



UNIVERSIDADE DE SÃO PAULO
CAMPOS PIRACICABA
Escola Superior de Agricultura "Luiz de Queiroz"

RELATÓRIO FINAL DE GEOESTATÍSTICA
Análise Geoestatística

Leila Sheila Silva Lisboa

Disciplina: LCE 5700 - Geoestatística
Prof. Paulo Justiniano

16 de novembro de 2011.

1 RESUMO

Este trabalho foi realizado com o intuito de fazer um estudo geoestatístico da área, objeto do trabalho. Neste estudo foram seguidos os passos dados na disciplina em questão: i) arrumar os dados; ii) realizar a análise exploratória dos dados; iii) Estimar os parâmetros e iv) realizar a interpolação espacial dos dados.

O atributo analisado foi a Soma de Base – SB. Os valores da SB, da capacidade de troca de cátions (T) e da porcentagem de saturação em bases (V) são parâmetros muito importantes no levantamento de solo. A soma de bases é obtida pela soma dos valores dos cátions trocáveis Ca, Mg, K e Na.

2 INTRODUÇÃO

A Geoestatística é um ramo da Estatística Espacial onde os dados consistem de finitas amostras de medições relativas a um fenômeno espacialmente contínuo (Diggle & Ribeiro Jr.). Estes dados devem possuir como característica valores observados (fenômeno de interesse) numa série discreta de localizações amostrais (coordenadas geográficas).

O termo Geoestatística foi criado por Georges Matheron para resolver problemas decorrentes da previsão espacial na indústria de mineração e com o avanço das metodologias hoje ela é usada em muitas áreas: minas, geofísica, geoquímica, floresta, ambiente, hidrologia, oceanografia, meteorologia, entre outros.

Neste estudo o atributo analisado foi a Soma de Base – SB, parâmetro muito importante em estudos sobre o levantamento de solo.

3 MATERIAIS E MÉTODOS

Este trabalho foi realizado no pacote computacional de análise estatística de dados - R, programa gratuito e de código aberto.

O dado utilizado, SB, é proveniente da análise de solo de uma propriedade estudada pelo professor Peterson da ESALQ, com 470 localizações georeferenciadas amostradas. Para este relatório o arquivo original denomina-se “Solos_SB.txt”, que apresenta 470 observações e 03 atributos: Coordenadas (atributo 1 e 2) e variável Soma de Base (atributo da coluna 3).

3.1 Análise Exploratória dos dados

O fenômeno analisado foi a “Soma de Base”, é um arquivo que se encontra no formato txt, o qual pode ser lido pelo pacote R. Este arquivo foi convertido para um objeto da classe Geodata, por possuir dois componentes principais: coordenadas e dados. A seguir este objeto foi plotado (Figura 01), pela função *plot()*, resultando em gráficos descritivos possibilitando a primeira análise exploratória dos dados. Por estes gráficos descritivos verificou-se a necessidade do teste de normalidade, pois, visualmente, notou-se que os dados não estavam padronizados.

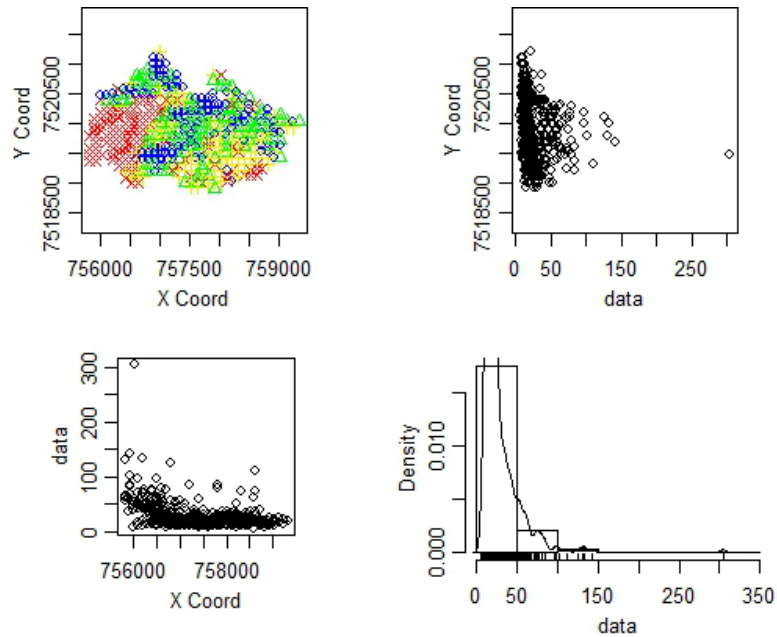


Figura 01 – Gráfico descritivo do padrão espacial da Soma de Base.

Através da função *summary()* foi possível também a análise exploratória dos valores descritivos (Tabela 01).

Tabela 01 – Resumo dos valores dos dados originais					
Number of data points: 470					
Coordinates summary					
	X	Y			
min	755805.2	7518918			
max	759331.2	7521204			
Distance summary					
	min	max			
	76.57719	3581.43271			
Data summary					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5.80	14.30	19.25	27.08	31.65	304.30

3.1.1 Teste de Normalidade

Foi realizado um teste de normalidade destes dados. Teste Shapiro-Wilk (Tabela 02), em que o p-value foi menor que 5%. Portanto, pode-se afirmar com um nível de significância de 5% que estes dados não provém de uma distribuição normal. Logo, foi necessária a transformação dos dados.

Tabela 02 – Teste de Normalidade

```
shapiro-wilk normality test
data: Sum_Base$data
w = 0.6229, p-value < 2.2e-16
```

3.1.2 Transformação dos dados

Foi realizada a requisição do pacote MASS - `require(MASS)` para que a função `boxcox` se torna-se disponível. A função `boxcox` é responsável por transformar os dados em uma distribuição normal. O gráfico gerado (Figura 02) mostra que o valor que maximiza a função é aproximadamente $\lambda = -0,6$.

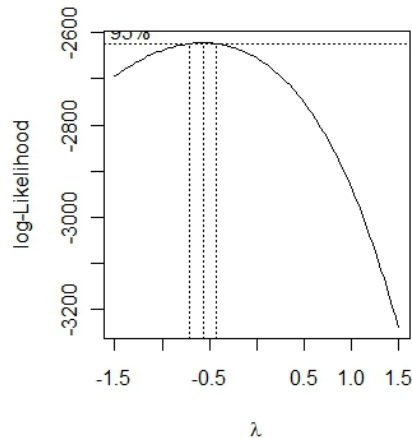


Figura 02 – Gráfico da transformação da variável Soma de Base pela função BoxCox.

3.1.3 Comparação entre os dados originais e os dados transformados

Após a transformação dos dados foi realizada uma comparação com os dados originais. A Figura 03 mostra os gráficos de círculos da variável estudada, onde os círculos representam os valores mensurados do atributo nas 470 localizações de uma determinada propriedade, enquanto que a Figura 04 mostra a comparação realizada através de histogramas.

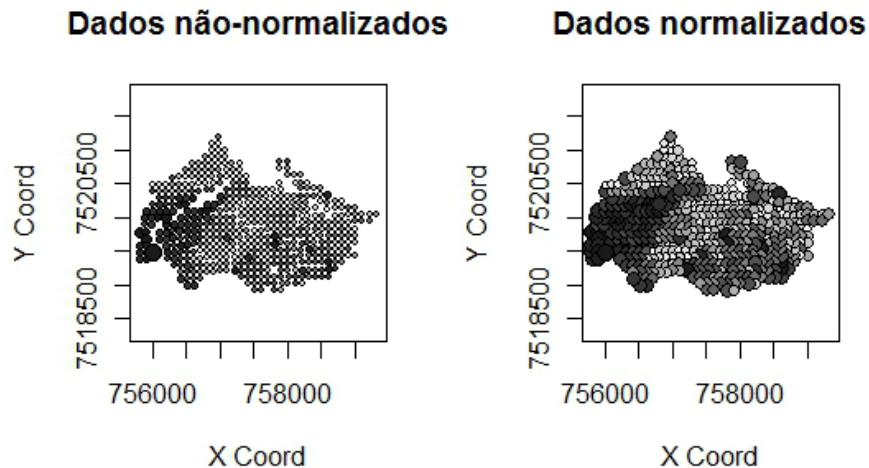


Figura 03 – Comparação dos dados originais e transformados através de gráficos de círculos pela função `points()`

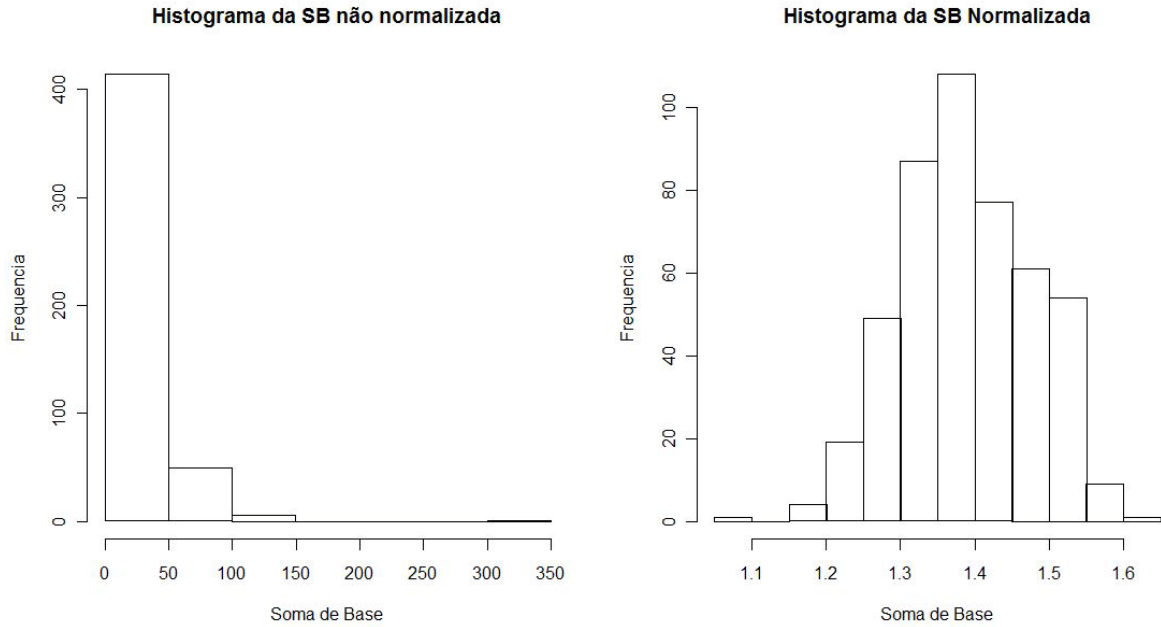


Figura 04 – Comparação dos dados originais e transformados através do histograma.

3.1.4 Análise exploratória nos dados transformados

A Figura 05 mostra os gráficos descritivos após os dados serem transformados.

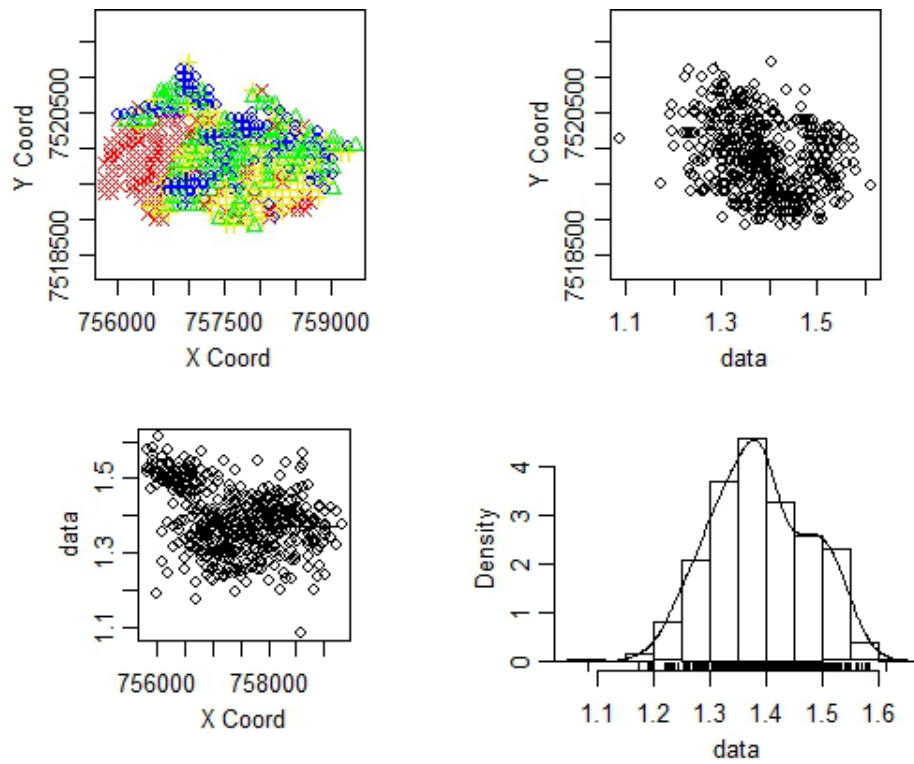


Figura 05 – Gráficos descritivos dos dados transformados

3.2 Métodos de estimação

Para detectar e quantificar as escalas da variabilidade espacial usou-se o semivariograma e fez-se um ajuste dos dados através da função *eyefit()*, usando a função *powered.exponential*, obtendo os parâmetros mostrados na Tabela 03.

Tabela 03 - Parâmetros estimados através do eyefit.

Parâmetro(Dist. Max.)	Covariância dos parâmetros (Cov.pars)			
	sigmasq	phi	tausq	kappa
3.443,68	0,0126	2.419,89	0,00157	0,6

3.2.1 Estimação pelo variograma

Modelo em função dos parâmetros de variabilidade e correlação pode ser ajustado através de métodos que ajustam um modelo de função de semivariância dos dados, sendo que os parâmetros de tal função correspondem aos parâmetros do modelo.

Os dados foram estimados através a função *variog()* de 02 maneiras: i) com média constante “trend=cte” e com tendência linear “trend=1st” (Figura 06).

O resultado do ajuste dos variogramas foi utilizado nos modelos exponencias, esféricos e gaussianos utilizando o método de estimação dos mínimos quadrados – OLS. Como os semivariogramas foram ajustados de 2 maneiras (com e sem tendência) e para o método de estimação foram utilizados três modelos, foram gerados (2 x 3) gráficos mostrando o comportamento dos dados (Figura 07).

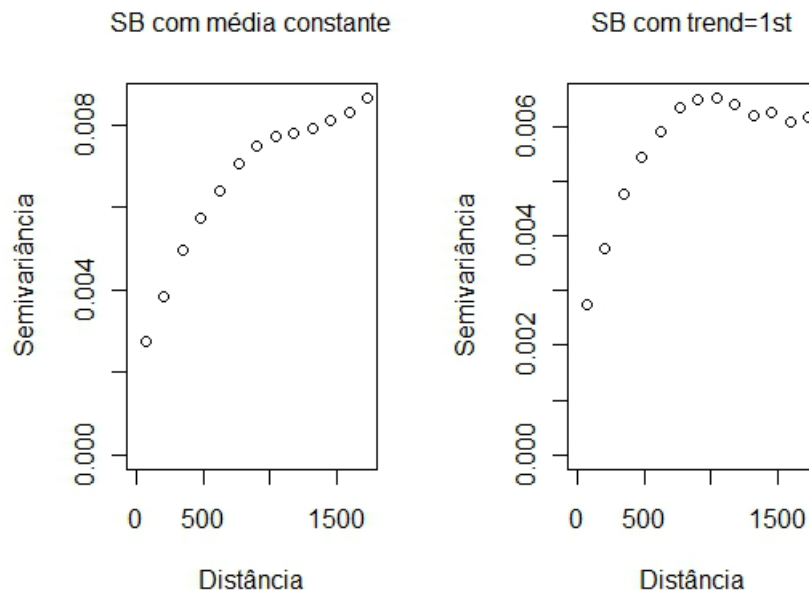


Figura 06 – Semivariogramas ajustados com trend=constante e trend=1st.

Depois foram especificados os modelos (Gaussiano, exponencial, esférico) para ver entre eles qual o melhor ajuste.

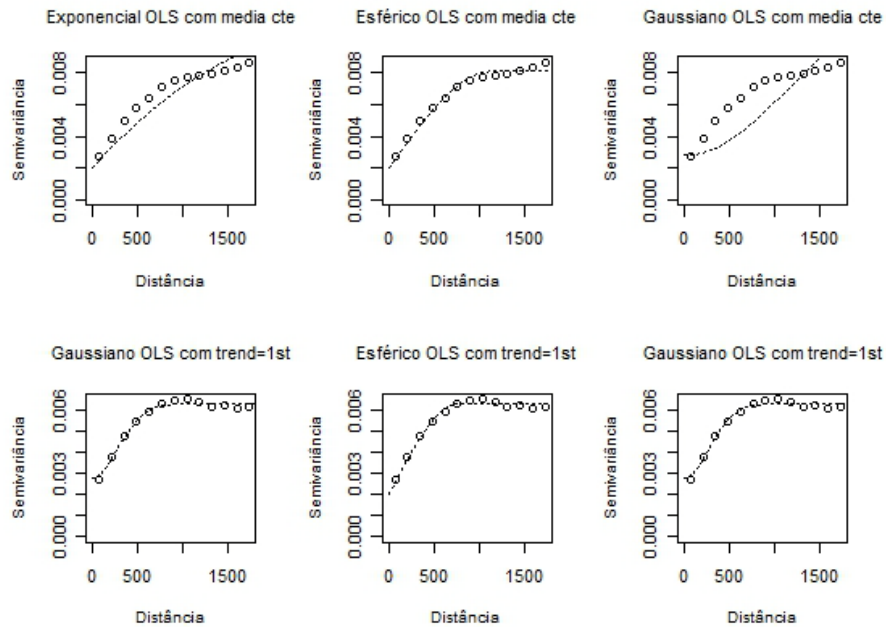


Figura 07 – Modelos gerados pelos métodos citados acima

Pela Figura 07 podemos concluir que:

- Com a média constante (trend=cte), o melhor modelo foi o esférico.

- Com a aplicação da tendência linear (trend=1st), ocorreram algumas dúvidas e não havia tempo para tentar refazer a análise, achei este resultado estranho, devo ter errado nos parâmetros.

Após esse passo tive dificuldades de passar adiante, e não tive tempo de verificar onde eu havia errado.

3.2.2 Estimação pela Máxima verossimilhança

Utilizou-se o modelo de máxima verossimilhança com os seguintes kappas: 0,2; 0,5; 1,5 e 3,5 com dois métodos: média constante e trend=1st. Esta estimação foi realizada através do método da máxima verossimilhança, *likfit()*. Logo oito modelos e o que apresentou o menor valor foi o modelo com kappa=3,5 e com trend=1st.

Foi utilizado o critério de AIC para ajudar na seleção do melhor modelo. O menor valor de AIC foi o modelo escolhido: st_ml3.5 (Figura 08).

```

> m10.2$AIC
[1] -1304.989
> m10.5$AIC
[1] -1301.058
> m11.5$AIC
[1] -1263.232
> m13.5$AIC
[1] -1160.475
> st_m10.2$AIC
[1] -1306.159
> st_m10.5$AIC
[1] -1297.76
> st_m11.5$AIC
[1] -1259.274
> st_m13.5$AIC
[1] -1157.362
>

```

Figura 08 – Resultado do critério de Akaike - AIC

3.3 Krigagem

Com o modelo escolhido o próximo passo foi realizar a Krigagem.

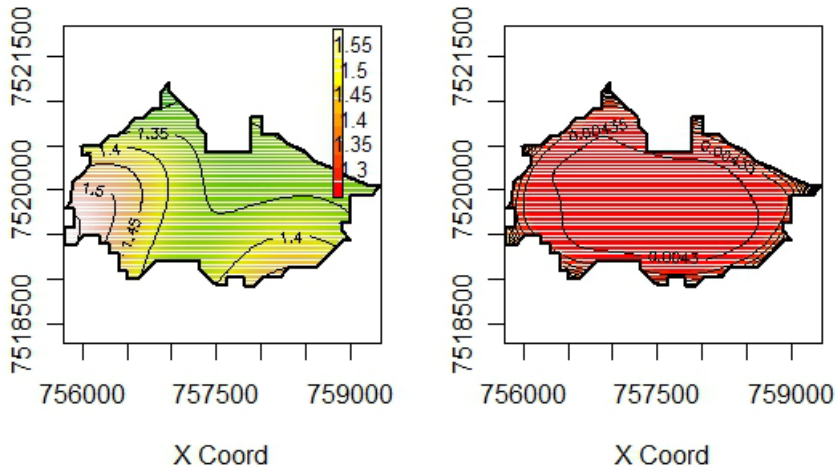


Figura 09 – Resultado da krigagem

3.3.4 Análise de resíduos

O teste de normalidade do Shapiro-Wilk resultou em um p-valor de 0,0001845, ou seja, os resíduos não estão numa distribuição normal;

Tabela – Teste de normalidade dos resíduos

```

shapiro-wilk normality test

data:  st_m13.5$model.components[, 3]
w = 0.9861, p-value = 0.0001845

```

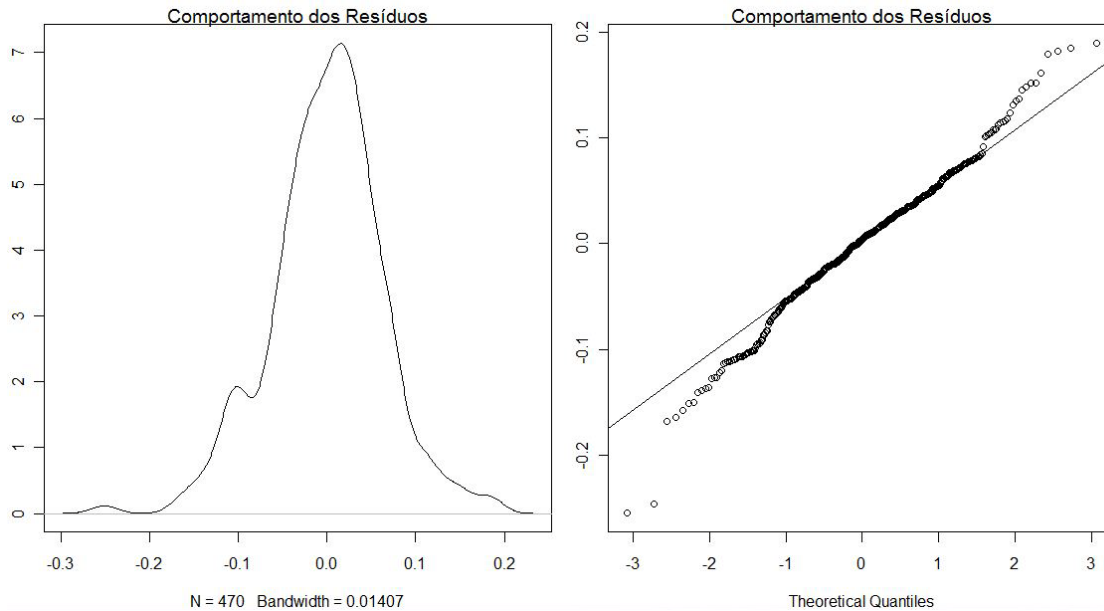



Figura – Densidade e gráfico dos quartis dos resíduos para o modelo final.

4 CONCLUSÃO

Este trabalho foi muito importante para mim, pois apesar de não ter conseguido discutir os resultados, reafirmei com esta experiência a importância da geoestatística para tentar perceber e analisar os padrões espaciais.

As dificuldades foram grandes, primeiro por tentar aprender um pouco a usar o pacote R e segundo pela base teórica necessária para a realização do trabalho.

5 REFERÊNCIA BIBLIOGRÁFICA

Diggle, Peter J. & Ribeiro Jr, Paulo Justiniano. *Model-based Geostatistics*. Series: Springer Series in Statistics. 2007, X, 230 p., Hardcover.

Fonseca, B.H.F. *Um estudo sobre estimação e predição em modelos geoestatísticos bivariados*. Dissertação (Mestrado). Piracicaba, 2008.