

2 - Geoestatística: Aspectos básicos de modelos para dados com referência pontual (Parte 1).

Prof. Vinícius D. Mayrink

EST171 - Estatística Espacial

Sala: 4073

Email: vdm@est.ufmg.br

1º semestre de 2024

Apresentamos aqui elementos essenciais de modelos espaciais para uma análise clássica de dados com referência pontual. O tópico abordado nesta seção é tradicionalmente conhecido como “Geoestatística”.

O conceito fundamental da teoria é um processo estocástico: $\{Y(\mathbf{s}); \mathbf{s} \in D\}$ sendo $D \subset \mathbb{R}^r$ um subconjunto fixo.

Séries temporais seguem esta configuração com $r = 1$.

No caso espacial teremos tipicamente:

- $r = 2$ (ex.: longitude e latitude);
- $r = 3$ (ex.: longitude, latitude e altitude acima do nível do mar).

Para $r > 1$, dizemos que temos um “*processo espacial*”.

Exemplo: $Y(\mathbf{s})$ poderia representar o nível de poluição no local \mathbf{s} . Os dados serão formados por realizações deste processo espacial consistindo em medidas obtidas em um conjunto finito de locais $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ onde há estações de monitoramento.

Exemplo: Locais de captura de *Scallops* (tipo de molusco) na costa de NY/NJ (EUA) em 1990. A variável aleatória $Y(s_i)$ representa o número de capturas registradas no local s_i (*script R*).

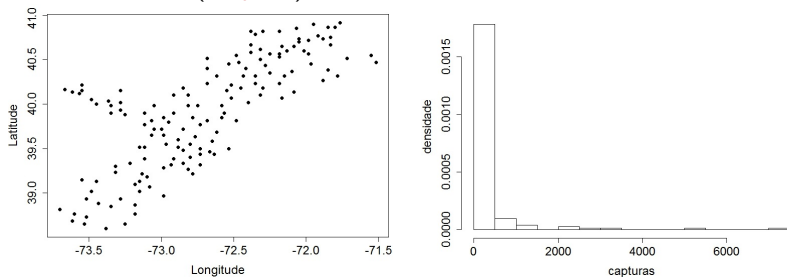


Fig.1 : Disposição espacial e histograma dos dados.

Sumário do número de capturas:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	8.0	30.0	274.6	115.2	7084.0

Gráfico de imagem com contornos e histograma para os dados $\log(n^\circ \text{ capturas} + 1)$ de cada local.

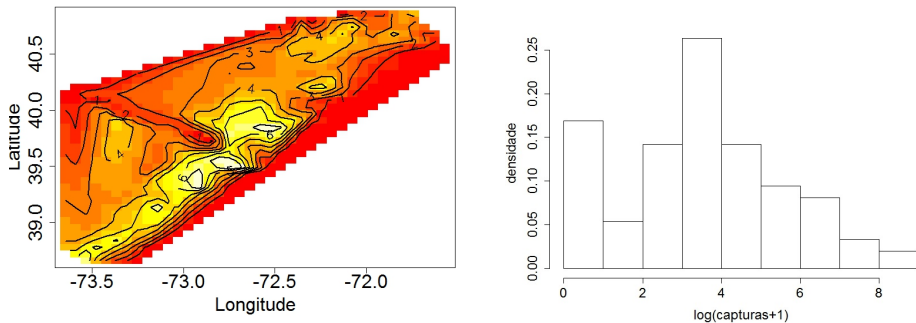


Fig.2 : Gráfico de imagem com contornos e histograma dos dados transformados.

Sumário dos dados de capturas transformados:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	2.197	3.434	3.483	4.756	8.866

Exemplo: Dados de elevação apresentados em Davis (1972); (*script R*).

$Y(s_i)$ representa a elevação da superfície no local s_i .

52 locais em uma região quadrada. Unidade de medida 10 *feet* $\approx 3.05m$.

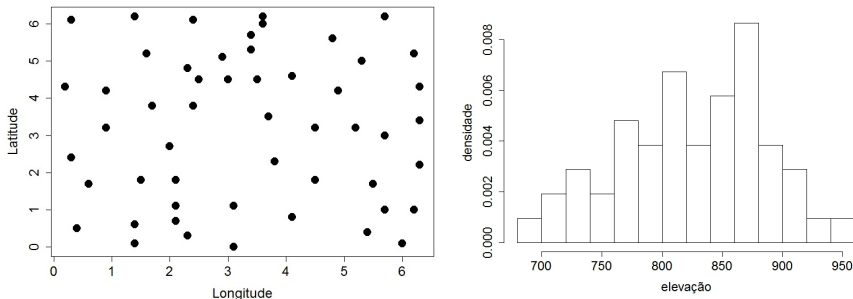


Fig.3 : Disposição espacial e histograma dos dados.

Sumário dos dados de elevação:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
690.0	787.5	830.0	827.1	873.0	960.0

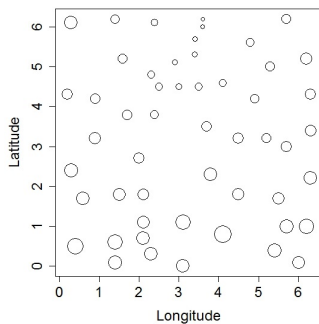
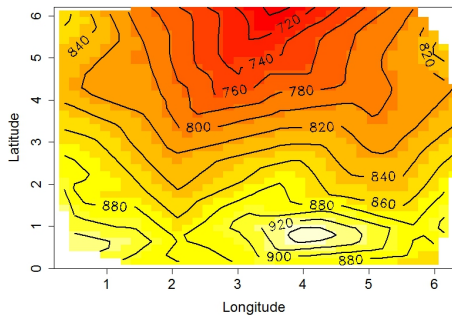


Fig.4 : Gráfico de imagem com contornos e *circle-plot* para os dados de elevação.

Exemplo: Contaminação residual por testes de armas nucleares (*script R*).

Dados coletados na ilha Rongelap (Arquipélago de Marshall, Pacífico); ver Diggle et al. (1998).

Para o sítio \mathbf{s}_i , $i = 1, \dots, 157$, $Y(\mathbf{s}_i)$ é a razão x_i/t_i sendo x_i a contagem de emissões de fótons pela radiação do Césio e t_i o tempo em segundos no qual a contagem x_i foi realizada.

Diferente dos dados de elevação, os $Y(\mathbf{s}_i)$ aqui estão sujeitos a um erro estatístico considerável.

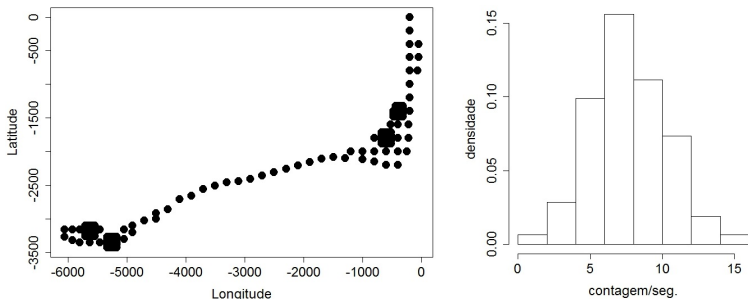


Fig.5 : Disposição espacial e histograma dos dados.

Sumário das contagens/seg.:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.250	5.890	7.475	7.604	9.363	15.100

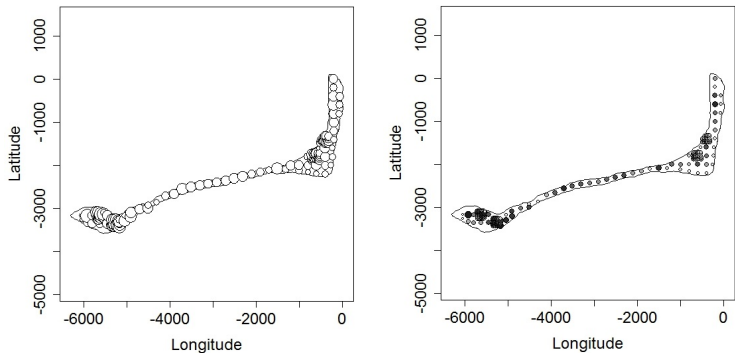


Fig.6 : Circle-plot para os dados de contaminação nuclear.

Exemplo: Dados de amostras do solo (*Soil data*, *script R*).

Amostras obtidas a uma profundidade de 0-20cm em $i = 1, \dots, 178$ locais dispostos em uma configuração de lattice regular incompleta com espaçamento de 50m entre os pontos.

Os dados foram coletados por pesquisadores da PESAGRO e EMBRAPA-Solos RJ/Brasil (Capeche, 1997).

Variáveis respostas (em $\text{mmol}_c/\text{dm}^3$):

$Y_1(\mathbf{s}_i)$ = quantidade de cálcio e $Y_2(\mathbf{s}_i)$ = quantidade de magnésio .

A região de estudo foi particionada em sub-regiões que passam por diferentes tratamentos:

- Sub-região 1: Tipicamente inundada durante períodos de chuva. Ela não é usada como área experimental devido a sua variação de elevação. Aqui o padrão de cálcio e magnésio representariam uma variação espacial natural.
- Sub-região 2: Recebeu fertilizante no passado e é usada para plantação de arroz. Entre as três partições, foi aquela que mais recentemente recebeu cálcio para neutralizar o efeito do alumínio; por isso observa-se maiores níveis de cálcio em suas amostras.
- Sub-região 3: Recebeu fertilizante no passado e é usada como área experimental.

Temos duas covariáveis:

$X_1(\mathbf{s}_i)$ = elevação do terreno e $X_2(\mathbf{s}_i)$ = indicadora da sub-região.

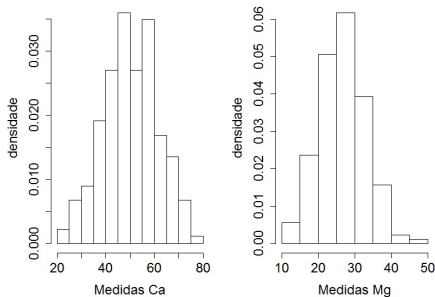


Fig.7 : Histogramas das medidas de Cálcio e Magnésio.

Sumário da resposta “Quantidade de Cálcio”:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
21.00	43.00	50.50	50.68	58.00	78.00

Sumário da resposta “Quantidade de Magnésio”:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
11.00	23.00	27.00	27.34	32.00	46.00

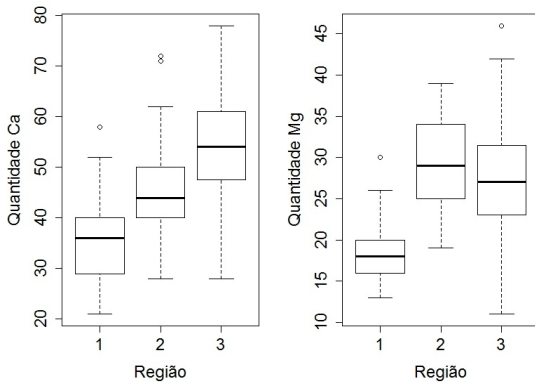


Fig.8 : Box-plots para os dados de solo: comparando as regiões.

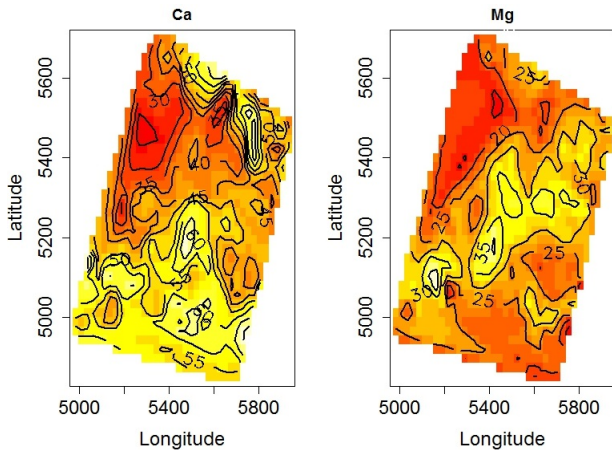


Fig.9 : Gráficos de imagem com contornos para os dados de solo.

Processo Gaussiano

Processo estocástico bastante usado para modelar dados com dependência temporal, espacial ou espaço-temporal.

Propriedades:

- 1 Um processo Gaussiano é completamente determinado pela especificação de suas funções média e de covariâncias (primeiro e segundo momentos do processo).
- 2 O melhor preditor de um Processo Gaussiano em um local não observado é uma função linear dos valores observados (em muitos casos, essas funções lineares podem ser calculadas facilmente usando fórmulas recursivas).

Um processo estocástico $\{Y(s), s \in D\}$ é um Processo Gaussiano se para qualquer escolha de índices distintos $s_1, \dots, s_m \in D$, o vetor aleatório $\mathbf{Y} = [Y(s_1), \dots, Y(s_m)]'$ tem distribuição Normal Multivariada $N_m[\boldsymbol{\mu}, \boldsymbol{\Sigma}]$ com $E(\mathbf{Y}) = \boldsymbol{\mu} = [\mu(s_1), \dots, \mu(s_m)]'$ e matriz de covariâncias $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{Y}, \mathbf{Y})$.

$$f(\mathbf{Y}) = (2\pi)^{-m/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{Y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right\}.$$

Para simplificar a ideia, restringiremos nossa atenção à situação temporal (dimensão $r = 1$). Assim, no caso contínuo temos usualmente $s \in [0, \infty)$ e no caso discreto $s \in \{0, 1, 2, \dots\}$.

As funções média e de covariâncias do Processo Gaussiano são definidas como segue: $\mu(s) = E[Y(s)]$ e $C(s_i, s_j) = \text{Cov}[Y(s_i), Y(s_j)]$.

Uma série temporal Gaussiana $\{Y(s)\}$ é dita fracamente estacionária se:

- 1 $E[Y(s)] = \mu$ (média independente de s);
- 2 $\text{Cov}[Y(s+h), Y(s)] = C(s+h, s) \equiv C(h)$ (covariância independente de s , $\forall h$).

Em um Processo Gaussiano estacionário $\{Y(s)\}$ temos:

- 1 $Y(s) \sim N[\mu, C(0)]$ para todo s ;
- 2 $[Y(s+h), Y(s)]' \sim N_2 \left[\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} C(0) & C(h) \\ C(h) & C(0) \end{pmatrix} \right]$

A ideia acima pode ser estendida para vetores maiores.

A função de covariâncias é um elemento crucial na modelagem visto que incorpora nossas suposições sobre a superfície (função) que desejamos aprender.

Uma escolha, bastante popular na literatura, é a função de covariâncias Gaussiana (ou exponencial quadrática): $C(s_i, s_j) = \sigma^2 \exp \left\{ -\phi^2 \|s_i - s_j\|^2 \right\}$.

Em particular, assuma $\sigma^2 = 1$ (teremos função de correlação) e $\phi^2 = 1/(2\kappa^2)$ e reescreva: $C(s_i, s_j) = \exp \left\{ -\frac{1}{2\kappa^2} \|s_i - s_j\|^2 \right\}$, sendo $(s_i, s_j) \in D$, $\kappa =$ parâmetro de comprimento-escala característico e $\|\mathbf{s}\|$ a norma Euclidiana do vetor \mathbf{s} .

Se os pontos s_i e s_j são muito próximos no espaço \mathbb{R}^r , então $C(s_i, s_j) \approx 1$. Quanto maior a distância entre estes pontos, menor será a similaridade entre as observações $Y(s_i)$ e $Y(s_j)$; teremos $C(s_i, s_j) \approx 0$.

O parâmetro de comprimento-escala κ controla o quão próximos os pontos s_i e s_j devem ser para que sejam considerados associados.

A função exponencial quadrática é estacionária e isotrópica (definição adiante). Além disso, ela é infinitamente diferenciável, significando que um Processo Gaussiano com esta configuração apresentará derivadas quadrado médio (*mean square derivatives*) de todas as ordens, sendo então um processo suave.

A especificação da estrutura de covariância de um processo espacial $\{Y(s)\}$ afeta diretamente a suavidade da superfície que gera o processo.

A descrição matemática da suavidade de uma superfície é feita por: continuidade e diferenciabilidade.

Entretanto, para superfícies geradas estocasticamente precisamos definir um tipo especial de continuidade e diferenciabilidade. A seguir considere, por conveniência de notação, um espaço unidimensional $r = 1$.

Continuidade quadrado médio: Um processo estocástico $\{Y(s)\}$ é quadrado médio contínuo se $E[\{Y(s+h) - Y(s)\}^2] \rightarrow 0$ conforme $h \rightarrow 0$

Diferenciabilidade quadrado médio: Um processo estocástico $\{Y(s)\}$ é quadrado médio derivável se

$$E \left[\left\{ \frac{Y(s+h) - Y(s)}{h} - Y'(s) \right\}^2 \right] \rightarrow 0, \text{ conforme } h \rightarrow 0.$$

sendo $Y'(s) = \lim_{h \rightarrow 0} \frac{Y(s+h) - Y(s)}{h}$ a derivada quadrado médio.

Diferenciabilidade quadrado médio de ordem maior é definida sequencialmente a partir daí; $Y(s)$ é duas vezes quadrado médio diferenciável se $Y'(s)$ é quadrado médio diferenciável, e assim por diante.

A especificação da função de covariâncias implica que um Processo Gaussiano estabelece uma distribuição sobre funções; nossa variável aleatória é a função desconhecida.

Exemplo - previsão de observações sem erro de medida: Considere os seguintes vetores aleatórios \mathbf{Y} (contemplando n valores observados) e \mathbf{Y}_* (representando n_* valores não observados a serem preditos) relacionados a uma superfície (função) que desejamos estudar. Especificação *a priori* para a superfície:

$$(\mathbf{Y}, \mathbf{Y}_*)' \sim N_{n+n_*} \left[\mathbf{0}, \begin{pmatrix} C(\mathbf{Y}, \mathbf{Y}) & C(\mathbf{Y}, \mathbf{Y}_*) \\ C(\mathbf{Y}_*, \mathbf{Y}) & C(\mathbf{Y}_*, \mathbf{Y}_*) \end{pmatrix} \right].$$

Note que $C(\mathbf{Y}, \mathbf{Y}_*)$ denota uma matriz $(n \times n_*)$ de covariâncias.

Distribuição *a posteriori* da superfície; precisamos restringir esta distribuição *a priori* para que ela englobe apenas as superfícies que passam pelos pontos observados.

Felizmente, em termos probabilísticos esta operação é extremamente simples, correspondendo a condicionar a distribuição *a priori* do Processo Gaussiano nos dados observados.

$$(\mathbf{Y}_* | \mathbf{Y}) \sim N_{n_*} [C(\mathbf{Y}_*, \mathbf{Y})C(\mathbf{Y}, \mathbf{Y})^{-1}\mathbf{Y}, C(\mathbf{Y}_*, \mathbf{Y}_*) - C(\mathbf{Y}_*, \mathbf{Y})C(\mathbf{Y}, \mathbf{Y})^{-1}C(\mathbf{Y}, \mathbf{Y}_*)].$$

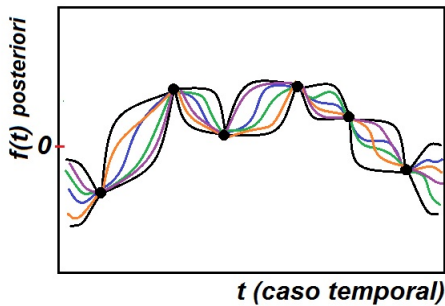
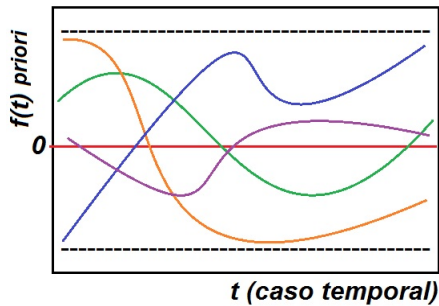


Fig.11 : Comparação das informações a priori e a posteriori.

A variação estocástica em uma quantidade física nem sempre é bem descrita pela distribuição Normal. Algumas vezes, precisamos aplicar uma transformação aos dados originais para que a suposição de Normalidade seja utilizada. Para variáveis resposta positivas, uma classe útil de transformações é dada pela família Box-Cox (Box e Cox, 1964):

$$Y^* = \begin{cases} (Y^\lambda - 1)/\lambda & \text{para } \lambda \neq 0; \\ \log(Y) & \text{para } \lambda = 0. \end{cases}$$

O parâmetro λ é estimado através de um gráfico de $\hat{\lambda}$ vs. log-verossimilhança. A estimativa será o valor que maximizar a função log-verossimilhança (Normal) para os dados transformados. Nesta abordagem, EMVs e sua teoria assintótica podem ser usados para estimar demais parâmetros do modelo (verossimilhança perfilada).

Estacionaridade:

Suponha que nosso processo espacial apresente média $\mu(\mathbf{s}) = E[Y(\mathbf{s})]$ e que a variância de $Y(\mathbf{s})$ existe para todo $\mathbf{s} \in D$.

O processo $Y(\mathbf{s})$ é dito *Gaussiano* se: para qualquer $n \geq 1$ e qualquer conjunto de sítios $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ temos $\mathbf{Y} = [Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)]'$ com distribuição Normal Multivariada.

O processo é dito “*fortemente estacionário*” ou “*estritamente estacionário*” se: para qualquer $n \geq 1$, qualquer conjunto de locais $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ e qualquer $\mathbf{h} \in \mathbb{R}^r$, a distribuição de $\{Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)\}$ é a mesma de $\{Y(\mathbf{s}_1 + \mathbf{h}), \dots, Y(\mathbf{s}_n + \mathbf{h})\}$.

O processo é dito “*fracamente estacionário*” ou “*estacionário de 2ª ordem*” se $\mu(\mathbf{s}) \equiv \mu$ (processo com média constante) e $\text{Cov}[Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h})] = C(\mathbf{h})$ para todo $\mathbf{h} \in \mathbb{R}^r$ tal que \mathbf{s} e $\mathbf{s} + \mathbf{h}$ caem ambos dentro de D .

Pela estacionaridade fraca, a relação entre os valores do processo em quaisquer dois locais pode ser sumarizada pela função de covariâncias $C(\mathbf{h})$. Esta função depende apenas do vetor de separação \mathbf{h} .

Nomenclatura e notação:

- $C(\mathbf{h})$ = função de covariância ou covariograma;
- $\rho(\mathbf{h})$ = função de correlação associada a $C(\mathbf{h})$.

Em qualquer processo estacionário, temos $\rho(-\mathbf{h}) = \rho(\mathbf{h})$ (função simétrica em \mathbf{h}).

Assumindo a existência de todas as variâncias, temos:

Estacionaridade forte \Rightarrow Estacionaridade fraca.

O contrário não é sempre verdade (mas é válido para Processos Gaussianos).

Variogramas:

Suponha que $E[Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})] = 0$ e defina

$$E[(Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s}))^2] = \text{Var}[Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})] = 2\gamma(\mathbf{h}).$$

Esta expressão faz sentido quando a parte esquerda depende apenas de \mathbf{h} e não depende de uma particular escolha de \mathbf{s} . Se este é o caso, dizemos que o processo é intrinsecamente estacionário.

A função $2\gamma(\mathbf{h})$ é chamada de variograma e $\gamma(\mathbf{h})$ é dito semivariograma.

Note que a estacionaridade intrínseca requer apenas o primeiro e o segundo momentos das diferenças $Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})$. Não diz nada a respeito da distribuição conjunta de uma coleção de variáveis $Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)$ e, portanto, não fornece uma verossimilhança.

Relação entre a função de covariância $C(\mathbf{h})$ e o variograma $2\gamma(\mathbf{h})$.

$$\begin{aligned} 2\gamma(\mathbf{h}) &= \text{Var}[Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})] \\ &= \text{Var}[Y(\mathbf{s} + \mathbf{h})] + \text{Var}[Y(\mathbf{s})] - 2\text{Cov}[Y(\mathbf{s} + \mathbf{h}), Y(\mathbf{s})] \\ &= C(\mathbf{0}) + C(\mathbf{0}) - 2C(\mathbf{h}) \\ &= 2[C(\mathbf{0}) - C(\mathbf{h})] \end{aligned}$$

Então: $\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$. Sobre esta igualdade temos:

- Dado C seremos capazes de determinar γ .
- Se o processo espacial é ergódico, então $C(\mathbf{h}) \rightarrow 0$ conforme $\|\mathbf{h}\| \rightarrow \infty$, sendo $\|\mathbf{h}\|$ = comprimento do vetor de separação \mathbf{h} . Isto significa que a covariância entre dois pontos se aproxima de zero conforme esses pontos são mais distantes no espaço.
- Aplicando $\lim_{\|\mathbf{h}\| \rightarrow \infty}$ em ambos os lados de $\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$, temos $\lim_{\|\mathbf{h}\| \rightarrow \infty} \gamma(\mathbf{h}) = C(\mathbf{0})$.

- Usando a variável auxiliar \mathbf{u} para evitar confusão, temos

$$C(\mathbf{h}) = C(\mathbf{0}) - \gamma(\mathbf{h}) = \left[\lim_{\|\mathbf{u}\| \rightarrow \infty} \gamma(\mathbf{u}) \right] - \gamma(\mathbf{h}).$$

Então $C(\mathbf{h})$ é bem definido se $\lim_{\|\mathbf{u}\| \rightarrow \infty} \gamma(\mathbf{u})$ existe. Neste caso, o processo é fracamente (ou 2ª ordem) estacionário.

Temos, portanto, uma forma de determinar a função de covariância C a partir de um semivariograma γ .

Estacionaridade fraca \Rightarrow estacionaridade intrínseca.

O contrário pode não ser verdade.

Um variograma válido deve satisfazer a condição de ser “negativo definido”. Desta forma, para qualquer conjunto de locais $\{s_1, \dots, s_n\}$ e qualquer conjunto de constantes $\{a_1, \dots, a_n\}$ tal que $\sum_{i=1}^n a_i = 0$, temos $\gamma(\mathbf{h})$ é válido se $\sum_i \sum_j a_i a_j \gamma(s_i - s_j) \leq 0$.

Apesar de termos a igualdade $\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$, não existe relação entre este resultado ($\gamma(\mathbf{h})$ satisfazendo condição de negativo definido) e a condição positiva definida para funções de covariâncias (que veremos mais adiante).

$$\sum_i \sum_j a_i a_j \gamma(s_i - s_j) = \sum_i \sum_j a_i a_j (1/2) \text{Var}[Y(s_i) - Y(s_j)] =$$

$$= \frac{1}{2} \sum_i \sum_j a_i a_j E[(Y(s_i) - Y(s_j))^2] = \frac{1}{2} E \left[\sum_i \sum_j a_i a_j (Y(s_i) - Y(s_j))^2 \right]$$

$$= \frac{1}{2} E \left[\sum_i \sum_j a_i a_j Y(s_i)^2 \right] - \frac{1}{2} E \left[2 \sum_i \sum_j a_i a_j Y(s_i) Y(s_j) \right] + \frac{1}{2} E \left[\sum_i \sum_j a_i a_j Y(s_j)^2 \right]$$

$$= \frac{1}{2} E \left[\sum_i a_i Y(s_i)^2 \sum_j a_j \right] - \frac{1}{2} E \left[2 \sum_i \sum_j a_i a_j Y(s_i) Y(s_j) \right] + \frac{1}{2} E \left[\sum_j a_j Y(s_j)^2 \sum_i a_i \right]$$

$$= \frac{1}{2} E \left[\sum_i a_i Y(s_i)^2(0) \right] - \frac{1}{2} E \left[2 \sum_i \sum_j a_i a_j Y(s_i) Y(s_j) \right] + \frac{1}{2} E \left[\sum_j a_j Y(s_j)^2(0) \right]$$

$$= -E \left[\sum_i \sum_j a_i a_j Y(s_i) Y(s_j) \right] = -E \left[\sum_i a_i Y(s_i) \right]^2 \leq 0.$$

Isotropia

Se o semivariograma $\gamma(\mathbf{h})$ depende do vetor de separação apenas através de seu comprimento (distância entre pontos) $\|\mathbf{h}\|$, dizemos que o processo é isotrópico (c.c. é anisotrópico; dependeria também da direção de \mathbf{h}).

Para um processo isotrópico, $\gamma(\mathbf{h})$ é uma função real com argumento univariado que pode ser escrita como $\gamma(\|\mathbf{h}\|)$.

Seja Q uma matriz ortogonal [i.e. $Q' = Q^{-1}$, $Q'Q = I$, $\det(Q) = +1$ ou -1].

A transformação $Q\mathbf{h}$ representa uma rotação de \mathbf{h} ; comprimento $\|Q\mathbf{h}\| = \|\mathbf{h}\|$.

Para um processo isotrópico devemos ter $C(\mathbf{h}) = C(Q\mathbf{h})$ ou $\gamma(\mathbf{h}) = \gamma(Q\mathbf{h})$.

Processo homogêneo: é intrinsecamente estacionário e isotrópico.

Processos isotrópicos são populares dado sua simplicidade, interpretabilidade e o fato de que vários candidatos para C e γ (com formulação simples) estarem disponíveis.

Denotando $\|\mathbf{h}\|$ por t (para simplificar notação) mostramos alguns exemplos a seguir...

Semivariograma Linear:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t & \text{se } t > 0, \tau^2 > 0, \sigma^2 > 0. \\ 0 & \text{caso contrário.} \end{cases}$$

- $\gamma(t) \rightarrow \infty$ conforme $t \rightarrow \infty$, então não temos um processo fracamente estacionário (o processo será intrinsecamente estacionário);
- $C(t)$ e $\rho(t)$ não existem.

Semivariograma esférico:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 & \text{se } t \geq 1/\phi. \\ \tau^2 + \sigma^2 \left\{ \frac{3\phi t}{2} - \frac{1}{2}(\phi t)^3 \right\} & \text{se } 0 < t \leq 1/\phi. \\ 0 & \text{caso contrário.} \end{cases}$$

- Variograma válido para dimensões $r = 1, 2$ e 3 ;
- O processo será estacionário;
- $C(t) = \begin{cases} 0 & \text{se } t \geq 1/\phi; \\ \sigma^2 \left[1 - \frac{3}{2}\phi t + \frac{1}{2}(\phi t)^3 \right] & \text{se } 0 < t \leq 1/\phi; \\ \tau^2 + \sigma^2 & \text{caso contrário.} \end{cases}$

- Lembrete: Para X_1 e X_2 variáveis aleatórias, $Corr(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)} \sqrt{Var(X_2)}}$;
- Função de correlação esférica entre dois pontos com distância t :

$$\rho(t) = C(t)/C(t=0^+) = \sigma^2 \left[1 - \frac{3}{2}\phi t + \frac{1}{2}(\phi t)^3 \right] / \sigma^2 = 1 - \frac{3}{2}\phi t + \frac{1}{2}(\phi t)^3$$
 para $0 < t \leq 1/\phi$ (0 para $t > 1/\phi$).

O semivariograma esférico tem bastante popularidade principalmente devido ao fato de oferecer uma clara ilustração de três características tradicionais de variogramas: *nugget*, *sill* e *range*.

- Por definição $\gamma(0) = 0$; $\gamma(0^+) \equiv \lim_{t \rightarrow 0^+} \gamma(t) = \tau^2$ chamado de nugget.
- $\lim_{t \rightarrow \infty} \gamma(t) = \tau^2 + \sigma^2$ chamado de sill.
 Neste caso, $sill - nugget = \sigma^2$ chamado de sill parcial.
- O valor $t = 1/\phi = R$ no qual $\gamma(t)$ atinge primeiro seu nível mais alto (*sill*) é chamado de range.
- ϕ é um parâmetro de decaimento (algumas vezes chamado incorretamente de *range* na literatura).

Semivariograma Exponencial:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2[1 - \exp\{-\phi t\}] & \text{se } t > 0, \\ 0 & \text{caso contrário.} \end{cases}$$

- Variograma válido em todas as dimensões;
- O *sill* é atingível apenas assintoticamente ($R = 1/\phi$ é infinito);
- Neste caso, usamos a noção de *range efetivo* (distância a partir da qual não resta correlação espacial). Para $t > 0$ temos:

$$C(t) = \lim_{u \rightarrow \infty} \gamma(u) - \gamma(t) = \tau^2 + \sigma^2 - [\tau^2 + \sigma^2(1 - \exp\{-\phi t\})] = \sigma^2 \exp\{-\phi t\}$$

Entretanto, $\gamma(t = 0^+) \approx \tau^2$ e $C(t = 0^+) \approx \sigma^2$; usando $\gamma(\mathbf{h}) = C(0) - C(\mathbf{h})$ temos $C(0) = \tau^2 + \sigma^2$ para $\|\mathbf{h}\| = 0^+$.

$$C(t) = \begin{cases} \tau^2 + \sigma^2 & \text{se } t = 0. \\ \sigma^2 \exp\{-\phi t\} & \text{se } t > 0. \end{cases}$$

- $\rho(t) = \frac{C(t)}{C(t=0^+)} = \sigma^2 \exp\{-\phi t\} / \sigma^2 = \exp\{-\phi t\}$ para $t > 0$;
- Note que $\rho(t = 0^+) = 1$ e $\rho(t = \infty) = 0$;
- *Range* efetivo t_0 = distância t na qual $\rho(t) = 0.05$.
Para $\exp\{-\phi t_0\} = 0.05$ temos $t_0 \approx 3/\phi$ visto que $\log(0.05) \approx -3$.
- O *nugget* τ^2 é visto como “variância do efeito não espacial”.
O *sill* parcial σ^2 é visto como “variância do efeito espacial”.

Semivariograma Gaussiano:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2[1 - \exp\{-\phi^2 t^2\}] & \text{se } t > 0, \\ 0 & \text{caso contrário.} \end{cases}$$

- Fornece um processo espacial bastante suave;
- $C(t) = \begin{cases} \sigma^2 \exp\{-\phi^2 t^2\} & \text{se } t > 0, \\ \tau^2 + \sigma^2 & \text{caso contrário.} \end{cases} \quad \rho(t) = \frac{C(t)}{C(t=0^+)} = \exp\{-\phi^2 t^2\} \text{ para } t > 0.$

Semivariograma Exponencial Potência:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2[1 - \exp\{-|\phi t|^p\}] & \text{se } t > 0, \\ 0 & \text{caso contrário.} \end{cases}$$

- Para $0 < p \leq 2$ temos uma família de variogramas válidos;
- Semivariogramas Exponencial e Gaussiano são casos especiais deste;
- $C(t) = \begin{cases} \sigma^2 \exp\{-|\phi t|^p\} & \text{se } t > 0, \\ \tau^2 + \sigma^2 & \text{caso contrário.} \end{cases} \quad \rho(t) = \frac{C(t)}{C(t=0^+)} = \exp\{-|\phi t|^p\} \text{ para } t > 0.$

Semivariograma Racional Quadrático:

$$\gamma(t) = \begin{cases} \tau^2 + \frac{\sigma^2 t^2}{\phi + t^2} & \text{se } t > 0, \\ 0 & \text{caso contrário.} \end{cases}$$

- $C(t) = \begin{cases} \sigma^2 \left(1 - \frac{t^2}{\phi + t^2}\right) & \text{se } t > 0, \\ \tau^2 + \sigma^2 & \text{caso contrário.} \end{cases} \quad \rho(t) = \frac{C(t)}{C(t=0^+)} = 1 - \frac{t^2}{\phi + t^2} \text{ para } t > 0.$

Semivariograma Wave:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left(1 - \frac{\text{sen}(\phi t)}{\phi t}\right) & \text{se } t > 0, \\ 0 & \text{caso contrário.} \end{cases}$$

- Exemplo de variograma que não é monotonicamente crescente.

- $C(t) = \begin{cases} \sigma^2 \left(\frac{\text{sen}(\phi t)}{\phi t}\right) & \text{se } t > 0, \\ \tau^2 + \sigma^2 & \text{caso contrário.} \end{cases} \quad \rho(t) = \frac{\text{sen}(\phi t)}{\phi t} \text{ para } t > 0.$

Semivariograma Potência:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t^\lambda & \text{se } t > 0, \\ 0 & \text{caso contrário.} \end{cases}$$

- Generalização do semivariograma linear;
- Para $0 \leq \lambda < 2$ é válido, intrinsecamente (não fracamente) estacionário;
- $C(t)$ e $\rho(t)$ não existem.

Semivariograma da classe Matérn:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{(2\phi t \sqrt{\nu})^\nu}{2^{\nu-1} \Gamma(\nu)} K_\nu(2\phi t \sqrt{\nu}) \right] & \text{se } t > 0, \\ \tau^2 & \text{caso contrário.} \end{cases}$$

- Bastante versátil e originalmente proposto em Matérn (1960, 1986);
- Esta parametrização foi proposta em Handcock e Wallis (1994);
- Parametrização em Diggle e Ribeiro (2007) e *geoR*: Use $\phi^* = 1/(2\phi\sqrt{\nu})$;

- $K_\nu(\cdot)$ é a função Bessel modificada de ordem ν (computacionalmente tratável e disponível no pacote *geoR*).
- $C(t) = \begin{cases} \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (2\phi t \sqrt{\nu})^\nu K_\nu(2\phi t \sqrt{\nu}) & \text{se } t > 0. \\ \tau^2 + \sigma^2 & \text{se } t = 0. \end{cases}$
- $\rho(t) = \frac{1}{2^{\nu-1}\Gamma(\nu)} (\phi t)^\nu K_\nu(\phi t)$ para $t > 0$.
- $\nu > 0$ é um parâmetro de suavização:
 - Em 2 dimensões, o maior inteiro $\leq \nu$ indica o número de vezes que realizações do processo serão quadrado médio diferenciável.
 - $\nu = 1/2 \Rightarrow$ Exponencial;
 - $\nu \rightarrow \infty \Rightarrow$ Gaussian;
 - $\nu = 3/2 \Rightarrow$ formas fechadas convenientes para $C(t)$ e $\gamma(t)$;
 - $C(t) = \sigma^2(1 + \phi t) \exp\{-\phi t\}$ se $t > 0$ e $C(t) = \tau^2 + \sigma^2$ c.c.
 - $\gamma(t) = \tau^2 + \sigma^2[1 - (1 + \phi t) \exp\{-\phi t\}]$ se $t > 0$ e $\gamma(t) = \tau^2$ c.c.

Exemplo:

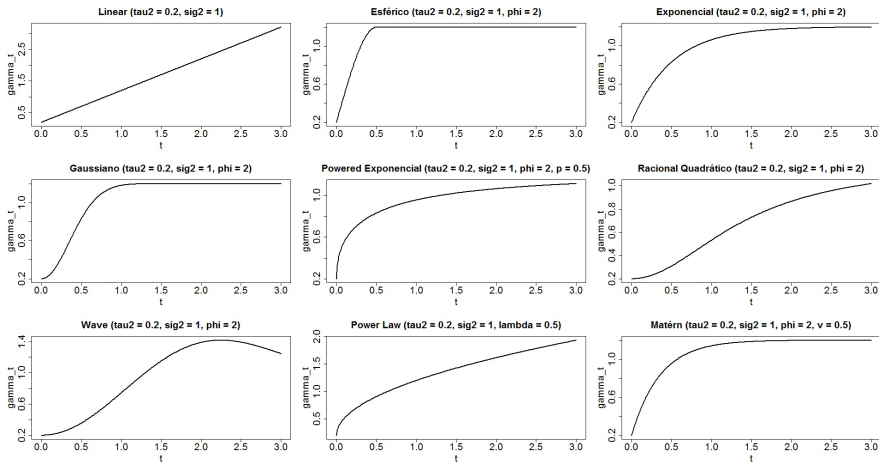


Fig.12 : Semivariogramas. No modelo linear temos $\gamma(t) \rightarrow \infty$ conforme $t \rightarrow \infty$, portanto, o processo gerado não é fracamente estacionário (entretanto é intrinsecamente estacionário). O nugget é igual a τ^2 , mas o sill e o range são ambos infinitos.

Realização de um processo espacial Gaussiano: Função de correlação *Matern* com parâmetro de suavização: $\nu = 0.5$ (linha sólida), $\nu = 1.5$ (tracejada) e $\nu = 2.5$ (pontilhada).

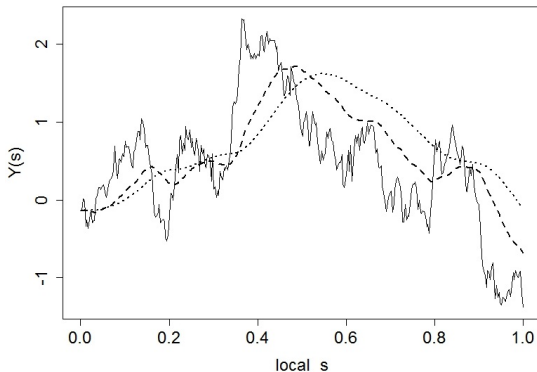


Fig.13 : Simulação (processo espacial Gaussiano).

Realização de um processo espacial Gaussiano: Função de correlação *Exponencial Potência* com parâmetro de suavização: $p = 0.7$ (linha sólida), $p = 1$ (tracejada) e $p = 2$ (pontilhada).

$p = 0.7$ e $p = 1 \Rightarrow$ processo quadrado médio contínuo (não diferenciável).
 $p = 2$ pode gerar uma estrutura de correlação muito simplificada (pobre); este caso equivalente à correlação Matern com $\nu \rightarrow \infty$,

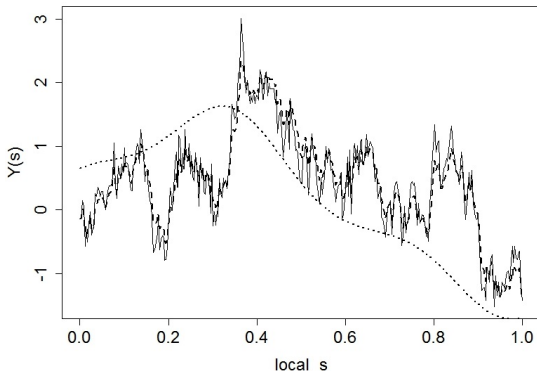


Fig.14 : Simulação (processo espacial Gaussiano).

Para que $C(\mathbf{h})$ seja uma função de covariâncias válida, ela precisa ser positiva definida, ou seja, $C(\mathbf{h}) \equiv \text{Cov}(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h}))$ é tal que para qualquer conjunto finito de sítios $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ e quaisquer constantes reais $\{a_1, \dots, a_n\}$,

$$\text{Var} \left[\sum_i a_i Y(\mathbf{s}_i) \right] = \sum_i \sum_j a_i a_j \text{Cov}(Y(\mathbf{s}_i), Y(\mathbf{s}_j)) = \sum_i \sum_j a_i a_j C(\mathbf{s}_i - \mathbf{s}_j) \geq 0,$$

sendo a desigualdade estritamente maior se $a_i \neq 0$ para pelo menos um $i = 1, \dots, n$.

Verificar a condição de positiva definida acima seria trabalhoso. Felizmente, temos o teorema a seguir que fornece uma condição necessária e suficiente para que $C(\mathbf{h})$ seja positiva definida.

Teorema de Bochner: $C(\mathbf{h})$ é uma função de covariância (válida) positiva definida se e somente se $C(\mathbf{h}) = \int \cos(\mathbf{w}'\mathbf{h}) G(d\mathbf{w})$ sendo G uma medida em \mathbb{R}^r limitada, positiva e simétrica em torno de zero.

A demonstração não será discutida neste curso...

As famílias de covariogramas apresentadas anteriormente (Esférica, Exponencial, Exponencial Potência, Gaussiana, Racional Quadrática, Wave e Matérn) são positivas definidas (válidas) e flexíveis para as aplicações de geoestatística.

Construindo funções de covariâncias (ou de correlação) válidas.

Misturas: Se C_1, \dots, C_m são funções de covariâncias (covariogramas) válidas em \mathbb{R}^r e se $\sum_{j=1}^m p_j = 1$ com $p_j > 0$, então $C(\mathbf{h}) = \sum_{j=1}^m p_j C_j(\mathbf{h})$ também é um covariograma válido em \mathbb{R}^r .

Produtos: Se C_1, \dots, C_m são covariogramas válidos em \mathbb{R}^r , então $\prod_{i=1}^m C_i$ é um covariograma válido em \mathbb{R}^r .

Convolução: Se C_1 e C_2 são covariogramas válidos em \mathbb{R}^r , então $C_{12}(\mathbf{h}) = \int C_1(\mathbf{h} - \mathbf{t}) C_2(\mathbf{t}) d\mathbf{t}$ é um covariograma válido em \mathbb{R}^r .

É possível mostrar que se C_1 e C_2 são isotrópicos, então C_{12} também será isotrópico.

Ajuste do modelo de variograma.

Tendo visto uma grande seleção de modelos paramétricos para variogramas, como iremos escolher um deles para um determinado conjunto de dados? Os dados podem mostrar diferenças entre eles?

O que tem sido feito na literatura é avaliar o gráfico do semivariograma empírico (Matheron, 1963) que é uma estimativa não paramétrica do semivariograma real. Iremos comparar o formato exibido no gráfico com aqueles dos modelos teóricos. O semivariograma empírico é dado por:

$$\hat{\gamma}(t) = \frac{1}{2|N(t)|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(t)} [Y(\mathbf{s}_i) - Y(\mathbf{s}_j)]^2,$$

sendo $N(t)$ um conjunto de pares de pontos tais que $\|\mathbf{s}_i - \mathbf{s}_j\| = t$, e $|N(t)|$ é o número de pares neste conjunto.

Na prática, as distâncias entre os pares de pontos são todas diferentes. Esta configuração não será útil.

Precisamos estabelecer um *grid* para agrupar os pontos em algumas categorias de distâncias.

Considere os intervalos: $I_1 = (0, t_1]$, $I_2 = (t_1, t_2]$, \dots , $I_K = (t_{K-1}, t_K]$ para um *grid* (podendo ser regular) $0 < t_1 < \dots < t_K$.

Tome o ponto central t_k^* como o valor representante de cada intervalo e defina $N(t_k^*) = \{(\mathbf{s}_i, \mathbf{s}_j) : \|\mathbf{s}_i - \mathbf{s}_j\| \in I_k\}$ para $k = 1, \dots, K$.

Assim como na construção de um histograma, devemos ter cuidado para escolher o número de intervalos K e o valor do limite final t_K . Recomenda-se intervalos I_k com amplitude para capturar pelo menos 30 pares de pontos.

Este método (semivariograma empírico) é uma estimação via método dos momentos. Seria análogo a estimar a variabilidade via variância amostral S^2 .

Existem razões para desconfiar que a estimativa do semivariograma não é boa.

- $\hat{\gamma}(t)$ é sensível a *outliers*;
- A média amostral das diferenças quadráticas $[Y(\mathbf{s}_i) - Y(\mathbf{s}_j)]^2$ pode ter comportamento ruim. Supondo $Y(\mathbf{s}_i) \sim \text{Normal}$, as diferenças quadráticas terão uma distribuição relacionada à χ_1^2 com forte assimetria;
- $\hat{\gamma}(t)$ usa diferenças $Y(\mathbf{s}_i) - Y(\mathbf{s}_j)$ ao invés da observação $Y(\mathbf{s}_i)$ em si;
- Os componentes da soma $\sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(t)} [Y(\mathbf{s}_i) - Y(\mathbf{s}_j)]^2$ serão dependentes dentro e entre os intervalos I_k .
- $|N(t_k^*)|$ não é o mesmo em cada intervalo.

Cressie e Hawkins (1980) propuseram um estimador mais robusto usando médias amostrais de $|Y(\mathbf{s}_i) - Y(\mathbf{s}_j)|^{1/2}$ (disponível em pacotes computacionais como o *geoR*);

Semivariograma empírico (*script R*).

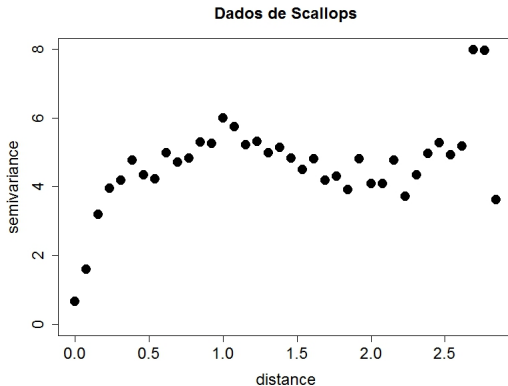


Fig.15 : Dados de Scallops (1990) transformados: $\log(n^{\circ} \text{capturas} + 1)$.

Semivariograma empírico (*script R*).

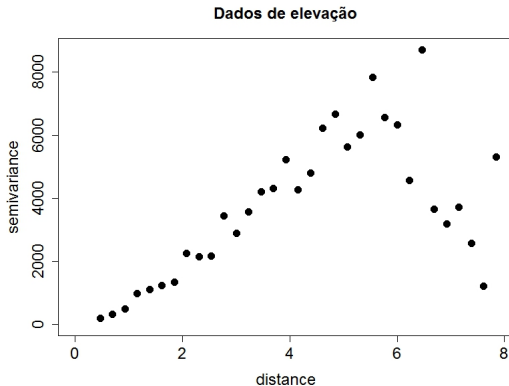


Fig.16 : Dados de elevação (originais, sem transformação).

Novamente, podemos fazer o gráfico de $\hat{\gamma}(t)$ e então comparar visualmente o resultado com o comportamento dos diversos semivariogramas teóricos. Por tentativa e erro escolheremos o *nugget*, *sill* e *range*.

Na prática, é difícil escolher visualmente o modelo teórico dado que vários dos candidatos irão parecer igualmente apropriados.

De maneira mais formal, poderíamos tratar este problema de estimação usando rotinas de maximização não-lineares para encontrar o *nugget*, *sill* e *range* que minimizam algum critério de ajuste (ex.: mínimos quadrados, mínimos quadrados ponderados ou mínimos quadrados generalizados).

Se tivermos um modelo assumindo uma distribuição para os dados, poderíamos usar Máxima Verossimilhança (ou Máxima Verossimilhança Restrita REML) para obter estimativas dos parâmetros. A abordagem Bayesiana também pode ser usada neste caso.

Anisotropia.

Em “Isotropia” a associação entre pontos depende apenas da distância entre eles. Em “Anisotropia” a associação depende do vetor de separação entre os locais, isto é, depende da distância e da direção (efeito direcional).

Revisão: Em duas dimensões, o ponto $\mathbf{h} = (h_1, h_2)'$ pode ser rotacionado como segue

$$\begin{pmatrix} h_1^* \\ h_2^* \end{pmatrix} = \begin{bmatrix} \cos(\eta) & -\sin(\eta) \\ \sin(\eta) & \cos(\eta) \end{bmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix},$$

sendo $\mathbf{h}^* = (h_1^*, h_2^*)'$ as coordenadas após a rotação. Os vetores \mathbf{h} e \mathbf{h}^* terão a mesma magnitude e serão separados por um ângulo η .

Esta definição pode ser estendida para dimensões maiores que 2.

A forma mais simples de efeito direcional na estrutura de covariâncias é a anisotropia geométrica. Ela surge quando uma estrutura de covariâncias estacionária é transformada via rotação e compressão dos pontos ao longo dos eixos.

Algebricamente, um modelo com anisotropia geométrica nas coordenadas (h_1, h_2) pode ser convertido para um modelo estacionário nas coordenadas (h_1^*, h_2^*) como segue:

$$(h_1^*, h_2^*) = (h_1, h_2) \begin{bmatrix} \cos(\eta) & -\sin(\eta) \\ \sin(\eta) & \cos(\eta) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \varphi^{-1} \end{bmatrix}$$

sendo η o ângulo de anisotropia e $\varphi > 1$ a razão de anisotropia.

Notação: $(h_1^*, h_2^*) = (h_1, h_2) \mathbf{R} \mathbf{D}$ sendo

\mathbf{R} = matriz de rotação e \mathbf{D} = a matriz diagonal acima.

Eixo principal: direção na qual a correlação decai mais lentamente com o aumento da distância.

Anisotropia geométrica: $C(\mathbf{s} - \mathbf{s}^*) = \sigma^2 \rho[(\mathbf{s} - \mathbf{s}^*)' \mathbf{B} (\mathbf{s} - \mathbf{s}^*)]$.

- \mathbf{B} é $(r \times r)$ simétrica positiva definida com $\rho(\cdot)$ sendo uma função de correlação válida em \mathbb{R}^r . Considere $\mathbf{h} = (\mathbf{s} - \mathbf{s}^*)$,

$$\begin{aligned}(\mathbf{h}^*)' &= \mathbf{h}' \mathbf{R} \mathbf{D} \\(\mathbf{h}^*)' \mathbf{h}^* &= \mathbf{h}' (\mathbf{R} \mathbf{D}) (\mathbf{R} \mathbf{D})' \mathbf{h} \\(\mathbf{h}^*)' \mathbf{h}^* &= \mathbf{h}' \mathbf{R} \mathbf{D} \mathbf{D}' \mathbf{R}' \mathbf{h} = \mathbf{h}' \mathbf{B} \mathbf{h} \\\rho(\|\mathbf{h}^*\|) &= \rho(\sqrt{\mathbf{h}' \mathbf{B} \mathbf{h}})\end{aligned}$$

- O parâmetro de *range*/decaimento ϕ está incorporado em \mathbf{B} .
Quando $r = 2 \Rightarrow$ 3 parâmetros de *range* (\mathbf{B}_{11} , \mathbf{B}_{22} e $\mathbf{B}_{12} = \mathbf{B}_{21}$) ao invés de 1; eles incluem a informação da direção ($r = 2$: horizontal e vertical);
- Contornos indicando uma associação constante (em C acima) serão elípticos; em particular, o contorno com $\rho(\cdot) = 0.5$ fornece um *range efetivo* em cada direção espacial.

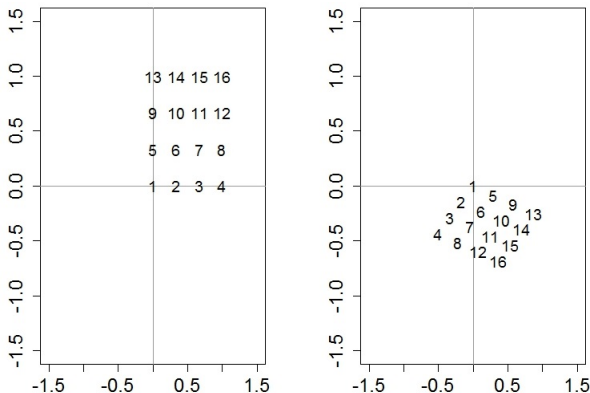


Fig.17 : Rotação de uma configuração de dados usando os parâmetros de anisotropia: $\eta = 2\pi/3 = 120^\circ$ e $\varphi = 2$. Painel esquerdo = locais originais, Painel direito = locais transformados em um espaço isotrópico.

É possível notar que cada ponto foi deslocado em 120° no sentido horário. Quanto maior φ , mais comprimidos ficarão os pontos transformados.