Estatística Aplicada a Recursos Hídricos

Docente: Rachid Muleia

(rachid.muleia@uem.mz)

Mestrado em Gestão de Recursos Hídricos - DGEO/UEM

Tema: Introdução a Geoestatística

Ano lectivo: 2023

Introdução

- Geo-estatística lida com a caracterização de dados espaciais e espacio-temporal;
- É uma ciência que surge da necessidade de modelação de recursos geológicos-caracterização espacial da concentração de metais em jazigos;
- Primeiros desenvolvimentos foram dados na África do Sul por D.G. Krige, um engenheiro de minas, e H.S. Sichel, um estaticista, por volta de 1950;
- Posteriormente, Georges Matheron, um engenheiro francês, expandiu o trabalho iniciado por D.G. Krige, dando origem a modelos de krigagem;

Porquê geoestatística

- Maior parte dos fenómenos/grandezas variam ao longo do espaço, e a geo-estatística ajuda a perceber e quantificar a variabilidade espacial de um dado fenómeno
- Estatística clássica assume independência das observações, enquanto que dados geo-referenciados apresentam dependência espacial
- Observações mais próximas apresentam maior dependência espacial, e as mais afastadas apresentam menor dependência espacial
- Ignorar a dependência espacial pode condicionar a validade das análises estatística

Fenómeno espacial- conceitos básicos

Fenómeno espacial

- Conjunto de todos os valores possíveis da variável de interesse, que define a distribuição e variabilidade espaciais dessa variável dentro de um dado domínio em 2D ou 3D. Em termos estatísticos, representa a população, que é o conjunto de todos valores da qual uma amostra pode ser extraída

Variável regionalizada

Toda variável distribuída no espaço é tida como "regionalizada"ou "espacial". Por exemplo:

- Concentração de fósforo nas machambas do vale do Infulene
- Níveis de precipitação na cidade de Maputo

Uma variável regionalizada pode ser vista como uma função $f(S)$ que assume um dado valor em cada ponto S, num espaço propriamente definido.

Cont.

Função aleatória

- Seja $Z(\mathrm{s}_i)$ o valor observado de uma variável de interesse em uma posição s_i. Este valor pode ser tido como uma realização particular de uma variável aleatória $Z(\mathrm{s}_i)$ em um ponto s_i.

- O conjunto de variáveis aleatórias $\{Z(\mathrm{s}) : \mathrm{s} \in R\}$, onde R é uma região de interesse (ex: campo de plantação), é tido como uma função aleatória. Por exemplo: $\{Z(\mathrm{s}) : \mathrm{s} \in R\}$, onde:

  ➜ $Z(\mathrm{s}) =$ valor observado no ponto $\mathrm{s} = (x_1, x_2)$ ou $\mathrm{s} = (x_1, x_2, x_3)$

  ➜ $R =$ Conjunto de todos pontos sobre a área de interesse

Tipo de dados

Seja s $\in R$ uma localização genérica num espaço euclidiano de dimensão d e seja $\{Z(\mathrm{s}) : \mathrm{s} \in R\}$ função aleatória espacial, em que Z denota o atributo de interesse

Dados geo-estatísticos (dados de superfície)

Dados geo-estatísticos surgem quando o domínio em estudo é um conjunto fixo e contínuo:

- $Z(\mathrm{s})$ pode ser observado em qualquer ponto do domínio D
- O domínio D é não-estocástico

Cont.

Dados de área/regionais (lattice data)

Os dados de área surgem, quando o domínio/região em estudo é de natureza discreta, e $Z(\mathrm{s})$ pode ser observado em locais fixos que possam ser devidamente enumerados. Os locais podem ser: Províncias, distritos, regiões. Para este tipo de dados, os dados, normalmente, aparecem de forma agregada

Dados de ponto padrão

Diferentemente dos dados geoestatísticos e regionais, nos dados de ponto padrão a região de interesse não é fixa, mas sim aleatória. Este tipo de dado surge quando o interesse reside em estudar/analisar os locais onde os eventos de interesse ocorrem

Estacionariedade

- Em geo-estatística o processo inferencial depende da estacionariedade da variável regionalizada
- Pode-se pensar da estacionariedade como sinónimo de homogeneidade da variável regionalizada na região em estudo
- Assume-se a estiocionariedade, pois a variável regionalizada só pode uma assumir uma única realização (isto contradiz o conceito de uma variável aleatória)
- Comportamento regular dos momentos de uma função aleatória sobre uma região ou intervalo de tempo.

Cont.

Estacionariedade estrita

- Uma função aleatória é estacionária de forma estrita, se a família de v.a's $Z(s_1), Z(s_2), \ldots, Z(s_k)$ e $Z(s_\mathbf{1} + h), Z(s_\mathbf{2} + h), \ldots, Z(s_k + h)$ para $\forall$ k e h a distribuição conjunta de probabilidades é a mesma.

- A distribuição de conjunta de probabilidades de $\{Z(s_1), Z(s_2), \ldots, Z(s_k)\}$ não depende de qualquer que seja a translação de h.

- A hipótese de estacionariedade estrista é bastante rigorosa, podendo se relaxar usando a hipótese de estacionariedade de segunda ordem.

Exemplo- estacionariedade estrita

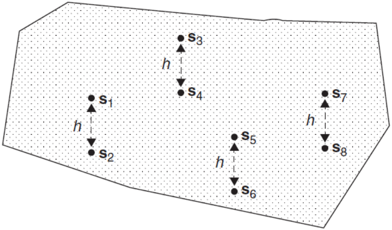


Figura: Quatro pares de pontos separados por uma distância h

No caso de estacionariedade estrita $\{Z(s_1), Z(s_2)\}$, $\{Z(s_3), Z(s_4)\}$ e $\{Z(s_5), Z(s_6)\}$ têm a mesma distribuição bivariada de probabilidades, pois a distância entre os pares $\{s_1, s_2\}$, $\{s_3, s_4\}$ e $\{s_5, s_6\}$ é a mesma

Estacionariedade de segunda ordem

- A hipótese de estacionariedade de segunda ordem exige apenas que os dois primeiros momentos da função aleatória estejam definidos.
  - ➜ $E(Z(s)) < \infty$ e $E(Z(s)) = \mu(s) = \mu$, e não de pende da posição s:
  - ➜ $C(Z(s), Z(s+h)) < \infty$,
  - ➜ $C(Z(s), Z(s+h)) = C(h), \forall \in R \quad \text{and} \quad h$
- A estacionariedade de segunda ordem pode ser interpretada como se a variável regionalizada assumisse valores que flutuam em volta de um valor constante (média), e a variação dessas flutuações fosse a mesma em todo o domínio

Cont.

- Se a covariância existe, e é finitia, então a variância está definida, e por sua vez é constance $V(Z(s)) = C(0)$,

- No caso de estacionariedade de segunda ordem, cumpre-se

$$
\begin{aligned}
\gamma(h) &= \frac{1}{2}\text{Var}\{Z(s+h) - Z(s)\} \\
&= \frac{1}{2}\Big\{\text{Var}[Z(s+h)] + \text{Var}[Z(s)] - 2C[Z(s+h), Z(s)]\Big\} \\
&= \frac{1}{2}C(0) + \frac{1}{2}C(0) - \frac{2}{2}C(h) \\
&= C(0) - C(h)
\end{aligned}
$$

Estacionariedade de intrínseca

- Uma função aleatória é intrinsecamente estacionária, se para qualquer translação h, as primeiras diferenças $Z(s+h) - Z(s)$ são estacionárias de segunda ordem.

  - ➜ $E\big(Z(s+h) - Z(s)\big) = 0$
  - ➜ $V\big(Z(s+h) - Z(s)\big) = 2\gamma(s)$

- A grandeza $2\gamma(s)$ é conhecida como variograma, e é um parâmetro de extrema importância em geo-estatística.

Variograma/Semivariograma

- Um dos principais atributos de dados espaciais é a autocorrelação espacial, isto é, observações mais próximas tendem a ser mais semelhantes do que observações mais distantes
- Na geoestatística, essa ideia de autocorrelação é quantificada por meio de uma função chamada semivariograma
- O semivariograma é uma função de um processo espacial com as seguintes propriedades :
  - $\gamma(-h) = \gamma(h)$ (a autocorrelação entre $Z(s)$ e $Z(u)$ é a mesma que a autocorrelação entre $Z(u)$ e $Z(s)$)
  - $\gamma(0) = 0$, visto que, $V(Z(s) - Z(s)) = 0$
  - $\gamma(h)/\|h\|^2 \to 0$, quando $\|h\| \to \infty$. $\|h\|$ representa distância euclidiana de um vector
  - $\gamma(\cdot)$ deve ser negativa definida, $\sum_{i=1}^{n} \sum_{j=1}^{m} a_i a_j \gamma(s_i - s_j) \leqslant 0$

Semivariograma experimental

- Na prática usa-se o semivariograma experimental para estudar a dependência espacial dos dados;
- Usa os dados observados da variável regionalizada para estimar variabilidade espacial do fenómeno em estudo;
- O estimador do semivariograma é dado por:

$$\hat{\gamma}(h) = \frac{1}{2\#N(h)} \sum_{N(h)} (Z(s_i + h) - Z(s_i))^2,$$

e é designado por estimador clássico. $\#N(h)$ representa o número de pares de observações que são separados por uma distância h;

Cont.

- A representação gráfica de $\hat{\gamma}(h)$ versus $\|h\|$ é designada por semivariograma experimental

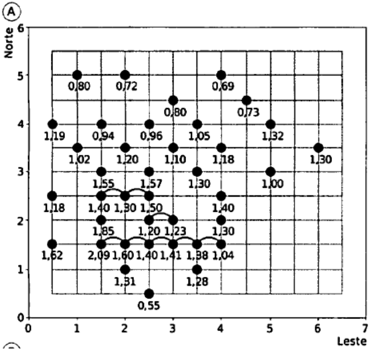
- O estimador do semivariograma $\hat{\gamma}(h)$ é um estimador não enviesado:

$$E(\hat{\gamma}(h)) = E\left(\frac{1}{2\#N(h)}\sum_{N(h)}\left(Z(\mathrm{s}_i + h) - Z(s_i)\right)^2\right) = \gamma(h)$$

Geralmente, na prática, o semivariograma é calculado para distâncias inferiores a metade do diâmetro do domínio de estudo. Isto porque, o número de pares diminui com a distância, e para distâncias maiores, o número de pares não é suficiente para produzir estimativas credíveis.

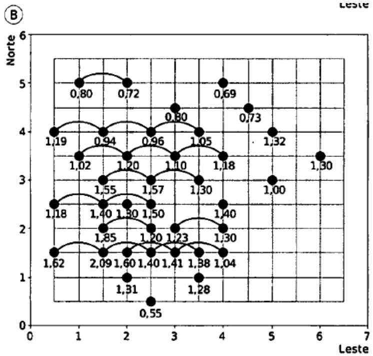
Exemplo- cálculo do semivariograma

Cálculo do semivariograma para $h = 0.5$



$$\gamma^*(0{,}5) = \frac{1}{2\times 8}\Big[(1{,}4-1{,}3)^2 + (1{,}3-1{,}5)^2 + (1{,}2-1{,}23)^2 + (2{,}09-1{,}6)^2 + (1{,}6-1{,}4)^2 + (1{,}4-1{,}41)^2 + (1{,}41-1{,}38)^2 + (1{,}38-1{,}04)^2\Big] = 0{,}028$$

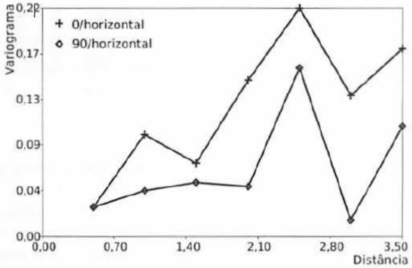
Cálculo do semivariograma para $h = 1.0$



$$\gamma^*(1{,}0) = \frac{1}{2\times 18}[(0{,}8-0{,}72)^2 + (1{,}19-0{,}94)^2 + (0{,}94-0{,}96)^2 + (0{,}96-1{,}05)^2 + (1{,}02-1{,}2)^2 + (1{,}2-1{,}1)^2 + (1{,}1-1{,}18)^2 + (1{,}55-1{,}57)^2 + (1{,}57-1{,}3)^2 + (1{,}18-1{,}4)^2 + (1{,}4-1{,}5)^2 + (1{,}85-1{,}2)^2 + (1{,}23-1{,}3)^2 + (1{,}62-2{,}09)^2 + (2{,}09-1{,}4)^2 + (1{,}6-1{,}41)^2 + (1{,}4-1{,}38)^2 + (1{,}41-1{,}04)^2] = 0{,}043$$

Cont.

Distância	Leste-oeste		Norte-sul	
	$\gamma(h)$	Np	$\gamma(h)$	Np
0,5	0,028	8	0,028	11
1,0	0,043	18	0,097	15
1,5	0,051	12	0,069	13
2,0	0,047	12	0,147	7
2,5	0,158	6	0,216	9
3,0	0,015	5	0,133	3
3,5	0,104	4	0,178	3



- A direção norte-sul apresenta maior variabilidade que a direção leste-oeste, significando que o comportamento é diferente conforme a direção pesquisada, o que indica, por sua vez, um fenómeno espacial anisotrópico.
- Os variogramas experimentais foram calculados até uma distância máxima igual a 3,5 m. A distância máxima em que se pode calcular o variograma experimental é chamada de campo geométrico e é igual à metade do diâmetro na direcção considerada

Características do semivariograma

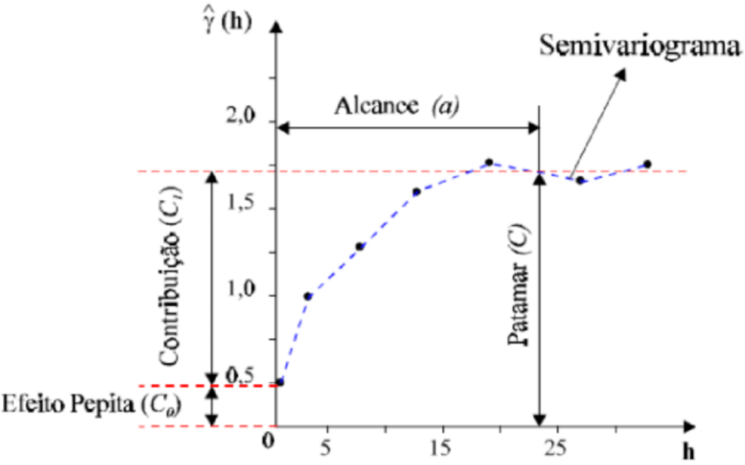


Figura: Parâmetros do semivariograma.

Cont.

- Efeito pepita- Valor do semivariograma para $h = 0$. Representa micro variações devido a erros de medição.
- Amplitude- Distância em que o variograma atinge o patamar, ou por outra, distância a partir da qual os dados não estão correlacionados.
- Patamar/soleira-Valor do semivariograma para uma distância igual a amplitude.

Estimação robusta do semivariograma

O estimador clássico do semivariograma apresenta algumas desvantagens:

- Extremamente sensível a valores atípicos;
- $(Z(s+h)-Z(s))^2 \sim 2\gamma(h)\chi_1^2$, consequentemente, $2\gamma(h)$ tem uma distribuição bastante assimétrica;

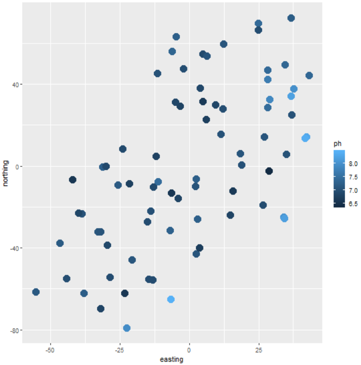
Como forma de remediar, Cressie and Hawkins (1980) propuseram :

$$2\bar{\gamma}(h) = \frac{1}{0.457 + 0.494/N(h)}\left\{\frac{1}{N(h)}\sum_{i=1}^{N(h)}\left[\left|Z(x_i+h)-Z(x_i)\right|\right]^{1/2}\right\}^4$$

Exemplo-DATA BREAK: Dados de pH da Smoky Mountain

- O pH nas águas do riacho pode afectar os organismos do riacho
- Mudança do pH pode ser um indicador de poluição
- $0 < \text{pH} < 14$, $\text{pH} < 7$-ácido, $\text{pH} > 7$-alcalino

	northing	easting	ph	elevation
1	-79.19	-22.53	7.91	0.244
2	-61.66	-55.39	7.14	0.375
3	-69.63	-31.93	6.81	0.567
4	-54.82	-44.38	6.97	0.512
5	-62.14	-37.94	7.21	0.408
6	-55.29	-14.60	6.94	0.512
7	-62.27	-23.30	6.59	0.506
8	-55.78	-13.17	6.94	0.488
9	-54.48	-28.70	6.94	0.488
10	-65.00	-6.62	8.43	0.232

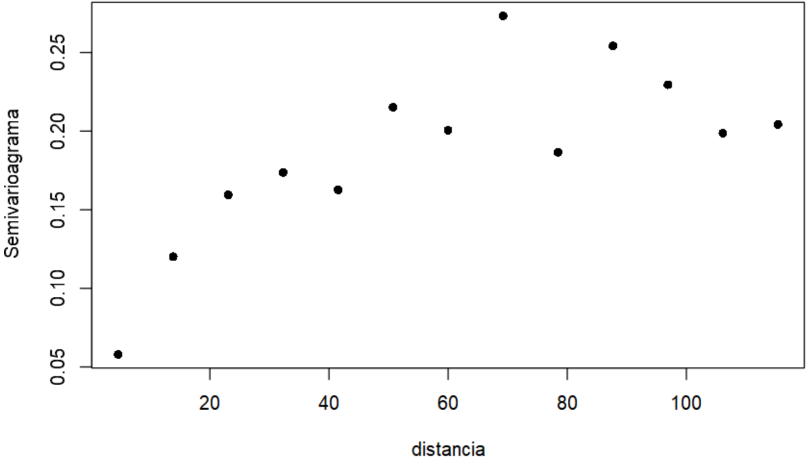


Exemplo - ilustração no R

```
> library(geoR)
--------------------------------------------------------------
 Analysis of Geostatistical Data
 For an Introduction to geoR go to http://www.leg.ufpr.br/geoR
 geoR version 1.9-2 (built on 2022-08-09) is now loaded
--------------------------------------------------------------

> path <- 'C:/Users/Rachid'
> ph_df <- read.table(paste(path, 'ph_data.txt',sep='/'),sep="",header=TRUE)
> ph_geo <- as.geodata(ph_df,coords.col=c(2,1),data.col=3)
> variogram_ph <- variog(ph_geo,max.dis=120)
variog: computing omnidirectional variogram
> variogram_ph
$u
 [1]   4.614145  13.842436  23.070726  32.299017  41.527308
 [6]  50.755598  59.983889  69.212179  78.440470  87.668760
[11]  96.897051 106.125341 115.353632

$v
 [1] 0.0579369 0.1201206 0.1594079 0.1737550 0.1626211
 [6] 0.2152035 0.2005385 0.2732983 0.1865881 0.2541853
[11] 0.2295052 0.1987260 0.2042285

$n
 [1]  42 189 258 339 311 288 286 258 193 174 134  98  79
```

Cont.

```
with(variogram_ph, plot(u, v, xlab ='distanica', ylab='Semivarioagrama', pch = 16))
```



Ajustamento do modelo do semivariograma

- Não ha garantia que o semivariograma experimental, $\hat{\gamma}(\cdot)$, seja definida negativa
- O não cumprimento da condição definida negativa não constitui nenhum problema se o interesse for apenas estudar a continuidade espacial. Contudo, caso se deseja fazer interpolações espaciais, a violação desta condição pode afectar consideralvemente a incerteza a volta das estimativas.
- Precisaremos encontrar um modelo de semivariograma teórico válido que tenha uma curvatura similar a do semivariograma empírico
- Existem vários modelos de semivariogramas. A nossa escolha irá se limitar nos modelos paramétricos (modelos com uma expressão matemática analiticamente definida).

Cont.

TAB. 2.6 Modelos teóricos de variogramas com patamar

Modelo	Equação
Esférico	$\begin{cases} \gamma(h) = C_0 + C\left[1{,}5\frac{h}{a} - 0{,}5\left(\frac{h}{a}\right)^3\right] \text{ para } h < a \\ \gamma(h) = C_0 + C \text{ para } h \geqslant a \end{cases}$
Exponencial	$\gamma(h) = C_0 + C\left[1 - \exp\left(-\frac{h}{a}\right)\right]$
Gaussiano	$\gamma(h) = C_0 + C\left[1 - \exp\left(-\left(\frac{h}{a}\right)^2\right)\right]$
Cúbico	$\begin{cases} \gamma(h) = C_0 + C\left[7\left(\frac{h}{a}\right)^2 - \frac{35}{4}\left(\frac{h}{a}\right)^3 + \frac{7}{2}\left(\frac{h}{a}\right)^5 - \frac{3}{4}\left(\frac{h}{a}\right)^7\right] \text{ para } h < a \\ \gamma(h) = C_0 + C \text{ para } h \geqslant a \end{cases}$
Pentaesférico	$\begin{cases} \gamma(h) = C_0 + C\left[\frac{15}{8}\left(\frac{h}{a}\right) - \frac{5}{4}\left(\frac{h}{a}\right)^3 + \frac{3}{8}\left(\frac{h}{a}\right)^5\right] \text{ para } h < a \\ \gamma(h) = C_0 + C \text{ para } h \geqslant a \end{cases}$
Efeito furo	$\gamma(h) = C_0 + C\left[1 - \frac{\operatorname{sen} \pi(h/a)}{\pi(h/a)}\right]$

Fonte: Olea (1999, p. 76-79).

Cont.

- Geralmente os semivariogramas, esférico, exponecial, e gaussiano, explicam a maior parte dos fenómenos espaciais

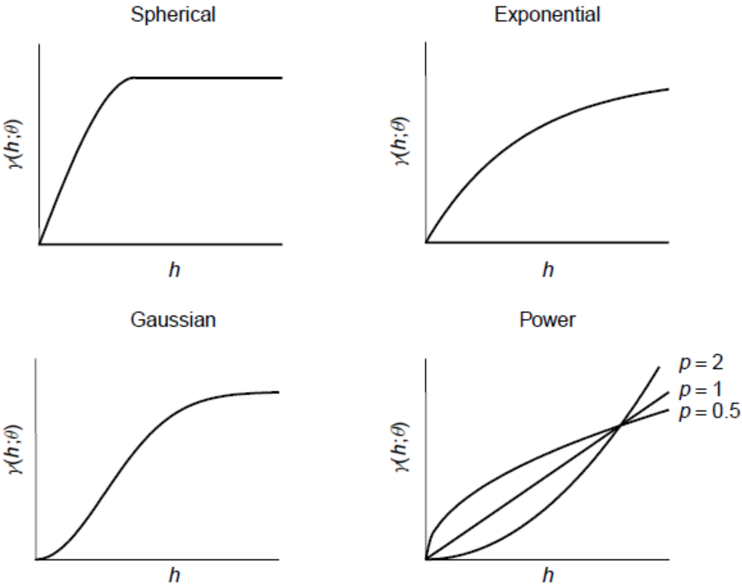
Semivariograma experimental

- A soleira é alcançada assimptoticamente quando $\|h\| \to \infty$
- Amplitude prática/efectiva $a' = 3a$ corresponde a distância na qual o valor do semivariogram é 95% da soleira, ou a distância para a qual a autocorrelação é igual a 0.05

Semivariograma gaussiano

- A soleira também é alcançada assimptoticamente
- Tem um comportamento parabólico na origem, indicativo de um processo espacial bastante suave

Cont.


Spherical
Exponential
Gaussian
Power
$\gamma(h;\theta)$
h
$p = 2$
$p = 1$
$p = 0.5$

Procedimento de selecção do semivariograma teórico.

- O ajustamento do variograma pode ser feito consoante o ajustamento manual, que consiste numa inspecção visual, ou usando procedimentos estatístico
- É recomendável que se use os dois métodos, manual e estatístico, onde primeiro se usa o método manual para seleccionar os modelos que melhor captam as características do semivariograma empírico. Depois, pode-se usar o procedemimento estatístico para seleccionar o melhor modelo dentre os pre-seleccionados
- Alguns autores afirmam que não é tão relevante se o ajustamento é feito de forma visual ou usando procedimentos estatístico. O que realmente importa é o tipo de continuidade e a hipótese de estacionariedade assumida.

Procedimento estatístico para selecção do modelo teórico

Existem vários procedimentos estatísticos que permitem estimar o conjunto de parâmentros (efeito pepita, patamar e a amplitude) dum semivariograma, dentre os quais, pode se destacar:

- Método dos mínimos quadrados
  - Método dos mínimos quadrados ordinários
  - Médodo dos mínimos quadrados ponderados
  - Método dos minimos quadrados generalizados
- Métodos baseados na maximização da função de verossimilhança
  - Método de máxima verossimilhança
  - Método de máxima verossimilhança restrita

Procedimento estatístico para selecção do modelo teórico

Mínimos quadrados ordinários- A estimação é feita minimizando a seguinte expressão:

$$\min \sum_h \left(\gamma(h) - \hat{\gamma}(h)\right)^2$$

Mínimos quadrados ponderados-Os quadrados dos resíduos são atribuídos um ponderador.

- Pode se usar número de pontos que distam h, isto é, $N(h)$

$$\min \sum_h N(h)\left(\gamma(h) - \hat{\gamma}(h)\right)^2$$

- MQP de Cressie tem como ponderador $\dfrac{N(h)}{\gamma(h)^2}$, logo, a estimação é feita usando:

$$\min \sum_h N(h)\left(\frac{\hat{\gamma}(h)}{\gamma(h)} - 1\right)^2$$

Exemplo - selecção do modelo teórico

```
#=======================Metodo dos minimos quadrados ordinario==========================
> # Modelo esferico
> fit1=variofit(variogram_ph,cov.model="sph",ini.cov.pars=c(0.23,60),
+ fix.nugget=FALSE,nugget=0,weights='equal',messages=FALSE)
> fit1
variofit: model parameters estimated by OLS (ordinary least squares):
covariance model is: spherical
parameter estimates:
  tausq sigmasq     phi
 0.0434  0.1786 60.3042
Practical Range with cor=0.05 for asymptotic range: 60.30417

variofit: minimised sum of squares = 0.0021
>
> # Modelo exponencial
> fit2=variofit(variogram_ph,cov.model="exponential",ini.cov.pars=c(0.23,60),
+ fix.nugget=FALSE,nugget=0,weights='equal',messages=FALSE)
> fit2
variofit: model parameters estimated by OLS (ordinary least squares):
covariance model is: exponential
parameter estimates:
  tausq sigmasq     phi
 0.0675  0.2077 59.9998
Practical Range with cor=0.05 for asymptotic range: 179.7434

variofit: minimised sum of squares = 0.0056
```

Cont.

```
#================== Metodo dos minimos quadrados ponderados ===========================
> # Modelo esferico
> fit3=variofit(variogram_ph,cov.model="sph",ini.cov.pars=c(0.23,100),
+ fix.nugget=FALSE,nugget=0,weights='npairs',messages=FALSE)
> fit3
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: spherical
parameter estimates:
  tausq sigmasq     phi
 0.0615  0.1622 65.9554
Practical Range with cor=0.05 for asymptotic range: 65.95537

variofit: minimised weighted sum of squares = 0.4801
>
> # Modelo exponencial
> fit4=variofit(variogram_ph,cov.model="exponential",ini.cov.pars=c(0.23,60),
+ fix.nugget=FALSE,nugget=0,weights='npairs',messages=FALSE)
> fit4
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: exponential
parameter estimates:
  tausq sigmasq     phi
 0.0279  0.2049 25.1745
Practical Range with cor=0.05 for asymptotic range: 75.4161

variofit: minimised weighted sum of squares = 0.607
```

Cont.

```
> # Modelo esferico
> fit31=variofit(variogram_ph,cov.model="sph",ini.cov.pars=c(0.23,100),
+ fix.nugget=TRUE,nugget=0,weights='npairs',messages=FALSE)
> fit31
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: spherical
fixed value for tausq =  0
parameter estimates:
sigmasq     phi
 0.2215 54.5730
Practical Range with cor=0.05 for asymptotic range: 54.57299

variofit: minimised weighted sum of squares = 0.9122
> #Modelo exponencial
> fit41=variofit(variogram_ph,cov.model="exponential",ini.cov.pars=c(0.23,60),
+ fix.nugget=TRUE,nugget=0,weights='npairs',messages=FALSE)
> fit41
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: exponential
fixed value for tausq =  0
parameter estimates:
sigmasq     phi
 0.2295 21.7987
Practical Range with cor=0.05 for asymptotic range: 65.30313

variofit: minimised weighted sum of squares = 0.6449
```

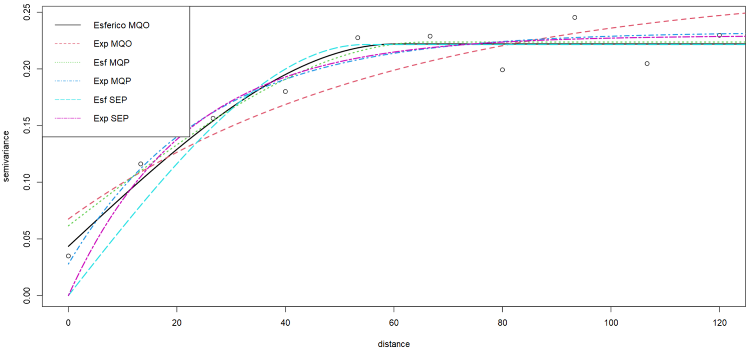
Cont.

Modelo	efeito pepita	soleira parcial	amplitude	SQR
		MQO		
Esferico	0.0434	0.1786	60.3042	0.0021
Exponencial	0.0675	0.2077	59.9998	0.0056
		MQP		
Esferico	0.0615	0.1622	65.9554	0.4801
Exponencial	0.0279	0.2049	25.1745	0.607
		Sem efeito pepita - MQP		
Esferico		0.2215	54.573	0.9122
Exponencial		0.2295	21.7987	0.6449

- O exercício do ajuste do modelo teórico aponta para o modelo esférico como sendo o melhor
- Os resultados apontam para um melhor ajuste quando se inclui o efeito pepita no modelo

Cont.

```
> var_fit=list(fit1,fit2,fit3,fit4,fit31,fit41)
> plot(variogram_ph)
> for(i in 1:6){
+ lines(var_fit[[i]],lty=i,col=i, lwd=2)
+ }
> legend('topleft',legend = c('Esferico MQO','Exp MQO', 'Esf MQP',
+ 'Exp MQP','Esf SEP','Exp SEP'), lty=1:6,col=1:6)
```



Cont.

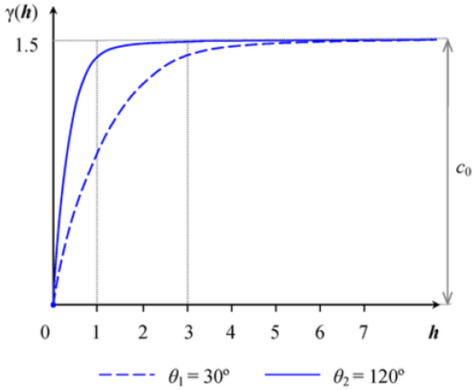
- As estimativas do semivariograma, isto é $\hat{\gamma}(h)$, estão correlacionadas e têm variâncias diferentes.
- MQO pressupões que as observaçoes sejam independentes com variância constante.
- Como alternativa, pode-se usar o método dos mínimos quadrados ponderados.
- Uma outra alternativa é o método dos mínimos quadrados generalizados (MQG).

Anisotropia

- Os modelos de semivariograma vistos até aqui são para fenómenos isotrópicos, isto é, que dependem apenas da separação (distância) entre as obseravações.
- Na prática, maior parte dos processos geoespaciais é de natureza anisotrópica, o que quer dizer que para além da distância, dependente também da direcção.
- A anisotropia pode ser identificada com base no semivariograma experimental, calculado para várias direcções.
- Ao detectar a presença de anisotropias, elas devem ser modeladas, ou seja, ajustadas a um modelo teórico de variograma. Na fase de modelagem, bem como na sua utilização para fins de estimativa e simulação, deve-se considerar a correção da anisotropia.

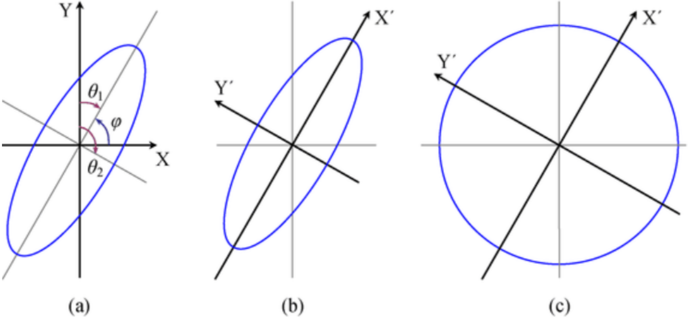
Anisotropia geométrica

- A anisotropia geométrica caracteriza-se pela existência de um único patamar e duas amplitudes diferentes



Cont

Correcção da anisotropia



Converter a anisotropia geométrica em um processo isotrópico

Cont

Para o caso, em que, os eixos da elipse não coincidem com os eixos do sistema cartesiano ortogonal, antes de calcular as distâncias transformadas, é preciso aplicar a rotação dos eixos, de tal maneira que estes coincidam.

$$\begin{pmatrix} \Delta x' \\ \Delta y' \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi/R & \cos\phi/R \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \tag{1}$$

,

onde $R = a_{min}/a_{max}$, $\Delta x = x_2 - x_1$ e $\Delta y = y_2 - y - 1$ e $\phi = 90^\circ - \theta_1$. As novas distâncias serão calculadas usando a fórmula:

$$\|h'\| = \sqrt{(\Delta x')^2 + (\Delta y')^2}$$

Os valores do semivariograma teórico são calculados com base nas distâncias isotropadas.

Cont

A fórmula apresentada no slide anterior é generalista. Todavia, para o caso onde as eixos da elipse coincidem com os eixos do sistema cartesiano ortogonal, as distâncias isotropadas podem ser calculados com base na seguinte fórmula:

$$\|h'\| = \sqrt{(\Delta x/a_x)^2 + (\Delta y/a_y)^2}$$

A amplitude do semivariograma nas várias direcções terá uma amplitude igual a 1

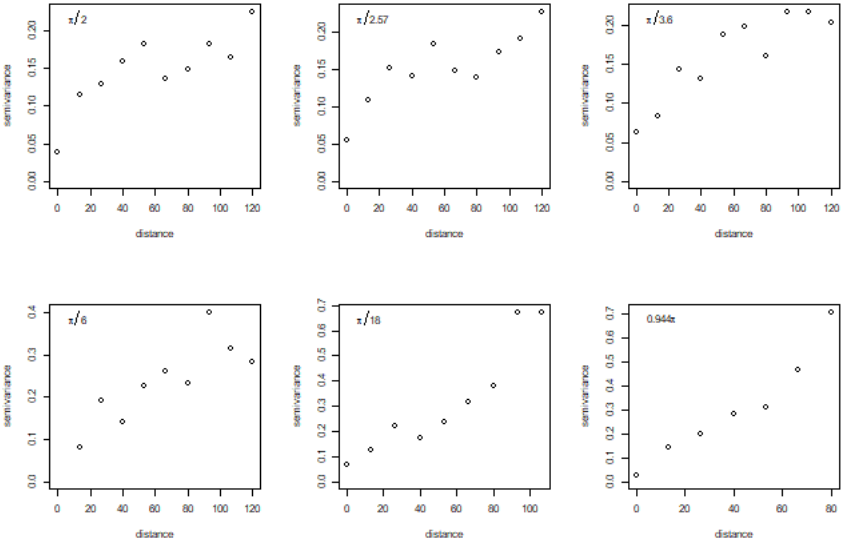
Cont

Se optarmos por transformar a anisotropia num semivariograma de referência (por exemplo, o de maior amplitude em vez do semivariograma com amplitude igual a 1), então a distância h "isotropizada"fica igual a :

$$\|h'\| = \sqrt{\Delta x \left(a_x/a_x\right)^2 + \Delta y \left(a_x/a_y\right)^2}$$

onde a_x é a amplitude se semivariograma de referência e $r_x = a_x/a_x$, $r_y = a_x/a_y$, são os factores de anisotropia nos 2 eixos principais.

Exemplo - pH do Smooky Mountain



```
> fit1=variofit(var1,cov.model="exponential",ini.cov.pars=c(0.23,60),
+ fix.nugget=FALSE,nugget=0.02,weights='npairs',messages=FALSE)
> fit2=variofit(var2,cov.model="exponential",ini.cov.pars=c(0.23,60),
+ fix.nugget=FALSE,nugget=0.02,weights='npairs',messages=FALSE)
> fit3=variofit(var3,cov.model="exponential",ini.cov.pars=c(0.23,60),
+ fix.nugget=FALSE,nugget=0.02,weights='npairs',messages=FALSE)
> fit4=variofit(var4,cov.model="exponential",ini.cov.pars=c(0.23,60),
+ fix.nugget=FALSE,nugget=0.02,weights='npairs',messages=FALSE)
> fit5=variofit(var5,cov.model="exponential",ini.cov.pars=c(0.23,60),
+ fix.nugget=FALSE,nugget=0.02,weights='npairs',messages=FALSE)
> fit6=variofit(var6,cov.model="exponential",ini.cov.pars=c(0.23,60),
+ fix.nugget=FALSE,nugget=0.02,weights='npairs',messages=FALSE)
> fit1; fit2
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: exponential
parameter estimates:
      tausq     sigmasq         phi
     0.0436    725.8673 139439.7379
Practical Range with cor=0.05 for asymptotic range: 417724.1

variofit: minimised weighted sum of squares = 0.9011
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: exponential
parameter estimates:
     tausq    sigmasq        phi
    0.0475   122.5758 31611.0551
Practical Range with cor=0.05 for asymptotic range: 94698.26

variofit: minimised weighted sum of squares = 1.0231
```

```
> fit3
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: exponential
parameter estimates:
  tausq sigmasq     phi
 0.0300  0.2817 58.4009
Practical Range with cor=0.05 for asymptotic range: 174.9536

variofit: minimised weighted sum of squares = 1.1912
> fit4
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: exponential
parameter estimates:
  tausq sigmasq     phi
 0.0667  0.1424 60.0115
Practical Range with cor=0.05 for asymptotic range: 179.7783

variofit: minimised weighted sum of squares = 0.2707
> fit5
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: exponential
parameter estimates:
  tausq sigmasq     phi
 0.0854  0.0980 60.0260
Practical Range with cor=0.05 for asymptotic range: 179.8217

variofit: minimised weighted sum of squares = 0.6764
> fit6
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: exponential
parameter estimates:
     tausq    sigmasq        phi
    0.0921    16.7275 10045.3723
```

Cont.

- Os semivariogramas experimentais apresentados na página 44 sugerem que o processo é anisotrópico;
- O modelo exponencial apresenta melhor ajuste na direcção $\pi/2$;
- Os semivariogramas experimentais nas direcções $\pi/6$, $\pi/18$ e 0.944π, apresentam uma tendência, isto é, a medida que a distância aumenta o semivariograma aumenta;
- Quando o processo apresenta uma tendência, antes de calcular o semivariograma é preciso remover a tendência;
- Sendo o processo anisotrópico, o ideal é ajustar um semivariograma anisotrópico;

Remoção da tendência (DRIFT)

- A presença notável da tendência faz com que o pressuposto da estacionariedade seja violado;
- A tendência torna as estimativas do semivariograma viciadas. Consequentemente, pode afectar negativamente o exercício de interpolação espacial.

Como identificar??

- A tendência pode ser identificada a partir do semivariograma experimental, através de um comportamento crescente acima da soleira.

Cont.

- A forma funcional da tendência, geralmente, é desconhecida;
- Se a tendência for linear, ajusta-se um modelo linear as dados, e extrai-se os resíduos. Ex:

$$Z(x, y) = \beta_0 + \beta_1 x + \beta_2 y$$

- A tendência pode, também, ser identificada a partir do diagrama de dispersão;

Exemplo smooky Mountain - remoção da tendência

```
> trend=lm(ph~easting+northing, data=ph_df)
> summary(trend)

Call:
lm(formula = ph ~ easting + northing, data = ph_df)

Residuals:
     Min       1Q   Median       3Q      Max
-1.03489 -0.23082 -0.05789  0.20716  1.16752

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.138924   0.045898 155.539  < 2e-16 ***
easting      0.009396   0.002217   4.239 6.57e-05 ***
northing    -0.002858   0.001462  -1.954   0.0546 .
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Residual standard error: 0.3975 on 72 degrees of freedom
Multiple R-squared:  0.2037,Adjusted R-squared:  0.1816
F-statistic: 9.209 on 2 and 72 DF,  p-value: 0.0002746

res=trend$residuals
res_df=cbind(ph_df,res)
res_geo=as.geodata(res_df,coords.col=c(2,1),data.col=5)
```

Cont.

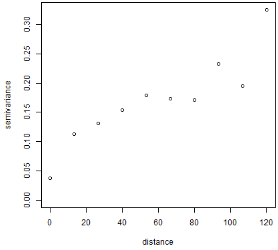


Figura: variograma experimental usando os resídous

```
> res_variog=variog(res_geo,uvec = seq(0,120,l=10),messages = FALSE, direction=0.944*pi)
> res_fit=variofit(res_variog,cov.model="exponential",ini.cov.pars=c(0.23,60),
+ fix.nugget=FALSE,nugget=0.02,weights='npairs',messages=FALSE)
> res_fit
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: exponential
parameter estimates:
   tausq  sigmasq      phi
  0.0906   0.4164 263.4559
Practical Range with cor=0.05 for asymptotic range: 789.2432

variofit: minimised weighted sum of squares = 0.2248
```

Cont.

```
> var_trend=variog(ph_geo,uvec = seq(0,120,l=10),messages = FALSE, direction=0.944*pi,trend ="1s
> fit_trend=variofit(var_trend,cov.model="exponential",ini.cov.pars=c(0.23,60),
+ fix.nugget=FALSE,nugget=0.02,weights='npairs',messages=FALSE)
> fit_trend
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: exponential
parameter estimates:
   tausq  sigmasq      phi
  0.0906   0.4164 263.4559
Practical Range with cor=0.05 for asymptotic range: 789.2432

variofit: minimised weighted sum of squares = 0.2248
>
> var_trend.quadra=variog(ph_geo,uvec = seq(0,120,l=10),messages = FALSE, direction=0.944*pi,tre
> fit_trend.quadra=variofit(var_trend.quadra,cov.model="exponential",ini.cov.pars=c(0.23,60),
+ fix.nugget=FALSE,nugget=0.02,weights='npairs',messages=FALSE)
> fit_trend.quadra
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: exponential
parameter estimates:
  tausq sigmasq     phi
 0.0958  0.0560 60.0282
Practical Range with cor=0.05 for asymptotic range: 179.8283

variofit: minimised weighted sum of squares = 0.4085
```

Anisotropia Zonal

Ocorre quando, para várias direcções, a soleira é diferente, mas apresenta o mesmo alcance/ amplitude. Geralmente, este tipo de anisotropia é difícil de encontrar na prática. O que tem acontecido, é encontrar uma combição da anisotropia geométrica e zonal.