

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

Prof. Vinícius D. Mayrink

EST171 - Estatística Espacial

Sala: 4073

Email: vdm@est.ufmg.br

1º semestre de 2024

Análise Exploratória de Dados (AED).

A AED é rotineiramente aplicada para avaliar conjuntos de dados em estudos de regressão e modelos lineares generalizados. Estas ferramentas são também apropriadas para analisar dados espaciais de referência pontual.

Alguns dos gráficos úteis em AED foram mostrados na “Parte 1” do conteúdo sobre geoestatística (rever slides). Apresentaremos aqui mais algumas opções.

Para dados contínuos, o primeiro passo é examinar a chamada “1ª lei da geoestatística”, ilustrada abaixo no contexto unidimensional.

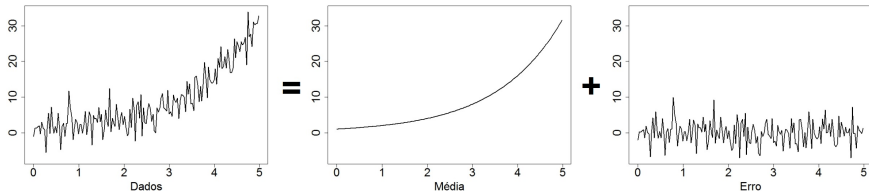


Fig.1: Os dados são descritos por um termo de “média” e um termo de “erro”. Média = comportamento global (1ª ordem). Erro = comportamento local (2ª ordem) capturado via função de covariância.

A 1ª lei também esclarece que:

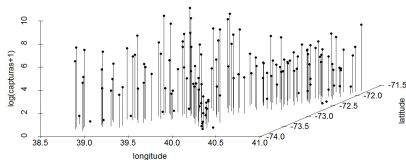
- 1 A associação espacial em $Y(\mathbf{s})$ corresponde a olhar para $E[(Y(\mathbf{s}) - \mu)(Y(\mathbf{s}^*) - \mu)]$, que é uma espécie de covariância em relação à média global μ . As observações seguem um padrão espacial.
- 2 A associação espacial nos resíduos $\epsilon(\mathbf{s})$ corresponde a olhar para $E[(Y(\mathbf{s}) - \mu(\mathbf{s}))(Y(\mathbf{s}^*) - \mu(\mathbf{s}^*))]$, que é uma espécie de covariância em relação à média local $\mu(\mathbf{s})$. A magnitude dos resíduos segue um padrão espacial.
- 3 Dado o local \mathbf{s} , temos a diferença: $[Y(\mathbf{s}) - \mu(\mathbf{s})] - [Y(\mathbf{s}) - \mu] = \mu - \mu(\mathbf{s})$ que não deve ser ignorada na análise.

Primeiramente devemos avaliar um mapa simples que mostra a disposição dos locais onde os dados foram coletados. Nesta avaliação, verifique o quão regular é o posicionamento dos pontos.

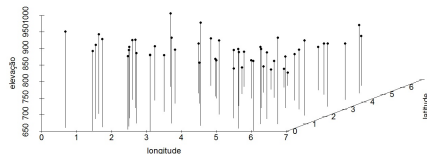
Gráficos não espaciais (ex.: histogramas e *boxplots*) podem ser usados para avaliar a existência de *outliers* e a necessidade de alguma transformação para alterar a magnitude dos dados.

Além do gráfico “imagem com contornos” visto anteriormente (rever slides), outras duas opções gráficas para explorar a estrutura espacial dos dados são mostradas a seguir. Cuidado! Estes gráficos baseados nos valores de $Y(\mathbf{s}_i)$ podem ser enganosos mostrando um padrão espacial fortemente explicado pela média $\mu(\mathbf{s}_i)$; estes gráficos podem esconder um padrão espacial relacionado à magnitude dos erros $\epsilon(\mathbf{s}_i)$.

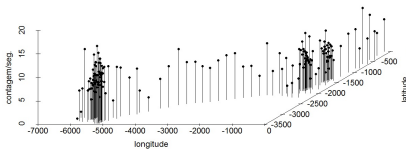
Dados Scallops



Dados Elevação



Dados Rongelap



Dados Solo

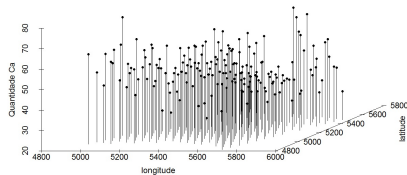
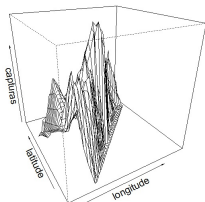
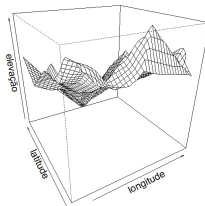


Fig.2: Drop line scatter plots.

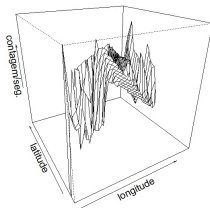
Dados Scallops



Dados Elevação



Dados Rongelap



Dados Solo

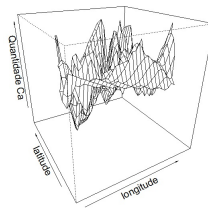


Fig.3: Surface plots.

Ao explorar $\mu(\mathbf{s}_i)$ podemos ter dois tipos de informação para o local \mathbf{s}_i : (1) as coordenadas geográficas expressas em longitude e latitude ou em *eastings* and *northings* via projeção, (2) Covariáveis medidas em \mathbf{s}_i e relevantes para explicar $Y(\mathbf{s}_i)$.

Quando a média do processo depende da informação geográfica $\mu(\mathbf{s}_i)$ dizemos que temos um superfície com tendência (*trend*).

Para dados tais que $\mathbf{s} \in \mathbb{R}^2$ (ex.: longitude e latitude) e podendo considerar um *grid* regular (*lattice* regular) sobre o mapa de pontos, podemos avaliar a tendência (*trend*) via *boxplots* construídos por linha ou por coluna. O próximo slide exhibe este tipo de gráfico para os dados de “elevação”.

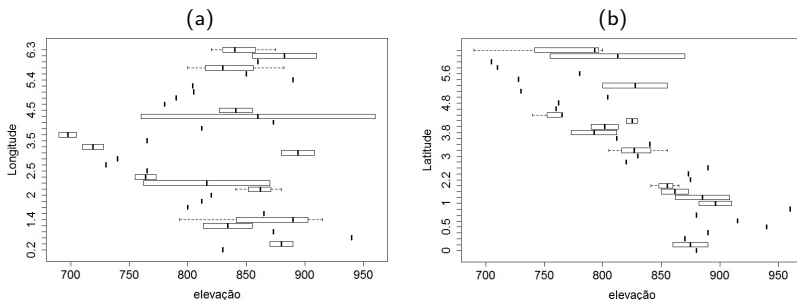


Fig.4: Boxplots por longitude ou por latitude para avaliar tendência. Lembrando que a variável resposta $Y(s_i)$ é a medida de elevação no local s_i , para $i = 1, \dots, 52$. A região de estudo tem formato retangular havendo, na maioria dos casos, mais de um sítio por longitude ou por latitude.

O Painel (b) sugere a existência de um padrão espacial (tendência) ao longo das latitudes (eixo y). O Painel (a) não indica um padrão espacial ao longo das longitudes (eixo x).

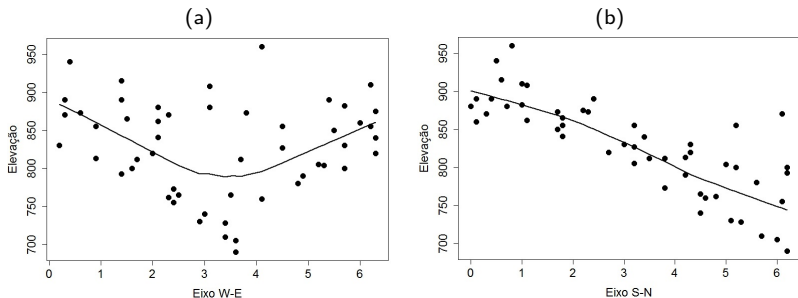


Fig.5: Elevação $Y(s_i)$ vs. coordenadas de s_i nos eixos $W-E$ e $S-N$ para avaliar tendência.

Os *scatter-plots* mostrados acima são um outro recurso gráfico para avaliação visual de tendência. O gráfico em (b) confirma uma tendência no eixo $S-N$. O gráfico em (a) sugere uma possibilidade de tendência não monótona para o eixo $W-E$ (maiores elevações parecem se concentrar em coordenadas nos extremos).

Retornando ao tópico de distinguir entre: uma variação espacial da média $\mu(\mathbf{s}_i)$ e a variação espacial de $Y(\mathbf{s}_i)$, em torno de uma média constante μ , explicada pela variação espacial dos erros $\epsilon(\mathbf{s}_i)$, considere os comentários:

- A melhor maneira de avaliar esta distinção seria trabalhar com replicações independentes do conjunto de dados (impossível);
- Na prática, assumimos formulações simples para $\mu(\mathbf{s}_i)$ (ex.: superfícies de tendência representadas por polinômios de pequena ordem) e utilizamos a estrutura de correlações de $Y(\mathbf{s}_i)$ para explicar padrões espaciais mais sutis.

Assuma o modelo $Y(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \epsilon(\mathbf{s}_i)$.

Algumas opções de modelagem da tendência:

- 1 $\mu(\mathbf{s}_i) = \mu$ (constante, sem tendência);
- 2 $\mu(\mathbf{s}_i) = \beta_0 + \beta_1 X_1(\mathbf{s}_i) + \beta_2 X_2(\mathbf{s}_i)$ (polinômio de 1º grau);
- 3 $\mu(\mathbf{s}_i) = \beta_0 + \beta_1 X_1(\mathbf{s}_i) + \beta_2 X_2(\mathbf{s}_i) + \beta_3 X_1(\mathbf{s}_i)^2 + \beta_4 X_2(\mathbf{s}_i)^2 + \beta_5 [X_1(\mathbf{s}_i)X_2(\mathbf{s}_i)]$ (polinômio de 2º grau);

sendo $X_1(\mathbf{s}_i)$ e $X_2(\mathbf{s}_i)$ as coordenadas espaciais do sítio \mathbf{s}_i .

Em geral, assumimos normalidade para os resíduos e estimamos os coeficientes β' s via máxima verossimilhança.

Para avaliar comportamentos considerados de menor escala nos dados (imperceptíveis em certas análises), alguns autores recomendam criar o semivariograma *cloud* (nuvem).

Este gráfico avalia $[Y(\mathbf{s}_i) - Y(\mathbf{s}_j)]^2$ versus $\|\mathbf{s}_i - \mathbf{s}_j\|$.

Em geral, o resultado final é uma nuvem de pontos “poluída” (muitos pontos, ruído alto), dificultando a avaliação de qualquer efeito.

Os semivariogramas empíricos (já estudados) acabam sendo preferidos na literatura, visto que reduzem esse ruído. Eles são bastante úteis para avaliar a existência de estruturas espaciais.

A sugestão é que o semivariograma *cloud* seja usado para avaliar os resíduos $[\epsilon(\mathbf{s}_i) - \epsilon(\mathbf{s}_j)]^2$ vs. $\|\mathbf{s}_i - \mathbf{s}_j\|$) a não ser que um modelo com média constante seja apropriado.

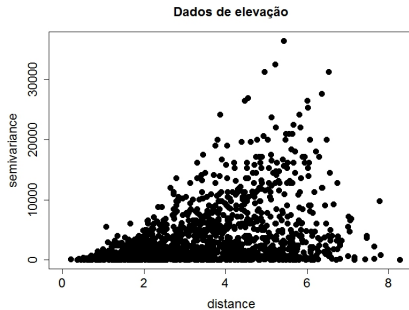
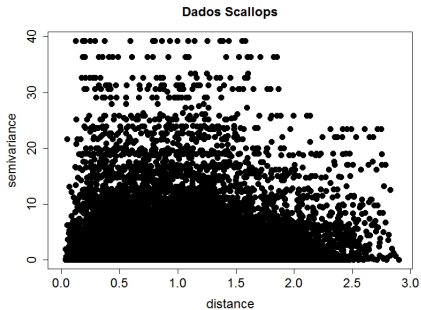
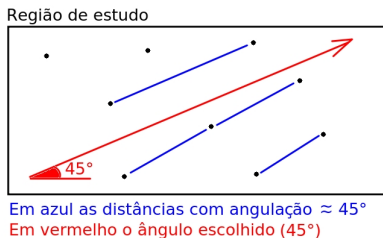


Fig.6: Semivariogramas *cloud*.

O conhecimento das características do processo espacial é valioso para a escolha das direções. Na prática, a escolha do n^o de ângulos e das direções é arbitrária.

Dado um ângulo η_i , podemos usar o semivariograma empírico para obter o correspondente semivariograma direcional. A ideia é construir o semivariograma empírico usando apenas as distâncias cuja angulação é $\approx \eta_i$ (uma tolerância é admitida para definir quais vetores de separação possuem ângulo $\approx \eta_i$).



Teoricamente, a anisotropia poderia ser examinada via semivariograma direcional; entretanto, na prática, determinar se o *sill*, *nugget* e/ou *range* variam com a direção é difícil.

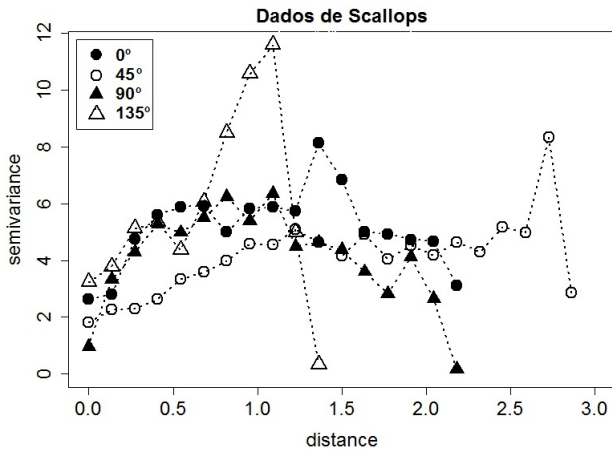


Fig.8: Semivariogramas direcionais para os dados de Scallops.

Comentários sobre o gráfico:

- Veja que a variabilidade/oscilação para 45° (paralela à costa) é menor do que nas demais direções.
- A variabilidade para 135° (perpendicular à costa) é bastante errática (pode haver anisotropia *sill*).
- Cuidado ao tirar conclusões via semivariogramas direcionais; eles não consideram o tamanho amostral (dados sumarizados no *grid* com perda de informação sobre a variabilidade).
- Semivariogramas direcionais construídos a partir de dados isotrópicos podem exibir as mesmas diferenças de magnitudes.
- É difícil tirar conclusões sobre a presença de anisotropia geométrica.

Contornos do semivariograma empírico (*ESC - Empirical Semivariogram Contour plot*).

Método mais informativo para verificar anisotropia.

Baseado em um gráfico de contornos representando a superfície em \mathbb{R}^2 do semivariograma empírico .

Também chamado: “Mapa de contorno de valores agrupados do variograma” ou “Gráfico isarítmico” (*isarithmic plot*).

Para cada $\binom{N}{2}$ pares de sítios em \mathbb{R}^2 , calcule as distâncias de separação h_x e h_y em cada eixo. Visto que o sinal de h_y depende da ordem arbitrária em que os sítios são comparados, force $h_y \geq 0$; isto é, faça $(-h_x, -h_y)$ quando $h_y < 0$.

Agregue estes pares dentro de sub-regiões retangulares B_{ij} e calcule o valor do semivariograma empírico para (i, j) como:

$$\gamma_{ij}^* = \frac{1}{2N_{B_{ij}}} \sum_{\{(k,l): (\mathbf{s}_k - \mathbf{s}_l) \in B_{ij}\}} [Y(\mathbf{s}_k) - Y(\mathbf{s}_l)]^2$$

sendo $N_{B_{ij}}$ o número de sítios dentro da sub-região B_{ij} .

Visto que forçamos $h_y \geq 0$ e não impomos restrição para h_x , devemos considerar o comprimento das sub-regiões ao longo do eixo y como metade do comprimento definido para o eixo x .

Também forçamos a sub-região do meio (eixo x) a ser centrada em 0.

Denotando o centro do (i, j) -ésima sub-região por (x_i, y_j) , um gráfico 3-dimensional de γ_{ij}^* vs. (x_i, y_j) fornecerá uma superfície para o semivariograma empírico. Um gráfico de contorno para uma versão suavizada (pacote R “akima”) desta superfície será o *ESC plot*.

O *ESC plot* ajuda a verificar desvios em termos de isotropia; para isotropia teremos “contornos circulares” e para anisotropia geométrica teremos “contornos elípticos”.

Uma possível desvantagem do *ESC plot* é a ocorrência de contagens esparsas (zeros) em sub-regiões dos extremos. Entretanto, estas sub-regiões podem ser removidas antes que a suavização seja aplicada.

Exercício: *ESC plot* para dados de Scallops.

Trabalhamos com dados coletados no nordeste dos EUA. Nesta região, temos 1ª latitude \neq 1ª longitude, podendo introduzir anisotropia artificial. Desta forma, emprega-se a projeção UTM (*Universal Transverse Mercator*) que transforma a configuração “longitude e latitude” para *km* nos eixos *East-West* (E-W) e *North-South* (N-S).

Os detalhes da conversão de coordenadas geográficas “longitude e latitude” para coordenadas em *km* UTM não serão tratadas neste curso. Considere o pacote “rgdal” (*script R*) para a conversão.

Tarefa: programar uma função no *software R* para o cálculo de γ_{ij}^* . Construa um gráfico de imagem com contornos (use pacote “akima”) mostrando o comportamento dos valores γ_{ij}^* ao longo das coordenadas *km* UTM. Utilize os dados de Scallops (1990) e avalie a ocorrência ou não de anisotropia.

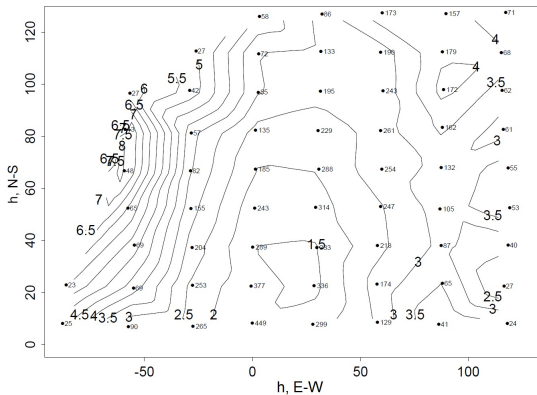


Fig.9: ESC plot para os dados de Scallops 1993.

- O gráfico mostra os centros das sub-regiões e os contornos;
- Esta configuração sugere anisotropia geométrica (contornos elípticos);
- O valores ESC na linha $h_y \approx 0$ fornecem uma alternativa ao usual semivariograma direcional para 0° .

Predição espacial clássica (*Kriging*).

A nomenclatura *Kriging* foi estabelecida por Matheron (1963) em referência a D.G. Krige (Engenheiro de Minas Sul-Africano) cujo trabalho sobre métodos empíricos para dados geoestatísticos inspirou uma abordagem mais geral.

Dadas observações de um campo aleatório $\mathbf{Y} = [Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)]'$ deseja-se prever $Y(\mathbf{s}_0)$ no sítio \mathbf{s}_0 onde não há observações.

Um preditor linear para $Y(\mathbf{s}_0)$ baseado em \mathbf{Y} é dado por: $\sum_i l_i Y(\mathbf{s}_i) + \delta_0$. Assumindo perda de erro quadrático, o melhor preditor linear deve minimizar $E\{[Y(\mathbf{s}_0) - (\sum_i l_i Y(\mathbf{s}_i) + \delta_0)]^2\}$ com a sua especificação de δ_0 e l_i .

Veja que:

$$\begin{aligned} E \left[Y(\mathbf{s}_0) - \left(\sum_i l_i Y(\mathbf{s}_i) + \delta_0 \right) \right] &= E[Y(\mathbf{s}_0)] - E \left[\sum_i l_i Y(\mathbf{s}_i) \right] + \delta_0 = \\ &= E[Y(\mathbf{s}_0)] - \sum_i l_i E[Y(\mathbf{s}_i)] + \delta_0 = \mu - \sum_i l_i \mu + \delta_0 = \mu - \mu \sum_i l_i + \delta_0 \end{aligned}$$

Sob estacionaridade intrínseca, o resultado anterior deve ser 0 (processo tem média constante). Para que isso ocorra, devemos ter $\sum_{i=1}^n l_i = 1$ e $\delta_0 = 0$.

Se denotarmos $a_0 = 1$ e $a_i = -l_i$ teremos:

$$E \left[Y(\mathbf{s}_0) - \left(\sum_{i=1}^n l_i Y(\mathbf{s}_i) + \delta_0 \right) \right]^2 = E \left[\sum_{i=0}^n a_i Y(\mathbf{s}_i) \right]^2,$$

com $\sum_{i=0}^n a_i = a_0 - \sum_{i=1}^n l_i = 1 - 1 = 0$.

Revisão: Um variograma válido deve satisfazer a condição de negativo definido, isto é, $\sum_i \sum_j a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j) = -E [\sum_i a_i Y(\mathbf{s}_i)]^2 \leq 0$.

Diante disso, temos:

$$E \left[Y(\mathbf{s}_0) - \left(\sum_{i=1}^n l_i Y(\mathbf{s}_i) + \delta_0 \right) \right]^2 = E \left[\sum_{i=0}^n a_i Y(\mathbf{s}_i) \right]^2 = - \sum_i \sum_j a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j).$$

Isto revela como o variograma surge no contexto de *Kriging* dentro da geoestatística.

Os l_i 's ótimos podem ser obtidos resolvendo a otimização indicada (com as restrições citadas) via multiplicador de Lagrange; eles serão funções de $\gamma(\mathbf{h})$. Este cálculo não será explorado neste curso; ver Cressie (1993), Seção 3.2, pag. 120.

Nenhuma suposição é necessária sobre a distribuição de $Y(\mathbf{s}_i)$.

Nossa única suposição (usada acima) é o modelo com estacionaridade intrínseca.

A seguir, iremos explorar o *Kriging* em um contexto de modelagem espacial supondo normalidade. Isso permitirá a inferência (Clássica e Bayesiana) tomando como base a verossimilhança. Focaremos em Processos Gaussianos e métodos baseados na verossimilhança para trabalhar com a função de covariância.

Kriging no contexto de Processos Gaussianos.

Seja \mathbf{s}_i um ponto no espaço \mathbb{R}^r e $\mathbf{Y} = [Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)]'$ um vetor aleatório de observações. Iniciamos com o seguinte modelo sem covariáveis (*Ordinary Kriging*):

$$\mathbf{Y} = \mu \mathbf{1} + \epsilon, \quad \text{sendo } \epsilon \sim N_n(\mathbf{0}, \Sigma),$$

Notação: μ é um escalar e $\mathbf{1}$ = vetor de 1's ($n \times 1$).

Para uma estrutura de covariâncias espacial sem efeito *nugget*, especificaremos:

$$\Sigma = \sigma^2 H(\phi)$$

sendo: $H(\phi)$ uma matriz ($n \times n$) de correlações com entrada $H(\phi)_{ij} = \rho(\phi; d_{ij})$,
 $\rho(\cdot)$ uma função de correlação válida em \mathbb{R}^r e
 $d_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$ a distância entre \mathbf{s}_i e \mathbf{s}_j .

Para um modelo com efeito *nugget* escreveremos: $\Sigma = \sigma^2 H(\phi) + \tau^2 I_n$,
sendo τ^2 a variância do efeito nugget e I_n uma matriz identidade de ordem n .

Quando covariáveis $\mathbf{X}(\mathbf{s}_i) = [1, X_1(\mathbf{s}_i), X_2(\mathbf{s}_i), \dots, X_p(\mathbf{s}_i)]'$, $i = 0, 1, \dots, n$, estão disponíveis temos:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \text{sendo } \boldsymbol{\epsilon} \sim N(\mathbf{0}, \Sigma).$$

sendo:

$$\mathbf{X} = \begin{bmatrix} 1 & X_1(\mathbf{s}_1) & X_2(\mathbf{s}_1) & \dots & X_p(\mathbf{s}_1) \\ 1 & X_1(\mathbf{s}_2) & X_2(\mathbf{s}_2) & \dots & X_p(\mathbf{s}_2) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_1(\mathbf{s}_n) & X_2(\mathbf{s}_n) & \dots & X_p(\mathbf{s}_n) \end{bmatrix} \quad \text{e } \boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'.$$

Novamente a especificação de Σ pode ser feita com ou sem o efeito *nugget*.

Esta configuração do procedimento com covariáveis é muitas vezes denominada *Universal Kriging*. Cuidado! alguns autores usam o termo “Universal” em referência à situação em que apenas a longitude e a latitude são assumidas como covariáveis.

O *Ordinary Kriging* pode ser visto como um caso particular do *Universal Kriging* com $\mathbf{X} =$ vetor coluna $\mathbf{1}$ e $\boldsymbol{\beta} =$ escalar μ .

O problema de previsão é configurado como segue:

Buscamos pela função $f(\mathbf{y})$ que minimiza a média quadrática do erro de previsão dada por $E[(Y(\mathbf{s}_0) - f(\mathbf{y}))^2|\mathbf{y}]$, isto é, trabalhamos com a distribuição condicional de $Y(\mathbf{s}_0)|\mathbf{y}$.

Para simplificar, denote somente aqui: $Y_0 = Y(\mathbf{s}_0)$, $f_y = f(\mathbf{y})$. Então:

$$\begin{aligned} E[(Y_0 - f_y)^2|\mathbf{y}] &= \\ &= E\{[Y_0 - E(Y_0|\mathbf{y}) + E(Y_0|\mathbf{y}) - f_y]^2|\mathbf{y}\} \\ &= E\{[Y_0 - E(Y_0|\mathbf{y})]^2 + 2[Y_0 - E(Y_0|\mathbf{y})][E(Y_0|\mathbf{y}) - f_y] + [E(Y_0|\mathbf{y}) - f_y]^2|\mathbf{y}\} \\ &= E\{[Y_0 - E(Y_0|\mathbf{y})]^2|\mathbf{y}\} + \\ &\quad + 2E\{Y_0 E(Y_0|\mathbf{y}) - Y_0 f_y - E^2(Y_0|\mathbf{y}) + E(Y_0|\mathbf{y}) f_y|\mathbf{y}\} + E\{[E(Y_0|\mathbf{y}) - f_y]^2|\mathbf{y}\} \\ &= E\{[Y_0 - E(Y_0|\mathbf{y})]^2|\mathbf{y}\} + \\ &\quad + 2E(Y_0|\mathbf{y})E(Y_0|\mathbf{y}) - 2E(Y_0|\mathbf{y})f_y - 2E^2(Y_0|\mathbf{y}) + 2E(Y_0|\mathbf{y})f_y + [E(Y_0|\mathbf{y}) - f_y]^2 \\ &= E\{[Y_0 - E(Y_0|\mathbf{y})]^2|\mathbf{y}\} + [E(Y_0|\mathbf{y}) - f_y]^2 \end{aligned}$$

Conclusão: $E[(Y(\mathbf{s}_0) - f(\mathbf{y}))^2|\mathbf{y}] = E\{[Y(\mathbf{s}_0) - E(Y(\mathbf{s}_0)|\mathbf{y})]^2|\mathbf{y}\} + [E(Y(\mathbf{s}_0)|\mathbf{y}) - f(\mathbf{y})]^2$.

Visto que o segundo termo é não-negativo, temos:

$$E[(Y(\mathbf{s}_0) - f(\mathbf{y}))^2 | \mathbf{y}] \geq E\{[Y(\mathbf{s}_0) - E(Y(\mathbf{s}_0) | \mathbf{y})]^2 | \mathbf{y}\} \quad \text{para qualquer } f(\mathbf{y}).$$

Se a igualdade é válida $\Leftrightarrow f(\mathbf{y}) = E(Y(\mathbf{s}_0) | \mathbf{y})$.

Portanto, neste caso, o preditor $f(\mathbf{y})$ que minimiza o erro será $E(Y(\mathbf{s}_0) | \mathbf{y})$

Este resultado é bastante intuitivo do ponto de vista Bayesiano, visto que $f(\mathbf{y})$ é apenas a média *a posteriori* de $Y(\mathbf{s}_0)$.

Após identificar o formato do melhor preditor, considere agora a sua estimação. Utilizaremos primeiro uma situação não realística, na qual todos os parâmetros da população (β , σ^2 , ϕ e τ^2) são conhecidos.

Considere a seguinte configuração da distribuição Normal Multivariada:

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \sim N \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \right], \quad \text{sendo } \Omega_{21} = \Omega'_{12}.$$

Usando a teoria relacionada à distribuição Normal Multivariada, é possível mostrar que $p(\mathbf{Y}_1|\mathbf{Y}_2)$ é Normal com a seguinte média e variância:

$$\begin{aligned}E[\mathbf{Y}_1|\mathbf{Y}_2] &= \boldsymbol{\mu}_1 + \Omega_{12}\Omega_{22}^{-1}(\mathbf{Y}_2 - \boldsymbol{\mu}_2), \\ \text{Var}[\mathbf{Y}_1|\mathbf{Y}_2] &= \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}.\end{aligned}$$

No contexto de *Kriging*, assumimos $\mathbf{Y}_1 = Y(\mathbf{s}_0)$ e $\mathbf{Y}_2 = \mathbf{y}$, significando que: $\Omega_{11} = \sigma^2 + \tau^2$, $\Omega_{22} = \Sigma = \sigma^2 H(\phi) + \tau^2 I_n$ e $\Omega_{12} = \boldsymbol{\gamma}'$, sendo $\boldsymbol{\gamma}' = [\sigma^2 \rho(\phi; d_{01}), \dots, \sigma^2 \rho(\phi; d_{0n})]$.

Substituindo estes valores nas fórmulas acima para média e variância, obtemos:

$$\begin{aligned}E[Y(\mathbf{s}_0)|\mathbf{y}] &= \mathbf{X}(\mathbf{s}_0)' \boldsymbol{\beta} + \boldsymbol{\gamma}' \Sigma^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \\ \text{Var}[Y(\mathbf{s}_0)|\mathbf{y}] &= \sigma^2 + \tau^2 - \boldsymbol{\gamma}' \Sigma^{-1} \boldsymbol{\gamma}.\end{aligned}$$

Comentários:

- Supomos que o valor das covariáveis em $\mathbf{X}(\mathbf{s}_0)$ foram de fato observados no novo sítio \mathbf{s}_0 ;
- Se considerarmos a previsão em sítios com observações $Y(\mathbf{s}_i)$, surge a pergunta: $E[Y(\mathbf{s}_i)|\mathbf{y}] = Y(\mathbf{s}_i)$? Resposta: Sim (se $\tau^2 = 0$, i.e. sem *nugget*), Não (se $\tau^2 > 0$).

Em um cenário mais realístico (parâmetros do modelo são desconhecidos) os resultados são modificados como segue.

O preditor $f(\mathbf{y})$ será dado por: $\hat{f}(\mathbf{y}) = \mathbf{X}(\mathbf{s}_0)' \hat{\beta} + \hat{\gamma}' \hat{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta})$, sendo:

$$\hat{\gamma} = [\hat{\sigma}^2 \rho(\hat{\phi}; d_{01}), \dots, \hat{\sigma}^2 \rho(\hat{\phi}; d_{0n})]',$$

$$\hat{\Sigma} = \hat{\sigma}^2 H(\hat{\phi}) \text{ (sem } \textit{nugget}) \text{ ou } \hat{\Sigma} = \hat{\sigma}^2 H(\hat{\phi}) + \hat{\tau}^2 I_n \text{ (com } \textit{nugget}),$$

$\hat{\beta} = \hat{\beta}_{WLS} = (\mathbf{X}' \hat{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\Sigma}^{-1} \mathbf{y}$ é o estimador usual de mínimos quadrados ponderados para β (WLS - *Weighted Least Squares*).

Diggle e Ribeiro (2007) definem “*simple kriging*” como: a busca por $f(\mathbf{y})$ que minimiza a média quadrática do erro de previsão sob um modelo estacionário Gaussiano, entretanto, ignoramos aqui a incerteza dos parâmetros, isto é, estimativas de todos os parâmetros do modelo são inseridas nas equações de previsão como se fossem os valores reais. O livro ressalta que ignorar essa incerteza pode ser um problema.

No *Ordinary Kriging* teremos: $\hat{\beta}_{WLS} = \hat{\mu} = (\mathbf{1}' \hat{\Sigma}^{-1} \mathbf{1})^{-1} \mathbf{1}' \hat{\Sigma}^{-1} \mathbf{y}$.

Substituindo a expressão de $\hat{\beta}_{WLS}$ em $\hat{f}(\mathbf{y})$ teremos $\hat{f}(\mathbf{y}) = \boldsymbol{\lambda}'\mathbf{y}$, ou seja, o preditor *Kriging* pode ser escrito como uma combinação linear das observações em \mathbf{y} . Neste caso, $\boldsymbol{\lambda}'$ é um vetor ($1 \times n$) contendo pesos (*kriging weights*). Este resultado é mostrado a seguir...

$$\begin{aligned}
 \hat{f}(\mathbf{y}) &= \mathbf{X}(\mathbf{s}_0)' \hat{\beta} + \hat{\gamma}' \hat{\Sigma}^{-1} (\mathbf{y} - \mathbf{X} \hat{\beta}) \\
 &= \mathbf{X}(\mathbf{s}_0)' (\mathbf{X}' \hat{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\Sigma}^{-1} \mathbf{y} + \hat{\gamma}' \hat{\Sigma}^{-1} [\mathbf{y} - \mathbf{X} (\mathbf{X}' \hat{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\Sigma}^{-1} \mathbf{y}] \\
 &= \left[\mathbf{X}(\mathbf{s}_0)' (\mathbf{X}' \hat{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\Sigma}^{-1} \right] \mathbf{y} + \left[\hat{\gamma}' \hat{\Sigma}^{-1} \right] \mathbf{y} - \left[\hat{\gamma}' \hat{\Sigma}^{-1} \mathbf{X} (\mathbf{X}' \hat{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\Sigma}^{-1} \right] \mathbf{y} \\
 &= \left[\mathbf{X}(\mathbf{s}_0)' (\mathbf{X}' \hat{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\Sigma}^{-1} + \hat{\gamma}' \hat{\Sigma}^{-1} - \hat{\gamma}' \hat{\Sigma}^{-1} \mathbf{X} (\mathbf{X}' \hat{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\Sigma}^{-1} \right] \mathbf{y} \\
 &= \boldsymbol{\lambda}' \mathbf{y} \quad \text{sendo} \quad \boldsymbol{\lambda} = \hat{\Sigma}^{-1} \hat{\gamma} + \hat{\Sigma}^{-1} \mathbf{X} (\mathbf{X}' \hat{\Sigma}^{-1} \mathbf{X})^{-1} (\mathbf{X}(\mathbf{s}_0) - \mathbf{X}' \hat{\Sigma}^{-1} \hat{\gamma}).
 \end{aligned}$$

Se $\mathbf{X}(\mathbf{s}_0)$ não é observado, podemos estimá-lo conjuntamente com $Y(\mathbf{s}_0)$ através de algoritmo com iterações envolvendo:

$$\boldsymbol{\lambda} = \hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}(\mathbf{X}(\mathbf{s}_0) - \mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\gamma}}) \quad \text{e} \quad \hat{\mathbf{X}}(\mathbf{s}_0) = \mathbf{X}'\boldsymbol{\lambda}.$$

Esta última expressão foi obtida conforme mostrado abaixo:

$$\boldsymbol{\lambda} = \hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}(\mathbf{X}(\mathbf{s}_0) - \mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\gamma}})$$

Multiplicando ambos os lados por $\mathbf{X}' \dots$

$$\mathbf{X}'\boldsymbol{\lambda} = \mathbf{X}' \left[\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}(\mathbf{X}(\mathbf{s}_0) - \mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\gamma}}) \right]$$

$$\mathbf{X}'\boldsymbol{\lambda} = \mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\gamma}} + \mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}(\mathbf{X}(\mathbf{s}_0) - \mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\gamma}})$$

$$\mathbf{X}'\boldsymbol{\lambda} = \mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\gamma}} + I_{p+1}(\mathbf{X}(\mathbf{s}_0) - \mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\gamma}})$$

$$\mathbf{X}'\boldsymbol{\lambda} = \mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\gamma}} + \mathbf{X}(\mathbf{s}_0) - \mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\gamma}}$$

$$\mathbf{X}'\boldsymbol{\lambda} = \mathbf{X}(\mathbf{s}_0)$$

Teremos um algoritmo EM (*Expectation-Maximization*) com:

“Passo E” = cálculo de $\hat{\mathbf{X}}(\mathbf{s}_0)$ e “Passo M” = atualização de $\boldsymbol{\lambda}$.

No contexto clássico, estimativas via Máxima Verossimilhança Restrita (REML) são frequentemente inseridas nesta fórmula (argumenta-se que tem propriedades melhores).