

Análise de teores de nitrogênio no solo de três regiões da Mata Atlântica: uma abordagem geoestatística

Everton Batista da Rocha - LCE, ESALQ/USP¹

Resumo

A Mata Atlântica é um importante bioma situado na costa brasileira, de norte a sul. Ainda não se tem informações suficientes disponíveis sobre o funcionamento biogeoquímico da Floresta Atlântica na faixa litorânea ao longo do gradiente altitudinal, o que é vital em programas de preservação e restauração de sua fauna e flora. O objetivo principal deste trabalho foi estudar o comportamento do teor de nitrogênio neste bioma na Floresta Ombrófila Densa de Terras Baixas a 100m de altitude, na Floresta Ombrófila Densa Submontana a 400m de altitude e na Floresta Ombrófila Densa Montana a 1000m de altitude. E utilizando de métodos geoestatísticos de análise, um procedimento no programa *R2.13.1* foi implementado, considerando o pacote *geoR*. Observa-se que existe diferença no teor de nitrogênio nas três regiões, confirmando a hipótese de que as regiões mais altas possuem um maior teor de nitrogênio.

1 Introdução

De acordo com Myers *et al.* (2000) e Ribeiro *et al.* (2009), a Mata Atlântica situada na costa brasileira, originalmente ocupava uma área de 148.194.638 ha, abrangendo muitos tipos de ecossistemas tropicais e possuindo uma das maiores biodiversidades e endemismos do mundo. Martins (2010) *apud* Ribeiro *et al.* (2009), elucida que desde 1500, quando teve início a colonização da costa brasileira pelos portugueses, a Floresta Atlântica foi submetida a diversos tipos de exploração, e após 500 anos da chegada dos portugueses e diversos ciclos agrícolas como a cana de açúcar, café, algodão, pastagem, e ultimamente a pressão da urbanização, muito da Floresta Atlântica foi perdida, visto que restam apenas 11,73% (16.377.472 ha) da vegetação original, distribuída em fragmentos de tamanhos e estados de conservação diferentes. Ainda com base em Martins (2010), é importante ressaltar que embora a faixa litorânea da Mata Atlântica seja próxima a grandes centros urbanos como São Paulo e Rio de Janeiro, a região

¹Doutorando em Estatística e Experimentação Agronômica: e.batista.rocha@usp.br

mais densamente povoada do Brasil, incluindo grandes universidades e centros de pesquisa, paradoxalmente sabe-se muito menos sobre a floresta do que a floresta Amazônica, situada em área remota com capacidade científica reduzida se comparada com a região sudeste do Brasil.

Na literatura, não há informações suficientes disponíveis sobre o funcionamento biogeoquímico da Floresta Atlântica na faixa litorânea. E sabe-se que o domínio de uma estratégia de ciclagem de nutrientes é crucial nos programas de regeneração da floresta. Isto é especialmente importante com base no decreto que regulamenta a Lei da Mata Atlântica (lei 11.428 de 22 de Dezembro de 2006), assinado no governo do Presidente Luís Inácio Lula da Silva, para aumentar a área de preservação da floresta de 7% para 20%.

A primeira etapa para se compreender completamente o funcionamento biogeoquímico básico de uma floresta tropical é avaliar seus tipos de solos e fertilidade, uma vez que este reservatório é o principal provedor de nutrientes para as plantas, e conseqüentemente, para o ecossistema inteiro. Neste estudo foi investigado o teor de nitrogênio no solo em três diferentes tipos de florestas, distribuídas ao longo do gradiente altitudinal da Floresta Atlântica, com o objetivo de comparação entre elas.

Em ecossistemas florestais a ciclagem de nutrientes ocorre envolvendo o processo de transferência de minerais acumulados na biomassa vegetal para o solo. De acordo com Gama-Rodrigues (2004), o processo de ciclagem de nutrientes pode ser avaliado de acordo com o objetivo do estudo e é composto pelo ciclo geoquímico, biogeoquímico e bioquímico. No ciclo geoquímico as principais entradas de nutrientes ocorrem via intemperismo, adições atmosféricas, fixação biológica de nitrogênio e fertilização. A saída desses nutrientes ocorre via lixiviação, fixação pela fase mineral do solo, escoamento superficial, erosão, denitrificação, volatilização por queima de resíduos e colheita. O ciclo bioquímico envolve a translocação de nutrientes de tecidos velhos e fotossinteticamente inativos para os tecidos novos da planta e com grande atratividade de crescimento. O ciclo biogeoquímico inicia-se com o processo de absorção e acúmulo de nutrientes na biomassa, sua alocação nos diferentes compartimentos da planta e renovação das raízes, incorporação do mesmo no solo mediante a decomposição e lixiviação da serapilheira acumulada e conclui-se com a reabsorção do nutriente pela planta. Dentre os ciclos biogeoquímicos, estão o do nitrogênio (N). As variações nas concentrações deste nutriente, somado a outros nutrientes, entre folheto e serapilheira, são indicadoras da intensidade de ciclagem biogeoquímica dos nutrientes da serapilheira.

De acordo com Martins (2010) *apud* Feger e Raspe (1998), Vogel, Schumacher e Trüby (2007), o estudo da ciclagem de nutrientes (ciclo biogeoquímico) desempenha um papel importante no conhecimento das condições e dinâmica dos processos internos dos ecossistemas naturais, que

auxiliam no entendimento das rápidas mudanças provocadas pela exploração vegetal do meio ambiente.

A hipótese deste trabalho é que a relatividade da abundância ou pobreza de determinado nutriente variará ao longo do gradiente altitudinal, visto que ecossistemas mais antigos tendem a ser menos limitados por nitrogênio que ecossistemas jovens, portanto, espera-se que os maiores teores de nitrogênio sejam observados nas maiores altitudes, onde os sistemas são mais antigos que a Mata de Restinga (Figura 1).

Para este estudo, para a análise de dados, considerou-se o uso de técnicas geoestatísticas, dada a forma de dependência nos dados, de forma a levar em consideração a dependência espacial do atributo em estudo (teor de nitrogênio) na análise. Com base nestas técnicas, para a análise, um procedimento computacional foi implementado no programa *R2.13.1*, utilizando o pacote *geoR*.

2 Material e Métodos

Com quase 315.000 ha, numa extensão que vai desde a divisa de São Paulo com o Rio de Janeiro até Itariri, no sul do estado, o Parque Estadual da Serra do Mar (PESM), criado em 1977 através do Decreto Estadual nº10.251,17 de 30 de agosto de 1977 (posteriormente alterado pelo Decreto Estadual nº 13.313 de 06 de março de 1979), representa a maior porção contínua preservada de Mata Atlântica do Brasil. No seu limite norte, o PESM é administrado por Núcleos, que são bases instaladas em áreas de domínio do estado. No presente trabalho as áreas de estudo ficarão limitadas aos Núcleos Picinguaba e Santa Virgínia do Parque Estadual da Serra do Mar. Para facilitar a apresentação dos resultados do presente trabalho foi adotado a seguinte classificação das áreas estudadas: **Região 1** - Floresta Ombrófila Densa de Terras Baixas (100m de altitude); **Região 2** - Floresta Ombrófila Densa Submontana (400m de altitude); **Região 3** - Floresta Ombrófila Densa Montana (1000m de altitude) (Figura 1)

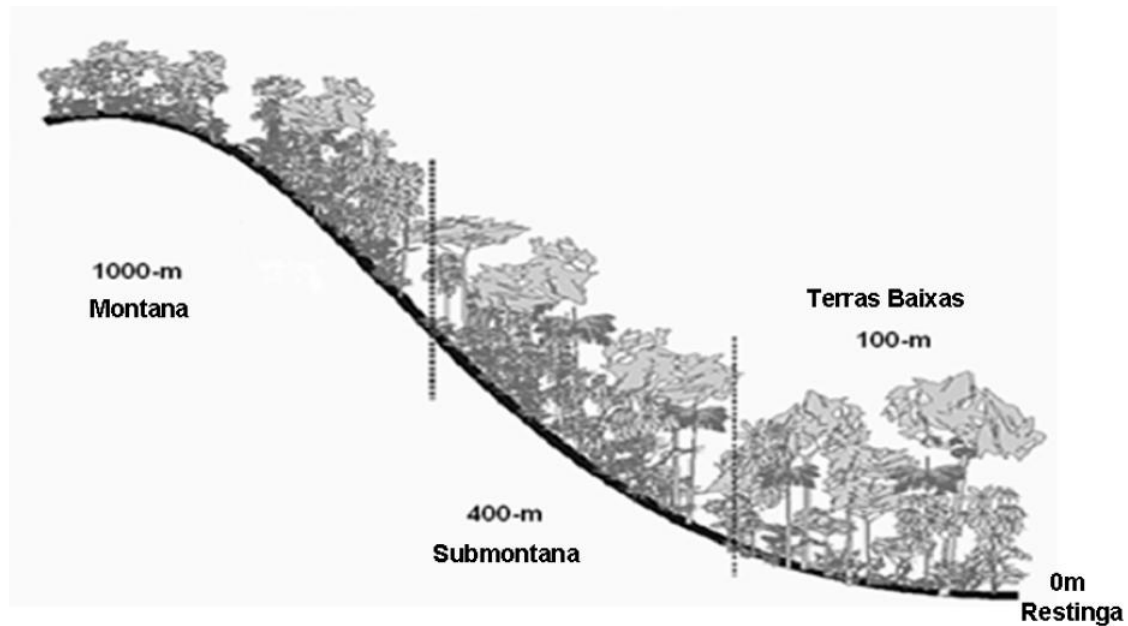


Figura 1: Gradiente altitudinal entre as áreas estudadas no Parque Estadual da Serra do Mar, Estado de São Paulo.

Na latitude das três áreas de estudo, as parcelas para a condução do experimento de decomposição da serapilheira e amostragem do solo, através da qual foram determinados os teores de nitrogênio, foram alocadas nas seguintes fitofisionomias e altitudes (Figura 2): Parcelas B, e E na Floresta Ombrófila Densa de Terras Baixas a 100m de altitude, Parcelas G e J na Floresta Ombrófila Densa Submontana a 400m de altitude e Parcelas K e N na Floresta Ombrófila Densa Montana a 1000m de altitude. As demais parcelas demarcadas na Figura 2 foram consideradas para caracterização mineralógica, química e física das regiões, que não são tratadas neste trabalho. A localização das parcelas do estudo foram realizadas com base nas diferenças altitudinais entre as mesmas (toposequências). No trecho amostral selecionado em cada topossequência foram alocadas parcelas de 100×100 m, totalizando 1 ha, subdivididas em sub-parcelas contíguas de 10×10 m ($100m^2$), de forma a manter a mesma altitude. Essa alocação da parcela maior e das sub-parcelas em cada área foi feita por equipe especializada de topografia, usando teodolito de precisão. Tanto a parcela como as sub-parcelas foram delimitadas com estacas permanentes. A parcela e as subparcelas foram georreferenciadas por uma equipe de topografia, de modo a possibilitar estudos de longo prazo e monitoramento contínuo nessas áreas, considerando as coordenadas U.T.M.. Todos os procedimentos de implantação das parcelas e da marcação e mapeamento dos indivíduos seguiram os protocolos propostos pelo RAINFOR (Rede Amazônica de Inventários Florestais) ver [http : //www.geog.leeds.ac.uk/projects/raifor/](http://www.geog.leeds.ac.uk/projects/raifor/) para estabelecimento e monitoramento de parcelas permanentes em florestas tropicais.

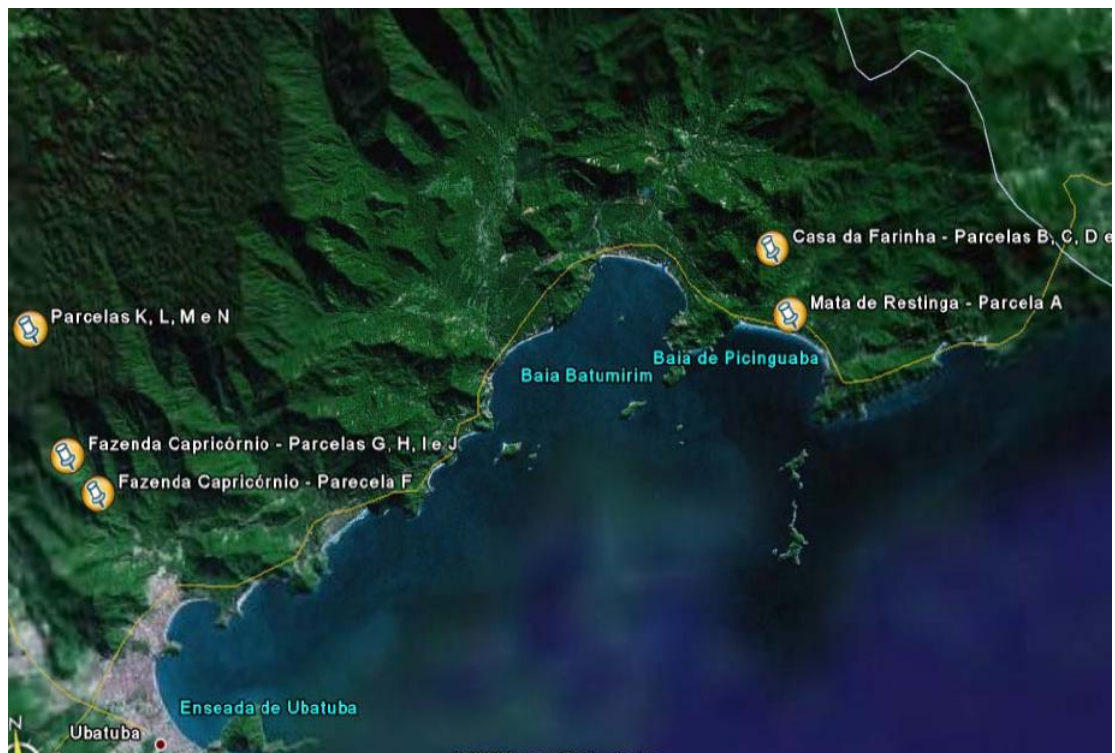


Figura 2: Distribuição das áreas amostrais do Parque Estadual da Serra do Mar, no Estado de São Paulo.

Foram feitas 32 tradagens nas camadas 0-5; 5-10; 10-20; 20-30; 30-40; 40-50; 50-75; 75-100cm para coleta de amostras deformadas (Figura 3). Note que na Figura 3 há a demarcação de apenas 16 tradagens ($T1 \dots, T16$), pois esta refere-se apenas uma parcela, por exemplo, na Região 1, a parcela B, que foi alocada na Floresta Ombrófila Densa de Terras Baixas, com 16 tradagens e o mesmo na parcela E, totalizando as 32 tradagens nesta região.

T1			T2			T3			T4
T5			T6			T7			T8
T9			T10			T11			T12
T13			T14			T15			T16

Figura 3: Localização dos pontos de coleta no interior da parcela de $100 \times 100m$. T1 ao T16: coleta para a determinação do teor de nitrogênio no solo.

O espaçamento das tradagens foi de 30cm, disposto na forma de grade quadrada. Para cada uma das profundidades e horizontes, foram apresentados valores médios de atributos químicos e granulométricos, e assim, também determinado o teor de nitrogênio.

Diante da estrutura de dados é de se esperar uma dependência espacial nas observações. Seguindo a hipótese do estudo, o objetivo é verificar se existe diferença entre os teores de nitrogênio nas três regiões em estudo. Para isto será considerado o ajuste de modelos geoestatísticos considerando diferentes cenários.

Do ponto de vista univariado o formato básico de um conjunto de dados geoestatístico é,

$$(x_i, y_i) : i = 1, \dots, n \quad (1)$$

em que x_i identifica a localização espacial, tipicamente em um espaço de duas dimensões, e y_i é um valor escalar associado com a localização x_i . É comum nomear y_i como variável medida, resposta ou atributo.

Para se certificar de que o conjunto de dados é uma estrutura de dados de geoestatística, no sentido de ter uma dependência espacial, neste trabalho é considerada uma análise exploratória espacial (ou canônica). Nesta análise é realizada uma visualização dos pontos na área de estudo, inspeção gráfica da distribuição das observações, que espera-se ser normal, e em caso contrário adota-se a Transformação de Box-Cox. Essa análise inicial é útil também na detecção de valores outliers e até mesmo informações errôneas no conjunto de dados. Através do variograma, é realizado o Teste de Monte-Carlo, para verificação de significância do parâmetro de dependência espacial, cuja hipótese nula é $H_0 : \phi = 0$, ou seja, H_0 : *os dados não possuem uma dependência espacial*. Neste caso, para a realização de uma análise geoestatística, deve-se observar a rejeição desta hipótese.

Considerando a estrutura de dados de geoestatística em (1), na ausência de qualquer variável explanatória, o modelo geoestatístico precisa incluir um valor real que represente uma “flutuação” dos valores y_i em torno de uma média, em função de sua localização, x_i , dito processo estocástico espacial gaussiano, $S(x)$. Um processo espacial gaussiano, $\{S(x) : x \in \mathbb{R}^2\}$, é um processo estocástico com a propriedade de que para qualquer coleção de localizações x_1, \dots, x_n em que cada $x_i \in \mathbb{R}^2$, a distribuição conjunta de $S = \{S(x_1), \dots, S(x_n)\}$ é uma normal multivariada. Desta forma, qualquer processo deste tipo é completamente especificado em termos das suas funções de média, $\mu(x) = E[S(x)]$, e função de covariância, $\gamma(x, x') = Cov\{S(x), S(x')\}$. Neste contexto, o variograma de um processo estocástico espacial $S(x)$ é dado pela função,

$$V(x, x') = \frac{1}{2} Var\{S(x) - S(x')\} \quad (2)$$

As especificações em (1) e (2) referem-se ao caso univariado, entretanto neste estudo, uma vez que os teores de nitrogênio foram observados em 3 regiões e deseja-se verificar a diferença entre eles, levando em consideração a dependência espacial, a estrutura de dados é multivariada (trivariada), de forma que as ocorrências em cada uma das regiões será considerada como uma variável. Modelos geoestatísticos multivariados são relevantes quando duas ou mais diferentes variáveis respostas são medidas em localizações espaciais dentro de uma região contínua espacial.

Um processo espacial d -dimensional é uma coleção de variáveis aleatórias $Y(x) = \{Y_1(x), \dots, Y_d(x)\}$, em que $x \in \mathbb{R}^2$. Então, a função de covariância de $Y(x)$ é uma matriz funcional $d \times d$ de valores $\Gamma(x, x')$, cujo (j, k) -ésimo elemento é

$$\gamma_{jk}(x, x') = Cov\{Y_j(x), Y_k(x')\}.$$

Para a construção de um modelo gaussiano estacionário, por exemplo, para dados trivariados ($Y_{ij} : i = 1, \dots, n_j, j = 1, 2, 3$) medidas as localizações x_{ij} , deve-se primeiro especificar um modelo para um processo estacionário gaussiano não observado $\{S(x) = (S_1(x), S_2(x), S_3(x)) : x \in \mathbb{R}^2\}$, com média zero e variâncias $\sigma_j^2 = Var\{S_j(x)\}$ e estrutura de correlação determinada pelas 6 funções $\rho_{11}(u) = Corr\{S_1(x), S_1(x - \mu)\}$, $\rho_{22}(u) = Corr\{S_2(x), S_2(x - \mu)\}$, $\rho_{33}(u) = Corr\{S_3(x), S_3(x - \mu)\}$, $\rho_{12}(u) = Corr\{S_1(x), S_2(x - \mu)\}$, $\rho_{13}(u) = Corr\{S_1(x), S_3(x - \mu)\}$ e $\rho_{23}(u) = Corr\{S_2(x), S_3(x - \mu)\}$, em que $u = x - x'$.

A mais simples suposição que se faz sobre os dados Y_{ij} é que $Y_{ij} = S_j(x_{ij})$, ou seja, o sinal em qualquer localização x pode ser observado sem erros. Quando os dados estão sujeitos a erros de medição, a simples suposição é que Y_{ij} são mutuamente independentes dado $S(\cdot)$ e normalmente distribuídos,

$$Y_{ij} \sim N\{\mu_j(x_{ij}) + S_j(x_{ij}), \tau_j^2\} : i = 1, \dots, n_j; j = 1, 2, 3 \quad (3)$$

Sob este modelo, cada dimensão da variável resposta separadamente segue um modelo gaussiano univariado, enquanto a dependência entre as três dimensões de resposta é modelada indiretamente através da estrutura do processo não observado $S(\cdot)$. De acordo com Diggle e Ribeiro Jr. (2007), a suposição de independência condicional em (3) leva a interpretação de que o parâmetro τ_j^2 representa as variações dos erros de medição em uma das dimensões da resposta. No caso dos termos de erro serem espacialmente independentes, a média e estrutura

de covariância dos dados, Y_{ij} , são dadas por,

$$\begin{aligned} E[Y_{ij}] &= \mu_j(x_{ij}) \\ Var\{Y_{ij}\} &= \tau_j^2 + \sigma_j^2 \end{aligned}$$

e, para $(i, j) \neq (i', j')$

$$Cov\{Y_{ij}, Y_{i'j'}\} = \sigma_j \sigma_{j'} \rho_j j' (\|x_{ij} - x_{i'j'}\|).$$

3 Resultados e Discussão

Considere inicialmente verificar a distribuição dos pontos onde foram amostrados os teores de nitrogênio, nas três regiões, seguindo o esquema apresentado na Figura (3). Nas Figuras 4, 5 e 6 são apresentadas as distribuições dos pontos amostrais nas regiões 1, 2 e 3, respectivamente. Nota-se que há indícios de valores outliers nos dados referentes às regiões 1 e 3, todavia estes devem permanecer na análise, pois sua grandeza é significativa na interpretação biogeoquímica do solo.

Verifica-se ainda que há indícios de padrão espacial, devido o comportamento dos pontos. Para a verificação deste indício, será considerado o teste de Monte-Carlo, baseado no variograma, considerando todos os dados, das três regiões, devido ao baixo número de observações. Uma vez que será considerado todos os pontos, para a obtenção do variograma, no cálculo da distância máxima foi considerado metade do valor da diferença entre a distância máxima e mínima, para a região que apresentou o menor valor para a distância máxima (veja Tabela 1). Desta forma espera-se não penalizar o variograma por utilizar todas as regiões. Caso esta medida não fosse considerada, na função variograma seria considerada a distância *entre* pontos da região 1, 2 e 3, e não há interesse na mesma, mas sim nas distâncias *dentro* (*dentre*) de cada uma das regiões. E ainda, dado o espaço de distância entre as regiões, isso influenciaria na detecção de dependência espacial nos dados, e conseqüentemente no teste de Monte-Carlo.

Antes de aplicar o teste de Monte-Carlo verificou-se a normalidade dos dados para as regiões, e notou-se que as regiões 1 e 3 não apresentavam normalidade (os valores de p , para o teste de Shapiro-Wilk, para as regiões 1, 2 e 3, respectivamente, foram: 0,0005, 0,4742 e 0,0024). Aplicou-se então a transformação de Box-Cox, $\log()$, nos dados das 3 regiões. Nas regiões 1 e 3 para corrigir a não normalidade e na região 2 para manter todos os dados na mesma escala. Após a transformação, uma análise gráfica dos dados apresentou resultados mais satisfatórios (Figura 7). E ainda, aplicando-se novamente o teste de Shapiro-Wilk para os dados, na escala

transformada (log), obteve-se para as regiões 1, 2 e 3 os respectivos valores de p , 0,4265, 0,9635 e 0,8124, não rejeitando claramente a hipótese de normalidade dos dados.

Corrigida a não normalidade dos dados, obteve-se o variograma para os dados e aplicou-se o teste de Monte-Carlo, baseado no variograma. Nota-se que o gráfico apenas do variograma, mesmo considerando a aproximação da Tabela 1, não é totalmente conclusivo, a respeito da dependência espacial (Figura 8 (a)), entretanto, com base no teste de Monte-Carlo (Figura 8 (b)), verifica-se que a hipótese de não dependência espacial é rejeitada, ou seja, rejeita-se a hipótese de que $\phi = 0$, pois, nota-se que a maior parte dos pontos estão fora do “envelope”. Desta forma, para os dados em estudo, métodos geoestatísticos podem ser aplicados.

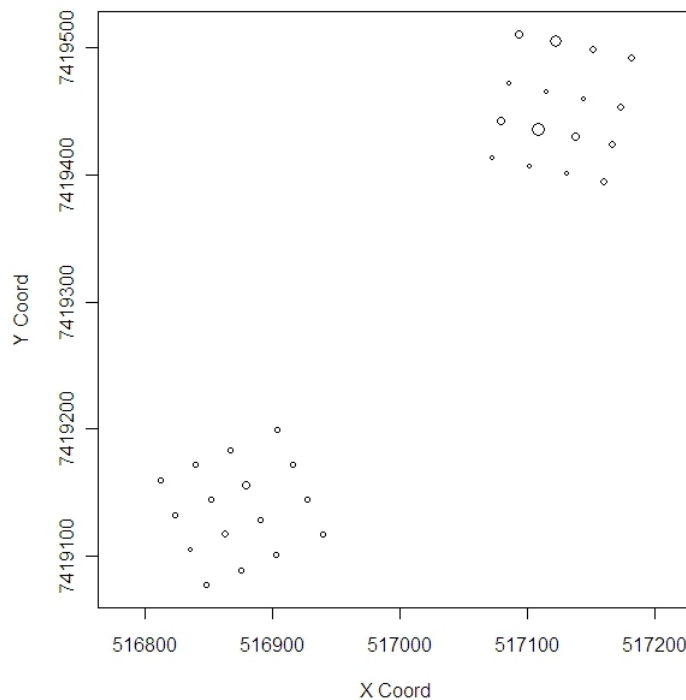


Figura 4: Distribuição dos pontos amostrais nas parcelas B e E, na Floresta Densa de Terras Baixas a 100m de altitude.

Tabela 1: Distâncias para cada uma das regiões e distância máxima considerada no variograma.

Região	Distância		Valor ²
	Mínima	Máxima	
1	29,99584	532,27623	$(532,27623-29,99584)/2$
2	27,75599	1070,2664	—
3	29,99424	926,75381	—

²Valor considerado para o cálculo da distância máxima no variograma.

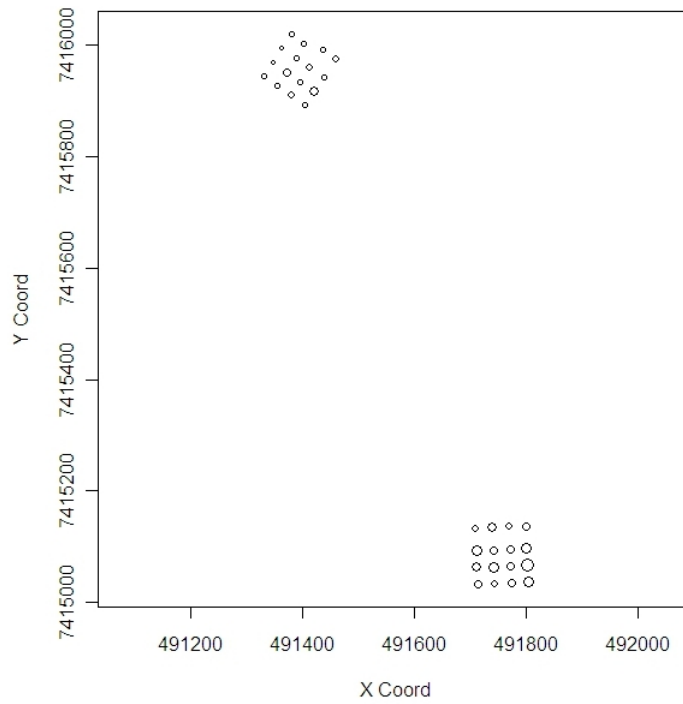


Figura 5: Distribuição dos pontos amostrais nas parcelas G e J, na Floresta Ombrófila Densa Submontana a 400m de altitude.

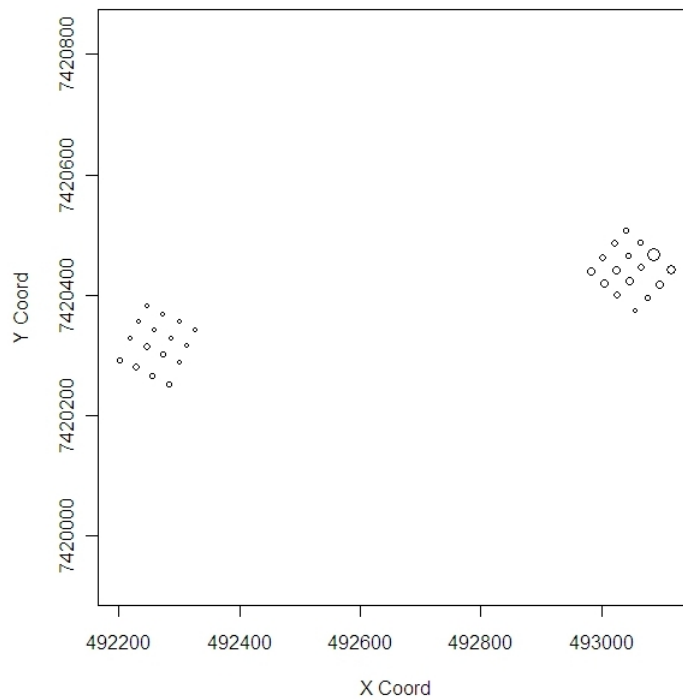


Figura 6: Distribuição dos pontos amostrais nas parcelas K e N, na Floresta Ombrófila Densa Montana a 1000m de altitude.

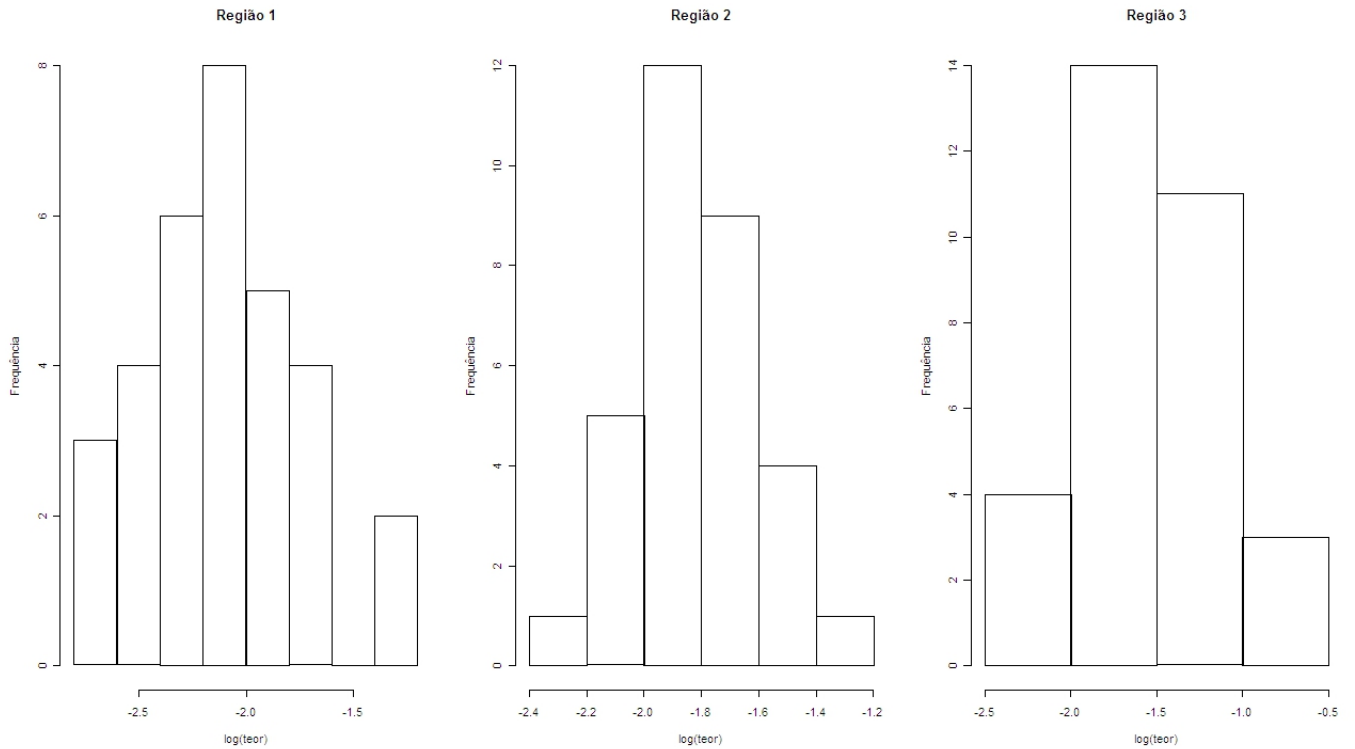


Figura 7: Histograma para os dados com a transformação de Box-Cox, log, para as três regiões.

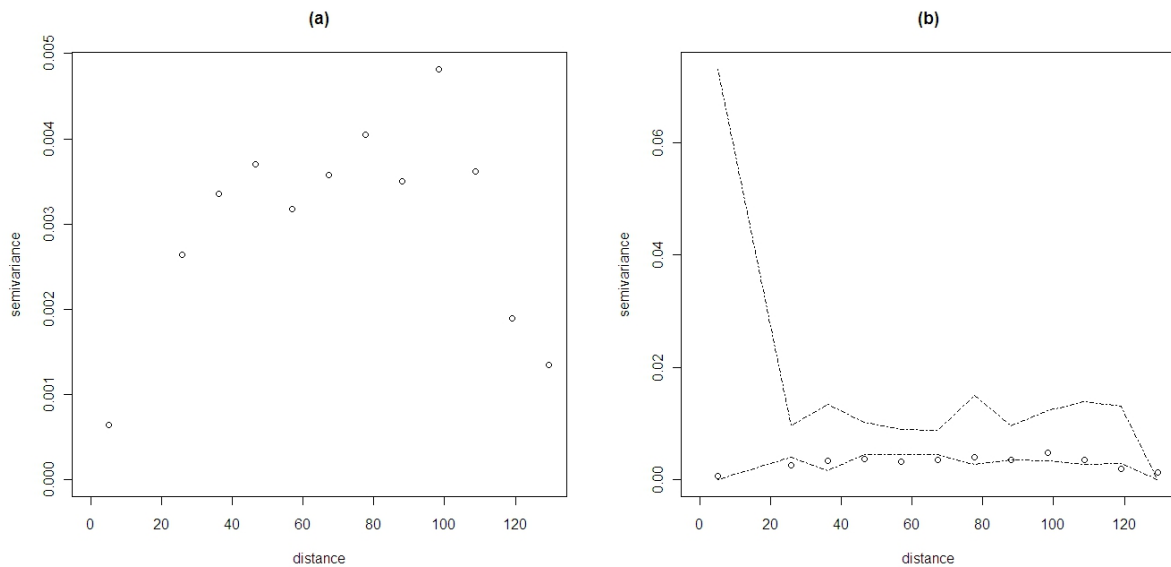


Figura 8: (a) Variograma dos dados; (b) Teste de Monte-Carlo, baseado no variograma.

Uma vez que há dependência espacial, parte-se agora para o objetivo principal do estudo, verificar se há diferença no teor de nitrogênio, nas regiões. Do ponto de vista univariado, caso não se considerasse a dependência espacial, este problema resumiria-se ao ajuste de modelos de ANOVA, considerando a região como fator. A idéia, do ponto de vista geoestatístico é estender a análise dos modelos ANOVA para o contexto de dados com dependência espacial, considerando

esta dependência. Para se verificar se existe diferença entre as três regiões serão considerados quatro cenários e para cada um deles será feito o ajuste de um modelo. O ajuste considera o modelo apresentado em (3), assim,

$$Y_{ij} \sim N \left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \begin{bmatrix} \Sigma_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Sigma_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Sigma_3 \end{bmatrix} \right)$$

A estrutura de covariância anterior considera independência entre as áreas, dada a distância entre elas que é suficientemente grande para não se considerar uma relação entre as observações nas áreas (veja a Figura 2). Assim, sua estrutura Σ_j fica totalmente especificada pelas funções ρ_{11} , ρ_{33} e ρ_{33} , definidas anteriormente.

Os cenários de análise são dados com base em considerações no vetor de médias, na estrutura de covariância e no parâmetro de dependência espacial, da seguinte forma:

1. Cenário 1

$$\text{Modelo 1} \Rightarrow \begin{cases} \mu & \text{comum} \\ \sigma^2 & \text{comum} \\ \phi & \text{comum} \end{cases}$$

2. Cenário 2

$$\text{Modelo 2} \Rightarrow \begin{cases} \mu_j \neq \mu_{j'}, & \text{para pelo menos um par } j \neq j', \text{ em que } j = 1, 2, 3 \\ \sigma^2 & \text{comum} \\ \phi & \text{comum} \end{cases}$$

3. Cenário 3

$$\text{Modelo 3} \Rightarrow \begin{cases} \mu_j \neq \mu_{j'}, & \text{para pelo menos um par } j \neq j', \text{ em que } j = 1, 2, 3 \\ \sigma_j^2 \neq \sigma_{j'}^2, & \text{para pelo menos um par } j \neq j', \text{ em que } j = 1, 2, 3 \\ \phi & \text{comum} \end{cases}$$

4. Cenário 4

$$\text{Modelo 4} \Rightarrow \begin{cases} \mu_j \neq \mu_{j'}, & \text{para pelo menos um par } j \neq j', \text{ em que } j = 1, 2, 3 \\ \sigma_j^2 \neq \sigma_{j'}^2, & \text{para pelo menos um par } j \neq j', \text{ em que } j = 1, 2, 3 \\ \phi_j \neq \phi_{j'}, & \text{para pelo menos um par } j \neq j', \text{ em que } j = 1, 2, 3 \end{cases}$$

Considerando, estes quatro cenários, o problema se resume a escolher o melhor modelo para descrever os dados. Para este caso, como a resposta é multivariada, implementou-se uma

generalização da *function* utilizada dentro do pacote *geoR* para a execução do comando *likfit*, considerando que a resposta é de dimensão 3 (veja Apêndice). Considerou-se também a função *optim* para o processo de maximização da verossimilhança.

Tabela 2: Valores do log-verossimilhança para os modelos ajustados aos diferentes cenários.

Modelos	log-verossimilhança
Modelo 1	-3109,554
Modelo 2	-3189,954
Modelo 3	-1193,657
Modelo 4	-1191,035

A escolha do melhor modelo dar-se-á com base no maior valor do log-verossimilhança. De acordo com a Tabela 2, este valor ocorre para o Modelo 4, entretanto, verifica-se que a diferença do log-verossimilhança do Modelo 3 e 4 não ultrapassa 4 unidades, e como o número de parâmetros é diferente, opta-se pelo modelo mais parcimonioso para descrever o comportamento dos dados, o Modelo 3. Desta forma, há diferença no teor médio de nitrogênio de cada região. Para este ajuste, tem-se $\hat{\phi} = 1,0000$, que mesmo sendo baixo, tem interpretação, uma vez que a hipótese $H_0 : \phi = 0$ foi rejeitada pelo teste de Monte Carlo (Figura 8 (b)), sendo que o mesmo é comum para as três regiões.

Na Tabela 3 tem-se os valores das médias ajustadas pelo modelo considerando a escala logarítmica, dada a transformação de Box-Cox, em seguida, os valores na escala original, que foram obtidos aplicando a função exponencial, e as médias amostrais obtidas diretamente dos dados. Note que os valores ajustados (na escala original) são próximos das médias amostrais.

Obteve-se ainda para cada uma das regiões o gráfico box-plot (Figura 9). Observa-se que a média aumenta, com o aumento da altitude da região e que há uma heterocedasticidade de variância entre as regiões, o que é concordante com o modelo adotado. E verifica-se também a presença de valores outliers, que dada a sua importância na interpretação biogeoquímica do solo foram mantidos na análise.

Tabela 3: Médias ajustadas e média amostral para as três regiões.

Região	Médias Ajustadas		Média Amostral
	Escala logarítmica	Escala original	
1	-2,1285	0,1190	0,1264
2	-1,8232	0,1615	0,1648
3	-1,5434	0,2136	0,2330

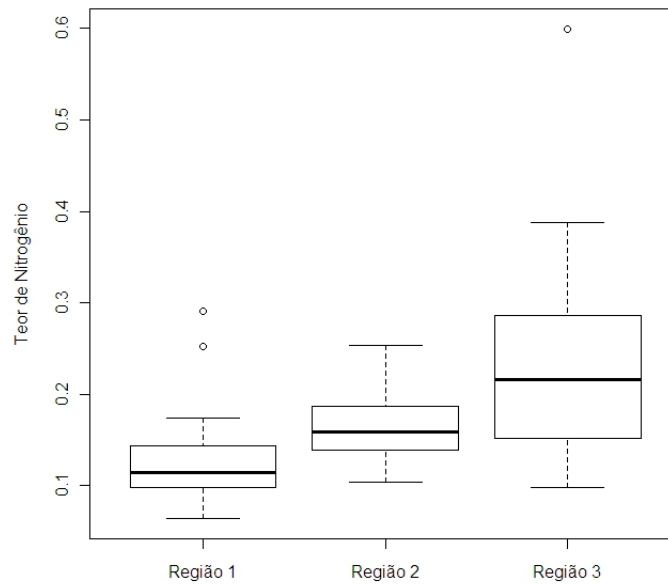


Figura 9: Box-plot para as três regiões.

4 Conclusões

Confirma-se a hipótese de que a relatividade da abundância ou pobreza de determinado nutriente varia ao longo do gradiente altitudinal, em que ecossistemas mais antigos tendem a ser menos limitados por nitrogênio que ecossistemas jovens, portanto, verifica-se maiores teores de nitrogênio nas maiores altitudes, onde os sistemas são mais antigos que a Mata de Restinga.

E ainda, os teores de nitrogênio aumentam com a elevação da altitude, assim como a variabilidade.

Mesmo com o baixo valor para a estimativa do parâmetro ϕ , ele é significativo, de modo que a dependência espacial deve ser considerada na análise.

5 Referências

GAMA-RODRIGUES, A. C. *Ciclagem de nutrientes em sistemas agroflorestais na região tropical: funcionalidade e sustentabilidade*. In: MULLER, M. W.; GAMA-RODRIGUES, A. C.; BRANDÃO, I. C. S. F. L.; SERÔDIO, M. H. C. F. (Ed.). *Sistemas agroflorestais, tendência da agricultura ecológica nos trópicos: Sustento da vida e sustento da vida*. Ilhéus: SBSAF; CE-PLAC; UENF, 2004. p. 64-84.

MARTINS, S. C. *Caracterização dos solos e serapilheira ao longo do gradiente altitudinal da Mata Atlântica, estado de São Paulo*. 2010. Tese (Doutorado em Ciências, Área de Concentração: Química na Agricultura e no Ambiente). Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2010.

MYERS, N; MITTERMEIER, R. A.; MITTERMEIER, C. G.; FONSECA, G. A. B.& KENT, J. *Biodiversity hotspots for conservation priorities.*, **Nature**, London, v. 403, p. 852-858, 2000.

RIBEIRO, M. C.; METZGER, J. P.; MARTENSEN, A. C.; PONZONI, F. J.& HIROTA, M. M. *The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implication for conservation.* **Biological Conservation**, Barking, v. 142, p. 1141-1153, 2009.

RIBEIRO, P. J. & DIGGLE, P. J. *Model-based Geostatistics*. Springer, 2007, New York.

6 Apêndice

Programa implementado em R

```
#-----
#Limpando a memória
#-----
rm(list=ls(all=TRUE))
getwd()

#-----
#Carregando o pacote
#-----
require(geoR)

#-----
#Entrada de dados
#-----

solo<-read.csv("banco1.csv", header = TRUE, sep = ";",dec=",") #todas as três áreas
head(solo)
View(solo)

solo1<-as.geodata(solo,coords.col=c(1,2), data.col=3) #geodata da região 1
View(solo1)
solo2<-as.geodata(solo,coords.col=c(4,5), data.col=6) #geodata da região 2
View(solo2)
```

```

solo3<-as.geodata(solo,coords.col=c(7,8), data.col=9)          #geodata da região 3
View(solo3)

solos<-read.csv("banco2.csv", header = TRUE, sep = ";", dec=",") #todas as regiões
head(solos)
View(solos)
solot<-as.geodata(solos,coords.col=c(1,2), data.col=3)        #geodata de todas as regiões
View(solot)

#-----
#Visualização dos pontos
#-----

points(solo1)    #região 1
points(solo2)    #região 2
points(solo3)    #região 3
points(solot)    #todas as regiões

#alternativamente

points.geodata(solo1)    #visualização dos dados da região 1
points.geodata(solo2)    #visualização dos dados da região 2
points.geodata(solo3)    #visualização dos dados da região 3
points.geodata(solot)    #visualização dos dados de todas as regiões

#todos juntos
par(mfrow=c(3,1))
points(solo1,main="Região 1");points(solo2,main="Região 2");points(solo3,main="Região 3")

#-----
#Análise Exploratória
#-----

summary(solo1)    #summary da região 1
summary(solo2)    #summary da região 2
summary(solo3)    #summary da região 3
summary(solot)    #summary de todos os dados

var(solo$N1)      #variância da variável resposta para a região 1
var(solo$N2)      #variância da variável resposta para a região 2
var(solo$N3)      #variância da variável resposta para a região 3
var(solos$N)      #variância da variável resposta para todos os dados

plot(solo1)      #plot da região 1
plot(solo2)      #plot da região 2
plot(solo3)      #plot da região 3
plot(solot)      #plot de todos os dados

```



```

shapiro.test(solo$N1) #região 1
shapiro.test(solo$N2) #região 2
shapiro.test(solo$N3) #região 3
shapiro.test(solos$N) #todas as regiões

#-----
#Transformação Box-Cox
#-----

library(MASS)
boxcox(solo1)      #transformando os dados da região 1
plot(solo1,lam=0)  #visualização os dados da região 1

boxcox(solo2)      #transformando os dados da região 2
plot(solo2,lam=0)  #visualização os dados da região 2

boxcox(solo3)      #transformando os dados da região 3
plot(solo3,lam=0)  #visualização os dados da região 3

par(mfrow=c(1,3))
z1<-solo$N1
z2<-solo$N2
z3<-solo$N3
z11<-log(z1)          #dados transformados devido a ausência de normalidade
z22<-log(z2)          #dados transformados devido a ausência de normalidade
z33<-log(z3)          #dados transformados devido a ausência de normalidade
hist(z11,main="Região 1",xlab="log(teor)", ylab="Frequência" )
hist(z22,main="Região 2",xlab="log(teor)", ylab="Frequência" )
hist(z33,main="Região 3", xlab="log(teor)", ylab="Frequência" )
shapiro.test(z11);shapiro.test(z22);shapiro.test(z33) #teste para os dados na escala transformada

#-----
#Variograma
#-----

#dados das 3 regiões
#par(mfrow=c(1,2))
v <- variog(solot, max.dist=(532.27623-29.99584)/2)
v
v$n
plot(v,main="(a)")

v.mc <- variog.mc.env(solot, obj.variog=v)      #Teste de Monte-Carlo
plot(v, env=v.mc,main="(b)")

#-----
#Maximização da Verossimilhança

```

```

#-----

#coordenadas
coords1<-cbind(solo$Cx1,solo$Cy1);head(coords1) #coordenadas dos dados da região 1
coords2<-cbind(solo$Cx2,solo$Cy2);head(coords2) #coordenadas dos dados da região 2
coords3<-cbind(solo$Cx3,solo$Cy3);head(coords3) #coordenadas dos dados da região 3

#distância dos pontos
d1<-dist(as.matrix(coords1));d1 #matriz de distâncias dos dados da região 1
d1<-as.vector(d1); d1 #distâncias dos dados da região 1
length(d1) #comprimento do vetor

d2<-dist(as.matrix(coords2));d2 #matriz de distâncias dos dados da região 2
d2<-as.vector(d2); d2 #distâncias dos dados da região 2
length(d2) #comprimento do vetor

d3<-dist(as.matrix(coords3));d3 #matriz de distâncias dos dados da região 3
d3<-as.vector(d3); d3 #distâncias dos dados da região 3
length(d3) #comprimento do vetor

#Para a estrutura de covariância
lower2matrix <- function(x){
    n <- (1+sqrt(1+8*length(x)))/2
    M <- diag(n)
    M[lower.tri(M)] <- x
    M <- M + t(M) - diag(n)
}

#Criando o vetor de dados
y1<-solo$N1
y2<-solo$N2
y3<-solo$N3
y11<-log(y1) #dados transformados devido a ausência de normalidade
y22<-log(y2) #dados transformados devido a ausência de normalidade
y33<-log(y3) #dados transformados devido a ausência de normalidade
y<-c(y11,y22,y33)
y
length(y)

#Maximização/Modelos

#-----
# Modelo 1
#
# mu1=mu2=mu3

```

```

# sigma comum
# phi comum
#-----

#pars=c(mu,nu1,nu2,nu3,sig,tau1,tau2,tau3,a),dados=list(n,n1,n2,n3,y,coords1,coords2,coords3)

l11<-function(pars,dados){
  cov11<-cov.spatial(dados$coords1,cov.model="matern",cov.pars=c(1,pars[9]),kappa=pars[2])
  c11 <- pars[5] * lower2matrix(cov11)+pars[6]*diag(dados$n1)
  cov22<-cov.spatial(dados$coords2,cov.model="matern",cov.pars=c(1,pars[9]),kappa=pars[3])
  c22 <- pars[5] * lower2matrix(cov22)+pars[7]*diag(dados$n2)
  cov33<-cov.spatial(dados$coords3,cov.model="matern",cov.pars=c(1,pars[9]),kappa=pars[4])
  c33 <- pars[5] * lower2matrix(cov33)+pars[8]*diag(dados$n3)
  SIGMA<-rbind(cbind(c11,matrix(0,nrow(c11),ncol(c11)),matrix(0,nrow(c11),ncol(c11))),
  cbind(matrix(0,nrow(c22),ncol(c22)),c22,matrix(0,nrow(c22),ncol(c22))),
  cbind(matrix(0,nrow(c33),ncol(c33)),matrix(0,nrow(c33),ncol(c33)),c33))
  detSIGMA<-determinant(SIGMA,log=T)$modulus
  Mu<-c(rep(pars[1],dados$n))
  Mu<-as.matrix(Mu,ncol=1)
  res<-dados$y - Mu
  HQ<-sum(res*solve(SIGMA,res))
  l1<- -0.5*(dados$n*detSIGMA-HQ)
  return(l1)
}

lista<-list(n=96,n1=32,n2=32,n3=32,y=y, coords1=d1,coords2=d2,coords3=d3)
ini1=c(mu=0,nu1=1,nu2=1,nu3=1,sig=1,tau1=1,tau2=1,tau3=1,a=1)
l11(ini1, dados=lista)
ajuste1 <- optim(par=ini1, fn=l11, dados=lista, method="BFGS")
#ajuste1 <- optim(par=ini1, fn=l11, dados=lista, method="CG")

#-----

# Modelo 2
#
# mu1;mu2;mu3 (diferentes)
# sigma comum
# phi comum
#-----

#pars=c(mu1,mu2,mu3,nu1,nu2,nu3,sig,tau1,tau2,tau3,a),dados=list(n,n1,n2,n3,y,coords1,coords2,coords3)

l12<-function(pars,dados){
  cov11<-cov.spatial(dados$coords1,cov.model="matern",cov.pars=c(1,pars[11]),kappa=pars[4])
  c11 <- pars[7] * lower2matrix(cov11)+pars[8]*diag(dados$n1)
  cov22<-cov.spatial(dados$coords2,cov.model="matern",cov.pars=c(1,pars[11]),kappa=pars[5])
  c22 <- pars[7] * lower2matrix(cov22)+pars[9]*diag(dados$n2)
  cov33<-cov.spatial(dados$coords3,cov.model="matern",cov.pars=c(1,pars[11]),kappa=pars[6])
  c33 <- pars[7] * lower2matrix(cov33)+pars[10]*diag(dados$n3)
  SIGMA<-rbind(cbind(c11,matrix(0,nrow(c11),ncol(c11)),matrix(0,nrow(c11),ncol(c11))),
  cbind(matrix(0,nrow(c22),ncol(c22)),c22,matrix(0,nrow(c22),ncol(c22))),

```

```

        cbind(matrix(0,nrow(c33),ncol(c33)),matrix(0,nrow(c33),ncol(c33)),c33))
        detSIGMA<-determinant(SIGMA,log=T)$modulus
        Mu<-c(rep(pars[1],dados$n1),rep(pars[2],dados$n2),rep(pars[3],dados$n3))
        Mu<-as.matrix(Mu,ncol=1)
        res<-dados$y - Mu
        HQ<-sum(res*solve(SIGMA,res))
        ll<- -0.5*(dados$n*detSIGMA-HQ)
        return(ll)
    }

lista<-list(n=96,n1=32,n2=32,n3=32,y=y, coords1=d1,coords2=d2,coords3=d3)
ini2=c(mu1=-2,mu2=-1.5,mu3=-1.8,nu1=1,nu2=1,nu3=1,sig=1,tau1=1,tau2=1,tau3=1,a=1)
ll2(ini2, dados=lista)
#ajuste2 <- optim(par=ini2, fn=ll2, dados=lista, method="BFGS")
ajuste2 <- optim(par=ini2, fn=ll2, dados=lista, method="CG")

#-----
# Modelo 3
#
# media diferente
# sigma diferente
# phi comum
#-----

#pars=c(mu1,mu2,mu3,nu1,nu2,nu3,sig1,sig2,sig3,tau1,tau2,tau3,a),dados=list(n,n1,n2,n3,y,coords1,coords2,coords3)

ll3<-function(pars,dados){
    cov11<-cov.spatial(dados$coords1,cov.model="matern",cov.pars=c(1,pars[13]),kappa=pars[4])
    c11 <- pars[7] * lower2matrix(cov11)+pars[10]*diag(dados$n1)
    cov22<-cov.spatial(dados$coords2,cov.model="matern",cov.pars=c(1,pars[13]),kappa=pars[5])
    c22 <- pars[8] * lower2matrix(cov22)+pars[11]*diag(dados$n2)
    cov33<-cov.spatial(dados$coords3,cov.model="matern",cov.pars=c(1,pars[13]),kappa=pars[6])
    c33 <- pars[9] * lower2matrix(cov33)+pars[12]*diag(dados$n3)
    SIGMA<-rbind(cbind(c11,matrix(0,nrow(c11),ncol(c11)),matrix(0,nrow(c11),ncol(c11))),
        cbind(matrix(0,nrow(c22),ncol(c22)),c22,matrix(0,nrow(c22),ncol(c22))),
        cbind(matrix(0,nrow(c33),ncol(c33)),matrix(0,nrow(c33),ncol(c33)),c33))
    detSIGMA<-determinant(SIGMA,log=T)$modulus
    Mu<-c(rep(pars[1],dados$n1),rep(pars[2],dados$n2),rep(pars[3],dados$n3))
    Mu<-as.matrix(Mu,ncol=1)
    res<-dados$y - Mu
    HQ<-sum(res*solve(SIGMA,res))
    ll<- -0.5*(dados$n*detSIGMA-HQ)
    return(ll)
}

lista<-list(n=96,n1=32,n2=32,n3=32,y=y, coords1=d1,coords2=d2,coords3=d3)
ini3=c(mu1=-2,mu2=-1.5,mu3=-1.8,nu1=1,nu2=1,nu3=1,sig1=0.3,sig2=0.2,sig3=0.4,tau1=1,tau2=1,tau3=1,a=1)
ll3(ini3, dados=lista)
#ajuste3 <- optim(par=ini3, fn=ll3, dados=lista, method="BFGS")
ajuste3 <- optim(par=ini3, fn=ll3, dados=lista, method="CG")

```

```

#-----
# Modelo 4
#
# media diferente
# sigma diferente
# phi diferente
#-----

#pars=c(mu1,mu2,mu3,nu1,nu2,nu3,sig1,sig2,sig3,tau1,tau2,tau3,a1,a2,a3),dados=list(n,n1,n2,n3,y,coords1,coords2,coords3)

ll4<-function(pars,dados){
  cov11<-cov.spatial(dados$coords1,cov.model="matern",cov.pars=c(1,pars[13]),kappa=pars[4])
  c11 <- pars[7] * lower2matrix(cov11)+pars[10]*diag(dados$n1)
  cov22<-cov.spatial(dados$coords2,cov.model="matern",cov.pars=c(1,pars[14]),kappa=pars[5])
  c22 <- pars[8] * lower2matrix(cov22)+pars[11]*diag(dados$n2)
  cov33<-cov.spatial(dados$coords3,cov.model="matern",cov.pars=c(1,pars[15]),kappa=pars[6])
  c33 <- pars[9] * lower2matrix(cov33)+pars[12]*diag(dados$n3)
  SIGMA<-rbind(cbind(c11,matrix(0,nrow(c11),ncol(c11)),matrix(0,nrow(c11),ncol(c11))),
  cbind(matrix(0,nrow(c22),ncol(c22)),c22,matrix(0,nrow(c22),ncol(c22))),
  cbind(matrix(0,nrow(c33),ncol(c33)),matrix(0,nrow(c33),ncol(c33)),c33))
  detSIGMA<-determinant(SIGMA,log=T)$modulus
  Mu<-c(rep(pars[1],dados$n1),rep(pars[2],dados$n2),rep(pars[3],dados$n3))
  Mu<-as.matrix(Mu,ncol=1)
  res<-dados$y - Mu
  HQ<-sum(res*solve(SIGMA,res))
  ll<- -0.5*(dados$n*detSIGMA-HQ)
  return(ll)
}

lista<-list(n=96,n1=32,n2=32,n3=32,y=y, coords1=d1,coords2=d2,coords3=d3)
ini4=c(mu1=-2,mu2=-1.5,mu3=-1,nu1=1,nu2=1,nu3=1,sig1=0.3,sig2=0.2,sig3=0.4,tau1=1,tau2=1,tau3=1,a1=0.5,a2=0.5,a3=0.5)
ll4(ini4, dados=lista)
#ajuste4 <- optim(par=ini4, fn=ll4, dados=lista, method="BFGS")
ajuste4 <- optim(par=ini4, fn=ll4, dados=lista, method="CG")

#-----
#Comparação dos modelos
#-----

#log-verossimilhança
ll1(ini1, dados=lista)
ll2(ini2, dados=lista)
ll3(ini3, dados=lista)
ll4(ini4, dados=lista)

#convergência dos modelos
ajuste1[4]

```

```

ajuste2[4]
ajuste3[4]
ajuste4[4]

#estimativas
ajuste1[1];mu1<-exp(-3.489347);mu1
ajuste2[1];mu1<-exp(-2.130972);mu2<-exp(-1.829300);mu3<-exp(-1.538551);mu1;mu2;mu3
ajuste3[1];mu1<-exp(-2.128540);mu2<-exp(-1.823206);mu3<-exp(-1.543412);mu1;mu2;mu3
ajuste4[1];mu1<-exp(-2.128539);mu2<-exp(-1.823183);mu3<-exp(-1.543417);mu1;mu2;mu3

#-----
#Box-plot dos dados na escala original
#-----

N<-solos$N
R<-c("Região 1","Região 1","Região 1","Região 1","Região 1","Região 1","Região 1","Região 1",
     "Região 1","Região 1","Região 1","Região 1","Região 1","Região 1","Região 1","Região 1",
     "Região 1","Região 1","Região 1","Região 1","Região 1","Região 1","Região 1","Região 1",
     "Região 2","Região 2","Região 2","Região 2","Região 2","Região 2","Região 2","Região 2",
     "Região 2","Região 2","Região 2","Região 2","Região 2","Região 2","Região 2","Região 2",
     "Região 2","Região 2","Região 2","Região 2","Região 2","Região 2","Região 2","Região 2",
     "Região 2","Região 2","Região 2","Região 2","Região 2","Região 2","Região 2","Região 2",
     "Região 3","Região 3","Região 3","Região 3","Região 3","Região 3","Região 3","Região 3",
     "Região 3","Região 3","Região 3","Região 3","Região 3","Região 3","Região 3","Região 3",
     "Região 3","Região 3","Região 3","Região 3","Região 3","Região 3","Região 3","Região 3",
     "Região 3","Região 3","Região 3","Região 3","Região 3","Região 3","Região 3","Região 3")
length(N);length(R)
boxplot(N~R,ylab="Teor de Nitrogênio")

```