser transformado em um problema univariado espacial (previsão de viagens utilizando técnicas geoestatísticas - krigagem ordinária – considerando apenas a autocorrelação espacial das variáveis relativas a viagens).

No entanto, o objetivo aqui não seria propor a substituição dos métodos tradicionais para previsão de viagens e sim, apenas apresentar uma nova ferramenta capaz de prever com certa precisão taxas de viagens. Futuramente, espera-se que a aplicação conjunta dos métodos seja promissora para análise de geração de viagens urbanas, pois há diversos fatores que influenciam na previsão de viagens como taxa de empregos, população, características socioeconômicas dos indivíduos ou

impedância de viagem (como tarifa, por exemplo). No entanto, tais variáveis geralmente possuem autocorrelação espacial. Assim, justifica-se a proposta de aplicação futura conjunta de métodos tradicionais e conceitos geoestatísticos.

As próximas seções apresentam o embasamento teórico relacionado à geoestatística, principais conceitos, descrevendo em seguida os objetivos, todo o tratamento espacial realizado, assim como previsão das viagens e teste de qualidade dos modelos de estimação.

2.GEOESTATÍSTICA

Na área de geologia, por exemplo, a variabilidade espacial de algumas características do solo vem sendo uma das preocupações de pesquisadores praticamente desde o início do século. No entanto, procedimentos utilizados inicialmente se baseavam na estatística clássica para caracterizar ou descrever a distribuição espacial da característica em estudo partindo da hipótese de que as variações das características do solo de um local para outro eram aleatórias.

No entanto, a partir da década de 50 (Krigé, 1951), concluiu-se que somente a estatística clássica seria insuficiente para explicar fenômenos onde se observa variabilidade espacial. Desta forma, seria necessário levar em consideração a distância entre as observações. Conclui-se, assim, que a localização de diferentes objetos é de extrema importância para análise de dados espaciais. Surge o conceito da geoestatística, que leva em consideração a localização geográfica e a dependência espacial.

Fazendo uma ligação entre estudos geológicos (uma das principais aplicações geoestatísticas) e Planejamento de Transportes, pode-se afirmar que embora a análise de variáveis que caracterizem solos e indivíduos (comportamento de viajantes urbanos) seja bastante diferente, existe um fator convergente entre tais estudos: a dependência espacial das variáveis envolvidas.

Estudos relacionados a viagens urbanas, por exemplo, são influenciados por atributos eminentemente geográficos, tais como: adensamentos residenciais e de atividades socioeconômicas, proximidade entre zonas, cobertura espacial da rede de transporte, impedância de viagem na malha rodoviária, etc. (Loureiro et al, 2006).

Assim, é compreensível a consideração do atributo espacial na previsão de viagens no meio urbano, já que se supõe que modelos espaciais traduzem melhor o padrão de comportamento individual relacionado a viagens. Neste trabalho será aplicada a geoestatística como técnica de auxílio para análise espacial de dados. Algumas adaptações ao conjunto de dados devem ser consideradas a fim de tratá-los como superfícies contínuas. Conceitos básicos geoestatísticos são descritos em seguida.

2.1 Características das variáveis regionalizadas

As *variáveis regionalizadas* (VR), tratadas através da estatística espacial, apresentam em comum o fato de dependerem fortemente da localização espacial (Soares, 2000). As propriedades são intermediárias entre uma variável totalmente aleatória e uma totalmente determinística. Apresentam continuidade de ponto para ponto, mas as mudanças são tão complexas que não são possíveis de descrever por nenhuma função determinística.

- **Suporte:** As variáveis regionalizadas estão intimamente ligadas ao *suporte*, termo habitualmente entendido como o volume e orientação espacial de uma *amostra*. Uma *amostra* em geoestatística é uma porção de material coletado em cada ponto (localizado por suas coor-

denadas geográficas). Os materiais são coletados através do *suporte*. O *suporte* influencia fortemente os valores da variável, determinando, quando aumenta, a diminuição da sua dispersão.

Embora uma VR seja espacialmente contínua, não é normalmente possível saber o seu valor em todos os locais. Geralmente, os seus valores são conhecidos apenas através de *amostras* que são recolhidas em localizações específicas. O tamanho, forma, orientação e arranjo espacial destas amostras constituem o *suporte* da variável regionalizada.

No caso do trabalho em questão, claramente não há uniformidade do suporte, pois trabalha-se com centróides (coordenadas – latitude e longitude) das 389 zonas de tráfego da RMSP. Assim, cada ponto representa medidas completamente heterogêneas de áreas. A fim de confrontar os problemas relativos aos dados há duas adaptações propostas:

A) Tratamento com centróides de zonas considerando valores médios das variáveis no seu centróide (Goovaerts, 2008); B) Considerar as variáveis relativas a viagens (taxas de atração e produção de viagens) como sendo VR.

- **Continuidade:** Apesar do aspecto errático evidenciado pela VR, existe uma certa continuidade espacial. Uma VR, $Z(x)$, é contínua se

$$\lim_{x \rightarrow x_0} E\left\{[Z(x) - Z(x_0)]^2\right\} = 0 \quad (i)$$

Na realidade, a continuidade raramente existe, ou seja, aquele limite tende para um valor $C_0 \neq 0$, designado por *efeito pepita (nugget effect)*. Este efeito deriva de dois fatores: possível existência de erros de amostragem, e existência de micro-regionalizações que a escala de amostragem não permite detectar.

$$\lim_{x \rightarrow x_0} E\left\{[Z(x) - Z(x_0)]^2\right\} = C_0 \neq 0 \quad (ii)$$

- **Anisotropia:** A VR pode apresentar uma continuidade espacial diferenciada segundo direções distintas.

2.2 Variogramas

A continuidade espacial da VR pode ser medida por uma quantidade positiva que traduz as diferenças dos valores da variável medidos em pontos distanciados de h . A função que traduz a semivariância em função das distâncias entre os pontos, designa-se por variograma. A semivariância é uma medida do grau de dependência espacial entre as amostras e é calculada através de:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2 \quad (iii)$$

$\gamma(h)$ é a semivariância para o intervalo h ,

$N(h)$ é o número de pares de dados que distam h entre eles,

$Z(x_i + h) - Z(x_i)$ é o valor da diferença de um par de dados que distam h entre eles.

O variograma é a ferramenta básica, que permite descrever quantitativamente a variação no espaço de um fenómeno regionalizado. A natureza estrutural de um conjunto de dados (assumido pela variável regionalizada) é definida a partir da comparação de valores tomados simul-

taneamente em dois pontos, segundo uma determinada direção. É a etapa anterior à estimação geoestatística (krigagem).

Para a construção do variograma experimental são consideradas as seguintes situações: (1) São examinados os possíveis pares de pontos; (2) Quando a distância entre pares é zero, o valor em cada ponto é comparado com ele próprio, e logo as diferenças são 0 e o valor da semivariância também zero; (3) Se a distância h é pequena, os pontos a serem comparados são muito semelhantes e estão bastante relacionados entre eles, pelo que o valor da semivariância é reduzido; (4) À medida que aumenta a distância entre os pontos a serem comparados, maior é a semivariância, até que, a partir de determinada distância, a amplitude ou alcance (*range*), a semivariância estabiliza num valor, o patamar (*sill*), que é igual à variância dos dados da população considerada, significando que já não existe qualquer relação entre os pares de dados considerados a esta distância.

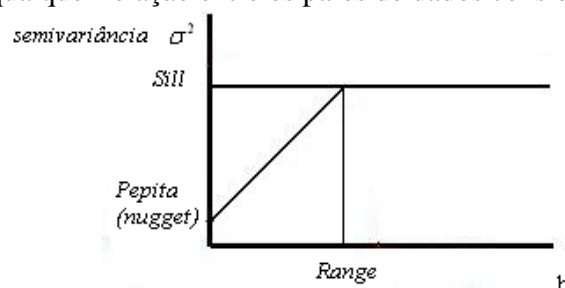


Figura 1 Forma característica de um semivariograma

Após a construção dos variogramas experimentais, estes são ajustados a modelos matemáticos conhecidos. É habitual encontrar dois tipos de variogramas elementares, variogramas com e sem patamares. Entre os variogramas com patamares o mais comum é o esquema esférico que apresenta um comportamento linear junto à origem, existindo ainda o variograma exponencial e o gaussiano. Quanto aos variogramas sem patamar é usual encontrar uma classe de funções que dependem de h^a , em que o variograma linear é um caso particular.

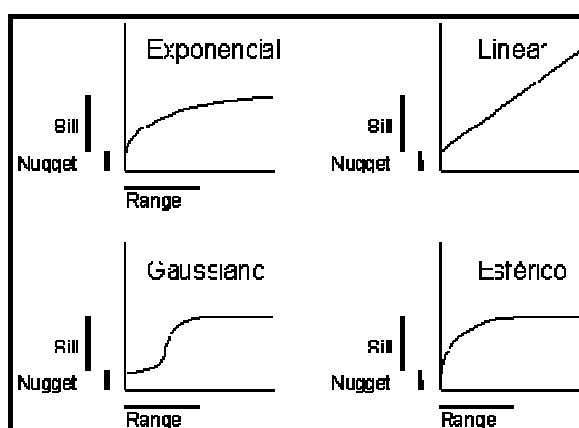


Figura 2- Modelos matemáticos aplicados a variogramas

Sendo o variograma uma função que depende da direção, é natural que apresente andamento diferente conforme a direção, refletindo a anisotropia da continuidade espacial da VR. Se o variograma apresenta uma forma semelhante em todas as direções do espaço, só dependendo de h , diz-se que a estrutura do fenômeno é isotrópica, sem direções privilegiadas de variabilidade.

2.3 Estimação Geoestatística – Krigagem

A técnica de krigagem assume que os dados recolhidos de uma determinada população se encontram correlacionados no espaço. Isto é, se num aterro de resíduos tóxicos e perigosos a concentração de zinco é $z(x)$ num determinado ponto de coordenada x , é muito provável que se encontrem resultados muito próximos de $z(x)$ quanto mais próximos se estiver do ponto x (princípio da geoestatística). Porém, a partir de determinada distância de x , certamente não se encontrarão valores relacionados com $z(x)$ porque a correlação espacial deixará de existir.

Na krigagem, o procedimento é semelhante ao de qualquer interpolação, exceto que aqui os pesos são determinados a partir de uma análise espacial, baseada no variograma. Além disso, a krigagem fornece, em média, *estimativas não tendenciosas e com variância mínima*.

3. OBJETIVOS

O principal objetivo do trabalho é apresentar a geoestatística como ferramenta viável para análise de geração de viagens. O trabalho se propõe a: **(3.1)** Fazer uma análise exploratória das variáveis relativas a viagens, detectando padrões espaciais na região de estudo; **(3.2)** Estimar taxas de produção de viagens por modo de transporte e motivo de viagem; e **(3.3)** Estimar taxas de atração de viagens por modo de transporte e motivo de viagem.

4. DESCRIÇÃO DO TRABALHO

Este trabalho tem como principal proposta estimar taxas de viagens (atração e produção) em diversas coordenadas, considerando os valores das variáveis nos centróides das Zonas de Tráfego da RMSP. Nos centróides devem ser considerados valores médios das taxas de viagens por área. Através da Krigagem Ordinária é possível estimar valores em n coordenadas diferentes (a depender dos parâmetros adotados na krigagem) para tais variáveis.

No presente projeto, os dados tratados, após seleção de variáveis para análise, serão sujeitos a interpretação geoestatística, mediante a utilização do *software geoMS (Geostatistical Modelling Software)*, desenvolvido pelo Instituto Superior Técnico da Universidade Técnica de Lisboa.

O *geoMs* é um *software* direcionado para analisar dados de geoestatística que apresentam dependência espacial. É composto por vários módulos que, no seu conjunto, são necessários para responder aos objetivos propostos. Os procedimentos geoestatísticos utilizados, ilustrados na Figura 3 serão descritos nas próximas seções.

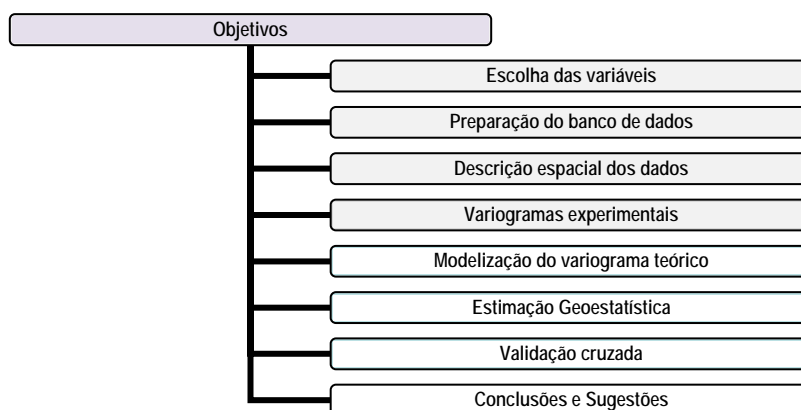


Figura 3- Etapas do trabalho

5. ESCOLHA DAS VARIÁVEIS RELATIVAS A VIAGENS E PREPARAÇÃO DO BANCO DE DADOS DE ENTRADA

Considerando como premissa básica a correlação espacial para previsão de viagens (em termos de taxas) foram adotadas apenas variáveis relativas a viagens (produção ou atração) por centróide de zona de tráfego (389 no total). Foram escolhidas 14 variáveis (limitação do *software* utilizado).

Os dados de entrada são apresentados na Tabela 1. O banco de dados apresenta um total de 389 Ids (zonas de tráfego). Na tabela são apresentadas as dez primeiras zonas, Id da zona de tráfego, coordenadas (em metros) e variáveis: (1) **prodcol** (taxa média de produção de viagens por transporte coletivo); (2) **prodaut** (taxa média de produção de viagens por automóvel); (3) **prodapé** (taxa média de produção de viagens a pé); (4) **prodtrabind** (taxa média de produção de viagens com motivo trabalho na indústria); (5) **prodtrabcom** (taxa média de produção de viagens com motivo trabalho no comércio); (6) **prodtrabserv** (taxa média de produção de viagens com motivo trabalho no setor de serviços); (7) **prodeduc** (taxa média de produção de viagens com motivo educação); (8) **atracol** (taxa média de atração de viagens por transporte coletivo); (9) **atraaut** (taxa média de atração de viagens por automóvel); (10) **atraapé** (taxa média de atração de viagens a pé); (11) **atratrabind** (taxa média de atração de viagens com motivo trabalho na indústria); (12) **atratrabcom** (taxa média de atração de viagens com motivo trabalho no comércio); (13) **atratrabserv** (taxa média de atração de viagens com motivo trabalho no setor de serviços); (14) **atraeduc** (taxa média de atração de viagens com motivo educação).

Tabela 1 Banco de dados de entrada

| ID | LONGIT (Y) | LATIT(X) | prodcol | prodaut | prodape | prodtrabind | prodtrabcom | prodtrabserv |
|----|------------|-----------|---------|---------|-------------|-------------|--------------|--------------|
| | (m) | (m) | (%) | (%) | (%) | (%) | (%) | (%) |
| 1 | -46633691 | -23548404 | 1.15 | 0.27 | 0.27 | 0.08 | 0.47 | 0.37 |
| 2 | -46629525 | -23543746 | 0.68 | 0.21 | 0.19 | 0.06 | 0.33 | 0.19 |
| 3 | -46632324 | -23553721 | 0.29 | 0.10 | 0.19 | 0.04 | 0.24 | 0.25 |
| 4 | -46641581 | -23551130 | 0.40 | 0.17 | 0.26 | 0.11 | 0.29 | 0.26 |
| 5 | -46642640 | -23543989 | 0.97 | 0.25 | 0.36 | 0.02 | 0.48 | 0.41 |
| 6 | -46638355 | -23538673 | 0.57 | 0.25 | 0.25 | 0.04 | 0.43 | 0.40 |
| 7 | -46634387 | -23530932 | 0.43 | 0.17 | 0.18 | 0.04 | 0.15 | 0.19 |
| 8 | -46620484 | -23546585 | 0.39 | 0.18 | 0.22 | 0.10 | 0.30 | 0.16 |
| 9 | -46615414 | -23560147 | 0.23 | 0.18 | 0.08 | 0.05 | 0.09 | 0.18 |
| 10 | -46619454 | -23570800 | 0.14 | 0.15 | 0.06 | 0.03 | 0.08 | 0.10 |
| ID | prodeduc | atracol | atraaut | atraape | atratrabind | atratrabcom | atratrabserv | atraeduc |
| | (%) | (%) | (%) | (%) | (%) | (%) | (%) | (%) |
| 1 | 0.11 | 1.19 | 0.30 | 0.24 | 0.16 | 1.38 | 1.83 | 0.14 |
| 2 | 0.07 | 0.65 | 0.21 | 0.24 | 0.15 | 1.74 | 0.77 | 0.08 |
| 3 | 0.10 | 0.33 | 0.10 | 0.16 | 0.06 | 0.21 | 0.58 | 0.10 |
| 4 | 0.12 | 0.39 | 0.17 | 0.28 | 0.14 | 0.39 | 0.70 | 0.08 |
| 5 | 0.16 | 0.97 | 0.26 | 0.35 | 0.10 | 1.51 | 1.55 | 0.06 |
| 6 | 0.13 | 0.63 | 0.22 | 0.24 | 0.18 | 0.93 | 1.24 | 0.01 |
| 7 | 0.13 | 0.41 | 0.19 | 0.18 | 0.34 | 0.89 | 0.57 | 0.26 |
| 8 | 0.13 | 0.42 | 0.17 | 0.21 | 0.43 | 1.20 | 0.49 | 0.18 |
| 9 | 0.06 | 0.23 | 0.18 | 0.10 | 0.62 | 0.32 | 0.37 | 0.05 |
| 10 | 0.06 | 0.14 | 0.15 | 0.06 | 0.20 | 0.12 | 0.16 | 0.08 |

6. DESCRIÇÃO ESPACIAL DOS DADOS

Esta etapa do trabalho tem como objetivo visualizar o modo como cada variável se dispersa no espaço, definindo-se as principais características de cada variável: anisotropias e descontinuidades. Através da visualização do arranjo espacial dos dados experimentais, se faz o planejamento dos parâmetros fundamentais para o cálculo dos variogramas: direções, classes de ângulos e distâncias.

Foram visualizados os mapas temáticos das 14 variáveis a fim de definir previamente o comportamento espacial de cada uma delas. De forma geral, tanto atração quanto produção de viagens possuem um padrão espacial bastante similar: valores aumentando da periferia para o centro com alguns pontos altos de valores de ambas variáveis nas regiões do ABCD Paulista e nas zonas de Carapicuíba, Jardim Veloso, Santo Antonio e Guarulhos.

Tais regiões não são centrais, no entanto apresentam taxas de empregos no setor de comércio e indústria muito altas, isto pode justificar valores altos tanto de atração e produção de viagens por modo de transporte motorizado ou motivo de trabalho no comércio, por exemplo. A Figura 4 apresenta o mapa temático da variável produção de viagens por transporte coletivo na RMSP.

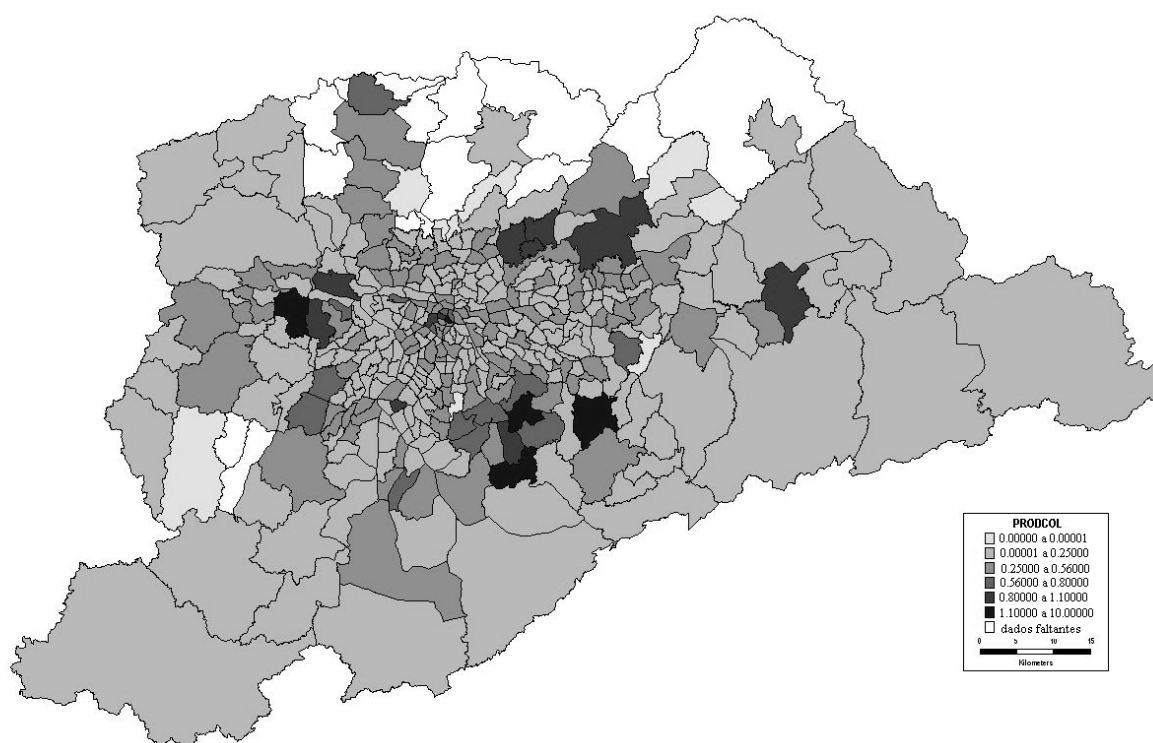


Figura 4 Mapa temático – viagens por transporte coletivo (produção)

7. CÁLCULO DOS VARIOGRAMAS EXPERIMENTAIS

O cálculo dos variogramas para as 14 variáveis foi efetuado segundo cinco direções: (0°, 45°, 90°, -45 e omnidireccional). No variograma omnidireccional é atribuído o mesmo peso a amostras que estão à mesma distância (h) do ponto, embora em diferentes direções. No variograma segundo determinada direção apenas se usam os pares de amostras alinhadas nestas direções estudadas.

O variograma experimental permite caracterizar o comportamento espacial das variáveis regionalizadas, identificando direções preferenciais no espaço (anisotropia), ou não (isotropia). Para o caso das variáveis analisadas no presente trabalho, observa-se que existe aparentemente comportamento espacial similar para todas as direções (isotropia). A partir de qualquer direção analisada as variáveis relativas à produção e atração de viagens aumentam de valor da periferia para o centro. Foi calculado um total de 70 variogramas.

A Figura 5 apresenta, por exemplo, os variogramas experimentais para as direções 45° e -45° da variável taxa de produção de viagens por automóvel. No eixo y é representada a semivariância enquanto que no eixo x as distâncias. Os pontos representam os dados e a reta representa a variância dos dados da população considerada (*Sill*). Em ambos os casos observa-se uma forma aproximada de parábola, ou seja, a dependência espacial cresce até certo ponto atingindo o máximo (periferia para o centro) e depois de certa distância diminui até que seja nula. Também são apresentados os variogramas omnidirecional e na direção 90°, onde se pode verificar uma tendência de ajuste para o modelo esférico.

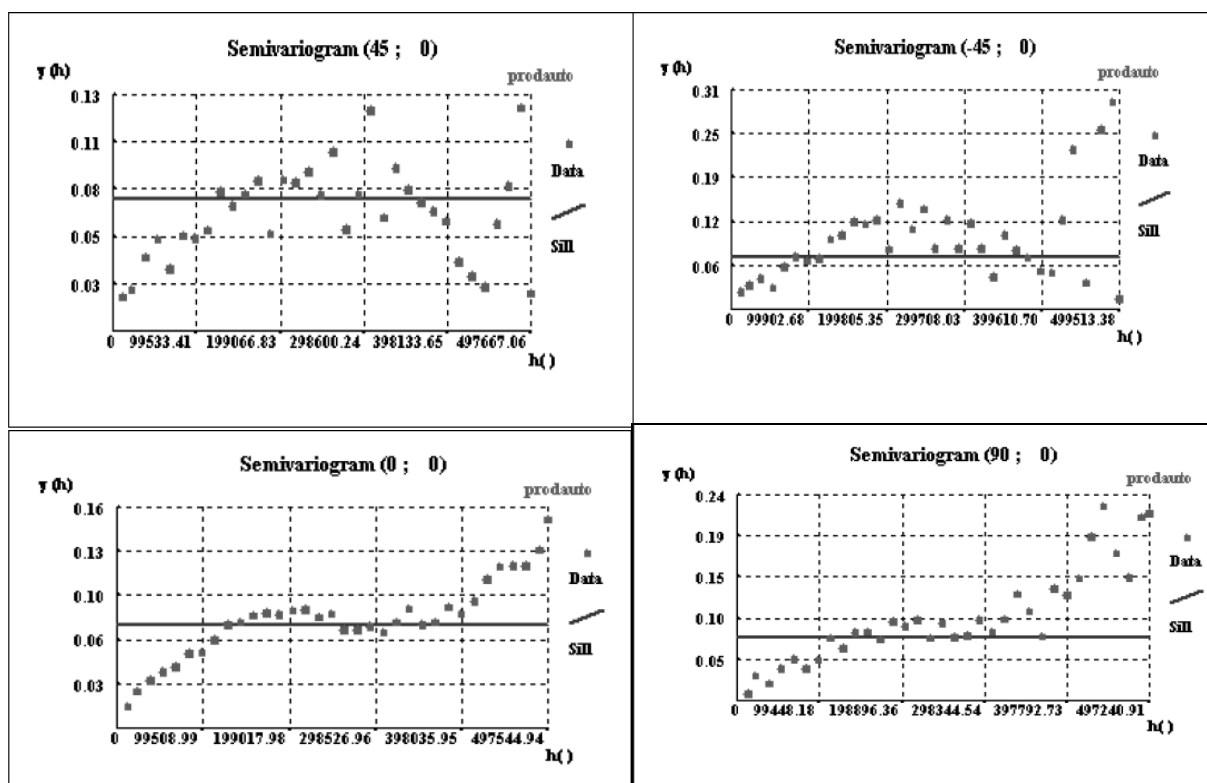


Figura 5 Variogramas experimentais (produção de viagens por automóveis) nas direções 45° - 45° omnidirecional (0) e 90°

8. MODELAGEM DOS VARIOGRAMAS TEÓRICOS

Depois de escolhidos os parâmetros das direções a estudar, é necessário transpor toda essa informação para uma função geral representativa. Sendo assim, é necessário o ajuste dos variogramas experimentais a uma função geral, sendo esta etapa extremamente importante no processo de estimação. Desta forma, para a krigagem, são considerados os parâmetros das curvas ajustadas dos variogramas.

Não sendo possível ajustar um modelo por qualquer tipo de função, existem funções interpoladoras que têm um número reduzido de parâmetros (sendo a sua escolha limitada às que fornecem soluções estáveis). Exemplos de tipos de modelos foram citados anteriormente na subseção 2.2.

Para as 14 variáveis, procedeu-se ao ajuste do variograma teórico (*do tipo esférico*), definindo-se a sua estrutura: *efeito de pepita ou nugget* (C_0), *patamar ou Sill* (C_1) e *amplitude ou range* (a).

Concluído o ajustamento do variograma teórico para todas as variáveis e direções, é possível selecionar qual a direção principal (a que corresponde uma maior continuidade espacial e a que foi ajustada com maior amplitude) e secundária (a que corresponde a menor amplitude). Para o caso das variáveis analisadas no trabalho, é possível afirmar que as mesmas têm comportamento isotrópico.

Para a variável *Produção de viagens por automóvel*, por exemplo, observam-se os seguintes valores dos parâmetros: • Variograma esférico; • Efeito pepita igual a zero $Co = 0$; • Patamar ($C1 = 0.072$); • Amplitude ($a = 159253.46$). A Figura 6 apresenta os variogramas modelados para a mesma variável nas direções 45° - 45° omnidirecional e 90° . A Tabela 2, em seguida, apresenta os parâmetros da modelagem para as demais variáveis.

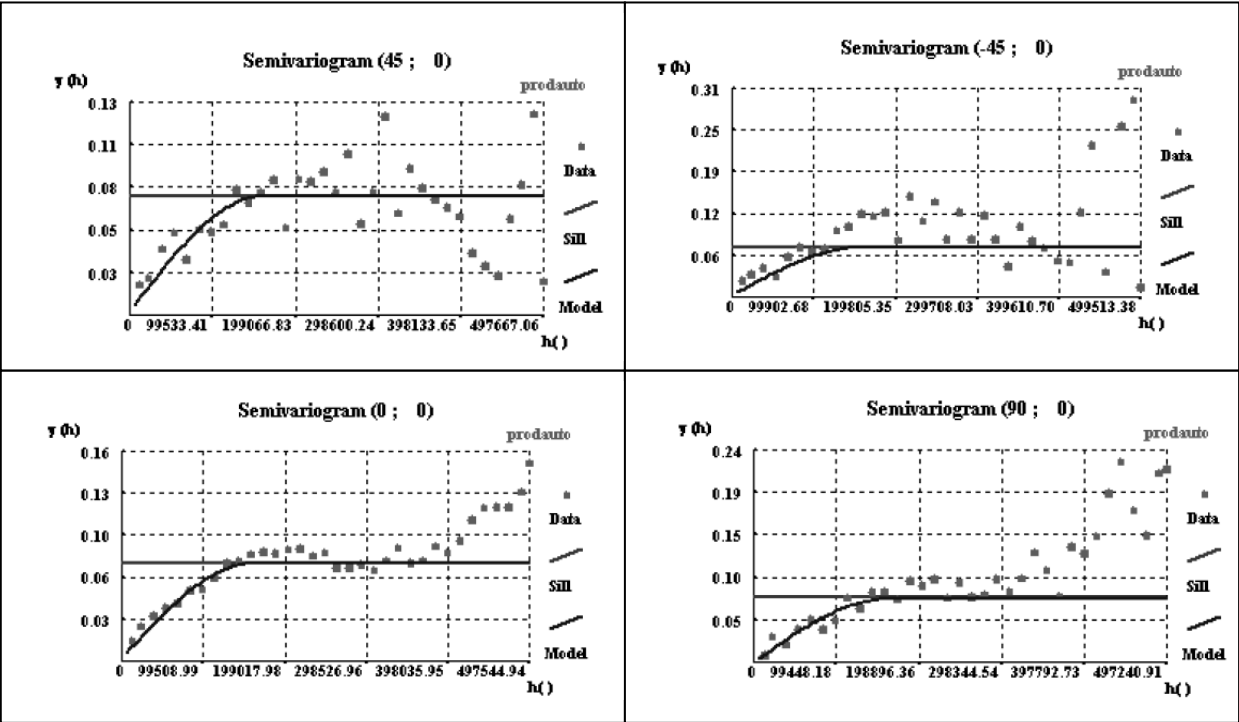


Figura 6 Variogramas modelados (produção de viagens por automóveis) nas direções 45° - 45° omnidirecional (0) e 90°

Tabela 2 Parâmetros dos variogramas modelados para as demais variáveis

| Variáveis | Comportamento | Tipo de curva | Efeito Pepita (Co) | Patamar ($C1$) | Amplitude(a) |
|--|---------------|---------------|------------------------|------------------|------------------|
| Produção de viagens por transporte coletivo | Isotrópico | Esférico | 0,065 | 0,162 | 139000 |
| Produção de viagens a pé | | | 0,066 | 0,15 | 141000 |
| Produção de viagens com motivo Trabalho na Indústria | | | 0 | 0,145 | 179000 |
| Produção de viagens com motivo Trabalho no comércio | | | 0,015 | 0,054 | 159000 |
| Produção de viagens com motivo Trabalho serviços | | | 0,034 | 0,055 | 134000 |
| Atração de viagens por automóvel | | | 0,015 | 0,06 | 159000 |
| Atração de viagens por transporte coletivo | | | 0,065 | 0,142 | 138000 |
| Atração de viagens a pé | | | 0,03 | 0,137 | 139000 |
| Atração de viagens com motivo Trabalho na Indústria | | | 0,073 | 0,018 | 99500 |
| Atração de viagens com motivo Trabalho no comércio | | | 0,022 | 0,065 | 158000 |
| Atração de viagens com motivo Trabalho serviços | | | 0,033 | 0,055 | 132000 |

9. ESTIMAÇÃO GEOESTATÍSTICA - INTERPOLAÇÃO POR KRIGAGEM ORDINÁRIA

A *krigagem* (Soares, 2000) é um interpolador exato e estocástico, baseado na taxa de alteração da variância dos pontos observados ao longo do espaço. São estimados os valores de um atributo (no caso do trabalho em questão das 14 variáveis relativas a viagens) em locais não amostrados a partir de dados amostrados na mesma área.

Para estimação geoestatística são considerados os parâmetros dos modelos teóricos dos variogramas e os parâmetros da malha de pontos ou blocos a estimar. Utilizou-se a interpolação por *Krigagem Ordinária* (Davis, 1987) em locais não amostrados, permitindo assim gerar mapas de distribuição espacial da variável analisada considerando a RMSP. A Tabela 3 mostra alguns valores encontrados para a variável produção de viagens por automóvel para 10 pontos. Foram encontrados um total de 40.000 valores de pontos conforme número de blocos adotados na malha (200x200). As duas primeiras colunas apresentam os valores das coordenadas (longitude e latitude respectivamente), a terceira coluna é referente à coordenada Z (no caso em questão igual a zero por se tratar de um problema bidimensional), a última coluna traz os valores estimados da variável nos pontos localizados nas respectivas coordenadas.

Tabela 3 Valores estimados em dez pontos

| | coordenadas (Y, X,Z) | | | valores |
|----------|----------------------|--------------|------|--------------|
| | X | Y | Z | interpolados |
| ponto 1 | -47032916.00 | -23956128.00 | 0.00 | 0.056133 |
| ponto2 | -47026932.00 | -23956128.00 | 0.00 | 0.053658 |
| ponto 3 | -47020952.00 | -23956128.00 | 0.00 | 0.052271 |
| ponto 4 | -47014968.00 | -23956128.00 | 0.00 | 0.054425 |
| ponto 5 | -47008988.00 | -23956128.00 | 0.00 | 0.056563 |
| ponto 6 | -47003004.00 | -23956128.00 | 0.00 | 0.058684 |
| ponto 7 | -46997020.00 | -23956128.00 | 0.00 | 0.060784 |
| ponto 8 | -46991040.00 | -23956128.00 | 0.00 | 0.062856 |
| ponto 9 | -46985056.00 | -23956128.00 | 0.00 | 0.061905 |
| ponto 10 | -46979076.00 | -23956128.00 | 0.00 | 0.063515 |

A Figura 7 ilustra a krigagem realizada na RMSP. Na figura são considerados valores da variável *Produção de viagens por automóvel* em alguns pontos (centróides de algumas zonas de tráfego) do município de São Paulo. Em itálico e acinzentados estão os valores estimados através da krigagem, em preto e negrito os valores originais nos centróides. Tal procedimento de análise espacial que segue uma sequência de etapas (análise espacial dos dados, modelagem dos variogramas e estimação geoestatística) permite estimar produção e atração de viagens em pontos diversos a partir dos valores dos centroides, levando-se em conta apenas a autocorrelação espacial das variáveis.

10. VALIDAÇÃO CRUZADA

A validação cruzada, ou teste do ponto fictício, é uma técnica que, para cada ponto, estima sequencialmente o valor da variável regionalizada, considerado desconhecido, calculando o erro de estimação real para cada ponto.

Assim é gerada uma tabela de resultados de pontos conhecidos (374 zonas de tráfego no total, pois foram excluídas aquelas cujos valores eram ausentes) com valores observados, estimados, coordenadas nas 374 zonas de tráfego e variância.

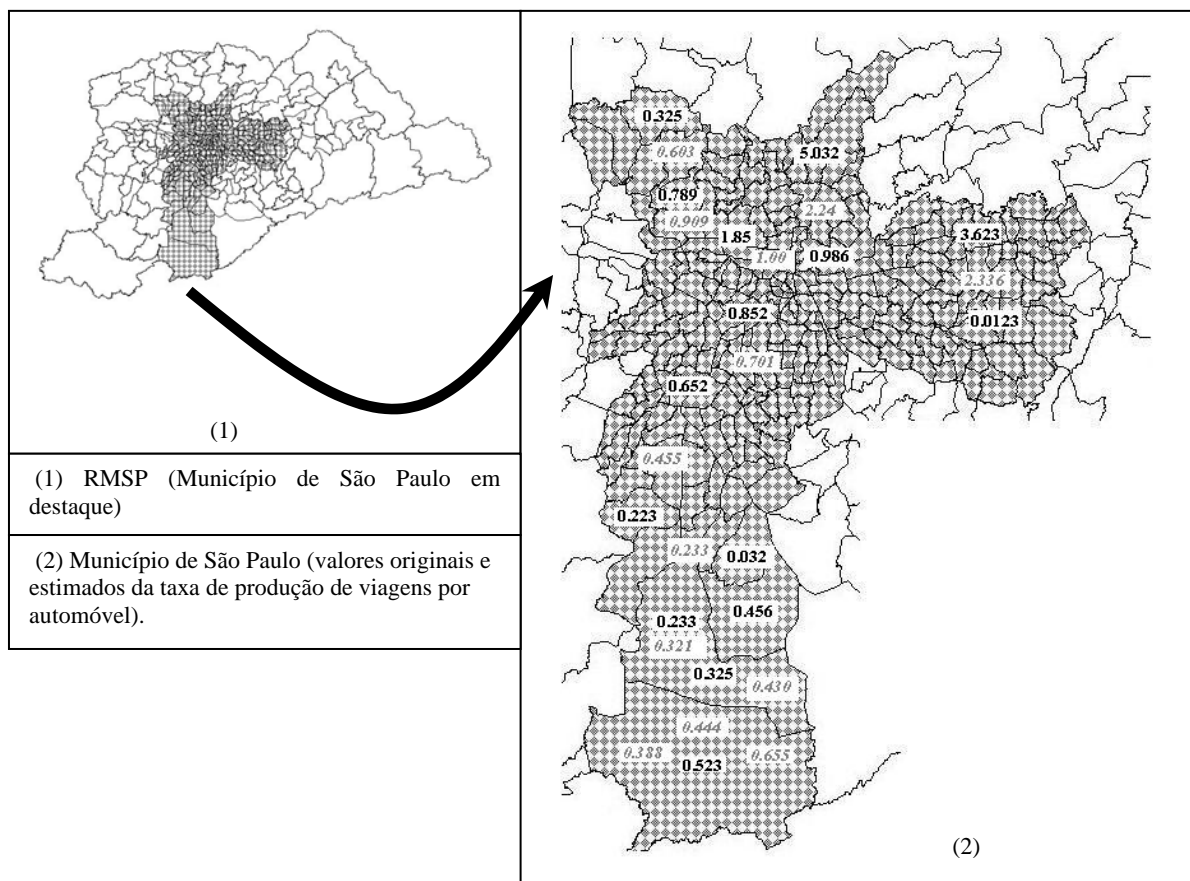


Figura 7 Krigagem ordinária na RMSP – exemplo de aplicação

A fim de averiguar a qualidade dos modelos foram calculados os seguintes parâmetros para a variável *Produção de viagens por automóvel*: coeficiente de correlação (0.47); média dos erros (0.0005); média do quadrado dos erros (0.06).

A Tabela 4 sintetiza os resultados da validação cruzada para todas as variáveis. Considerando os parâmetros adotados para mensurar a qualidade dos modelos, todos os modelos estimados foram considerados razoáveis, sendo krigagem uma ferramenta viável para os objetivos do trabalho, ou seja, estimação de taxas de viagens em diversas coordenadas geográficas.

Tabela 4 Resultados – Validação cruzada

| Variável | coeficiente de correlação | média dos erros | média dos quadrados dos erros | Variável | coeficiente de correlação | média dos erros | média dos quadrados dos erros |
|---|---------------------------|-----------------|-------------------------------|--|---------------------------|-----------------|-------------------------------|
| Produção de viagens por auto | 0.47 | 0.0005 | 0.06 | Atração de viagens por auto | 0.47 | 0.0003 | 0.059 |
| Produção de viagens a pé | 0.45 | 0.0002 | 0.07 | Atração de viagens a pé | 0.48 | 0.0004 | 0.21 |
| Produção de viagens por transporte coletivo | 0.38 | 0.0009 | 0.151 | Atração de viagens por transporte coletivo | 0.39 | 0.0005 | 0.08 |
| Produção de viagens trabalho na indústria | 0.37 | 0.0008 | 0.142 | Atração de viagens trabalho na indústria | 0.52 | -0.001 | 0.12 |
| Produção de viagens trabalho no comércio | 0.3 | 0.0038 | 0.065 | Atração de viagens no comércio | 0.48 | -0.008 | 0.07 |
| Produção de viagens trabalho no serviços | 0.39 | 0.0008 | 0.13 | Atração de viagens no setor de serviços | 0.4 | 0.0009 | 0.15 |

11. CONSIDERAÇÕES FINAIS E PROPOSTAS PARA TRABALHOS FUTUROS: GEOESTATÍSTICA + MÉTODOS TRADICIONAIS

Através do presente trabalho pode-se concluir que a geoestatística é uma ferramenta adequada a ser aplicada a estudos relativos ao Planejamento de Transportes, havendo ainda diversas possibilidades de utilização a depender dos objetivos a serem atingidos.

A partir do *software* GeoMS foram realizadas todos os procedimentos geoestatísticos. Na etapa de descrição espacial de dados, através de mapas temáticos pôde-se detectar padrões espaciais. As variáveis cresciam da periferia para o centro com alguns pontos altos em regiões ou zonas de tráfego específicas.

Na etapa de cálculo dos variogramas experimentais e definições dos parâmetros para modelagem dos variogramas teóricos pôde-se observar na maioria das vezes, variogramas similares e valores dos parâmetros definidos também muito parecidos entre as variáveis. Tal fato é completamente justificável considerando que todas as variáveis analisadas são correlacionadas entre si. Todos os casos foram isotrópos e definidos variogramas esféricos.

Através da estimação geoestatística através de Krigagem Ordinária foi possível estimar os valores das variáveis relativas a viagens em 40.000 pontos desconhecidos e, finalmente, através da validação cruzada mensurar a qualidade dos modelos obtidos. Tais modelos foram considerados razoáveis através dos parâmetros adotados.

Vale ressaltar que o verdadeiro intuito do presente trabalho é mostrar a viabilidade do uso de estimação geoestatística para análise de geração de viagens. No entanto, problemas de transportes consideram uma infinidade de fatores além da correlação espacial.

A partir dos resultados obtidos com a geoestatística univariada pode-se propor futuramente alguns trabalhos considerando uso de geoestatística multivariada principalmente no intuito de analisar simultaneamente modelos tradicionais de geração de viagens e modelos geoestatísticos. Pode-se propor futuramente, por exemplo, uma *krigagem com deriva externa* - *KDE* (método alternativo onde uma variável secundária é linearmente correlacionada com a variável de interesse).

Neste caso, pode-se fazer inicialmente uma regressão linear multivariada sendo produção de viagens a variável dependente e considerando como independentes variáveis como população, emprego, renda, etc. Para aplicação da KDE os valores da regressão seriam a variável secundária e produção de viagens seria a variável primária.

AGRADECIMENTO: A Fundação para Ciência e Tecnologia (Portugal), pelo apoio financeiro à pesquisa. Ao Metrô-SP pela cessão dos dados.

REFERÊNCIAS BIBLIOGRÁFICAS

- Davis, B. (1987). Uses and Abuses of Cross-Validation in Geostatistics. 1997, *Mathematical Geology*, vol. 19, nº 3, pp 249-258.
- Goovaerts, P. (2008). Geostatistical analysis of health data: state-of-the-art and perspectives. In: Soares A. et al. (eds) *geoENV VI - Geostatistics for Environmental Applications*. Springer-Verlag, Berlin, Germany, 3-22.
- Krige, D. G. (1951) A statistical approach to some basic mine evaluation problems on the witwatersrand. *J. Chem. Metall. Min. Soc. S. Afri., Johannesburg*, v. 52, p. 151-163.
- Loureiro, C.F.G.; Silva, H.N.; Carvalho, L.E.X. (2006) Metodologia de análise de regressão geograficamente ponderada aplicada ao fenômeno das viagens intermunicipais. *Panorama Nacional de Pesquisa em Transportes*, XX Anpet, v.1, p.479-491.
- Soares, A. (2000) – *Geoestatística para as Ciências da Terra e do Ambiente*. IST Press, 206p.

Cira Souza Pitombo cirapitombo@gmail.com

Antonio Jorge Sousa ajsousa@ist.utl.pt

CERENA-Centre for Natural Resources and the Environment, Instituto Superior Técnico.

Av. Rovisco Pais, 1049-001 Lisboa, PORTUGAL