

UNICAMP

UNIVERSIDADE ESTADUAL DE
CAMPINAS

Instituto de Matemática, Estatística e
Computação Científica

LEONARDO UCHÔA PEDREIRA

**Modelo de Kriging Multivariado para
Mapeamento de Dados Composicionais**

Campinas

2021

Leonardo Uchôa Pedreira

Modelo de Kriging Multivariado para Mapeamento de Dados Composicionais

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Estatística.

Orientador: Guilherme Vieira Nunes Ludwig

Este trabalho corresponde à versão final da Dissertação defendida pelo aluno Leonardo Uchôa Pedreira e orientada pelo Prof. Dr. Guilherme Vieira Nunes Ludwig.

Campinas

2021

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

P343m Pedreira, Leonardo Uchôa, 1995-
Modelo de Kriging multivariado para mapeamento de dados
composicionais / Leonardo Uchôa Pedreira. – Campinas, SP : [s.n.], 2021.

Orientador: Guilherme Vieira Nunes Ludwig.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de
Matemática, Estatística e Computação Científica.

1. Análise espacial (Estatística). 2. Krigagem. 3. Estimção de máxima
verossimilhança. 4. Geologia - Métodos estatísticos. I. Ludwig, Guilherme
Vieira Nunes, 1985-. II. Universidade Estadual de Campinas. Instituto de
Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Multivariate Kriging model for compositional data mapping

Palavras-chave em inglês:

Spatial analysis (Statistics)

Kriging

Maximum likelihood estimation

Geology - Statistical methods

Área de concentração: Estatística

Titulação: Mestre em Estatística

Banca examinadora:

Guilherme Vieira Nunes Ludwig [Orientador]

Mariana Rodrigues Motta

Alessandro Samuel Rosa

Data de defesa: 24-03-2021

Programa de Pós-Graduação: Estatística

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0003-1545-0576>

- Currículo Lattes do autor: <http://lattes.cnpq.br/9301162741707828>

**Dissertação de Mestrado defendida em 24 de março de 2021 e aprovada
pela banca examinadora composta pelos Profs. Drs.**

Prof(a). Dr(a). GUILHERME VIEIRA NUNES LUDWIG

Prof(a). Dr(a). MARIANA RODRIGUES MOTTA

Prof(a). Dr(a). ALESSANDRO SAMUEL ROSA

A Ata da Defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-Graduação do Instituto de Matemática, Estatística e Computação Científica.

Agradecimentos

Agradeço primeiramente aos meus pais e familiares por cada uma das decisões que tomaram até o presente momento. Sem eles e tais decisões não haveria maneira alguma de me permitir chegar onde estou, fato do qual me orgulho muito. Em especial agradeço ao meu pai, Luiz Campeão, que jamais economizou esforços para meu ingresso e formação em uma faculdade de tamanho calibre.

Em segundo lugar agradeço à UNICAMP por fomentar minha formação como pessoa. Ela foi a responsável por fornecer meios para que ocorressem alguns dos mais importantes aprendizados que tive na vida e assim, tornar-me quem sou. É curioso inclusive o tamanho do orgulho que carrego por esta instituição, pois confesso que de todas das lágrimas que derramei durante todos os últimos quase 8 anos, várias foram de orgulho e agradecimento por ter feito parte de sua história. O que me deixa triste, porém, é o fato de jamais poder retribuir tudo que me foi proporcionado por ela, pois não há número que cubra o valor que atribuo ao que veio dela.

Em terceiro lugar agradeço às minhas duas outras famílias, sem as quais não haveria a menor possibilidade ter passado por tantos desafios que me foram postos. Agradeço então a Lucas Ricomini Bertagnoli, Caio Rosa Bertagnoli e Sônia Bertagnoli, juntamente com Hugo Calegari, Cida Calegari e Antônio Calegari. Nos momentos mais difíceis destes últimos anos, estas pessoas estavam do meu lado sempre que precisei e não precisei e a tais, devo tanto quanto é minha alegria de ter-los conhecidos. Em especial agradeço ao Lucas, sô Xems - meu genuíno irmão -, pelo grande presente de me ter proporcionado dois anos de convênio em uma academia, quando minha condição financeira não permitia tal luxo. Tal fortuna não se resume só a momentos de exercício físico, pois ela é na verdade uma história de aprendizado e crescimento pessoal, condensada em quatro anos de dedicação a tornar-me o mais forte que posso.

Em quarto lugar agradeço ao Guilherme Ludwig por me ensinar, mais uma vez, a lição da perseverança. Em vários momentos minha fraqueza mental mostrou-se ser meu maior inimigo, dificultando muito o desafio de compor esta dissertação de mestrado. Estes foram os momentos que o Guilherme melhor me auxiliou, seguido de todas as revisões de códigos e esclarecimentos a cerca de teoria estatística e matemática. Ele, sem dúvida, é um excelente orientador devido a sua infindável paciência e imensa educação. Tenho muito orgulho de ser o primeiro aluno de mestrado dele e espero que estes últimos quase três anos, em que fomos aluno e professor, possam ser descritos como a coleção de divertidas e produtivas conversas e que o sirvam como boas direções para futuras orientações.

Agradeço também a cada um de meus colegas e amigos de graduação, pós-

graduação e caronas. Eles foram muito importantes para minha formação, pois são parte de várias histórias e risadas das quais lembro-me até hoje. São estas pessoas que formaram um ambiente amigável, divertido e de solidariedade quando os desafios da graduação, pós-graduação e da vida chegavam. Em especial, agradeço aos meus colegas mais próximos e que são pessoas por quem tenho muito apreço. São eles e elas: Hugo Calegari, Matheus Marques, Caio Augusto, Rafael Rodrigues, Diserée, Lara Acrani, Guilherme Pazian, Henrique Capatto, Murilo Razzoli, Gabriel Freitas, Paulo Rogério, Daniel, Júlia, Natália, Ricardo, Rodolfo Lindemute, Marcelo Rubio, Daniel Ifanger, Ananda, Nicholas, Fernando, Christian Porcel, Fabiana, Helio e Benedito. Pelos mesmos motivos, agradeço meus amigos e amigas da Copa do IMECC e biblioteca do IMECC, com especial carinho a: Dona Zefa, Luan Mattos, Gilberto, Márcia, Ana, Reginaldo, Rosimeire e Silvânia.

Além de colegas tais histórias não estão completas sem os excelentes professores e professoras que tive e, a eles e elas, tenho muito que agradecer. Cada um teve sua importância em minha formação, mas alguns estão mais próximos, aos quais tenho mais ainda o que agradecer. São eles: Guilherme Ludwig, Rafael Maia, Eduardo Abreu, Benilton Carvalho e Joao Frederico. Ao Rafael Maia agradeço muito por todas as oportunidades de usar e praticar conhecimento de estatística por meio de consultoria assistida, ele foi muito importante para que quebrar a barreira que existia em mim entre teoria e prática. Ao Benilton, agradeço muito por vários ensinamentos não estatísticos, aqueles que dizem respeito à vida. Ao João Frederico, Joni, sou grato por sua simpatia, alegria e educação, que certamente torna o IMECC um lugar muito melhor de estar. Meus agradecimentos ao Eduardo Abreu e ao Guilherme Ludwig, por sua vez, são infundáveis, já que ambos me ensinaram e re-ensinaram o significado de perseverança. Perseverança, persistência é a maior dívida dos meus últimos 8 anos. Muito obrigado Guilherme e Eduardo Abreu.

Por fim agradeço aos meus grandes amigos do Residencial Flat Itaici (RFI), pois nunca colecionei tantas boas memórias na minha vida quanto aquelas que me vêm quando lembro dos quatro anos em que morei lá. Por este motivo, e apesar de todas as adversidades, considero estes quatro anos os melhores anos da minha vida. Sem as pessoas a seguir, jamais poderia ter chegado onde estou com a alegria que tenho. Muito obrigado Fausto, Glaciele, Zé, Regina, Alexandre, Martina, Fernando, Lupércio, Paula, Akson, Wellington, Erick, Roberto, Thais, Denis, Souza e Marley.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, assim como do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, processo nº 134683/2018-8.

Resumo

Em certas localidades do Estado de Minas Gerais foram coletadas do solo quantidades de areia, argila e silte. Tais quantidades foram transformadas em proporções, ou composições. Com o objetivo de usar as amostras coletadas para criar mapas de tais composições em uma região próxima utilizou-se co-Krigagem com função de covariância Matérn bivariada. Os parâmetros da função de covariância foram estimados por máxima verossimilhança e em seguida fez-se a co-Krigagem para obter os mapas de predição composicional, junto com seus erros de predição aproximados. Por fim, foram obtidas predições por co-Krigagem em dados ainda não utilizados, na mesma região, e comparados com os de um classificador K vizinhos mais próximos, treinado com validação cruzada do tipo “deixe-um-de-fora”. Os resultados mostram desempenhos preditivos similares, mas a co-Krigagem fornece mapas contínuos e suaves dos erros de predição para cada composição.

Palavras-chave: estatística espacial, dados composicionais, co-Krigagem, Matérn bivariada, estimação por máxima verossimilhança.

Abstract

In certain locations of Minas Gerais's State, quantities of sand, clay and silt were collected from the soil. Such quantities were transformed into proportions, or compositions. In order to use the gathered samples to create maps of such compositions in a nearby region, co-Kriging with a bivariate Matérn covariance function was used. The parameters of the covariance function were estimated by maximum likelihood and then co-Kriging was performed to obtain the compositional prediction maps, together with their approximate prediction errors. Finally, co-Kriging predictions were obtained from a dataset not used in estimation at the same region, and compared with those of a nearest neighbor classifier, trained with leave-one-out cross-validation procedure. The results show similar predictive performances, but co-Kriging provides smooth continuous maps of the prediction errors for each composition.

Keywords: spatial statistics, co-kriging, bivariate Matérn, compositional data, maximum likelihood estimation.

Lista de ilustrações

Figura 1	– Exemplo de diagrama ternário	21
Figura 2	– Esquerda: Mapa com as classificações associadas às composições de areia, argila e silte em uma região de Minas Gerais. Nota-se que nessa amostra há uma grande concentração de argila. As classificações (“Etiqueta”) são fornecidas pela Tabela 5, disponível na Seção E do Apêndice. Direita: Diagrama ternário com os dados da região de Minas Gerais.	22
Figura 3	– Dependência entre dois pontos (supostas observações) imposta pela estacionariedade. Sendo a covariância função somente de um vetor, ela é caracterizada por direção, sentido e tamanho.	24
Figura 4	– Gráficos de contornos para dependência entre dois pontos (supostas observações) nos casos de isotropia e anisotropia. Sendo a covariância função somente da distância entre as observações, não é mais preciso saber sua direção e sentido e a dependência é esférica. Já para a anisotropia considera-se direção e sentido, o que nos permite mais maleabilidade da formulação da co-dependência.	25
Figura 5	– processo Gaussiano Simulado simulado com função de covariância Matérn $\sigma^2 = 3, 1/a = 0.3, \nu = 1.0$	27
Figura 6	– Variograma Empírico obtido a partir de uma amostra de 85 pontos do processo gaussiano simulado na Figura 5 com função de covariância Matérn $\sigma^2 = 3, 1/a = 0.3, \nu = 1.0$	28
Figura 7	– Processo de Wiener simulado em $[0, 2]$ com início em $t = 0$. Este é um processo de Wiener padrão.	29
Figura 8	– Ilustração de uma Krigagem. A partir de medições feitas em certas coordenadas (bolinhas pretas), deseja-se realizar a predição para uma nova localização em que ainda não se realizou medição (marcado com \times).	31
Figura 9	– Krigagem baseada em uma amostra aleatória de 85 pontos, obtida a partir do processo simulado ilustrado na Figura 5	34
Figura 10	– Variância da Krigagem baseada em uma amostra aleatória de 85 pontos, obtida a partir do processo simulado ilustrado na Figura 5	35
Figura 11	– Função de covariância Matérn bivariada. Aqui $\sigma_1^2 = 1, \sigma_2^2 = 3, a = 1, \rho = 0.8, \nu_1 = 0.5, \nu_2 = 1.5, \nu_{12} = 1.0$	39
Figura 12	– Simulação de um processo Gaussiano com função de covariância Matérn bivariada. Aqui $\sigma_1^2 = 1, \sigma_2^2 = 3, a = 1, \rho = 0.9, \nu_1 = 0.5, \nu_2 = 1.5, \nu_{12} = 1.0$	40
Figura 13	– Simulação de um processo Logístico-Gaussiano com função de covariância Matérn bivariada. Aqui $\sigma_1^2 = 1, \sigma_2^2 = 3, a = 1, \rho = 0.9, \nu_1 = 0.5, \nu_2 = 1.5, \nu_{12} = 1.0$	41

Figura 14 – Covariograma de pontos sorteados aleatoriamente que seguem um processo Logístico-Gaussiano simulado com função de covariância Matérn bivariada. Aqui $\sigma_1^2 = 1, \sigma_2^2 = 3, a = 1, \rho = 0.8, \nu_1 = 0.5, \nu_2 = 1.5, \nu_{12} = 1.0$	42
Figura 15 – Curvas de níveis para combinações dois-a-dois de parâmetros da função de covariância Matérn bivariada, onde o asterisco representa os valores verdadeiros dos parâmetros que são $\sigma_1^2 = 1, \sigma_2^2 = 1, a = 2, \rho = 0.5, \nu_1 = 0.5, \nu_2 = 0.5, \nu_{12} = 1.0$. É possível perceber que para os casos a versus σ^2 , σ_1^2 versus σ_2^2 a log-verossimilhança reconhece muito bem os valores verdadeiros. Entretanto para σ_1^2 versus ρ , há um deslocamento do máximo em relação ao verdadeiro parâmetro.	44
Figura 16 – Histogramas dos parâmetros estimados da função de covariância Matérn bivariada, onde as barras verticais representam os valores verdadeiros dos parâmetros, que são $\sigma_1^2 = 1, \sigma_2^2 = 1, a = 2, \rho = 0.5, \nu_1 = 0.5, \nu_2 = 0.5, \nu_{12} = 1.0$. Nota-se que	46
Figura 17 – Processo simulado e predição composicional. Esta é a continuação do exemplo simulado na Figura 13. O gráfico de cima é o processo simulado, enquanto que o de baixo é a Krigagem simples feita com função de covariância Matérn bivariada de parâmetros estimados $\sigma_1^2 = 0.4228, \sigma_2^2 = 0.6736, a = 2.4587, \rho = 0.2890$. Delas é possível perceber que as partes estruturais do processo foram capturadas e que as partes de variação local foram suavizadas.	47
Figura 18 – Predição composicional e erro padrão da predição. Esta é a continuação do exemplo simulado na Figura 13 com parâmetros $\sigma_1^2 = 1, \sigma_2^2 = 3, a = 1, \rho = 0.9, \nu_1 = 0.5, \nu_2 = 1.5, \nu_{12} = 1.0$. Ao comparar os dois Gráficos é possível perceber que o erro padrão é menor (de fato é zero) nos locais amostrados e que aumenta a medida que nos afastamos destes pontos.	49
Figura 19 – Mapa com as classificações associadas às composições de areia, argila e silte em uma região de Minas Gerais. Nota-se que nessa amostra há uma grande concentração de argila, de acordo com as classificações fornecidas pela Tabela 5	51
Figura 20 – Diagrama ternário das composições de areia, argila e silte em uma região de Minas Gerais. Nota-se que na maioria das amostras a composição dominante é de argila. Também há casos de amostras com bastante areia, mas somente cinco amostras em que silte é responsável ao menos 40% do todo. Isto traduz-se ao fato de que as composições de argila e areia são mais dispersas, ao passo que silte varia entre próximo de zero e 40%.	52

Figura 21 – Covariograma dos dados de solo CTB0562. Pode-se notar que o vario- grama cruzado gira em torno do 0, que sugere os processos individuais como independentes. Também nota-se que os ambos variogramas diretos apresentam efeito de pepita em torno de 0.75 e ambos σ_1^2 σ_2^2 em torno de 1, o que indica que há muita variação de micro-escala. Por fim pode-se perceber que o alcance prático de dependência está em torno de 0.4 para o segundo processo composicional log(argila/silte) e para o primeiro parece ser 0.8.	53
Figura 22 – Krigagem trivariada de composição do solo, feita com função de covari- ância Matérn Bivariada de parâmetros estimados iguais a $\sigma_1^2 = 2.200$, $\sigma_2^2 = 2.228$, $a = 1.960$ e $\rho = 0.685$	55
Figura 23 – Erro de predição da Krigagem composicional do solo, feita com função de covariância Matérn Bivariada de parâmetros estimados iguais a $\sigma_1^2 = 2.200$, $\sigma_2^2 = 2.228$, $a = 1.960$ e $\rho = 0.685$	56
Figura 24 – Classificação predita do solo, obtida das categorias criadas pela SIBCS 5 a partir de Krigagem feita com função de covariância Matérn Bivariada de parâmetros estimados iguais a $\sigma_1^2 = 2.200$, $\sigma_2^2 = 2.228$, $a = 1.960$ e $\rho = 0.685$	57
Figura 25 – Mapa da região de Uberlândia em Minas Gerais. Amostras utilizadas para estimar o modelo estão representadas pelos pontos pretos, enquanto que as novas amostras (utilizadas para quantificar a qualidade preditiva do modelo) são os números.	58
Figura 26 – Diagrama ternário dos dados na região de Uberlândia (MG) que foram usados para ajustar o modelo.	59
Figura 27 – Diagrama ternário dos dados na região de Uberlândia (MG) que foram usados para determinar o poder preditivo.	60
Figura 28 – Mapa da predição da classificação de solo de acordo com o critério SIBCS, realizada pelo algoritmo KNN com $K = 5$ vizinhos mais próximos.	62
Figura 29 – Processo gaussiano simulado com $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $\rho = 0.2$, $\nu_1 = 0.5$, $\nu_2 =$ 0.5 e $a = 1.0000, 0.0991, 0.0521, 0.0354$ o que levam a, respectivamente $s = 1.0000, 10.0948, 19.1895, 28.2843$, onde $s = 28.2843$ é a diagonal da caixa em que o processo foi simulado.	92
Figura 30 – Processo gaussiano simulado com $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $a = 1$, $\nu_1 = 0.5$, $\nu_2 =$ 0.5 e $\rho = -0.8, -0.2, 0.2, 0.8$	93
Figura 31 – Processo gaussiano simulado com $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $\rho = 0.2$, $a = 1$, $\nu_1 =$ 0.5 , e $nu_2 = 1, 1.5, 2.0$	94
Figura 32 – Diagrama ternário representativo da classificação de solos da SiBCS. As setas indicam a direção de início das proporções (0% → 100%) para facilitar orientação.	96

Lista de tabelas

Tabela 1	– Comparativo entre a classificação do solo para as composições observadas e suas predições pela Krigagem. Notam-se 5 divergências entre as classificações, destacadas pelo negrito, sendo que a maioria dos erros (4 dos 5 erros) é por classificar um terreno arenoso como médio arenoso.	61
Tabela 2	– Classificação obtida a partir da Krigagem contra categoria observada.	61
Tabela 3	– Taxas de acurácias nos dados CTB0809, para cada algoritmo KNN ajustado por validação cruzada “deixe um de fora” nos dados CTB0562. O melhor resultado foi obtido com $K = 5$, com acurácia de 59.1%.	62
Tabela 4	– Classificador criados com base em 5 vizinhos mais próximos contra classificação do solo observada. As linhas representam as categorias observadas em CTB0809, enquanto que as colunas representam as categorias preditas pelo algoritmo.	62
Tabela 5	– Abreviações do tipo de solo segundo o sistema SiBCS (EMBRAPA, 2020)	95
Tabela 6	– Quantidades observadas e preditas de areia, junto com o respectivo erro padrão de predição. O maior erro padrão é de 0.24, enquanto que o menor é de 0.12	97
Tabela 7	– Quantidades observadas e preditas de argila, junto com o respectivo erro padrão de predição. O maior erro padrão é de 0.37, enquanto que o menor é de 0.24	97
Tabela 8	– Quantidades observadas e preditas de argila, junto com o respectivo erro padrão de predição. O maior erro padrão é de 0.36, enquanto que o menor é de 0.15.	98
Tabela 9	– Taxas de acurácias nos dados CTB0809, para cada algoritmo KNN ajustado por validação cruzada “deixe um de fora” nos dados CTB0562. O melhor resultado foi obtido com $K = 5$, com acurácia de 59.1%.	99
Tabela 10	– Classificador criados com base em 5 vizinhos mais próximos contra classificação do solo observada. As linhas representam as categorias observadas em CTB0809, enquanto que as colunas representam as categorias preditas pelo algoritmo.	100
Tabela 11	– Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 1 vizinho mais próximo.	100
Tabela 12	– Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 2 vizinhos mais próximos.	101

Tabela 13 – Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 3 vizinhos mais próximos.	101
Tabela 14 – Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 4 vizinhos mais próximos.	102
Tabela 15 – Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 5 vizinhos mais próximos.	102
Tabela 16 – Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 6 vizinhos mais próximos.	103
Tabela 17 – Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 7 vizinhos mais próximos.	103
Tabela 18 – Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 8 vizinhos mais próximos.	104
Tabela 19 – Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 9 vizinhos mais próximos.	104
Tabela 20 – Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 10 vizinhos mais próximos.	105
Tabela 21 – Validação Cruzada Tipo “Deixe-um-de-fora” na Co-Krigagem em CTB0562106	

Sumário

	Introdução	16
1	ANÁLISE DE DADOS COMPOSICIONAIS	18
1.1	Dado Composicional e o Simplex	18
1.2	Diagrama Ternário	20
2	GEOESTATÍSTICA	23
2.1	Contextualização e Conceitos Básicos	23
2.2	Função de Covariância e o Variograma	25
2.3	O processo Gaussiano	29
2.4	Krigagem	31
2.5	Geoestatística Multivariada	35
3	GEOESTATÍSTICA E DADOS COMPOSICIONAIS	38
3.1	Geoestatística Multivariada para Dados Composicionais	38
3.2	Estimação da Função de Covariância Matérn bivariada Reduzida	42
3.3	Dificuldades da Otimização da Verossimilhança em Função da Co- variância	43
3.4	Predição Composicional e Transformações	46
4	APLICAÇÃO	50
	Aplicação	50
5	CONSIDERAÇÕES FINAIS	63
	REFERÊNCIAS	64
	APÊNDICE A – GRADIENTE DA LOG-VEROSSIMILHANÇA PARA O MODELO DE COVARIÂNCIA MATÉRN BIVA- RIADA	68
A.1	Fórmula geral da derivada da log-verossimilhança	68
A.2	Derivadas Parciais da Função de Covariância	69
	APÊNDICE B – AVALIAÇÃO ESTÁVEL DO GRADIENTE PARA O MODELO DE COVARIÂNCIA MATÉRN BIVA- RIADA	71

B.1	Inversa	71
B.2	Logaritmo do Determinante	72
B.3	Traço e Forma Quadrática	72
B.3.1	Caso em que $\theta_k = \sigma_1^2$	73
B.3.2	Caso em que $\theta_k = \sigma_2^2$	75
B.3.3	Caso em que $\theta_k = \rho$	76
B.3.4	Caso em que $\theta_k = a$	77
B.3.5	Caso em que $\theta_k = \tau_1^2$	78
B.3.6	Caso em que $\theta_k = \tau_2^2$	78
	APÊNDICE C – INFORMAÇÃO DE FISHER	79
C.1	Resultado Preliminares	79
C.2	Resultado Geral	81
C.3	Informação de Fisher para σ_1^2 e σ_2^2	83
C.4	Informação de Fisher para σ_1^2 e ρ	85
C.5	Informação de Fisher para σ_1^2 e a	86
C.6	Informação de Fisher para σ_2^2 e a	87
C.7	Informação de Fisher para σ_2^2 e ρ	88
C.8	Informação de Fisher para a e ρ	89
	APÊNDICE D – SIMULAÇÃO DE PROCESSOS GAUSSIANOS ES- PACIAIS COM ESTRUTURA MATÉRN BIVARI- ADA	91
	APÊNDICE E – SISTEMA BRASILEIRO DE CLASSIFICAÇÃO DE SOLOS	95
	APÊNDICE F – RESULTADOS DE KRIGAGEM PARA DADOS CTB0809	97
	APÊNDICE G – RESULTADOS DOS K VIZINHOS MAIS PRÓXI- MOS	99
	APÊNDICE H – VALIDAÇÃO CRUZADA TIPO “DEIXE-UM-DE- FORA” NA CO-KRIGAGEM EM CTB0562	106

Introdução

Agricultura e pecuária são duas atividades econômicas de especial importância para o Brasil que, segundo acompanhamento anual do CEPEA/USP (Centro de Estudos Avançados em Economia Aplicada), teve em 2019 aproximadamente 21,4% de seu PIB constituído em atividades do agronegócio (BARROS et al., 2020). De acordo com tais relatórios, este número é alto há anos (24.1% em 2018, 24.6% em 2017 e 24.9% em 2016) e muito deste desenvolvimento é fruto de avanços tecnológicos que aumentam a produtividade da agricultura (FIGUEIREDO; CORRÊA, 2006).

Uma das informações de interesse é a textura do solo (YOLCUBAL et al., 2004), uma característica fundamental do solo que impulsiona a produção agrícola e o manejo do campo. Vários fatores para o desenvolvimento ótimo de alimentos são influenciados por ela: sistema de drenagem, aeração, quantidade de matéria orgânica, níveis de pH e outros (LEPSCH; SILVA; ESPIRONELO, 1982; CARLESSO; SANTOS, 1999; SANTOS et al., 2008). Como exemplo sabe-se que terrenos siltosos são retentores de luz e umidade (WUEST et al., 2005), o que promove boas condições de fertilidade (TIESSEN; CUEVAS; CHACON, 1994) e por este motivo são interessantes para agricultura.

Dada sua importância neste cenário, estudar a distribuição espacial da composição do solo pode ser uma maneira de identificar áreas de maior fertilidade e diminuir a quantidade utilizada de fungicidas para cultivo de alimentos, similar ao que foi feito em Freitas et al. (2016). Entretanto os dados analisados são chamados de composicionais, por representarem partes de um todo (AITCHISON, 1986), de forma que é possível ajustá-los para que cada amostra composicional seja restrita à soma unitária e represente proporções de um todo. Porém, tal restrição impõe uma dificuldade para a krigagem, que não é uma abordagem adequada para realizar predições em casos que os dados são restritos, já que tais predições não reconheceriam tal estrutura (PAWLOWSKY-GLAHN; EGOZCUE, 2016).

Com teoria de dados composicionais é possível mapear unicamente as observações restritas para o domínio dos reais e em seguida utilizar geoestatística. Neste momento entra a necessidade de postular uma função de covariância multivariada que consiga compreender a estrutura de dependência espacial presente na amostra. A função utilizada aqui será a Matérn bivariada, proposta por Genton e Kleiber (2015) e estimada por máxima verossimilhança. Finalmente realiza-se a Krigagem, no domínio dos reais, e transforma-se as predições de volta ao domínio restrito para obter os mapas composicionais preditos pela co-Krigagem, junto com mapas de erros de predições aproximados.

Por fim, em um conjunto de dados ainda não utilizado, compara-se o desempe-

nho desta abordagem com o de um classificador por vizinhos mais próximos ([HASTIE; TIBSHIRANI; FRIEDMAN, 2009](#)), criado por validação cruzada do tipo “deixe-um-de-fora”.

1 Análise de Dados Composicionais

1.1 Dado Composicional e o Simplex

Dados composicionais como um ramo da estatística é uma área relativamente nova, cuja metodologia foi unificada por [Aitchison \(1986\)](#). Aqui utiliza-se o simplex para definir o espaço amostral de interesse. O simplex S^D é caracterizado na definição que segue, de forma que seus elementos $\mathbf{x} = (x_1, \dots, x_D)$ são chamados de composições e a quantidade D de componentes determina a dimensão de seu espaço vetorial.

Definição 1. *Um simplex de dimensão D , S^D , é definido por*

$$S^D = \left\{ (x_1, \dots, x_D) \in \mathbf{R}^D \mid x_1 > 0, \dots, x_D > 0; \sum_{i=1}^D x_i = 1 \right\}.$$

Não necessariamente a definição de simplex fornecida concorda com dados coletados de um experimento, pois a soma das componentes podem estar em unidades de medida não-normalizadas. Como exemplo temos que na camada superficial de solos encontram-se quantidades de areia, argila e silte. Neste caso é possível relatar suas quantidades em valores brutos ou considerar suas proporções relativas ao total areia + argila + silte. Suponha neste então que uma observação, de todas as medições feitas, foi $\mathbf{w} = (w_1, \dots, w_D)$ tal que $w_1 + \dots + w_D \neq 1$. Neste caso é importante padronizar os dados de forma a seguir a Definição 1, o que pode ser obtido se dividirmos as coordenadas pelo seu comprimento: $\mathbf{w}/\|\mathbf{w}\|_1$, i.e, $\mathbf{x} = \mathbf{w}/\|\mathbf{w}\|_1$, em que $\|\mathbf{w}\|_1 = w_1 + \dots + w_D$, posto que os valores de w_i são positivos, $i = 1, 2, \dots, D$. É então necessário distinguirmos \mathbf{w} (chamada de base) de \mathbf{x} (chamada de composição), pois a metodologia discutida aqui só poderá ser aplicada a \mathbf{x} . Aliás, uma base \mathbf{w} determina completamente seu tamanho $t = \|\mathbf{w}\|_1$ e sua composição \mathbf{x} , como diz a definição a seguir.

Definição 2. *Uma base \mathbf{w} de partes D é um vetor $D \times 1$ de componentes positivos (w_1, \dots, w_D) todos registrados na mesma escala de medição. Além disso cada base $w \in \mathbf{R}_+^D$ tem um tamanho $t = \sum_{i=1}^D w_i$ único e composição \mathbf{w}/t .*

Agora, um fato muito importante é que as estatísticas descritivas clássicas como média, variância e correlação têm comportamentos diferentes no simplex, como descrito em [Chayes \(1960\)](#), com mais detalhe em [Chayes \(1971\)](#) e [Aitchison \(1986\)](#). Alguns destes comportamentos são

1. Viés negativo.

2. Correlação Nula.

3. Restrição da Base.

O viés negativo acontece por causa da restrição da soma unitária e o motivo é ilustrado a seguir. Note que $\text{Var}\left(\sum_{i=1}^D x_i\right) = \text{Var}(1) = 0$, e se usarmos o operador de covariância, então

$$\begin{aligned} 0 &= \text{Cov}(x_1, 1) = \text{Cov}\left(x_1, \sum_{i=1}^D x_i\right) \\ &= \sum_{i=1}^D \text{Cov}(x_1, x_i) = \text{Var}(x_1) + \sum_{i=2}^D \text{Cov}(x_1, x_i), \end{aligned}$$

de onde concluímos que

$$\sum_{i=2}^D \text{Cov}(x_1, x_i) = -\text{Var}(x_1).$$

Esta igualdade quer dizer que ao menos uma covariância será negativa, o que não é necessariamente verdadeiro para os dados na forma \mathbf{w} . Vejamos um exemplo de não podermos estender essa conclusão para \mathbf{w} , ilustrado em [Aitchison \(1986, p.56\)](#):

$$\begin{aligned} \text{Corr}(w_1, w_2) &= 0.73, \text{Corr}(w_1, w_3) = 0.38, \text{Corr}(w_2, w_3) = 0.37, \\ \text{Corr}(x_1, x_2) &= -0.37, \text{Corr}(x_1, x_3) = -0.92, \text{Corr}(x_2, x_3) = -0.033. \end{aligned}$$

O fato de que $\text{Corr}(w_2, w_3) = 0.37$ e $\text{Corr}(x_2, x_3) = -0.033$ é decorrente da restrição da base, efeito da padronização feita para obtermos as proporções. Isto acontece já que a padronização é uma operação restritiva, pois adicionamos a condição $\sum_{i=1}^D x_i = 1$, para todo i . Além disso, o fenômeno da Correlação Nula, observado em $\text{Corr}(x_2, x_3) = -0.033$ é comum ao se realizar este tipo de análise. Ou seja, utilizar as estatísticas básicas sem as adaptar ao contexto composicional pode acarretar em confusões e inferências equivocadas. Para adaptá-las, uma forma proposta em [Aitchison \(1986\)](#) é interpretada como mapear os dados para outro espaço vetorial que não tenhamos a restrição da soma unitária e daí aplicá-las. Esta abordagem é a que permite a construção de uma linha de análise de dados composicionais, mas é possível fazer esta transformação de algumas formas como pelas transformações aditiva de log-razão (abreviada por alr) e isométrica de log-razão (abreviada por ilr). Aqui não é de interesse ampliar a teoria de dados composicionais, mas sim utilizar alguma destas transformações para migrar entre os domínios dos reais e do simplex. Entretanto é necessário escolher entre as diversas transformações existentes no campo de dados composicionais. Apesar de a transformação ilr poder em alguns cenários trazer melhores resultados ([WANG et al., 2020](#)), a escolha aqui é pela transformação aditiva de log-razão (alr) e o motivo disso é que ela é uma função mais matematicamente tratável

do que a ilr e também extensivamente estudada, com boas propriedades. A transformação aditiva de log-razão (alr) é definida como:

Definição 3. *Seja $d = D - 1$. A transformação alr é uma função bijetora de S^D para \mathbb{R}^d de forma*

$$z_{ij} = \log(x_{ij}/x_{iD}), i = 1, \dots, N \text{ e } j = 1, \dots, d,$$

com transformação inversa

$$\begin{aligned} x_{ij} &= \frac{e^{z_{ij}}}{1 + \sum_{j=1}^d e^{z_{ij}}}, i = 1, \dots, N \text{ e } j = 1, \dots, d, \\ x_{iD} &= \frac{1}{1 + \sum_{j=1}^d e^{z_{ij}}}. \end{aligned} \tag{1.1}$$

Segundo [Tjelmeland e Lund \(2003\)](#) esta transformação tem um apelo muito interessante por naturalmente sugerir a suposição de uma distribuição normal para os dados transformados. Neste caso, cada variável aleatória composicional $\mathbf{X} := (X_1, \dots, X_D)$ é distribuída como $\mathbf{X} \sim L^d(\boldsymbol{\mu}, \Sigma)$, de forma que a aplicação da função alr fornece $\mathbf{Z} := (Z_1, \dots, Z_d)$ tal que $\mathbf{Z} \sim N_d(\boldsymbol{\mu}, \Sigma)$, como formaliza a Definição 4 a seguir. Portanto caso seja razoável assumir tal hipótese, agora está disponível toda a teoria estatística desenvolvida em torno da normalidade multivariada, como testes de hipóteses, intervalos de confiança, modelos de regressão, etc. Também é bastante direto simular vetores aleatórios das composições, pois basta simular observações normais multivariadas e transportá-las ao simplex, via transformação alr .

Definição 4. *Um vetor aleatório composicional \mathbf{X} de D partes tem distribuição Normal-Logística Aditiva se o vetor aleatório $\mathbf{Z} = (\log(x_1/x_D), \dots, \log(x_d/x_D))$ de dimensão $d = D - 1$ é tal que $\mathbf{Z} \sim N_d(\boldsymbol{\mu}, \Sigma)$*

Com esta definição, temos toda a teoria de dados composicionais necessária para desenvolver a metodologia proposta nesse texto. Apesar de existir uma extensa quantidade de conhecimento criado para avançar a análise de dados composicionais, para a aplicação desejada aqui o indispensável é somente a ponte que liga os domínios dos números reais e simplex. Vamos agora brevemente comentar o diagrama ternário, uma ferramenta muito útil na análise exploratória de dados composicionais, para logo em seguida descrever a teoria de Geoestatística univariada e multivariada para, logo em seguida, integrar com a teoria de dados composicionais.

1.2 Diagrama Ternário

O diagrama ternário é uma maneira de visualizar o comportamento das composições que estejam no S^3 em um gráfico de duas dimensões. A Figura 1 abaixo ilustra qual

é a intuição para obter cada ponto do diagrama ternário, que é basicamente centrar os pontos na origem e projetá-los no simplex. Apesar disto, existe a desvantagem de que como o espaço do simplex projetado no diagrama é menor do que o \mathbb{R}^2 , rapidamente pode-se popular o diagrama ternário de forma a esconder os padrões de interesse quando existe uma grande quantidade de pontos, por isso ele é mais útil quando não há uma grande quantidade de pontos.

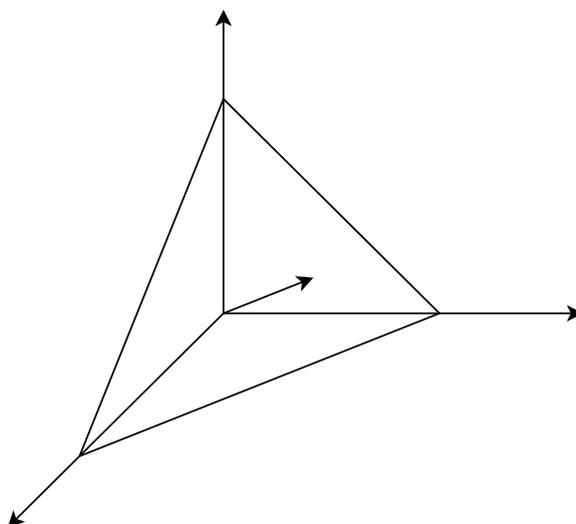


Figura 1 – Exemplo de diagrama ternário

Um conjunto de dados composicionais é tipicamente exibido na forma ilustrada na Figura 2. Os dados são proporções de areia, argila e silte em localizações observadas no leste de Minas Gerais e o norte de São Paulo e estão disponíveis no Repositório Brasileiro Livre para Dados Abertos do Solo (SAMUEL-ROSA et al., 2020) e corresponde ao item “CTB0562”. Dele foram extraídas as composições de areia, argila e silte da tabela “camada” e suas respectivas localizações, da tabela “observação”. Nesta aplicação, em alguns locais foram coletadas amostras em várias profundidades e, como forma de simplificar a interpretação dos resultados, foram escolhidas as composições mais próximas da superfície. No mapa da Figura 2 temos a região que foi considerada, em que os pontos coloridos representam as coordenadas das amostras e suas respectivas classificações, criadas de acordo com o Sistema Brasileiro de Classificação de Solos (SiBCS) (SOLOS, 2013) ¹.

¹ O anexo E contém o diagrama ternário que cria as classificações

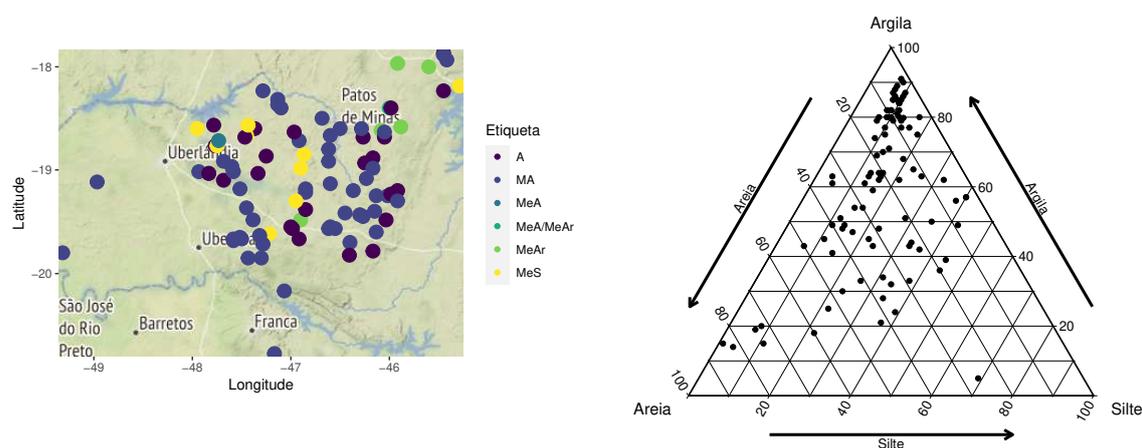


Figura 2 – Esquerda: Mapa com as classificações associadas às composições de areia, argila e silte em uma região de Minas Gerais. Nota-se que nessa amostra há uma grande concentração de argila. As classificações (“Etiqueta”) são fornecidas pela Tabela 5, disponível na Seção E do Apêndice. Direita: Diagrama ternário com os dados da região de Minas Gerais.

2 Geoestatística

2.1 Contextualização e Conceitos Básicos

A Geoestatística é uma área da Estatística Espacial que se preocupa em analisar observações geo-referenciadas para predizer ou simular observações em locais em que ainda não foram coletados dados. O coração do ferramental geoestatístico é composto pela caracterização da dependência espacial, i.e, de que forma e com que intensidade diferentes observações estão interligadas. Para isto, postula-se uma função que descreve tal dependência, a chamada função de covariância.

Para chegar ao coração, entretanto, é preciso compreender outras partes do corpo da geoestatística. Começamos por descrever o cenário em que usualmente aplica-se geoestatística. Considere que uma medição feita na localização $s \in B$ é $z(\mathbf{s})$, em que B é um subconjunto de \mathbb{R}^2 e representa a região de interesse. Modela-se esta realização do experimento por meio de um processo estocástico indexado no espaço, aqui $Z(\mathbf{s})$. Um processo estocástico $X(i)$ é uma coleção de variáveis aleatórias X (JAMES, 1996) indexadas por i , em que i é elemento de um conjunto E . Como exemplo o conjunto E pode ser o conjunto dos valores reais \mathbb{R} , dos inteiros \mathbb{N} ou um subconjunto destes dois. Assim um processo estocástico pode representar:

- O peso de um pessoa, conforme os dias: $X \in \mathbb{R}$ reais e $E \in \mathbb{N}$;
- O número de passos em uma caminhada, com o tempo: $X \in \mathbb{N}$ reais e $E \in \mathbb{R}_+$.

Em parentesco com o contexto colocado anteriormente, nosso processo estocástico será $Z(\mathbf{s})$, onde $Z \in \mathbb{R}$ e $B \subset \mathbb{R}^2$. O próximo passo é descrever o que é a distribuição espacial deste processo, ou seja, como mapear um evento indexado por coordenadas geográficas em probabilidade. Para isto, precisamos de uma definição formal de processo estocástico que, de acordo com Gaetan e Guyon (2010):

Definição 5. *Um processo estocástico que assume valores em E é um família $X = \{X(\mathbf{s}) : \mathbf{s} \in E\}$ de variáveis aleatórias em (Ω, F, P) com estado de espaços (E, \mathcal{E}) .*

Definição 6. *Para cada inteiro $n \geq 1$ e $(\mathbf{s}_1, \dots, \mathbf{s}_n) \in B^n$ e para cada $A_i \in \mathcal{E}$, a distribuição de $X(\mathbf{s}_1), \dots, X(\mathbf{s}_n)$ é*

$$P(A_1, \dots, A_n) = P(X(\mathbf{s}_1) \in A_1, \dots, X(\mathbf{s}_n) \in A_n) \quad (2.1)$$

e a distribuição espacial do processo é família de todas as distribuições multidimensionais de X .

O que temos definido até então é uma estrutura de base bastante geral para descrever processos espaciais mas faltam ainda ideias que permitam nos restringir a cenários matematicamente menos gerais, mas não realisticamente simplistas demais. Continuamos então com os conceitos de estacionariedade e isotropia para processos de segunda ordem.

Definição 7. Um processo aleatório é dito de Segunda Ordem se $\mathbb{E}(X^2(\mathbf{s})) < \infty$, para todo $\mathbf{s} \in B$. Além disso sua esperança $\mathbb{E}(X(\mathbf{s})) = \mu(\mathbf{s})$ é chamada função média e $C(\mathbf{s}_i, \mathbf{s}_j) = \text{Cov}(X(\mathbf{s}_i), X(\mathbf{s}_j))$ é a sua de função de covariância.

Definição 8. Um processo de segunda ordem é dito estacionário se $\mu(\mathbf{s}) = \mu$ e $C(\mathbf{s}_i, \mathbf{s}_j) = C(\mathbf{s}_i - \mathbf{s}_j)$.

Definição 9. Um processo de segunda ordem estacionário é dito isotrópico se sua função de covariância $C(\mathbf{s}_i - \mathbf{s}_j)$ depende somente de $\|\mathbf{s}_j - \mathbf{s}_i\|$

A definição de estacionariedade serve para representar o caso em que a esperança não muda conforme a localização, ou seja, para cada $\mathbf{s} \in B$, a média populacional do processo se mantém constante. Além disso há também suposição sobre a maneira em que as observações dependem entre si: a forma que as observações covariam é somente função da diferença entre as localizações, $\mathbf{s}_i - \mathbf{s}_j$. Isto nada nos diz sobre a força desta relação, mas sendo $\mathbf{s}_i - \mathbf{s}_j$ um vetor, temos que a interdependência entre as observações depende da direção e do tamanho da diferença entre suas coordenadas, como mostra a Figura 3 a seguir.

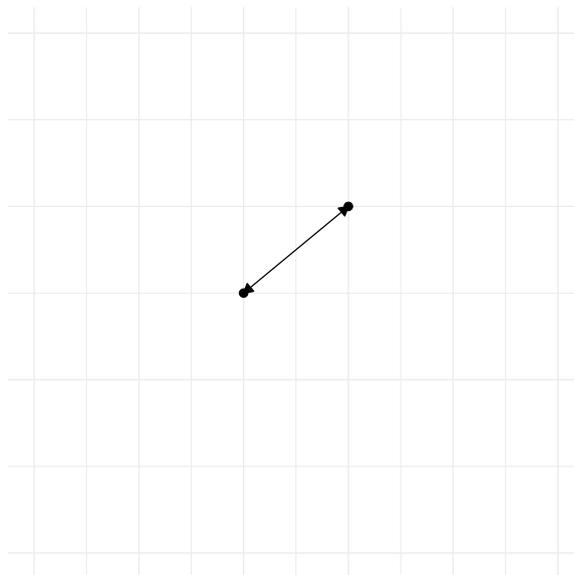


Figura 3 – Dependência entre dois pontos (supostas observações) imposta pela estacionariedade. Sendo a covariância função somente de um vetor, ela é caracterizada por direção, sentido e tamanho.

Isotropia, por sua vez, supõe que a covariância é função somente da distância entre as observações. Isto significa que removemos da Figura 3 anterior a direção e o sentido, e que nos é suficiente somente o comprimento do vetor. Outra maneira de enxergar esta implicação é que agora a forma da dependência é esférica. Entretanto se é sabido que a dependência não é esférica, como é caso das correntes de vento, é mais interessante considerar o caso anisotrópico, que leva em consideração direção e sentido. As duas situações estão ilustradas na Figura 4 abaixo, entretanto daqui em diante assume-se a situação de isotropia.

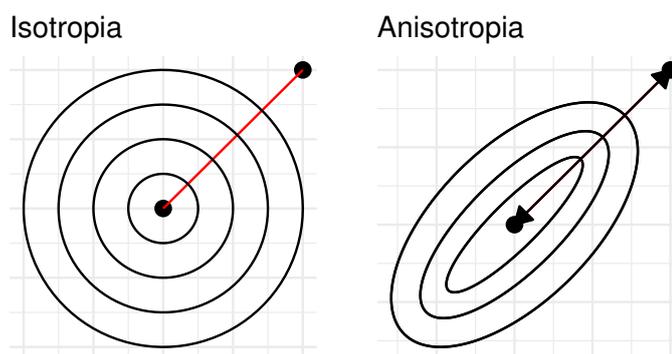


Figura 4 – Gráficos de contornos para dependência entre dois pontos (supostas observações) nos casos de isotropia e anisotropia. Sendo a covariância função somente da distância entre as observações, não é mais preciso saber sua direção e sentido e a dependência é esférica. Já para a anisotropia considera-se direção e sentido, o que nos permite mais maleabilidade da formulação da co-dependência.

2.2 Função de Covariância e o Variograma

Passamos agora a discutir sobre a força da dependência. Para duas medições $z(\mathbf{s}_1)$, $z(\mathbf{s}_2)$ no espaço, é bastante intuitivo e até razoável assumir em muitos casos que se elas estão perto, devem ser bem similares – no sentido de que os valores assumidos por $z(\mathbf{s}_1)$ e $z(\mathbf{s}_2)$ devem ser próximos. Enquanto que se elas estiverem longe, devem ser bem diferentes. Mas proximidade é um conceito vago, pois depende de um referencial, e cada análise trás consigo sua própria definição de “próximo”. Por isso não é o propósito aqui sanar este problema, mas sim fornecer meios de quantificar proximidade.

De volta ao pensamento das duas medições feitas, como podemos atribuir um valor à força das medições baseada nas distância entre elas? A solução é supor uma forma funcional para ela que leve em consideração a distância. Alguns exemplos, que podem ser encontrados em (WACKERNAGEL, 2013) e (CRESSIE, 1992), são os modelos

- Exponencial (CRESSIE, 1992):

$$C(\mathbf{h}) = \sigma^2 e^{-a\|\mathbf{h}\|^2}$$

- Racional Quadrático (CRESSIE, 1992; WILSON; ADAMS, 2013):

$$C(\mathbf{h}) = \sigma^2 e^{(1+a\|\mathbf{h}\|^2/2\beta)^\beta}$$

- Esférico (WACKERNAGEL, 2013):

$$C(\mathbf{h}) = \sigma^2 \left(1 - \frac{3a\|\mathbf{h}\|}{2} \frac{a^3\|\mathbf{h}\|^3}{2} \right) \mathbb{1}(0 \leq \|\mathbf{h}\| \leq a)$$

- Efeito de Pepita (CRESSIE, 1992):

$$C(\mathbf{h}) = \tau \mathbb{1}\{\mathbf{h} = \mathbf{0}\}$$

Dos exemplos citados a função de covariância “efeito de pepita” tem um papel distinto em geoestatística. O motivo é que ele pode ser utilizado conjuntamente com outras funções de covariância para amenizar o problema de que em processos geoespaciais normalmente não é possível modelar toda a incerteza presente nos dados, já que a resolução dos dados e sua natureza observacional costumam ser fatores limitantes para compreender toda a estrutura amostrada.

Além dos exemplos citados anteriormente, também existe a função de covariância Matérn que, para $h \in B$ é dada por

$$C(\mathbf{h}) = \sigma^2 \left(a\|\mathbf{h}\| \right)^\nu K_\nu \left(a\|\mathbf{h}\| \right).$$

em que $K_\nu(\cdot)$ é a Função de Bessel modificada do tipo II (MATHERON, 1963), ν é o parâmetro de suavidade e a^{-1} é o alcance de dependência. A Matérn engloba os modelos exponencial com $\nu = 1/2$ e Gaussiano com $\nu \rightarrow \infty$ também e por isso é especial interesse em Geoestatística. Abaixo temos a Figura 5, que mostra a realização de uma simulação de um processo Gaussiano¹ de média nula e função de covariância Matérn, em que $\sigma^2 = 3$, $1/a = 0.3$, $\nu = 1.0$. O parâmetro σ^2 representa a variabilidade e ilustra o quão dispersos são os valores, enquanto que ν confere a suavidade do processo e $1/a$ é representa a escala de dependência entre as observações.

¹ Feita a partir do pacote RandomFields (SCHLATHER et al., 2015)

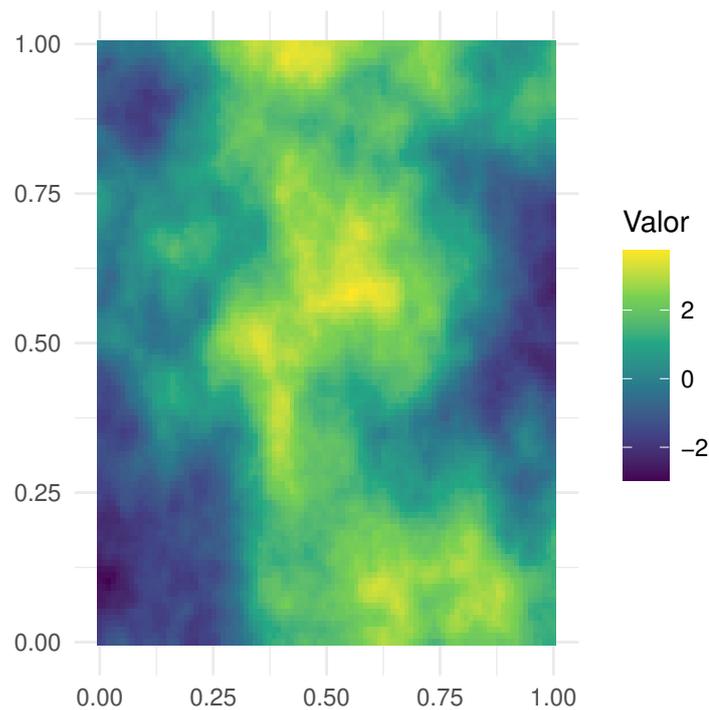


Figura 5 – processo Gaussiano Simulado simulado com função de covariância Matérn $\sigma^2 = 3, 1/a = 0.3, \nu = 1.0$

Mas apesar de termos encontrado uma maneira para incorporar a relação espacial entre as observações, não está claro como se escolhe entre os modelos citados. Para isto entra o variograma e a noção estacionariedade intrínseca.

Definição 10. $X(\mathbf{s})$ é um processo intrinsecamente estacionário se para cada $h \in B$ o processo $\Delta X^{(h)} = \{\Delta X^{(h)} = X(\mathbf{s} + \mathbf{h}) - X(\mathbf{s}) : \mathbf{s} \in B\}$ é estacionário de segunda ordem.

Definição 11. O semi-variograma γ de um processo X é

$$\gamma(\mathbf{h}) = \frac{1}{2} \text{Var}(X(\mathbf{s} + \mathbf{h}) - X(\mathbf{s})).$$

Analogamente $2\gamma(\mathbf{h})$ é chamado de variograma.

A noção de estacionariedade intrínseca vem para relaxar a hipótese de estacionariedade, que muitas vezes é uma hipótese muito forte. Aqui o processo gerado pelos incrementos $\Delta X^{(h)}$ é assumido como estacionário e sua variância é o variograma². O variograma e a função de covariância têm uma relação muito especial entre si: $\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$ e $C(\mathbf{0}) = \text{Var}(X(\mathbf{s}))$. Mas nem sempre é possível obter $C(\cdot)$ a partir de $\gamma(\cdot)$ pois em geral o variograma não é limitado. Ou seja, se reescrevermos a relação como $C(\mathbf{0}) = \gamma(\mathbf{h}) + C(\mathbf{h})$, percebe-se que implicitamente é permitido que $C(\mathbf{0})$ seja negativo, já que $\gamma(\mathbf{h})$ não é

² Note que a esperança do processo intrínseco é zero.

limitado. Entretanto, nos casos citados a relação entre eles é bem caracterizada (WACKERNAGEL, 2013). Para visualizarmos o variograma em ação, temos a Figura 6 que exibe o variograma empírico estimado para uma amostra aleatória de 85 localidades do processo Matérn simulado anteriormente na Figura 5. Nele temos que o variograma começa com o valor 0, o que representa que naquelas localidades não há efeito de pepita. Em seguida ele cresce rapidamente até o primeiro ponto de inflexão, que é aproximadamente 4, o que sugere 4 como valor para σ^2 e 0.5 para $1/a$.

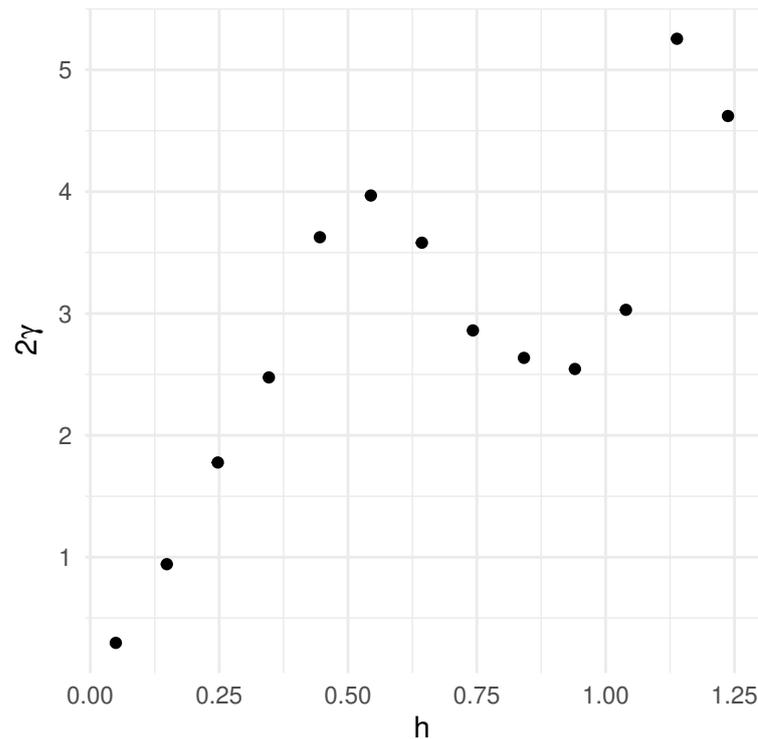


Figura 6 – Variograma Empírico obtido a partir de uma amostra de 85 pontos do processo gaussiano simulado na Figura 5 com função de covariância Matérn $\sigma^2 = 3$, $1/a = 0.3$, $\nu = 1.0$

Do exemplo anterior, vimos que o variograma representa uma ferramenta interessante na parte descritiva de uma modelagem geoestatística. Ele permitiu, sem ajustar modelos, observar algumas características intrínsecas do processo simulado e com isso, fornecer meios de auxiliar no processo de modelagem, como:

- Obter bons valores iniciais para parâmetros do modelo, como alcance de dependência a^{-1} ;
- Identificar se há o efeito de pepita (WACKERNAGEL, 2013);
- Escolher modelos de covariância. Exemplo: O modelo exponencial não é diferenciável na origem, mas o modelo gaussiano é. Então olhar o comportamento na origem do

variograma ajuda a diferenciar casos como este.

Mais adiante veremos que o variograma irá desempenhar um papel importante na estimação do modelo proposto para a aplicação em dados de solo.

2.3 O processo Gaussiano

O processo Gaussiano é um elemento chave na estatística espacial e já bastante reconhecido: (CRESSIE, 1992; STEIN, 2012; BANERJEE et al., 2008). Ele é um processo estocástico definido de forma que a combinação linear (STRANG, 1993) de quaisquer variáveis aleatórias do processo tem distribuição normal. Ou ainda, cada subcoleção finita de variáveis tem distribuição normal multivariada. Para o contexto espacial, podemos definir ele como:

Definição 12. (GAETAN; GUYON, 2010) $X(\mathbf{s})$ é um processo Gaussiano em B se para cada subconjunto $G \subset B$ e para cada sequência $a = (a(\mathbf{s}) : \mathbf{s} \in G, a(\mathbf{s}) \in \mathbb{R})$, a variável aleatória $\sum_{\mathbf{s} \in G} a(\mathbf{s})X(\mathbf{s})$ tem distribuição normal.

Vale a pena aqui comentar um pouco do porque de ser assim, antes de avançarmos no contexto desejado. A discussão que a seguir acompanha Mörters e Peres (2010), que é recomendado para se obter mais detalhes.

Primeiro é interessante conhecer o processo de Wiener, também conhecido por movimento Browniano. De forma muito ingênua e por meio de um exemplo, ele é uma sequência de variáveis aleatórias indexadas no tempo, $W(t)$, com distribuição normal. Na Figura 7 abaixo temos a ilustração deste processo obtida a partir de simulação e, logo em seguida, uma maneira de o definir.

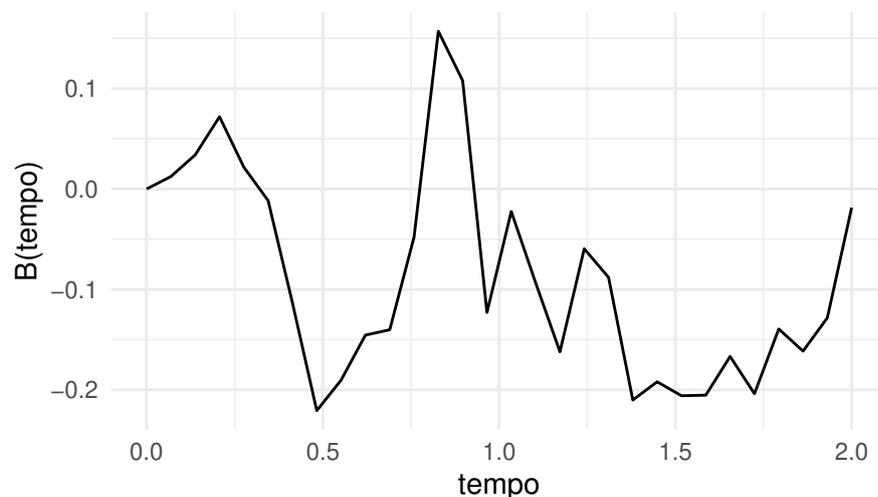


Figura 7 – Processo de Wiener simulado em $[0, 2]$ com início em $t = 0$. Este é um processo de Wiener padrão.

Definição 13. Um processo estocástico $\{W(t), t \geq 0\}$ que assume valores em \mathbb{R} é um processo de Wiener com começo em $x \in \mathbb{R}$ se $B(0) = x$ e se

(i) (Incrementos Independentes): Para quaisquer tempo $0 \leq t_1 \leq \dots \leq t_n$, as variáveis aleatórias

$$W(t_n) - W(t_{n-1}), \dots, W(t_2) - W(t_1), W(t_1)$$

são independentes

(ii) Para quaisquer t e h não-negativos,

$$W(t+h) - W(t) \sim N(0, h)$$

(iii) Com probabilidade um a função $B : t \mapsto W(t)$ é contínua

Além disso, define-se o processo de Wiener padrão por $\{W(t), t \geq 0, x = 0\}$.

Mas qual é exatamente a ligação entre um processo de Wiener e um processo Gaussiano? Um processo Gaussiano é aquele que $(X(t_1), \dots, X(t_n))$ tem distribuição normal multivariada, para qualquer conjunto de índices t_1, \dots, t_n ou que uma combinação linear delas é normal. Como para o processo de Wiener cada $W(t_1), \dots, W(t_n)$ tem a propriedade de incrementos independentes, uma combinação linear dos incrementos também é normal (assim como qualquer subcoleção dos incrementos $(W(t_n) - W(t_{n-1}), \dots, W(t_2) - W(t_1), W(t_1))$ é normal multivariada). A definição alternativa a seguir pode ajudar a esclarecer melhor a relação.

Definição 14. (MÖRTERS; PERES, 2010) Um processo estocástico $\{W(t), t \geq 0\}$ é um movimento Browniano se e somente

- $W(t)$ é um processo Gaussiano
- $\mathbb{E}(W(t)) = 0$ e $\text{Cov}(X(\mathbf{s}), X(t)) = \min(s, t)$, para todo $s, t \geq 0$
- Com probabilidade um a função $B : t \mapsto W(t)$ é contínua

Em outras palavras, um processo de Wiener é um processo Gaussiano com função média nula e função de covariância $\Sigma(s, t) = \text{Cov}(X(\mathbf{s}), X(t)) = \min(s, t)$, ou seja: $W(t) \stackrel{B}{=} PG(0, \Sigma(s, t))$. Para a Geoestatística, incorpora-se o processo Gaussiano de uma forma muito simples: como $B \subset \mathbf{R}^2$, para cada $s \in B$ temos que $Z(\mathbf{s}) \sim PG(\mu(\mathbf{s}), \Sigma)$, onde $\Sigma_{ij} = \text{Cov}(X(s_i), X(s_j))$ é uma função postulada. Com isso, a partir da definição 12 é possível obter a função densidade de probabilidade de cada variável aleatória:

$$f_B(x_B) = (2\pi)^{-\text{card}(B)/2} \det(\Sigma_B)^{-1/2} \exp \left\{ -\frac{1}{2} (x_B - \mu_B)^T \Sigma_B^{-1} (x_B - \mu_B) \right\}, \quad (2.2)$$

Isto abre portas para calcularmos a verossimilhança e a log-verossimilhança do processo para posterior estimação.

2.4 Krigagem

A Krigagem é uma técnica de regressão capaz de, a partir de medições obtidas somente em algumas localizações, realizar a predição de observações em novas coordenadas. A Figura 8 exemplifica o cenário habitual de uso da Krigagem, aquele em que a partir de medições feitas em certas coordenadas (bolinhas pretas), deseja-se realizar a predição para uma nova localização em que ainda não se realizou medição.

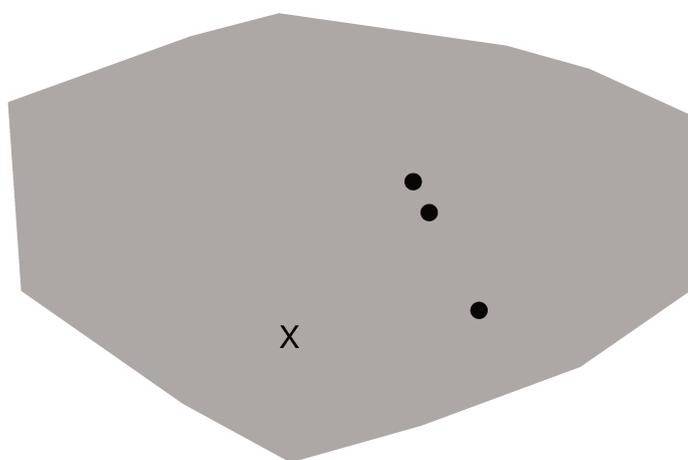


Figura 8 – Ilustração de uma Krigagem. A partir de medições feitas em certas coordenadas (bolinhas pretas), deseja-se realizar a predição para uma nova localização em que ainda não se realizou medição (marcado com \times).

Neste caso, como descrito em [Cressie e Wikle \(2015, p. 133\)](#) o modelo a ser considerado é $Z(\mathbf{s}) = \mu(\mathbf{s}) + \eta(\mathbf{s}) + \epsilon(\mathbf{s})$, em que $\mathbf{s} \in \mathbb{R}^2$ representa uma determinada lococoordenada geográfica (latitude e longitude) e

1. $\mu(\mathbf{s}) = \mathbb{E}(Z(\mathbf{s}))$ é chamada de variação determinística, ou de larga-escala.
2. $\eta(\mathbf{s})$ é chamado de efeito aleatório espacial, somente detectável a partir da resolução amostral.
3. $\epsilon(\mathbf{s})$ é a variação em micro-escala e inclui efeitos que não podem ser medidos na resolução amostral, além do erro instrumental.

Este modelo informa que é possível particionar $Z(\mathbf{s})$ em três partes, sendo $\mu(\mathbf{s})$ detectável e $\epsilon(\mathbf{s})$ não detectável, enquanto que $\eta(\mathbf{s})$ talvez seja detectável. Como a principal função da Krigagem é recuperar a variação determinística, ou seja, encontrar os

padrões espaciais de grande escala e, como consequência suavizar os efeitos aleatórios e de micro-escala, então podemos reescrever o modelo e considerar, para termos práticos, somente $Z(\mathbf{s}) = Z^*(\mathbf{s}) + \epsilon(\mathbf{s})$ em que $Z^*(\mathbf{s}) = \eta(\mathbf{s}) + \mu(\mathbf{s})$. Para isto, assume-se a estrutura de um processo Gaussiano para descrever este processo espacial, como o da Definição 12. Em seguida é necessário especificar as funções média e de covariância para descrever a dependência espacial e, uma vez feito isso, a Krigagem pode ser resumida como a esperança condicional de uma variável aleatória normal multivariada. Para a Figura 8, assume-se que os três pontos esféricos e o ponto \times têm distribuição conjunta normal multivariada e portanto, a Krigagem é a esperança condicional do ponto marcado com \times na Figura 8, dados os pontos marcados em bolinhas pretas. Portanto se $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_N))$ são as N amostras em N localizações e $\mathbf{Z}_0^* = (Z(\mathbf{s}_{N+1}), \dots, Z(\mathbf{s}_{N+N^*}))$ são aquelas outras N^* em que se deseja fazer a predição, assume-se que o processo é intrinsecamente estacionário de forma que

$$\begin{pmatrix} \mathbf{Z} \\ \mathbf{Z}_0^* \end{pmatrix} \sim \mathbf{N}_{N+N^*} \left(\begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}_0^* \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right), \quad (2.3)$$

onde $\boldsymbol{\mu}_0^* = \mathbb{E}(\mathbf{Z}_0^*)$, $\boldsymbol{\Sigma}_{22} = \text{Var}(\mathbf{Z}_0^*)$ e $\boldsymbol{\Sigma}_{12} = \text{Cov}(\mathbf{Z}, \mathbf{Z}_0^*)$. Assim a distribuição condicional do vetor aleatório $\mathbf{Z}_0^* | \mathbf{Z} = \mathbf{z}$ é normal multivariada com média e covariâncias dadas por

$$\boldsymbol{\mu}_{\mathbf{Z}_0^* | \mathbf{Z}} = \mathbb{E}(\mathbf{Z}_0^* | \mathbf{Z} = \mathbf{z}) = \boldsymbol{\mu}_0^* + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{z} - \boldsymbol{\mu}), \quad (2.4)$$

$$\bar{\boldsymbol{\Sigma}} = \text{Cov}(\mathbf{Z}_0^* | \mathbf{Z} = \mathbf{z}) = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{21}. \quad (2.5)$$

Mas como exatamente isto se encaixa em uma formulação geoespacial? Uma forma é: sendo fornecida uma amostra aleatória z_1, \dots, z_N , onde assume-se que $\mu(\mathbf{s}) = \mu$ (que recebe o nome de Krigagem simples), para realizar a predição em um par de coordenadas ainda não amostrado, define-se o melhor preditor linear de $Z^*(s_0)$, abreviado por Z_0^* , como a combinação linear $p(z, s_0) = \sum_{i=1}^N l_i z_i + m$ que minimiza $\mathbb{E}([Z_0^* - p(z, s_0)]^2)$, sujeita à restrição $\sum_{i=1}^n l_i = 1$. O resultado desta minimização é o estimador de mínimos quadrados generalizados de $\mu(\mathbf{s})$, que irá coincidir com a esperança condicional, como veremos a seguir. Ao supor que conhecemos a distribuição de Z, Z_0^*, ϵ , que o processo é intrinsecamente estacionário e que $\mu(\mathbf{s}) = \mu$ temos que o Lagrangiano do problema é

$$\underset{l_1, \dots, l_N, \lambda}{\text{argmin}} E \left[Z_0^* - \sum_{i=1}^N l_i z_i \right]^2 - 2\lambda \left(\sum_{i=1}^N l_i - z_i \right). \quad (2.6)$$

Agora já que $\sum_{i=1}^n l_i = 1$, e que $2\gamma(\mathbf{s}) = C(\mathbf{0}) - C(\mathbf{s})$,

$$\left[Z_0^* - \sum_{i=1}^N l_i z_i \right]^2 = - \sum_{i=1}^N \sum_{j=1}^N l_i l_j (z_i - z_j)^2 / 2 + \sum_{i=1}^N l_i (Z_0^* - z_i)^2 / 2$$

e portanto o problema torna-se

$$\operatorname{argmin}_{l_1, \dots, l_N, \lambda} - \sum_{i=1}^N \sum_{j=1}^N l_i l_j \gamma(\mathbf{s}_i - \mathbf{s}_j) + 2 \sum_{i=1}^N l_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) - 2\lambda \left(\sum_{i=1}^N l_i - 1 \right),$$

de forma que ao derivar em relação a l_1, \dots, l_N, λ , obtém-se que os parâmetros ótimos $l_1, \dots, l_N = \mathbf{1}$ são

$$\mathbf{l}^T = \left(\boldsymbol{\gamma} + \mathbf{1} \frac{1 - \mathbf{1}\Gamma^{-1}\boldsymbol{\gamma}^T}{\mathbf{1}\Gamma^{-1}\mathbf{1}^T} \right)^T \Gamma^{-1}$$

e

$$\lambda = - \frac{1 - \mathbf{1}\Gamma^{-1}\mathbf{1}^T}{\mathbf{1}\Gamma^{-1}\boldsymbol{\gamma}},$$

onde $\gamma_i = \operatorname{Cov}(\mathbf{s}_0, \mathbf{s}_i)$. Portanto

$$\begin{aligned} p(\mathbf{z}, \mathbf{s}_0) &= \mathbf{l}^T \mathbf{z} = \left(\boldsymbol{\gamma} + \mathbf{1} \frac{1 - \mathbf{1}\Gamma^{-1}\boldsymbol{\gamma}^T}{\mathbf{1}\Gamma^{-1}\mathbf{1}^T} \right)^T \Gamma^{-1} \mathbf{z} \\ &= \bar{\mathbf{z}} + \boldsymbol{\gamma}^T \Gamma^{-1} (\mathbf{z} - \mathbf{1}\bar{\mathbf{z}}) \end{aligned}$$

onde $\bar{\mathbf{z}} = \frac{\mathbf{1}\Gamma^{-1}\boldsymbol{\gamma}^T}{\mathbf{1}\Gamma^{-1}\mathbf{1}^T}$. Mas ao lembrar que para processos estacionários vale que $2\gamma(\mathbf{s}) = C(\mathbf{0}) - C(\mathbf{s})$, então podemos reescrever o resultado acima como

$$\begin{aligned} p(\mathbf{z}, \mathbf{s}_0) &= \mathbf{l}^T \mathbf{z} = \left(\boldsymbol{\gamma} + \mathbf{1} \frac{1 - \mathbf{1}\Sigma_Z^{-1}\Sigma_{Z_0, Z}}{\mathbf{1}\Sigma_Z^{-1}\mathbf{1}^T} \right)^T \Sigma_Z^{-1} \mathbf{z} \\ &= \bar{\mathbf{z}} + \Sigma_{Z_0, Z}^T \Sigma_Z^{-1} (\mathbf{z} - \mathbf{1}\bar{\mathbf{z}}) \end{aligned}$$

em que $\bar{\mathbf{z}} = \frac{\mathbf{1}\Sigma_Z^{-1}\mathbf{z}}{\mathbf{1}\Sigma_Z^{-1}\mathbf{1}}$, o que remete à distribuição condicional de uma variável aleatória normal multivariada $N_N(p(\mathbf{z}, \mathbf{s}_0), \Sigma_{Z_0, Z}^T \Sigma_Z^{-1} \Sigma_{Z_0, Z})$. Já a Krigagem universal permite uma estrutura mais flexível sobre $\mu(\mathbf{s})$, de forma que agora $\beta_0 + \sum_{k=1}^p \beta_k f_k(\mathbf{s}) + \psi(\mathbf{s})$ onde ψ é um processo intrinsecamente estacionário, f_1, \dots, f_p são funções conhecidas e β_0, \dots, β_p são desconhecidos. A flexibilidade aqui vem da capacidade de incorporar covariáveis ao modelo, desde que elas estejam disponíveis nas localidades em que ocorre a estimação e naquelas em deseja-se fazer a predição. Para mais detalhes, veja [Cressie \(1992\)](#). A Figura 9 mostra o resultado da Krigagem simples³ em uma malha de dimensões 100×100 , com parâmetros estimados por máxima verossimilhança ([DIGGLE; RIBEIRO; CHRISTENSEN, 2003](#)) a partir de 85 coordenadas sorteadas ao acaso, do processo gaussiano com função de covariância Matérn simulado na Figura 5 de parâmetros $\sigma^2 = 3, \nu = 1.0, 1/a = 0.3$. Dela é possível perceber que a Krigagem é capaz de recuperar os grandes padrões do processo original na Figura 5 de forma suavizada, isto é, padrões mais sutis não são discernidos pela Krigagem. Já a Figura 10 exhibe o erro de predição envolvido e dela nota-se que os pontos

³ Feito a partir do pacote `geor` ([Ribeiro Jr et al., 2020](#))

em que a estrutura não foi tão bem captada estão associados a erros maiores, também dela é possível observar que nos pontos amostrados o erro de predição é zero, o que é uma característica interessante da Krigagem e é explicada pelo fato de que a predição espacial em pontos amostrados é o próprio valor do ponto amostrado.

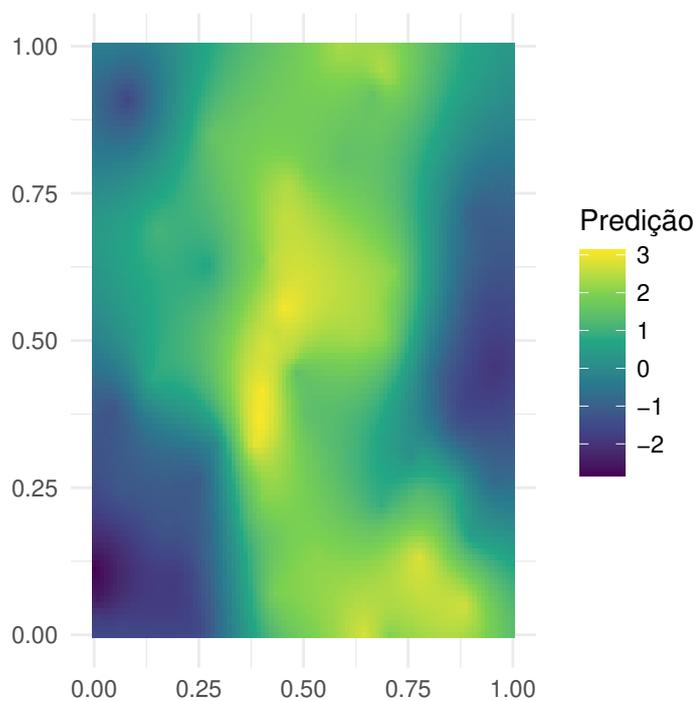


Figura 9 – Krigagem baseada em uma amostra aleatória de 85 pontos, obtida a partir do processo simulado ilustrado na Figura 5

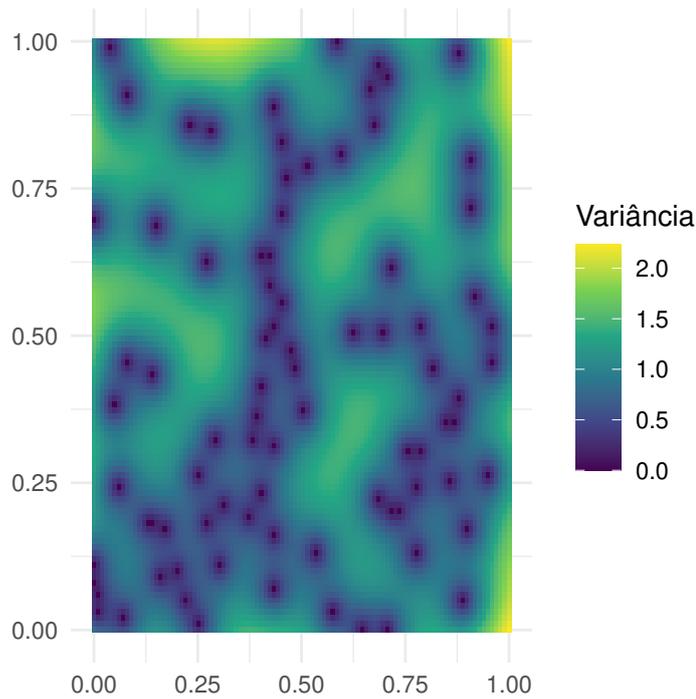


Figura 10 – Variância da Krigagem baseada em uma amostra aleatória de 85 pontos, obtida a partir do processo simulado ilustrado na Figura 5

2.5 Geoestatística Multivariada

Esta Seção é dedicada a estender alguns conceitos estabelecidos nas seções (2.1) e (2.2), com o intuito de fornecer ferramentas para lidar com respostas multivariadas. Aqui novamente temos que $\mathbf{s} \in \mathbb{R}^2$, onde B é um subconjunto de \mathbb{R}^2 e representa a região de interesse. Entretanto a diferença com o cenário univariado é que agora a variável de interesse é $\mathbf{z}(\mathbf{s})$ e assume valores em \mathbb{R}^k , $k \geq 2$. Em particular vamos considerar $k = 2$.

Novamente utiliza-se um processo estocástico $\mathbf{Z}^T(\mathbf{s}) = (Z_1(\mathbf{s}), Z_2(\mathbf{s}))$ para representar nossas observações $\mathbf{z}(\mathbf{s})$, mas dessa vez bivariado. Para tal processo é possível assumir estacionariedade de segunda ordem conjunta para as duas variáveis:

Definição 15. Para $\mathbf{s}, \mathbf{h} \in \mathbb{R}^2$ o processo bivariado $\mathbf{Z}^T(\mathbf{s}) = (Z_1(\mathbf{s}), Z_2(\mathbf{s}))$ é conjuntamente estacionário se

$$\mathbb{E}(Z_1(\mathbf{s} + \mathbf{h})) = \mathbb{E}(Z_1(\mathbf{s})) = \mu_1$$

$$\mathbb{E}(Z_2(\mathbf{s} + \mathbf{h})) = \mathbb{E}(Z_2(\mathbf{s})) = \mu_2$$

$$\mathbb{E}([Z_1(\mathbf{s} + \mathbf{h}) - \mu_1][Z_1(\mathbf{s}) - \mu_1]) = C_{11}(\mathbf{h})$$

$$\mathbb{E}([Z_2(\mathbf{s} + \mathbf{h}) - \mu_2][Z_2(\mathbf{s}) - \mu_2]) = C_{22}(\mathbf{h})$$

$$\mathbb{E}([Z_2(\mathbf{s} + \mathbf{h}) - \mu_2][Z_1(\mathbf{s}) - \mu_1]) = C_{12}(\mathbf{h})$$

$$\mathbb{E}([Z_1(\mathbf{s} + \mathbf{h}) - \mu_1][Z_2(\mathbf{s}) - \mu_2]) = C_{21}(\mathbf{h})$$

onde $C_{ii}(\mathbf{h})$ e $C_{ij}(\mathbf{h})$ são as funções de covariância direta e cruzada, respectivamente. Se organizarmos tais funções como

$$\boldsymbol{\Sigma}(\mathbf{h}) = \begin{pmatrix} C_{11}(\mathbf{h}) & C_{12}(\mathbf{h}) \\ C_{21}(\mathbf{h}) & C_{22}(\mathbf{h}) \end{pmatrix}, \quad (2.7)$$

e também definir um processo conjuntamente intrínseco:

Definição 16. Para $\mathbf{s}, \mathbf{h} \in \mathbb{R}^2$ o processo bivariado $\mathbf{Z}^T(\mathbf{s}) = (Z_1(\mathbf{s}), Z_2(\mathbf{s}))$ é conjuntamente intrínseco se

$$\begin{aligned} \mathbb{E}(Z_i(\mathbf{s}+\mathbf{h}) - Z_i(\mathbf{s})) &= 0, \forall i, j = 1, 2 \\ \text{Cov}\left[(Z_i(\mathbf{s}+\mathbf{h}) - Z_i(\mathbf{s})), (Z_j(\mathbf{s}+\mathbf{h}) - Z_j(\mathbf{s}))\right] &= 2\gamma_{ij}(h). \end{aligned}$$

Neste caso $2\gamma_{ii}(h)$ é chamado de variograma direto, enquanto que $2\gamma_{ij}(h)$ é definido como variograma cruzado. Com isso fica mais claro que definir um processo gaussiano bivariado, que será pelos moldes de $\mathbf{Z}(\mathbf{s}) \sim \mathbf{N}(\boldsymbol{\mu}(\mathbf{s}), \boldsymbol{\Sigma}(\mathbf{s}))$, em que $\boldsymbol{\mu}(\mathbf{s})^T = (\mu_1(\mathbf{s}), \mu_2(\mathbf{s}))$ e $\boldsymbol{\Sigma}(\mathbf{s})$ é dado por (2.7), de forma que para realizar uma Krigagem bivariada (ou co-Krigagem) é necessário especificar a função de covariância (agora multivariada) e o tipo de Krigagem (simples, ordinário, universal). Existem alguns tipos de funções de covariância multivariada (estacionárias e não-estacionárias) e algumas elas podem ser encontradas em [Genton e Kleiber \(2015\)](#), em especial a função de covariância Matérn Multivariada, que será descrita com mais detalhes no capítulo a seguir. O tipo de Krigagem também será de importante especificação pois é possível distinguir alguns cenários interessantes (além dos casos simples, ordinário e universal) que são a Isotopia e Heterotopia e que podem ser compreendidos com mais detalhes em [Wackernagel \(2013\)](#).

- Heterotopia: as medições das duas respostas foram feitas em diferentes localizações.
- Isotopia: as medições das duas respostas foram feitas nas mesmas localizações.

A situação mais intuitiva é a da isotopia, em que a resposta bivariada é medida na mesma localização geográfica. A Heteropia é o contrário, quando elas são medidas em localidades diferentes, o que implica em alguns problemas, como a construção de funções de covariância válidas e a dificuldade de se construir o variograma cruzado empírico (já que não temos informações no mesmo ponto geográfico não é possível calcular a covariância cruzada). Existem também os cenários em que há dados parcialmente heterotópicos e em que a medição de uma variável é tida em todo o domínio (variável auxiliar) enquanto deseja-se a predição para a outra que é presente somente em alguns locais. Para mais detalhes destes casos veja ([WACKERNAGEL, 2013](#)), pois aqui o caso de interesse é aquele em que cada medição bivariada é feita na mesma localização ou seja, em um ponto do

domínio ambas as respostas estão presentes. Portanto a relação multivariada entre as observações será escrita como

$$\begin{pmatrix} Z_1(\mathbf{s}) \\ Z_2(\mathbf{s}) \end{pmatrix} \sim \mathbf{N}_{2N} \left(\begin{pmatrix} \mu(\mathbf{s}) \\ \mu(\mathbf{s}) \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right), \quad (2.8)$$

de forma que o critério de ajuste será (2.6). Ou seja, estimador será

$$Z_{COK}^*(\mathbf{s}) = \sum_{i=1}^N l_i(\mathbf{s})Z(\mathbf{s}_i) + \sum_{j=1}^N l'_j(\mathbf{s})Y(\mathbf{s}'_j),$$

de forma que $\sum_{l_i} \lambda_{l_i}(\mathbf{u}) = 1$ e $\sum_{l_j} \lambda_{l_j}(\mathbf{u}) = 0$, o que leva a estimadores similares aos obtidos em (2.4) e (2.5), com as dimensões adaptadas pois agora a dimensão da distribuição normal multivariada será $2N$, ao invés de N .

3 Geoestatística e Dados Composicionais

3.1 Geoestatística Multivariada para Dados Composicionais

A natureza multidimensional dos Dados Composicionais exige uma contrapartida multidimensional da Estatística Espacial. Aqui faremos uso do ferramental geoestatístico multivariado desenvolvido nos capítulos anteriores da Seção 2.1 e uni-lo com as metodologias de dados composicionais para criar previsões espaciais multivariadas de composições.

Suponha que é observado em N localizações $\mathbf{s}_1, \dots, \mathbf{s}_N$ as quantidades $\mathbf{w}_i(\mathbf{s}) = (w_{i1}(\mathbf{s}), w_{i2}(\mathbf{s}), w_{i3}(\mathbf{s}))$ de areia, argila e silte. O primeiro passo é transformar as quantidades em proporções, por meio da normalização $\mathbf{z}_i(\mathbf{s}) = \mathbf{w}_i(\mathbf{s}) / \|\mathbf{w}_i(\mathbf{s})\|_1$, para $i = 1, \dots, N$. Como os dados agora moram no simplex postula-se que sua representação no \mathbb{R}^2 , criada pela transformação alr (1.1) (AITCHISON, 1986), é dada pelo processo logístico-Gaussiano (TJELMELAND; LUND, 2003).

Definição 17. Para $\mathbf{s} \in B$ seja $\mathbf{Z}^T(\mathbf{s}) = (Z_1(\mathbf{s}), \dots, Z_d(\mathbf{s}))$ um processo Gaussiano multivariado com função média $E(\mathbf{Z}(\mathbf{s})) = \boldsymbol{\mu}^T(\mathbf{s}) = (\mu_1(\mathbf{s}), \dots, \mu_d(\mathbf{s}))$ e função de covariância $\boldsymbol{\Sigma}(\mathbf{s}) = C_{ij}(\mathbf{s}), i, j = 1, \dots, d$. Se $\mathbf{X}^T(\mathbf{s}) = (X_1(\mathbf{s}), \dots, X_d(\mathbf{s}))$ é o vetor aleatório obtido ao aplicar a função $\text{alr}(\cdot)$ em $\mathbf{Z}(\mathbf{s})$, então diz-se que $\mathbf{X}(\mathbf{s})$ segue um Processo Logístico-Gaussiano parametrizado por $\boldsymbol{\mu}(\mathbf{s})$ e $\boldsymbol{\Sigma}(\mathbf{s})$.

Aqui a postulação de $\boldsymbol{\Sigma}(\mathbf{s})$ volta a ser o elemento chave da modelagem e é preciso escolher uma classe de funções de covariância para representar as covariâncias diretas e cruzadas. Para o caso bivariado uma classe desenvolvida é a Matérn bivariada, também chamada de Wittle-Matérn, definida em (GNEITING; KLEIBER; SCHLATHER, 2010). Ao assumirmos estacionariedade de segunda ordem o modelo completo é

$$\begin{aligned}
 C_{11}(\mathbf{h}) &= \sigma_1^2 M(\mathbf{h}|a_1, \nu_1) \\
 C_{22}(\mathbf{h}) &= \sigma_2^2 M(\mathbf{h}|a_2, \nu_2) \\
 C_{21}(\mathbf{h}) &= C_{12}(\mathbf{h}) = \rho_{12} \sigma_1 \sigma_2 M(\mathbf{h}|a_3, \nu_3) \\
 M(\mathbf{h}|\nu, a) &= \frac{2^{1-\nu} (a \|\mathbf{h}\|)^\nu K_\nu(a \|\mathbf{h}\|)}{\Gamma(\nu)}
 \end{aligned} \tag{3.1}$$

Enquanto que o modelo reduzido é

$$\begin{aligned}
 C_{11}(\mathbf{h}) &= \sigma_1^2 M(\mathbf{h}|a, \nu_1) \\
 C_{22}(\mathbf{h}) &= \sigma_2^2 M(\mathbf{h}|a, \nu_2) \\
 C_{21}(\mathbf{h}) = C_{12}(\mathbf{h}) &= \rho_{12} \sigma_1 \sigma_2 M(\mathbf{h}|a, (\nu_1 + \nu_2)/2) \\
 M(\mathbf{h}|\nu, a) &= \frac{2^{1-\nu} (a \|\mathbf{h}\|)^\nu K_\nu(a \|\mathbf{h}\|)}{\Gamma(\nu)}
 \end{aligned} \tag{3.2}$$

A seguir temos a Figura 11 que representa estrutura de dependência para o caso $\sigma_1^2 = 1, \sigma_2^2 = 3, a_1 = a_2 = 1, \rho = 0.8, \nu_1 = 0.5, \nu_2 = 1.5, \nu_{12} = 1.0$. Pode-se perceber que é possível ter controle sobre (1) a suavidade do processo em pequenas distâncias, através de ν ; (2) a rapidez de decaimento das caudas, o que representa o alcance de dependência, por meio de a ; (3) a dispersão em torno da média (neste caso nula), por meio de σ^2 .

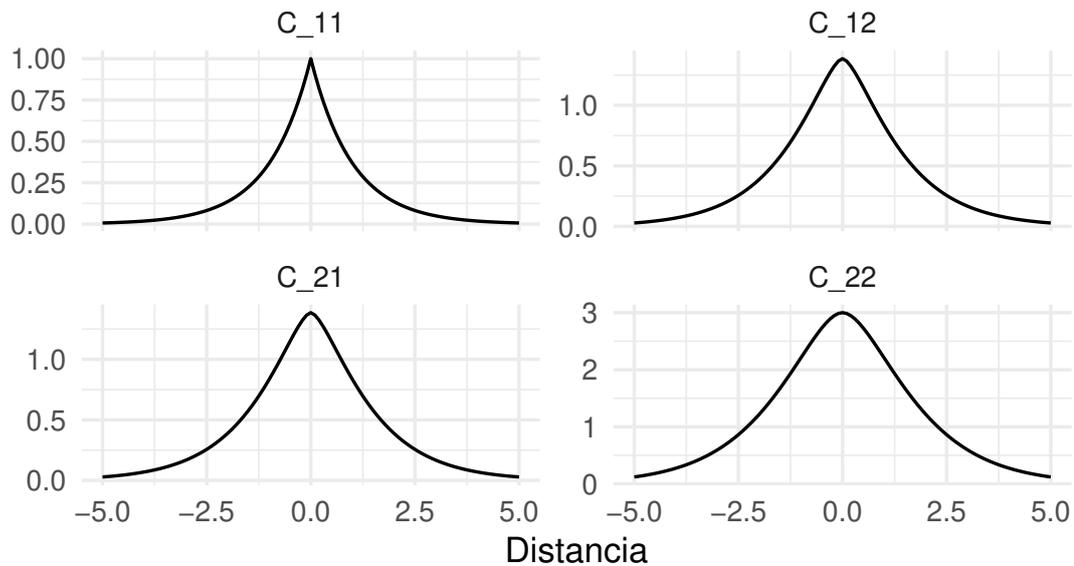


Figura 11 – Função de covariância Matérn bivariada. Aqui $\sigma_1^2 = 1, \sigma_2^2 = 3, a = 1, \rho = 0.8, \nu_1 = 0.5, \nu_2 = 1.5, \nu_{12} = 1.0$.

A simulação de um processo Gaussiano com vetor de funções médias nulo a partir para uma determinada função de covariância pode ser obtida por meio da biblioteca `RandomFields` (SCHLATHER et al., 2015) no pacote estatístico R (R Core Team, 2020). A Figura 12 a seguir mostra a simulação e dela é possível notar algumas das características das funções acima. Por exemplo uma característica é a dependência, em que localizações mais próximas têm valores parecidos com seus vizinhos. O mapa à esquerda da Figura 12, denominado processo 01, corresponde à realização de um processo com média zero e variância um, enquanto que o mapa à direita da Figura 12, corresponde à realização de um processo com média zero e variância três. A dependência espacial é a mesma ($a = 1$),

e a dependência entre os dois processos é alta ($\rho = 0.9$). Por exemplo, ambos exibem um “vale” com valores negativos na região $s_1 = -1, s_2 = 0$, e em ambos os processos, os valores próximos a esta região também são negativos.

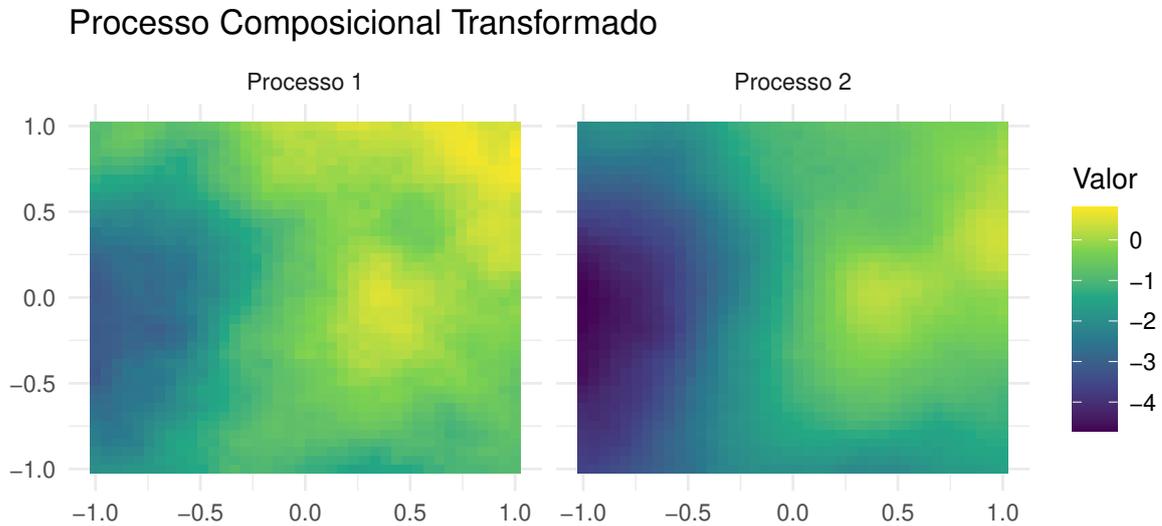


Figura 12 – Simulação de um processo Gaussiano com função de covariância Matérn bivariada. Aqui $\sigma_1^2 = 1, \sigma_2^2 = 3, a = 1, \rho = 0.9, \nu_1 = 0.5, \nu_2 = 1.5, \nu_{12} = 1.0$

Para simular um processo logístico-Gaussiano, como proposto em [Tjelmeland e Lund \(2003\)](#), pode-se utilizar a simulação ilustrada na Figura 12 acima e aplicar aos valores a alr inversa, o que fornece o processo logístico-Gaussiano no simplex. A Figura 13 mostra então como se comporta um processo logístico-Gaussiano com função de covariância Matérn bivariada com parâmetros $\sigma_1^2 = 1, \sigma_2^2 = 3, a = 1, \rho = 0.9, \nu_1 = 0.5, \nu_2 = 1.5, \nu_{12} = 1.0$, quando representado no simplex. Da figura podemos facilmente identificar a alta dependência presente nas composições e entre as composições devido à natureza restritiva dos dados composicionais.

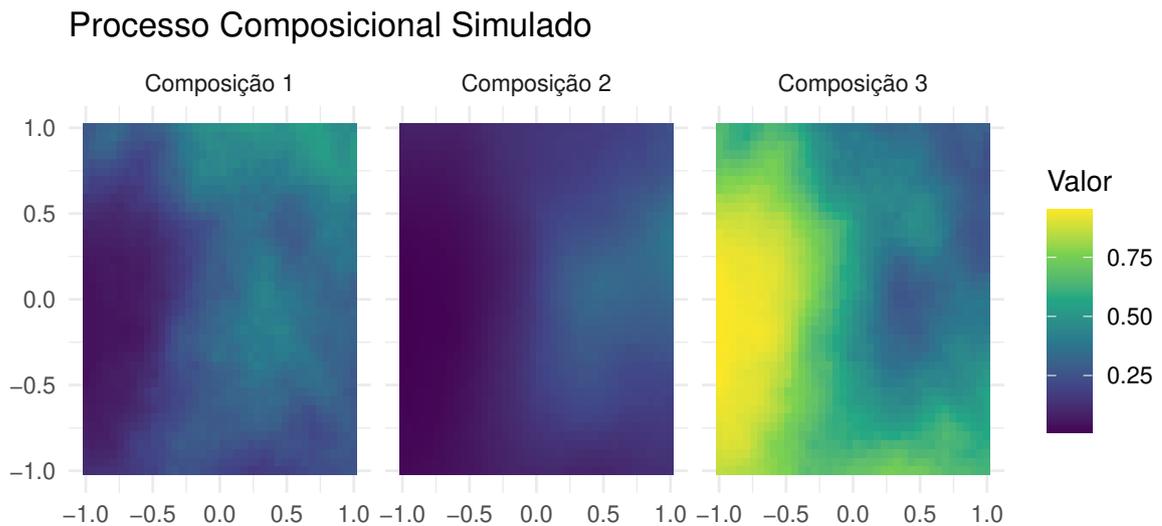


Figura 13 – Simulação de um processo Logístico-Gaussiano com função de covariância Matérn bivariada. Aqui $\sigma_1^2 = 1$, $\sigma_2^2 = 3$, $a = 1$, $\rho = 0.9$, $\nu_1 = 0.5$, $\nu_2 = 1.5$, $\nu_{12} = 1.0$.

Um processo Gaussiano nunca é observado continuamente. Em uma coleção finita de localizações espaciais, determinamos o covariograma associado ao processo Gaussiano simulado. A Figura 14 exibe tal covariograma, calculado a partir de 80 pontos aleatoriamente amostrados do processo ilustrado na figura 12, de forma que dele observa-se que o variograma direto do processo 02 sugere uma estagnação de crescimento para $\gamma(\mathbf{h}) \approx 3$ enquanto que o variograma direto do processo 01 indica $\gamma(\mathbf{h}) \approx 1.5$. Já para o alcance de dependência $1/a = 1$ é compatível com o alcance prático (*practical range*) definido por Diggle e Ribeiro (2007). Em particular, exibimos apenas as distâncias correspondentes a 1/3 da maior diagonal da região do estudo, devido à instabilidade do variograma nas caudas (PEBESMA, 2004).

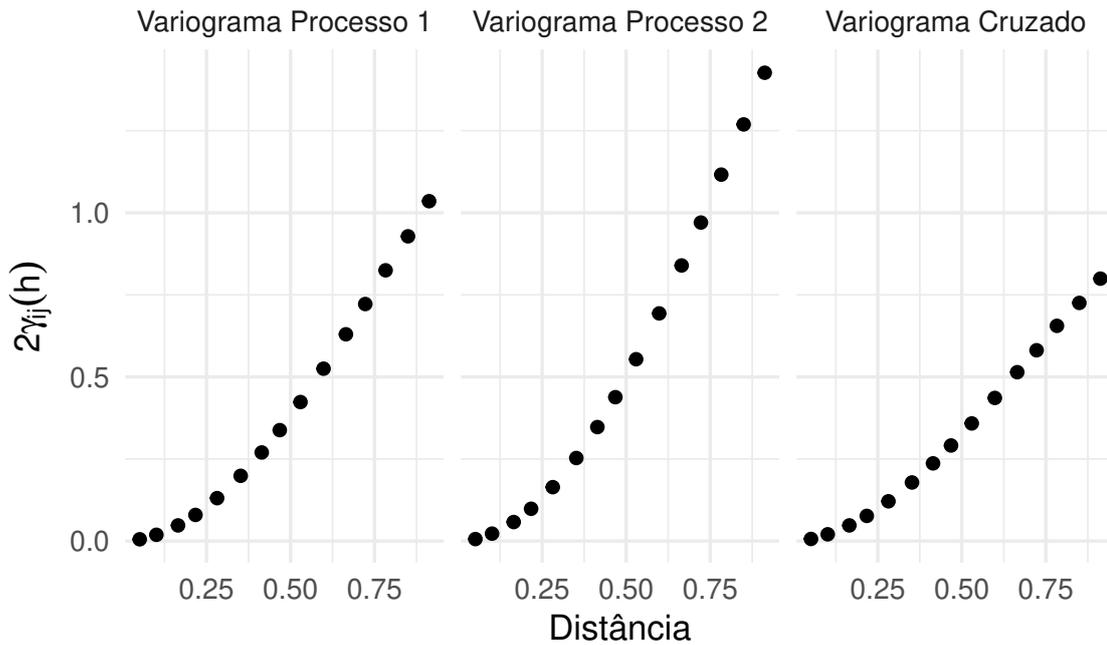


Figura 14 – Covariograma de pontos sorteados aleatoriamente que seguem um processo Logístico-Gaussiano simulado com função de covariância Matérn bivariada. Aqui $\sigma_1^2 = 1$, $\sigma_2^2 = 3$, $a = 1$, $\rho = 0.8$, $\nu_1 = 0.5$, $\nu_2 = 1.5$, $\nu_{12} = 1.0$.

3.2 Estimação da Função de Covariância Matérn bivariada Reduzida

Para estimar a função de covariância aqui será empregado o método da máxima verossimilhança, apesar de existirem outras abordagens como mínimos quadrados (GNEITING; KLEIBER; SCHLATHER, 2010) e amostragem por cadeias de Markov (TJELMELAND; LUND, 2003). Além disso não serão considerados na estimação a função média nem os parâmetros de suavidade – a média será otimizada por verossimilhança perfilada, e o parâmetro de diferenciabilidade ν será fixado por recomendação de Stein (2012), que adverte contra a estimação de derivadas de um processo que não é observado continuamente. Portanto o objetivo é obter $\boldsymbol{\theta} := (\sigma_1^2, \sigma_2^2, a, \rho)$. Neste caso ao fazer uso da densidade do processo Gaussiano (2.2) temos que a função objetivo é a log-verossimilhança

$$\ell(\boldsymbol{\theta}) \propto -\frac{1}{2} \left[\log(\det(\boldsymbol{\Sigma}_{\boldsymbol{\theta}})) + (\text{vec}(\mathbf{z}) - \text{vec}(\boldsymbol{\mu}))^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\text{vec}(\mathbf{z}) - \text{vec}(\boldsymbol{\mu})) \right],$$

onde $\mathbf{z}(\mathbf{s}) = (z_1(\mathbf{s}), z_2(\mathbf{s}))$, $\boldsymbol{\mu}(\mathbf{s}) = (\mu_1(\mathbf{s}), \mu_2(\mathbf{s}))$ e $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ é como dada em (3.2). Agora é possível usar métodos numéricos de otimização restrita para a obtenção de $\boldsymbol{\theta}$, como o L-BFGS (BYRD et al., 1995). O motivo de precisarmos fazer uso de tal algoritmo é que a função de covariância é não-linear em seus parâmetros e todos os elementos de $\boldsymbol{\theta}$ têm restrições: $\sigma_1^2 > 0$, $\sigma_2^2 > 0$, $a \geq 0$ e para ρ , segundo Gneiting, Kleiber e Schlather (2010), é

preciso que $|\rho| \leq \frac{(\nu_1\nu_2)^2}{(\nu_1 + \nu_2)/2}$ para o modelo reduzido. Se seguirmos [Mardia e Marshall \(1984\)](#) é possível inclusive aproveitar a estrutura de blocos de Σ_θ para calcular o gradiente da função mais rapidamente e de forma mais estável. Tais resultados estão presentes no Apêndice A.

3.3 Dificuldades da Otimização da Verossimilhança em Função da Covariância

Os parâmetros da função de covariância são de difícil estimação pelo método de máxima verossimilhança. A Figura 15 a seguir mostra as curvas de níveis da log-verossimilhança para dois parâmetros, quando os demais estão fixos, junto com os verdadeiros valores dos parâmetros assinalados pelo asterisco vermelho. Dela pode-se perceber que no gráfico em que temos σ_1^2 contra ρ há um leve deslocamento da verossimilhança na abscissa, de aproximadamente 0.5 unidades na escala de σ_1^2 , o que leva a um viés de estimação de σ_1^2 e σ_2^2 .

As curvas de nível de verossimilhança, sob suposições de regularidade, terão assintoticamente comportamento de uma densidade Gaussiana e as flechas, representadas pela direção e tamanho do gradiente, fornecem o rumo ao máximo. Neste gráfico um aspecto importante a se atentar é o comprimento das setas, que indicam o tamanho do passo a ser dado em cada iteração da otimização pelo algoritmo L-BFGS. Pode-se perceber que o passo é adaptável e muda de acordo com a proximidade entre o valor atual dos parâmetros e o centro da elipse, como pode ser visto no gráfico de a versus ρ e σ_1^2 versus ρ . Através de experimentos numéricos, decidimos manualmente diminuir o tamanho do passo por um fator de 10^{-6} . Isto foi feito porque para valores iniciais afastados do máximo, o gradiente é muito grande e o algoritmo de otimização diverge. Entretanto, quando o valor atual do algoritmo está relativamente próximo do máximo, o comprimento do gradiente torna-se praticamente inalterável naquela região, como pode ser visto nos gráficos em que há a versus σ_1^2 e σ_2^2 versus σ_1^2 . A consequência disto é que em regiões próximas do máximo o algoritmo tem dificuldades de se aproximar mais ainda do máximo global, o que aumenta a variabilidade da estimativa em torno do parâmetro verdadeiro, como pode ser visto na Figura 16 em que há re-amostragem das realizações com coordenadas fixas. Entretanto apesar das consequências providas por caminhar tão lentamente em direção ao máximo, esta estratégia foi o fator crítico que possibilitou uma estimação crível dos parâmetros.

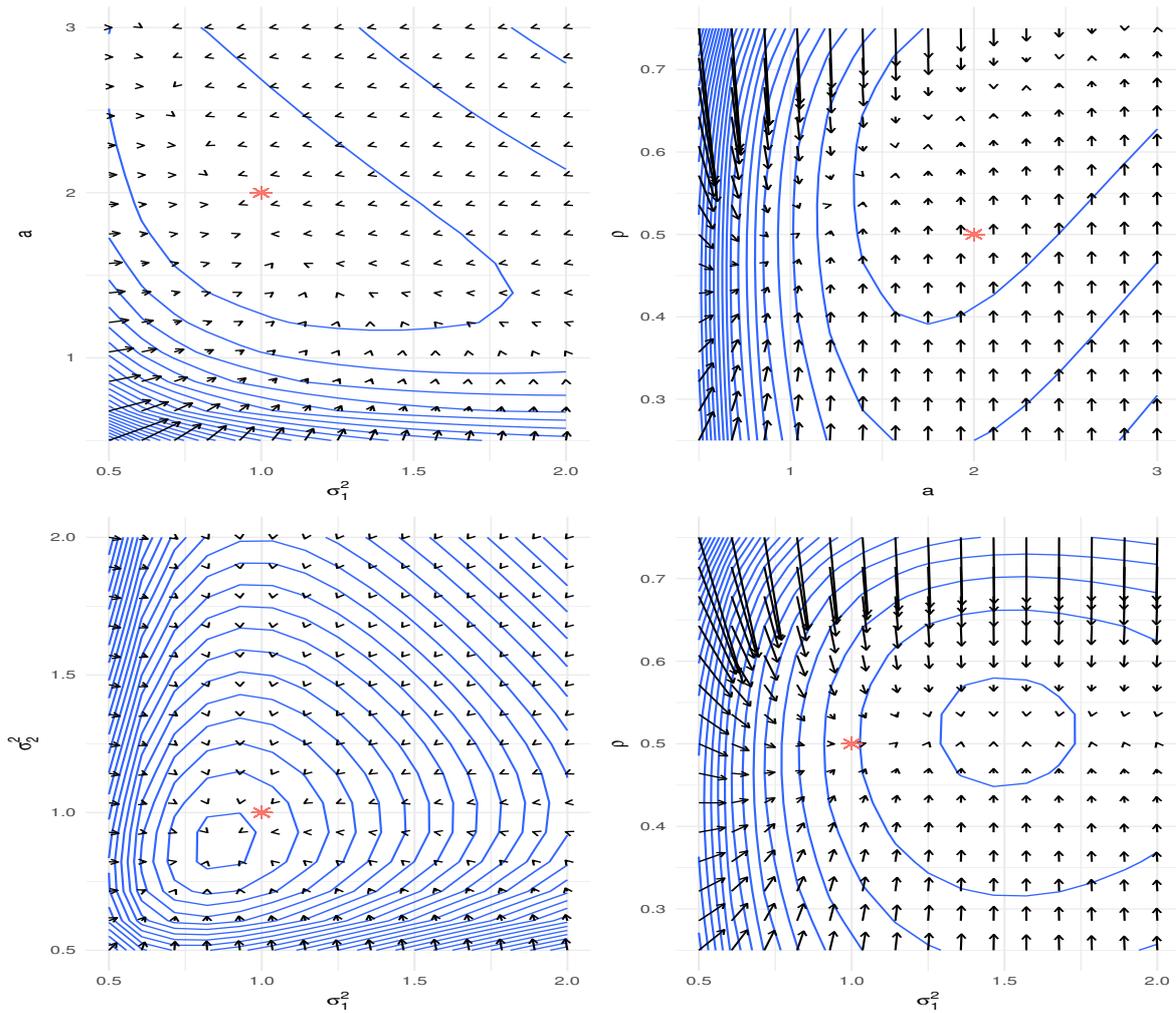


Figura 15 – Curvas de níveis para combinações dois-a-dois de parâmetros da função de covariância Matérn bivariada, onde o asterisco representa os valores verdadeiros dos parâmetros que são $\sigma_1^2 = 1, \sigma_2^2 = 1, a = 2, \rho = 0.5, \nu_1 = 0.5, \nu_2 = 0.5, \nu_{12} = 1.0$. É possível perceber que para os casos a versus σ^2, σ_1^2 versus σ_2^2 a log-verossimilhança reconhece muito bem os valores verdadeiros. Entretanto para σ_1^2 versus ρ , há um deslocamento do máximo em relação ao verdadeiro parâmetro.

O experimento de fixar coordenadas e re-amostrar as observações de uma normal multivariada nos permite obter a distribuição dos estimadores, dadas tais coordenadas. Ou seja,

1. Fixamos

os verdadeiros parâmetros;

N coordenadas;

os valores iniciais, passados ao otimizador, dos parâmetros $\sigma_1^2, \sigma_2^2, a$ e ρ ;

2. Repetimos 500 vezes

Simula-se $2N$ observações de uma normal multivariada com $\mu = (0, 0)$ e Σ dada pela Matérn bivariada;

Com o que foi fixo e simulado, estimamos os parâmetros pelo procedimento descrito na Seção 3.2;

3. Computamos o histograma das estimativas dos parâmetros σ_1^2 , σ_2^2 , a e ρ ;

A Figura 16 mostra uma realização deste experimento e dele pode-se perceber que em grande parte das vezes a estimação chega perto do valor desejado mas a distribuição das estimativas não está centrada no valor verdadeiro, o que indica um pequeno viés de estimação. Vamos examinar o efeito deste pequeno viés na interpolação espacial, na Seção 3.4.

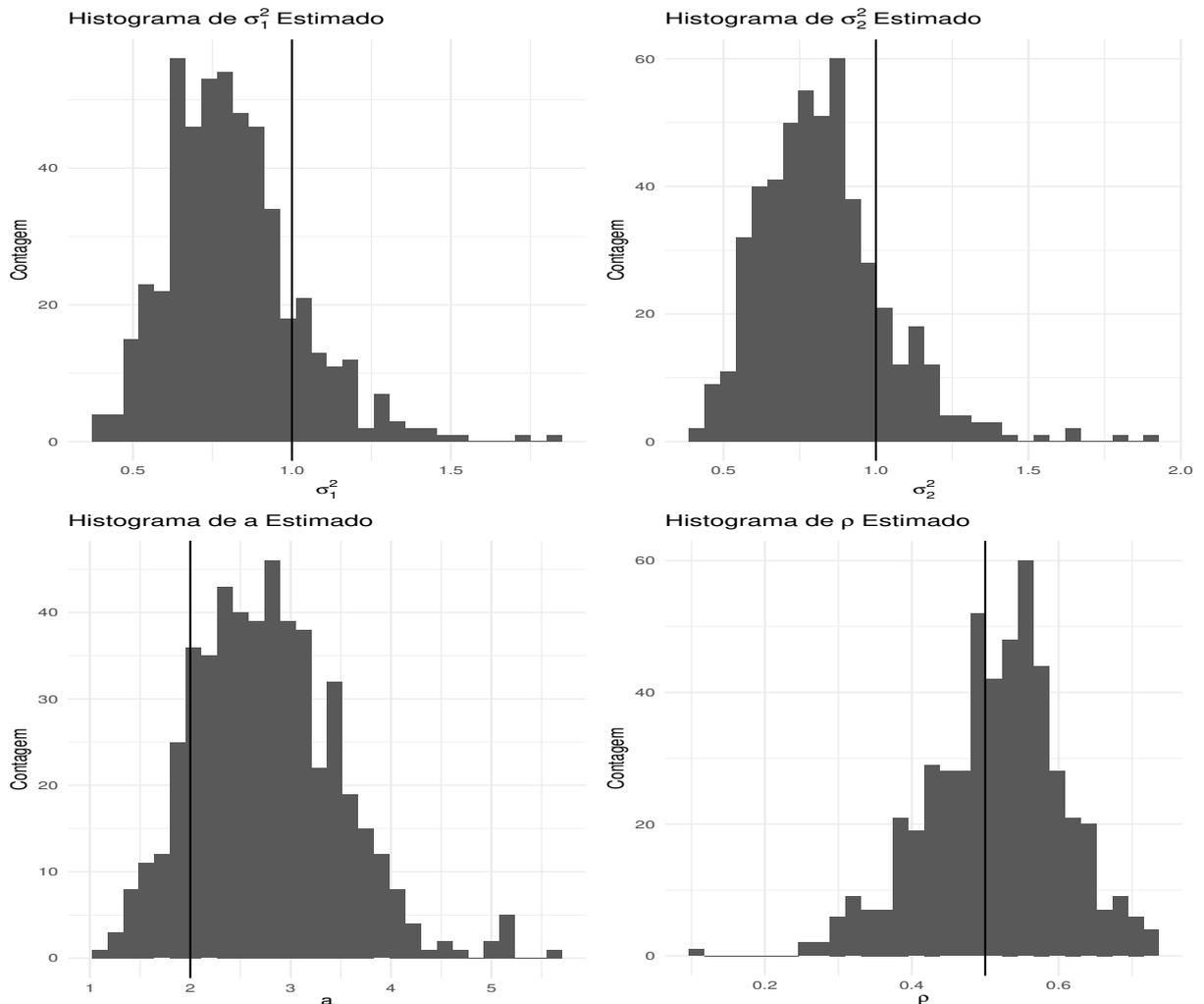


Figura 16 – Histogramas dos parâmetros estimados da função de covariância Matérn bivariada, onde as barras verticais representam os valores verdadeiros dos parâmetros, que são $\sigma_1^2 = 1, \sigma_2^2 = 1, a = 2, \rho = 0.5, \nu_1 = 0.5, \nu_2 = 0.5, \nu_{12} = 1.0$. Nota-se que

3.4 Predição Composicional e Transformações

Se voltarmos ao exemplo relativo às Figuras 13 e 12, é possível amostrar alguns pontos do processo simulado, estimar parâmetros da função de covariância Matérn bivariada e em seguida realizar Krigagem para obter mapa de previsão. A Figura 17 a seguir ilustra o caso em que foram amostrados 80 pontos do processo (com parâmetros $\sigma_1^2 = 1, \sigma_2^2 = 1, a = 2, \rho = 0.5$) para realizar a estimação e predição composicional. O gráfico de cima exibe o processo verdadeiro, enquanto que o de baixo representa o resultado da Krigagem simples, com parâmetros estimados $\sigma_1^2 = 0.4228, \sigma_2^2 = 0.6736, a = 2.4587, \rho = 0.2890$. Nota-se que a Krigagem simples permite recuperar boa parte da estrutura verdadeira, mas os detalhes mais sutis são perdidos devido à suavização, assim como no caso univariado ilustrado

pela Figura 9. Para chegar em predição a partir dos pontos amostrados, o procedimento realizado foi:

1. Utilizar a função alr para transportar os dados do simplex ao \mathbb{R}^2 ;
2. Estimar os parâmetros $\sigma_1^2, \sigma_2^2, a, \rho$ da função de covariância Matérn bivariada;
3. Com os parâmetros estimados, realizar a Krigagem na malha desejada;
4. Re-transportar os dados do \mathbb{R}^2 para o simplex por meio da transformação alr -inversa.

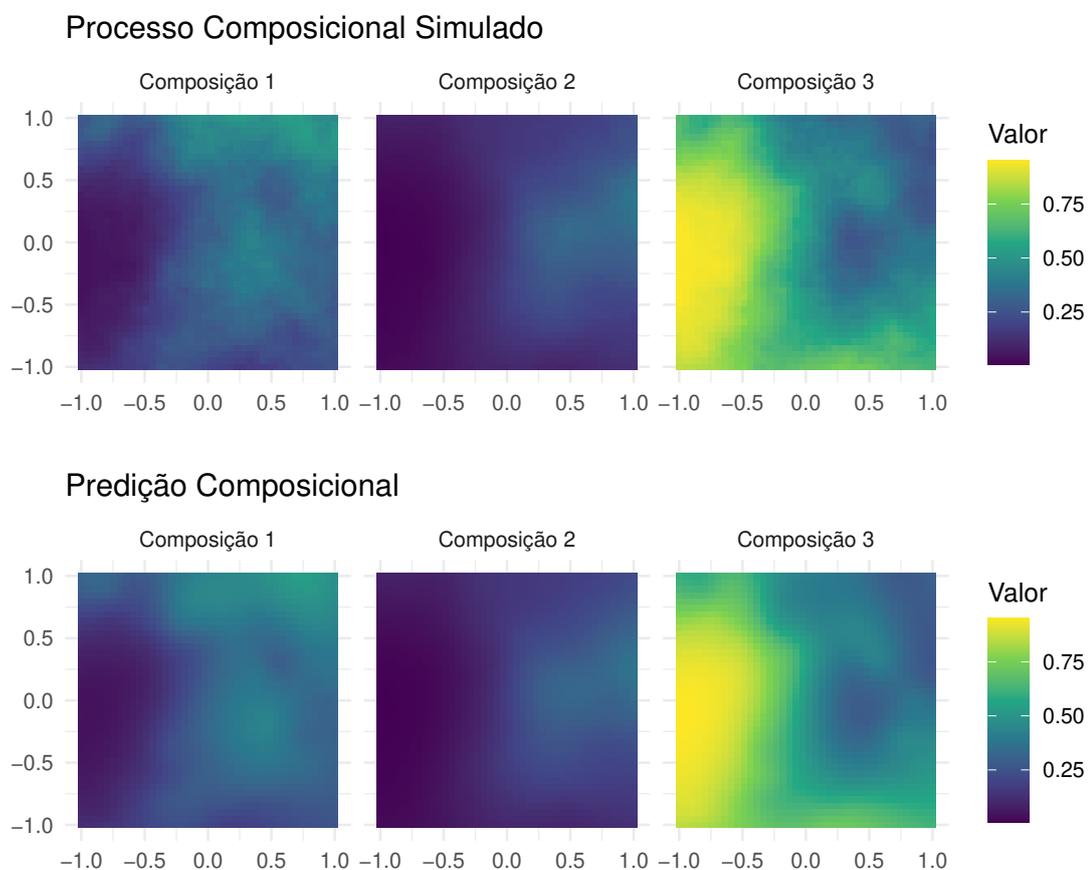


Figura 17 – Processo simulado e predição composicional. Esta é a continuação do exemplo simulado na Figura 13. O gráfico de cima é o processo simulado, enquanto que o de baixo é a Krigagem simples feita com função de covariância Matérn bivariada de parâmetros estimados $\sigma_1^2 = 0.4228, \sigma_2^2 = 0.6736, a = 2.4587, \rho = 0.2890$. Delas é possível perceber que as partes estruturais do processo foram capturadas e que as partes de variação local foram suavizadas.

Mapas de Krigagem usando funções de covariância com efeito de *nugget* causam um efeito de encolhimento (ou suavização) das composições (WAHBA, 1990; CRESSIE, 1992; HASTIE; TIBSHIRANI; FRIEDMAN, 2009). As proporções serão encolhidas para a

proporção média ($areia \approx 0.33, argila \approx 0.33, silte \approx 0.33$), e dificultar a estimação de composições que tenham uma categoria muito prevalente, como por exemplo ($areia = 0.05, argila = 0.90, silte = 0.05$).

Diretamente aplicar a transformação alr inversa aos elementos estimados da diagonal principal da função de covariância condicional das estimativas de Krigagem não é o suficiente para estimar o erro padrão da predição composicional. O motivo é que em localizações longe de dados observados, onde a incerteza é grande devido à postulação da função de covariância, duas composições serão de grande incerteza e uma, necessariamente, será de baixa incerteza devido à restrição simplicial de que a soma composições é unitária. Então para obter um erro padrão de predição crível pela aplicação da transformação alr, considere a variância do procedimento de Krigagem, dada por

$$\begin{aligned}\tilde{\boldsymbol{\mu}} &= \bar{\boldsymbol{\mu}} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{z}(\mathbf{s}) - \boldsymbol{\mu}^T \mathbf{1}) \\ \tilde{\boldsymbol{\Sigma}} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.\end{aligned}$$

Assuma primeiramente que $N = 1$, e $\tilde{\boldsymbol{\Sigma}}$ tem dimensão 2×2 . Se temos a função de alr inversa $\text{alrInv} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ definida por

$$\text{alrInv}(\mathbf{z}) = \text{alrInv}(z_1, z_2) = \left(\frac{e^{z_1}}{1 + e^{z_1} + e^{z_2}}, \frac{e^{z_2}}{1 + e^{z_1} + e^{z_2}}, \frac{1}{1 + e^{z_1} + e^{z_2}} \right),$$

então através da aproximação de Taylor de ordem 1

$$\begin{aligned}\text{Var}(\text{alrInv}(\mathbf{Z})) &\approx \text{Var}(\text{alrInv}(\tilde{\boldsymbol{\mu}}) + \nabla \text{alrInv}(\tilde{\boldsymbol{\mu}})^T (\mathbf{Z} - \tilde{\boldsymbol{\mu}})) \\ &= \text{Var}(\text{alrInv}(\tilde{\boldsymbol{\mu}}) + \nabla \text{alrInv}(\tilde{\boldsymbol{\mu}})^T \mathbf{Z} - \nabla \text{alrInv}(\tilde{\boldsymbol{\mu}})^T \tilde{\boldsymbol{\mu}}) \\ &= \text{Var}(\nabla \text{alrInv}(\tilde{\boldsymbol{\mu}})^T \mathbf{Z}) \\ &= \nabla \text{alrInv}(\tilde{\boldsymbol{\mu}})^T \tilde{\boldsymbol{\Sigma}} \nabla \text{alrInv}(\tilde{\boldsymbol{\mu}}).\end{aligned}$$

Para determinar $\nabla \text{alrInv}(\mathbf{z})$, observe que

$$\nabla f(z_1, z_2) = \begin{pmatrix} \frac{e^{z_1}(1 + e^{z_2})}{(1 + e^{z_1} + e^{z_2})^2} & -\frac{e^{z_1}e^{z_2}}{(1 + e^{z_1} + e^{z_2})^2} & -\frac{e^{z_1}}{(1 + e^{z_1} + e^{z_2})^2} \\ \frac{e^{z_1}e^{z_2}}{(1 + e^{z_1} + e^{z_2})^2} & \frac{e^{z_2}(1 + e^{z_1})}{(1 + e^{z_1} + e^{z_2})^2} & -\frac{e^{z_2}}{(1 + e^{z_1} + e^{z_2})^2} \\ -\frac{e^{z_1}}{(1 + e^{z_1} + e^{z_2})^2} & -\frac{e^{z_2}}{(1 + e^{z_1} + e^{z_2})^2} & \frac{1}{(1 + e^{z_1} + e^{z_2})^2} \end{pmatrix},$$

e conseqüentemente

$$\text{Var}(\text{alrInv}(\mathbf{Z})) \approx \begin{pmatrix} e^{\tilde{\mu}_1}(1 + e^{\tilde{\mu}_2}) & -e^{\tilde{\mu}_1}e^{\tilde{\mu}_2} \\ -e^{\tilde{\mu}_1}e^{\tilde{\mu}_2} & (1 + e^{\tilde{\mu}_1})e^{\tilde{\mu}_2} \\ -e^{\tilde{\mu}_1} & -e^{\tilde{\mu}_2} \end{pmatrix} \frac{\tilde{\boldsymbol{\Sigma}}}{(1 + e^{\tilde{\mu}_1} + e^{\tilde{\mu}_2})^4} \begin{pmatrix} e^{\tilde{\mu}_1}(1 + e^{\tilde{\mu}_2}) & -e^{\tilde{\mu}_1}e^{\tilde{\mu}_2} & -e^{\tilde{\mu}_1} \\ -e^{\tilde{\mu}_1}e^{\tilde{\mu}_2} & (1 + e^{\tilde{\mu}_1})e^{\tilde{\mu}_2} & -e^{\tilde{\mu}_2} \end{pmatrix}.$$

Para finalizar o exemplo então, vejamos como fica o erro de predição aproximado pela Figura 18. Dela pode-se perceber que nos locais amostrados (pontos em vermelho)

a variância é muito menor do que em locais não observados, o que é uma consideração importante pois nestes locais o erro de predição deve ser mínimo (como consequência da predição no local ser o próprio ponto). Nota-se também o efeito da restrição $\hat{p}_1 + \hat{p}_2 + \hat{p}_3 = 1$, sendo mais bem ilustrado na janela com abscisa no intervalo $(0, 0.5)$ e ordenada no intervalo $(-0.6, 0)$, que mostra que o efeito da dependência é carregado para o erro de predição também.

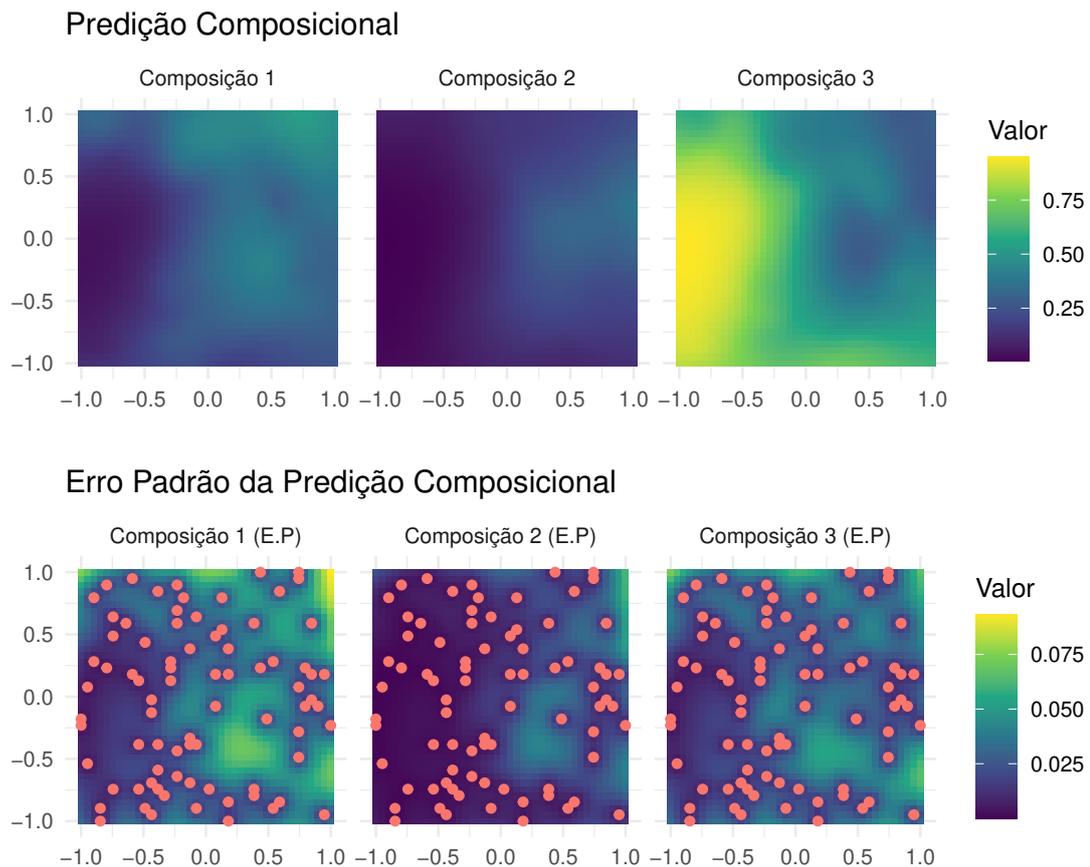


Figura 18 – Predição composicional e erro padrão da predição. Esta é a continuação do exemplo simulado na Figura 13 com parâmetros $\sigma_1^2 = 1, \sigma_2^2 = 3, a = 1, \rho = 0.9, \nu_1 = 0.5, \nu_2 = 1.5, \nu_{12} = 1.0$. Ao comparar os dois Gráficos é possível perceber que o erro padrão é menor (de fato é zero) nos locais amostrados e que aumenta a medida que nos afastamos destes pontos.

4 Aplicação

A metodologia desenvolvida aqui tem o propósito de criar predições para as proporções de areia, argila e silte em localizações ainda não observadas entre o leste de Minas Gerais e o norte do Estado de São Paulo. Os dados coletados estão disponíveis no Repositório Brasileiro Livre para Dados Abertos do Solo ([SAMUEL-ROSA et al., 2020](#)) e corresponde ao item “CTB0562”. Dele foram extraídas as composições de areia, argila e silte da tabela “camada” e suas respectivas localizações, da tabela “observação”. Em seguida fez-se um cruzamento das informações para obter os dados aqui utilizados. Um detalhe a se atentar nesta aplicação é que em alguns locais foram coletadas amostras em várias profundidades e, como forma de simplificar a interpretação dos resultados, foram escolhidas as composições mais próximas da superfície. No mapa da Figura 19 temos a região que foi considerada, em que os pontos coloridos representam as coordenadas das amostras e suas respectivas classificações, criadas de acordo com o Sistema Brasileiro de Classificação de Solos (SiBCS) ([SOLOS, 2013](#))¹.

¹ O anexo E contém o diagrama ternário que cria as classificações

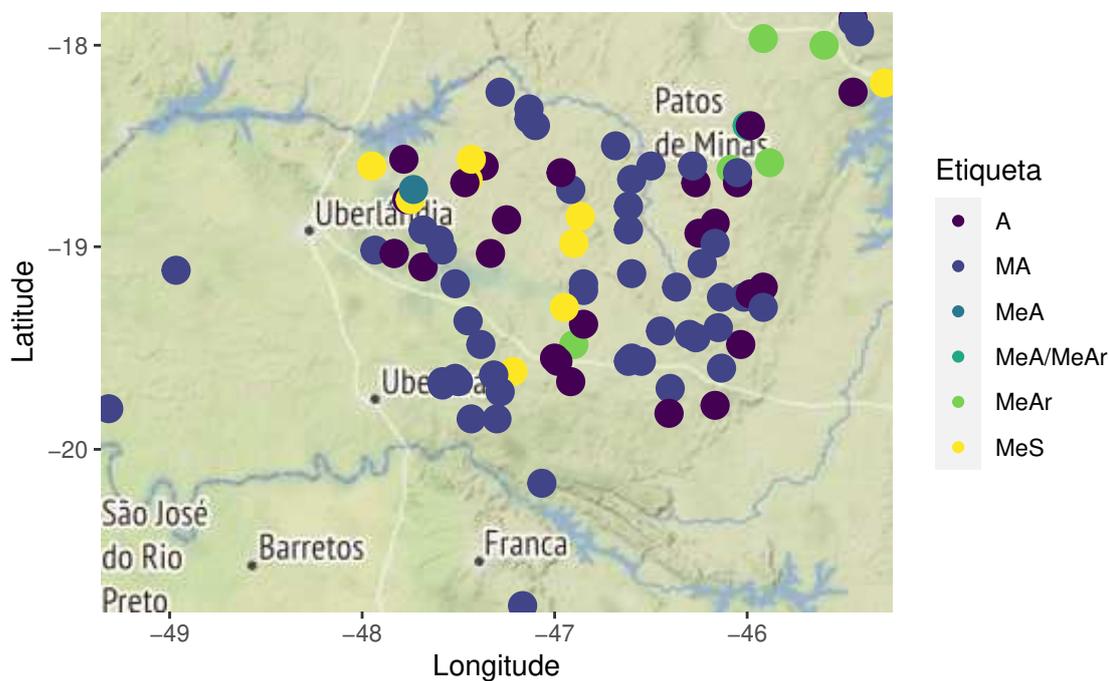


Figura 19 – Mapa com as classificações associadas às composições de areia, argila e silte em uma região de Minas Gerais. Nota-se que nessa amostra há uma grande concentração de argila, de acordo com as classificações fornecidas pela Tabela 5.

O diagrama ternário da Figura 20 a seguir ilustra como se distribuem as concentrações das composições amostradas que criaram as classificações do mapa na Figura 19. A característica que salta aos olhos é que nesta amostra há várias composições com grande concentração de argila, algumas com bastante areia mas somente uma com mais metade composta por silte e cinco mais de 40% de silte. Também é interessante notar que as composições de argila e areia são mais dispersas, ao passo que silte varia entre próximo de zero e 40%, o que leva a intuir que o mapa da predição terá mais contrastes (e consequentemente, incerteza) para as composições de argila e areia, ao passo que o mapa de silte provavelmente será mais uniforme com baixas proporções.

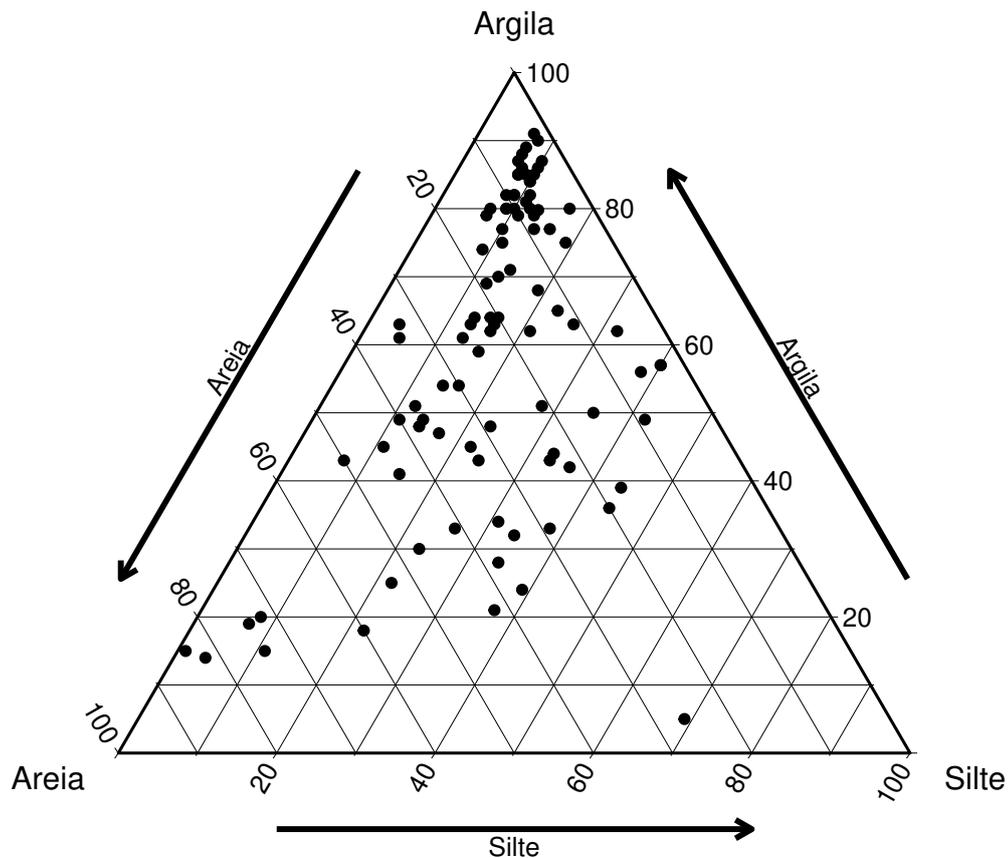


Figura 20 – Diagrama ternário das composições de areia, argila e silte em uma região de Minas Gerais. Nota-se que na maioria das amostras a composição dominante é de argila. Também há casos de amostras com bastante areia, mas somente cinco amostras em que silte é responsável ao menos 40% do todo. Isto traduz-se ao fato de que as composições de argila e areia são mais dispersas, ao passo que silte varia entre próximo de zero e 40%.

Tal mapa de predição entretanto é criado em algumas etapas, como resumido no fim da Seção 3.3, sendo a primeira delas a mudança do domínio simplex para \mathbb{R}^2 com a transformação aditiva de log-razão (alr). O covariograma empírico do processo bivariado obtido pela transformação alr é ilustrado na Figura 21 e dele é possível observar que: (1) o variograma cruzado gira em torno do 0, que sugere os processos individuais como independentes, enquanto que: (2) ambos variogramas diretos apresentam efeito de pepita em torno de 0.75 e ambos σ_1^2 σ_2^2 em torno de 1, o que indica que há muita variação de micro-escala e deve-se provavelmente ao fato de que a comparação cartográfica é entre cidades. Por fim pode-se perceber que: (3) o alcance prático de dependência está em torno de 0.4 para o processo composicional 02 $\log(\text{argila}/\text{silte})$, enquanto que para o processo 01 parece ser 0.8.

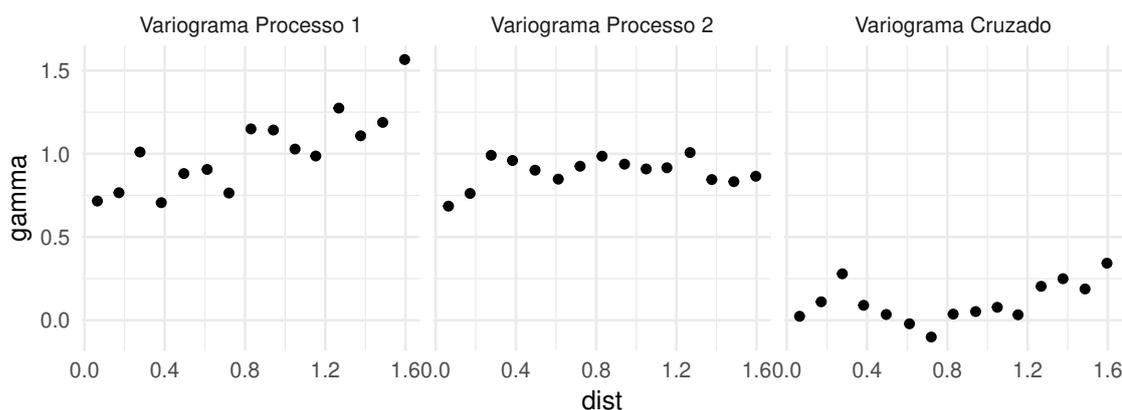


Figura 21 – Covariograma dos dados de solo CTB0562. Pode-se notar que o variograma cruzado gira em torno do 0, que sugere os processos individuais como independentes. Também nota-se que os ambos variogramas diretos apresentam efeito de pepita em torno de 0.75 e ambos σ_1^2 σ_2^2 em torno de 1, o que indica que há muita variação de micro-escala. Por fim pode-se perceber que o alcance prático de dependência está em torno de 0.4 para o segundo processo composicional $\log(\text{argila/silte})$ e para o primeiro parece ser 0.8.

Para gerar as predições, supõe-se que o processo gerador dos dados é um processo Gaussiano com função média constante e função de covariância Matérn bivariada, estima-se os parâmetros da função média e de covariância Matérn bivariada, para gerar as predições por meio da Krigagem e, por fim, basta trazê-las de volta ao Simplex. Para a estimação, foram usados como estimativas iniciais valores baseados na Figura 21 do variograma cruzado empírico, nomeadamente: $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $a = 0.5$ e $\rho = 0$. A Figura 22 a seguir mostra a Krigagem feita a partir dos parâmetros estimados, que foram $\sigma_1^2 = 2.200$, $\sigma_2^2 = 2.228$, $a = 1.960$ e $\rho = 0.685$. Dessa figura nota-se primeiramente que a predição nas redondezas dos pontos amostrados vai ao encontro das proporções observadas, o que é esperado. Em segundo lugar é interessante perceber que a “sensibilidade” da predição é boa e um exemplo disto está no mapa da predição de silte, pois no centro dele existem alguns pontos de alta concentração de silte rodeados por outros de baixa concentração e a predição consegue captar bem a relação de dependência entre eles, pois existem “ilhas” de maior concentração de silte que indicam que o valor estimado para alcance de dependência $1/a = 0.510$ (cerca de 56.6km)² é razoável.

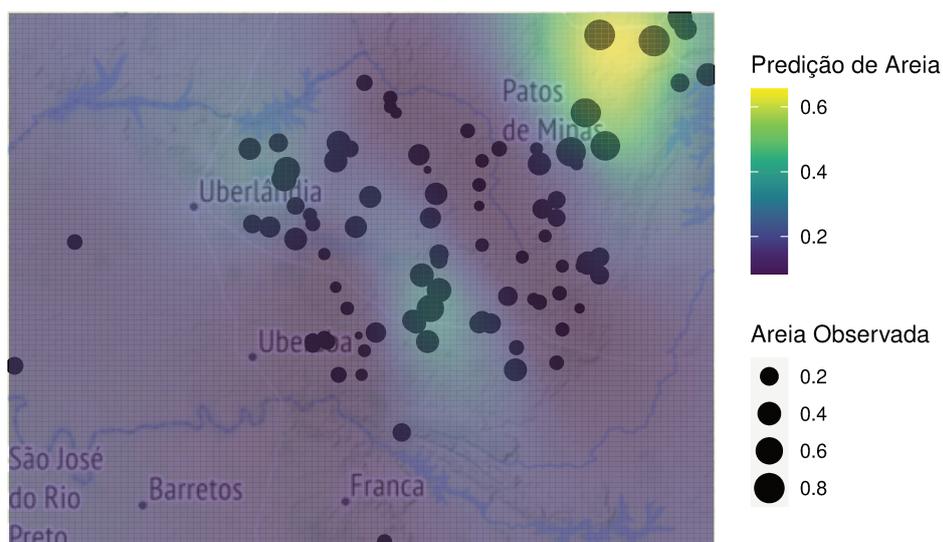
A Figura 23 ilustra os mapas dos erros padrão da Krigagem para cada composição. O primeiro aspecto interessante aqui é que a correlação simplicial entre as composições é carregada para o desvio padrão, mas não há imposição da restrição $E.P(\hat{p}_1) + E.P(\hat{p}_2) + E.P(\hat{p}_3) = 1$ pois caso houvesse ocorreria necessariamente de, em regiões com ausência de observações (como nos cantos inferiores e superiores esquerdos dos mapas), uma composição apresentar erro padrão pequeno já que o erro padrão de

² Já que 1 grau de latitude equivale a 111,045 quilômetros.

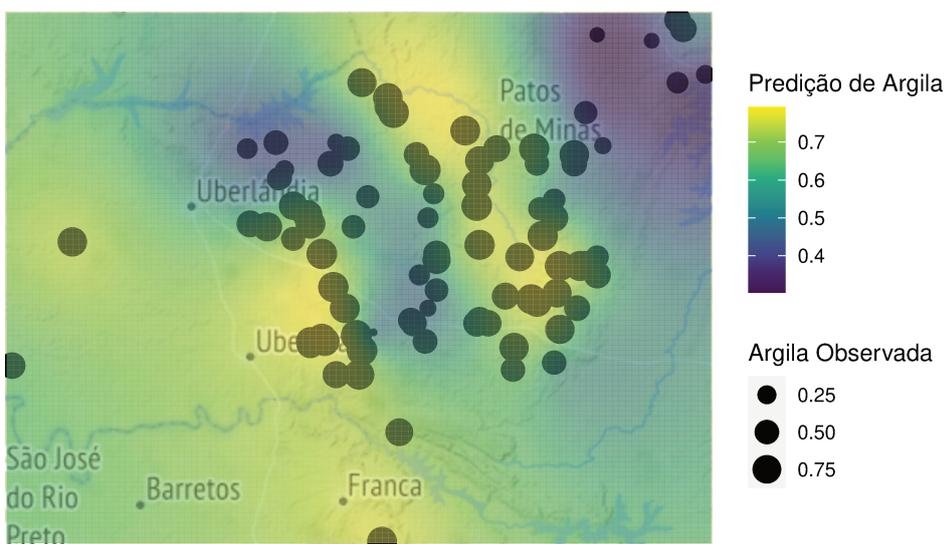
uma das composições provavelmente será alto devido à falta de amostras na região. Isto levaria à conclusão de que em tais locais, teríamos alta precisão das estimativas, o que dificilmente seria verdade pois não há pontos amostrados perto da região. Nota-se também que para a composição de areia, existe um trecho no canto superior direito que representa a maior incerteza a cerca da média, o que é curioso já que nesta região existem algumas amostras (de alta concentração de areia, inclusive), mas ainda não há explicação para essa situação. Ademais os mapas dos erros padrão portam-se como o de uma Krigagem univariada, que é o comportamento esperado.

Por fim a Figura 24 é um complemento do mapa de predição composicional e exhibe qual é a classificação do solo predita, de acordo com SiBCS (o Sistema Brasileiro de Classificação de Solos), e que permite em um só gráfico visualizar as classificações predominantes. Nota-se primeiramente que na predição que existem 4 das 8 classificações existentes sendo elas Argilosa (A), Muito Argilosa (MA), Média Argilosa (MeA) e Média Siltosa (MeS). Impressiona também o fato de não termos classificações Arenosas e este é um ponto delicado pois considere o seguinte vetor, que fornece a menor quantidade de argila predita na malha em que foi feita a Krigagem: ($areia = 0.656$, $argila = 0.302$, $silte = 0.041$). Segundo o diagrama ternário da Figura 24, para termos classificações arenosas é necessário uma baixa concentração de Argila (menor do que 20%) e não existe no mapa predito casos com menos do que 30%, o que não leva à classificações arenosas, mesmo que a proporção de areia seja a mais presente. Entretanto, esta classificação está sujeita a erros pois é condicionada ao erro padrão da predição (que para o exemplo ilustrado é erro padrão da areia = 0.315, erro padrão da argila = 0.235, erro padrão do silte = 0.049) e, conseqüentemente, ao intervalo de confiança associado ao valor nominal do erro do tipo I que se deseja controlar. Ou seja, a classificação do exemplo é Média Argilosa mas existe um intervalo de confiança que cobre a média que cujos valores cobertos permitem classificar o solo como Médio Arenoso.

Predição Composicional de Areia



Predição Composicional de Argila



Predição Composicional de Silte

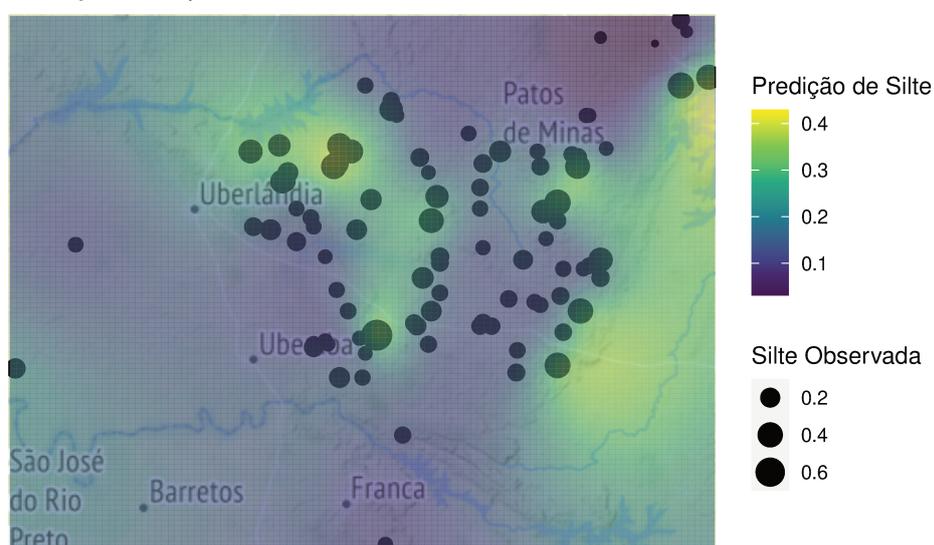
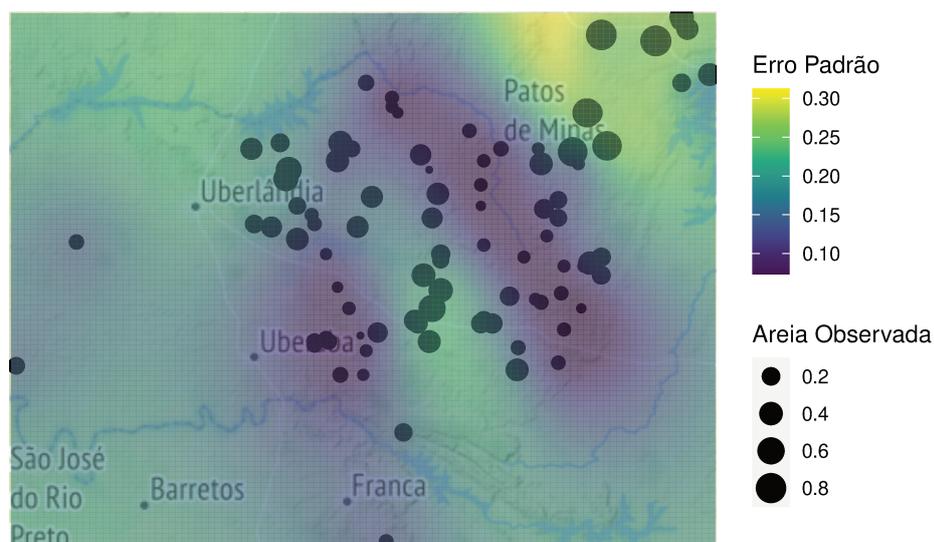
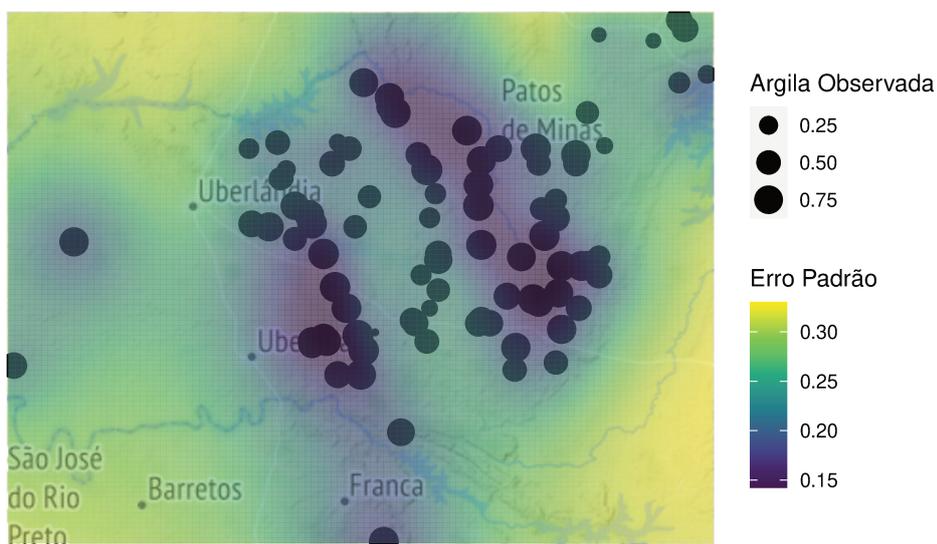


Figura 22 – Krigagem trivariada de composição do solo, feita com função de covariância Matérn Bivariada de parâmetros estimados iguais a $\sigma_1^2 = 2.200$, $\sigma_2^2 = 2.228$, $a = 1.960$ e $\rho = 0.685$.

Desvio Padrão da Predição – Areia



Desvio Padrão da Predição – Argila



Desvio Padrão da Predição – Silte

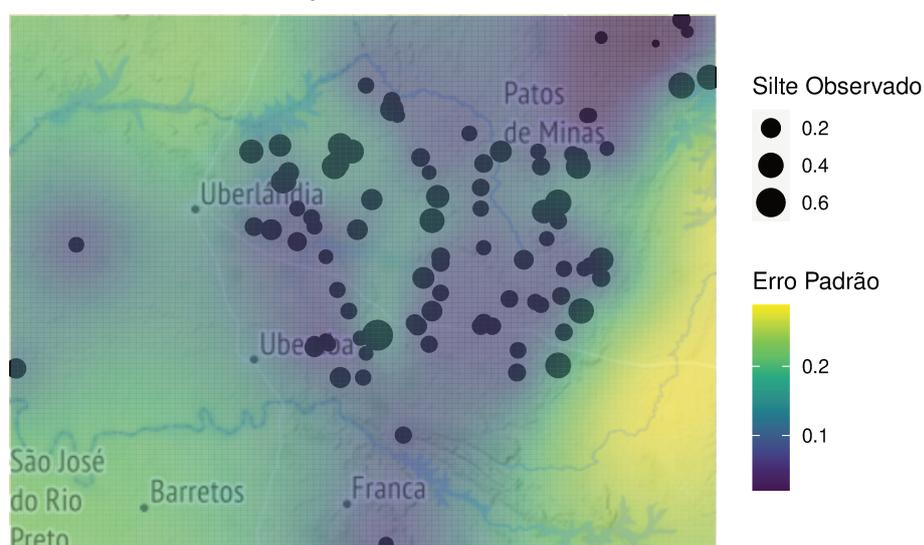


Figura 23 – Erro de predição da Krigagem composicional do solo, feita com função de covariância Matérn Bivariada de parâmetros estimados iguais a $\sigma_1^2 = 2.200$, $\sigma_2^2 = 2.228$, $a = 1.960$ e $\rho = 0.685$

Predição Composicional de Classificação de Solo

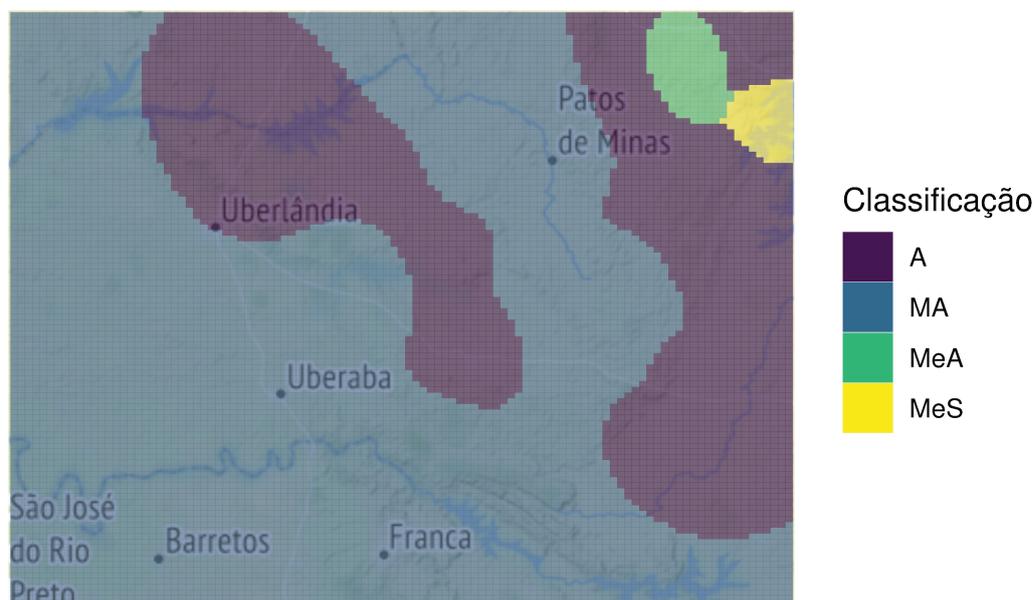


Figura 24 – Classificação predita do solo, obtida das categorias criadas pela SIBCS 5 a partir de Krigagem feita com função de covariância Matérn Bivariada de parâmetros estimados iguais a $\sigma_1^2 = 2.200$, $\sigma_2^2 = 2.228$, $a = 1.960$ e $\rho = 0.685$

A Figura 25 ilustra o mapa da região de Uberlândia em Minas Gerais, em que nele há o conjunto utilizado para estimar dos dados (CTB0562) e fornecer os mapas de predição anteriores (representadas pelos pontos pretos), junto com novas amostras obtidas através do item “CTB0809”³, que serão usados com o objetivo de quantificar a qualidade preditiva do modelo (representados pelos números). Note que novos pontos estão na região mais bem povoada do mapa, o que sugere que a predição composição destes pontos deve ter boa qualidade e baixo erro, exceto pelos pontos 2, 3 e 8 que estão em uma região complicada.

³ Somente aquelas que estão na região delimitada pelos dados CTB0562

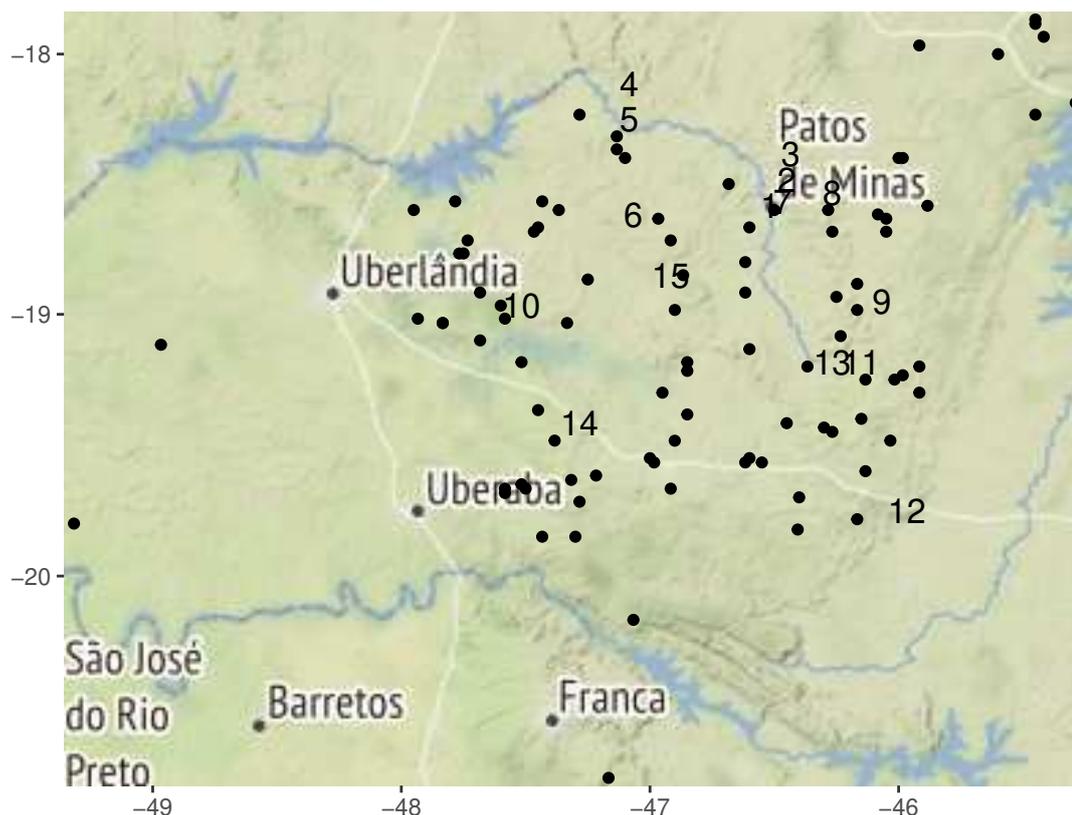


Figura 25 – Mapa da região de Uberlândia em Minas Gerais. Amostras utilizadas para estimar o modelo estão representadas pelos pontos pretos, enquanto que as novas amostras (utilizadas para quantificar a qualidade preditiva do modelo) são os números.

As Tabelas 6, 7 e 8 na Seção F do Apêndice mostram as quantidades observadas de areia, argila e silte no conjunto de dados CTB0809 e qual é a predição do modelo para tais composições, junto com seus respectivos erros padrão e identificadores (coluna chamada Id para identificação no mapa). As Figuras 26 e 27 mostram os diagramas ternários para as quantidades observadas e preditas das composições em CTB0809. Nota-se claramente que o Krigagem aglomerou as composições, sendo os pontos 3, 6 e 4 aqueles mais notáveis, assim como a composição de número 12, que teve (curiosamente) um efeito oposto. Tal aglomeração é consequência natural da Krigagem, que funciona por regressir à média, como um suavizador, (da mesma forma que uma regressão linear simples) e gerar mais

categorias como média-arenosa/siltosa/argilosa. A Tabela 1 a seguir exibe o comparativo entre a classificação do solo das composições observadas em CTB0809 e suas previsões pela Krigagem, enquanto que a Tabela 2 resume a Tabela 1. Notam-se cinco divergências entre as classificações, destacadas pelo **negrito**, sendo que a dos 5 erros ocorrem por classificar um terreno arenoso como médio arenoso, efeito da suavização, além do curioso décimo segundo ponto, que sofreu o efeito contrário.

Composições em CTB0809

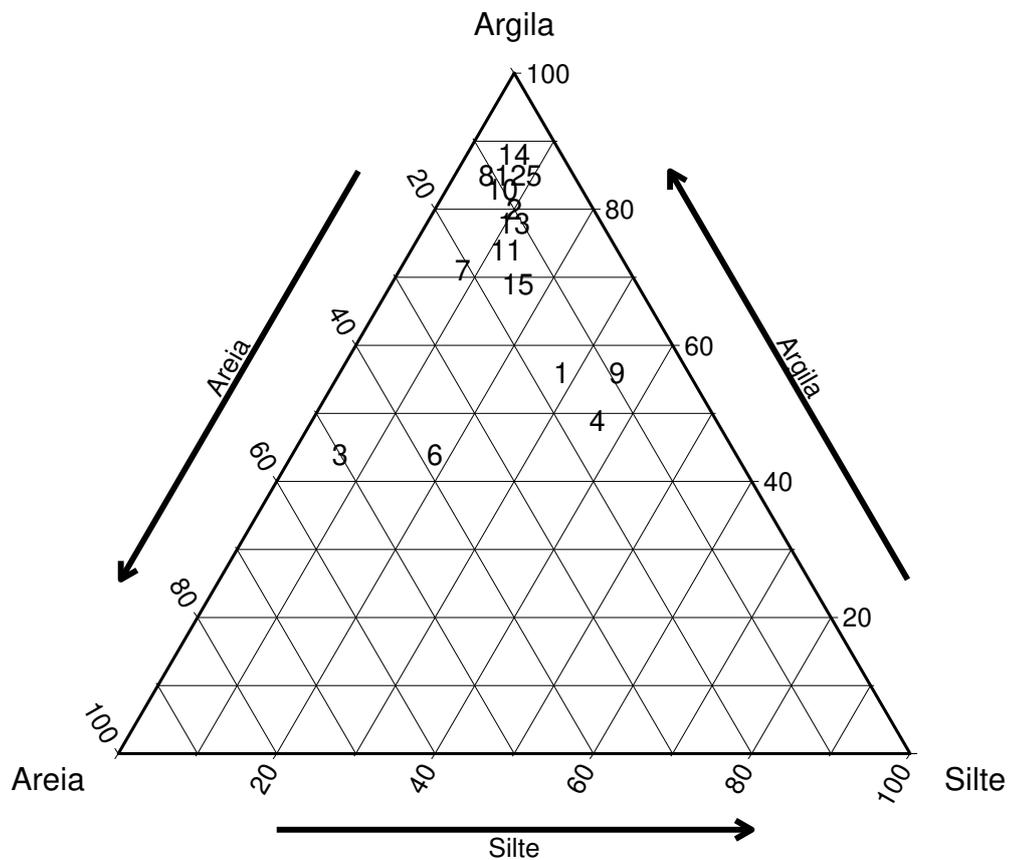


Figura 26 – Diagrama ternário dos dados na região de Uberlândia (MG) que foram usados para ajustar o modelo.

Predição das Composições em CTB0809

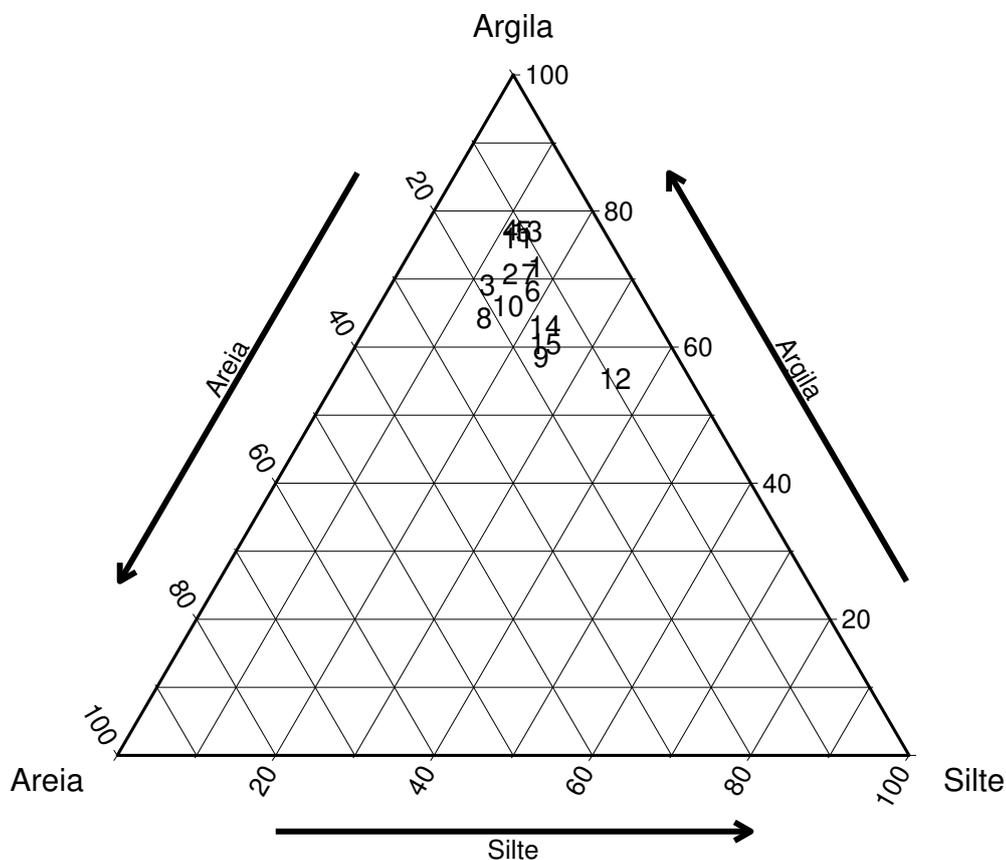


Figura 27 – Diagrama ternário dos dados na região de Uberlândia (MG) que foram usados para determinar o poder preditivo.

Para comparar o desempenho da Krigagem com outras técnicas de modelagem populares, estimou-se também o desempenho do classificador K vizinhos mais próximos (HASTIE; TIBSHIRANI; FRIEDMAN, 2009) para classificações de solo descritas no apêndice E e encontradas em CTB0562. O procedimento aqui utilizado para estimar o número de K foi usar a validação cruzada (HASTIE; TIBSHIRANI; FRIEDMAN, 2009) “deixe um de fora” (*leave one out*) no conjunto de dados CTB0562 e avaliar seu desempenho no conjunto de dados CTB0809. O motivo de se criar um classificador sobre as etiquetas obtidas da classificação SIBCS, ao invés de um regressor sobre as composições, é devido à hipótese de que o classificador terá um desempenho melhor do que o regressor, dado que há menos variabilidade nas etiquetas do que nas proporções. A Tabela 3 a seguir mostra quais foram as acurácias obtidas a partir da validação cruzada. A maior acurácia encontrada foi de 59.1% (contra 61,29%, como descrito no Apêndice H) o que forneceu como melhor classificador encontrado aquele que com $K = 5$. A matriz de confusão (HASTIE;

Tabela 1 – Comparativo entre a classificação do solo para as composições observadas e suas predições pela Krigagem. Notam-se 5 divergências entre as classificações, destacadas pelo negrito, sendo que a maioria dos erros (4 dos 5 erros) é por classificar um terreno arenoso como médio arenoso.

Id	Classificação Observada	Classificação Predita
1	A	MA
2	MA	MA
3	A	MA
4	A	MA
5	MA	MA
6	A	MA
7	MA	MA
8	MA	MA
9	A	A
10	MA	MA
11	MA	MA
12	MA	A
13	MA	MA
14	MA	MA
15	MA	MA

Tabela 2 – Classificação obtida a partir da Krigagem contra categoria observada.

	A	MA
A	1	4
MA	1	9

TIBSHIRANI; FRIEDMAN, 2009) gerada a partir da previsão do classificador nos dados de CTB0809 e as categorias observadas está na Tabela 4. Nota-se que seu desempenho é pouco superior à Krigagem pois o algoritmo KNN só acertou uma categoria a mais que a Krigagem. Além disso, não é possível obter para o algoritmo utilizado, uma medida de erro sobre a probabilidade da classificação, o que não permite gerar mapas de incerteza associada à predição. Por fim a Figura 28 exhibe como seria o mapa de classificações de solo predito pelo método com $K = 5$, que é bem diferente do mapa 22 descrito pela Krigagem .

Tabela 3 – Taxas de acurácias nos dados CTB0809, para cada algoritmo KNN ajustado por validação cruzada “deixe um de fora” nos dados CTB0562. O melhor resultado foi obtido com $K = 5$, com acurácia de 59.1%.

K	Acurácia
1	54.80
2	52.70
3	51.60
4	55.90
5	59.10
6	54.80
7	57.00
8	52.70
9	49.50
10	51.60

Tabela 4 – Classificador criados com base em 5 vizinhos mais próximos contra classificação do solo observada. As linhas representam as categorias observadas em CTB0809, enquanto que as colunas representam as categorias previstas pelo algoritmo.

	A	MA	MeA	MeA/MeAr	MeAr	MeS
A	1	4	0	0	0	0
MA	0	10	0	0	0	0

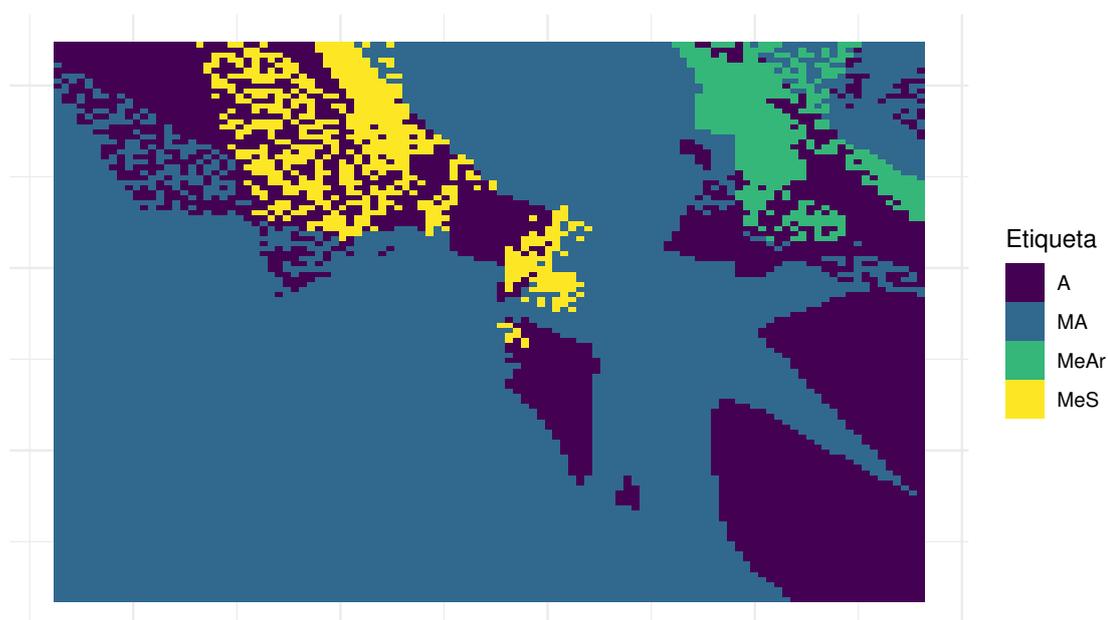


Figura 28 – Mapa da predição da classificação de solo de acordo com o critério SIBCS, realizada pelo algoritmo KNN com $K = 5$ vizinhos mais próximos.

5 Considerações Finais

A metodologia de co-krigagem com função de covariância Matérn Bivariada se mostrou promissora como forma de criar mapas de predição espacial em processos multivariados por permitir não só criar predições das composições mas por também fornecer mapas de imprecisão da estimativa composicional.

Entretanto a estimação dos parâmetros não foi simples sendo, certamente, a parte mais intrincada do projeto. O motivo é que só foi possível estimar os parâmetros quando alterou-se o tamanho do passo ao manualmente encolher as estimativas do gradiente por fatores de ordem 10^{-6} . Porém apesar de viável esta certamente não é uma forma elegante nem objetiva, o que abre vias de estudos para tornar a otimização dos parâmetros uma etapa mais conceitualmente concisa. Uma ideia para adaptativamente mudar o comprimento do gradiente é o pré-multiplicar pela inversa da Matriz de Informação de Fisher, de forma a amenizar a dificuldade de impor um fator de redução de tamanho único 10^{-6} para parâmetros de diferentes escalas. Outra opção é seguir a linha de [Martins, Bonat e Jr \(2016\)](#), que fornece como abordagem aproximar o gradiente por métodos computacionais e usar verossimilhança perfilada para estimar os parâmetros.

Outros futuros tópicos de pesquisa são:

1. Utilizar a transformação ilr ao invés da transformação alr , pois existem indícios ([WANG et al., 2020](#)) de melhora em poder preditivo ao utilizar a primeira função;
2. Implementar krigagem ordinária ao invés de krigagem simples para criar os mapas e quantificar a qualidade das predições;
3. Implementar na função de covariância Matérn bivariada o caso geral e compará-la com o caso completo, já que a restrição de que os alcances de dependências são os mesmos pode não ser adequada.
4. Utilizar a teoria citada em [Aitchison \(1986\)](#) para criar elipses de confiança para as proporções, de forma adaptada ao caso geoestatístico.

Referências

- ABRAMOWITZ, M.; STEGUN, I. A. *Handbook of mathematical functions*. [S.l.]: NBS, 1972. Citado 2 vezes nas páginas 70 e 77.
- AITCHISON, J. *The Statistical Analysis of Compositional Data*. London: Chapman and Hall, 1986. Citado 5 vezes nas páginas 16, 18, 19, 38 e 63.
- BANERJEE, S.; GELFAND, A. E.; FINLEY, A. O.; SANG, H. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B*, v. 70, n. 4, p. 825–848, 2008. Citado na página 29.
- BARROS, G. S. de C.; CASTRO, N. R.; MACHADO, G. C.; ALMEIDA, F. M. de S.; SILVA, A. F.; FACHINELLO, A. L. *PIB do agronegócio cresce 3,81% em 2019*. 2020. Disponível em: <https://www.cepea.esalq.usp.br/upload/kceditor/files/Cepea_PIB_CNA_2019.pdf>. Citado na página 16.
- BYRD, R. H.; LU, P.; NOCEDAL, J.; ZHU, C. A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, SIAM, v. 16, n. 5, p. 1190–1208, 1995. Citado na página 42.
- CARLESSO, R.; SANTOS, R. F. Disponibilidade de água às plantas de milho em solos de diferentes texturas. *Revista Brasileira de Ciência do Solo*, v. 23, n. 1, p. 17–25, 1999. Citado na página 16.
- CHAYES, F. On correlation between variables of constant sum. *Journal of Geophysical research*, v. 65, n. 12, p. 4185–4193, 1960. Citado na página 18.
- _____. *Ratio correlation: a manual for students of petrology and geochemistry*. Chicago, IL: University of Chicago Press, 1971. Citado na página 18.
- CRESSIE, N. *Statistics for Spatial Data*. Hoboken, NJ: Wiley, 1992. Citado 5 vezes nas páginas 25, 26, 29, 33 e 47.
- CRESSIE, N.; WIKLE, C. K. *Statistics for Spatio-Temporal Data*. Hoboken, NJ: Wiley, 2015. Citado na página 31.
- DIGGLE, P.; RIBEIRO, P. J. *Model-based geostatistics*. Springer, New York, 2007. Citado na página 41.
- DIGGLE, P. J.; RIBEIRO, P. J.; CHRISTENSEN, O. F. An introduction to model-based geostatistics. In: *Spatial Statistics and Computational Methods*. [S.l.]: Springer, 2003. p. 43–86. Citado na página 33.
- EMBRAPA. *Embrapa Solos*. 2020. Disponível em: <<https://www.embrapa.br/solos/sibcs>>. Citado 2 vezes nas páginas 11 e 95.
- FIGUEIREDO, N. M. S. d.; CORRÊA, A. M. C. J. Tecnologia na agricultura brasileira: indicadores de modernização no início dos anos 2000. Instituto de Pesquisa Econômica Aplicada (Ipea), 2006. Citado na página 16.

- FREITAS, A.; POZZA, E.; ALVES, M.; COELHO, G.; ROCHA, H.; POZZA, A. Spatial distribution of yellow sigatoka leaf spot correlated with soil fertility and plant nutrition. *Precision agriculture*, Springer, v. 17, n. 1, p. 93–107, 2016. Citado na página 16.
- GAETAN, C.; GUYON, X. *Spatial statistics and modeling*. New York: Springer, 2010. v. 90. Citado 2 vezes nas páginas 23 e 29.
- GENTON, M. G.; KLEIBER, W. Cross-covariance functions for multivariate geostatistics. *Statistical Science*, JSTOR, p. 147–163, 2015. Citado 2 vezes nas páginas 16 e 36.
- GNEITING, T.; KLEIBER, W.; SCHLATHER, M. Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, Taylor & Francis, v. 105, n. 491, p. 1167–1177, 2010. Citado 2 vezes nas páginas 38 e 42.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning: data mining, inference, and prediction*. [S.l.]: Springer Science & Business Media, 2009. Citado 6 vezes nas páginas 17, 47, 60, 61, 99 e 106.
- JAMES, B. R. *Probabilidade: um curso em nível intermediário*. Rio de Janeiro, RJ: IMPA, 1996. Citado na página 23.
- LEPSCH, I. F.; SILVA, N. M. da; ESPIRONELO, A. Relação entre matéria orgânica e textura de solos sob cultivo de algodão e cana-de-açúcar, no estado de são paulo. *Bragantia*, Instituto Agrônomo, v. 41, n. 1, p. 231–236, 1982. Citado na página 16.
- MAGNUS, J. R.; NEUDECKER, H. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. [S.l.]: Wiley, 1999. Citado na página 68.
- MARDIA, K. V.; MARSHALL, R. J. Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, Oxford University Press, v. 71, n. 1, p. 135–146, 1984. Citado na página 43.
- MARTINS, A. B. T.; BONAT, W. H.; JR, P. J. R. Likelihood analysis for a class of spatial geostatistical compositional models. *Spatial Statistics*, Elsevier, v. 17, p. 121–130, 2016. Citado na página 63.
- MATHERON, G. Principles of geostatistics. *Economic Geology*, v. 58, n. 8, p. 1246–1266, 1963. Citado na página 26.
- MOEYS, J. *soiltexture: Functions for Soil Texture Plot, Classification and Transformation*. [S.l.], 2018. R package version 1.5.1. Disponível em: <<https://CRAN.R-project.org/package=soiltexture>>. Citado na página 95.
- MÖRTERS, P.; PERES, Y. *Brownian motion*. [S.l.]: Cambridge University Press, 2010. v. 30. Citado 2 vezes nas páginas 29 e 30.
- PAWLOWSKY-GLAHN, V.; EGOZCUE, J. J. Spatial analysis of compositional data: A historical review. *Journal of Geochemical Exploration*, Elsevier, v. 164, p. 28–32, 2016. Citado na página 16.
- PEBESMA, E. J. Multivariable geostatistics in S: the *gstat* package. *Computers & Geosciences*, v. 30, p. 683–691, 2004. Citado na página 41.

- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2020. Disponível em: <<https://www.R-project.org/>>. Citado 2 vezes nas páginas 39 e 91.
- Ribeiro Jr, P. J.; DIGGLE, P. J.; SCHLATHER, M.; BIVAND, R.; RIPLEY, B. *geoR: Analysis of Geostatistical Data*. [S.l.], 2020. R package version 1.8-1. Disponível em: <<https://CRAN.R-project.org/package=geoR>>. Citado na página 33.
- SAMUEL-ROSA, A. et al. *FEBR: Repositório Brasileiro Livre para Dados Abertos do Solo*. 2020. [Http://coral.ufsm.br/febr/catalog/](http://coral.ufsm.br/febr/catalog/). Acessado: 2020-05-07. Citado 2 vezes nas páginas 21 e 50.
- SANTOS, F. C. dos; NOVAIS, R. F.; NEVES, J. C. L.; FOLONI, J. M.; ALBUQUERQUE FILHO, M. R. de; KER, J. C. Produtividade e aspectos nutricionais de plantas de soja cultivadas em solos de cerrado com diferentes texturas. *Revista Brasileira de Ciência do Solo*, v. 32, n. 5, p. 2015–2025, 2008. Citado na página 16.
- SCHLATHER, M.; MALINOWSKI, A.; MENCK, P. J.; OESTING, M.; STROKORB, K. Analysis, simulation and prediction of multivariate random fields with package RandomFields. *Journal of Statistical Software*, v. 63, n. 8, p. 1–25, 2015. Disponível em: <<http://www.jstatsoft.org/v63/i08/>>. Citado 4 vezes nas páginas 26, 39, 91 e 92.
- SEARLE, S. R. *Linear Models*. [S.l.]: John Wiley & Sons, 1971. Citado 2 vezes nas páginas 79 e 81.
- SEBER, G. A.; LEE, A. J. *Linear regression analysis*. [S.l.]: John Wiley & Sons, 2012. v. 329. Citado na página 79.
- SOLOS, E. Sistema brasileiro de classificação de solos. *Centro Nacional de Pesquisa de Solos: Rio de Janeiro*, 2013. Citado 2 vezes nas páginas 21 e 50.
- STEIN, M. L. *Interpolation of Spatial Data: some Theory for Kriging*. New York: Springer, 2012. Citado 2 vezes nas páginas 29 e 42.
- STRANG, G. *Introduction to linear algebra*. Wellesley, MA: Wellesley-Cambridge Press, 1993. v. 3. Citado na página 29.
- TIESSEN, H.; CUEVAS, E.; CHACON, P. The role of soil organic matter in sustaining soil fertility. *Nature*, Springer, v. 371, n. 6500, p. 783–785, 1994. Citado na página 16.
- TJELMELAND, H.; LUND, K. V. Bayesian modelling of spatial compositional data. *Journal of Applied Statistics*, Taylor & Francis, v. 30, n. 1, p. 87–100, 2003. Citado 4 vezes nas páginas 20, 38, 40 e 42.
- WACKERNAGEL, H. *Multivariate geostatistics: an introduction with applications*. New York: Springer, 2013. Citado 4 vezes nas páginas 25, 26, 28 e 36.
- WAHBA, G. *Spline models for observational data*. [S.l.]: SIAM, 1990. Citado na página 47.
- WANG, Z.; SHI, W.; ZHOU, W.; LI, X.; YUE, T. Comparison of additive and isometric log-ratio transformations combined with machine learning and regression kriging models for mapping soil particle size fractions. *Geoderma*, Elsevier, v. 365, p. 114214, 2020. Citado 2 vezes nas páginas 19 e 63.

WILSON, A.; ADAMS, R. Gaussian process kernels for pattern discovery and extrapolation. In: *International Conference on Machine Learning*. Atlanta, GA: JMLR: W&CP, 2013. p. 1067–1075. Citado na página 26.

WUEST, S. B.; CAESAR-TONTHAT, T.; WRIGHT, S. F.; WILLIAMS, J. D. Organic matter addition, n, and residue burning effects on infiltration, biological, and physical properties of an intensively tilled silt-loam soil. *Soil and Tillage Research*, Elsevier, v. 84, n. 2, p. 154–167, 2005. Citado na página 16.

YOLCUBAL, I.; BRUSSEAU, M. L.; ARTIOLA, J. F.; WIERENGA, P.; WILSON, L. G. Environmental physical properties and processes. In: ARTIOLA, J. F.; PEPPER, I. L.; BRUSSEAU, M. L. (Ed.). *Environmental Monitoring and Characterization*. Burlington, VT: Academic Press, 2004. p. 207–239. Citado na página 16.

APÊNDICE A – Gradiente da log-verossimilhança para o Modelo de Covariância Matérn Bivariada

A.1 Fórmula geral da derivada da log-verossimilhança

O modelo Matérn bivariado reduzido para processos Gaussianos e parametrizado por $\text{vec}(\boldsymbol{\mu}) \in \mathbb{R}^{2N}$ e $\boldsymbol{\theta} = (\sigma_1^2, \sigma_2^2, a, \rho, \tau_1^2, \tau_2^2)$, com $\sigma_1^2 > 0, \sigma_2^2 > 0, a > 0, 0 < \rho < 1$ e $\tau_1^2 > 0, \tau_2^2 > 0$ tem log-verossimilhança dada por

$$l(\boldsymbol{\mu}, \boldsymbol{\theta}) \propto -\frac{1}{2} \log(\det(\boldsymbol{\Sigma}_{\boldsymbol{\theta}})) - \frac{1}{2} (\text{vec}(\mathbf{z}) - \text{vec}(\boldsymbol{\mu}))^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\text{vec}(\mathbf{z}) - \text{vec}(\boldsymbol{\mu})), \quad (\text{A.1})$$

onde

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \begin{pmatrix} C_{11}(\mathbf{h}) & C_{12}(\mathbf{h}) \\ C_{21}(\mathbf{h}) & C_{22}(\mathbf{h}) \end{pmatrix}$$

é uma matriz $2N \times 2N$ com blocos

$$\begin{aligned} C_{11}(\mathbf{h}) &= \sigma_1^2 M(\mathbf{h} | \nu_1, a) + \tau_1^2 \mathbb{1}(\mathbf{h} = \mathbf{0}) \\ C_{22}(\mathbf{h}) &= \sigma_2^2 M(\mathbf{h} | \nu_2, a) + \tau_2^2 \mathbb{1}(\mathbf{h} = \mathbf{0}) \\ C_{21}(\mathbf{h}) &= C_{12}(\mathbf{h}) = \rho \sigma_1 \sigma_2 M\left(\mathbf{h} \mid \frac{\nu_1 + \nu_2}{2}, a\right) \\ M(\mathbf{h} | \nu, a) &= \frac{2^{1-\nu} (a \|\mathbf{h}\|)^{\nu} K_{\nu}(a \|\mathbf{h}\|)}{\Gamma(\nu)}, \end{aligned} \quad (\text{A.2})$$

para algum ν_1, ν_2 fixos positivos, e onde K_{ν} é a função de Bessel modificada de segunda ordem. Sem perda de generalidade, assumamos $\text{vec}(\boldsymbol{\mu}) = \mathbf{0}$, e definamos $\mathbf{y} = \text{vec}(\mathbf{z})$. Ao derivarmos a log-verossimilhança em relação a qualquer elemento (chamado θ por conveniência) de $\boldsymbol{\theta}$, temos a expressão geral da derivada da log-verossimilhança:

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \theta} = -\frac{1}{2} \left(\text{tr} \left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\theta}}}{\partial \theta} \right) - \mathbf{y}^T \left[\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\theta}}}{\partial \theta} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \right] \mathbf{y} \right) \quad (\text{A.3})$$

pois, usando as identidades de (MAGNUS; NEUDECKER, 1999),

$$\frac{\partial \log(\det(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}))}{\partial \theta} = \frac{1}{\log(\boldsymbol{\Sigma}_{\boldsymbol{\theta}})} \frac{\partial \det(\boldsymbol{\Sigma}_{\boldsymbol{\theta}})}{\partial \theta} = \frac{1}{\log(\boldsymbol{\Sigma}_{\boldsymbol{\theta}})} \log(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \text{tr} \left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\theta}}}{\partial \theta} \right) = \text{tr} \left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\theta}}}{\partial \theta} \right)$$

e

$$\frac{\partial \Sigma_{\theta}^{-1}}{\partial \theta} = -\Sigma_{\theta}^{-1} \frac{\partial \Sigma_{\theta}}{\partial \theta} \Sigma_{\theta}^{-1}.$$

Então se $\mathbf{r} = \Sigma_{\theta}^{-1} \mathbf{y}$, pela simetria da matriz de covariância, tem-se que

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \theta} = -\frac{1}{2} \left(\text{tr} \left(\Sigma_{\theta}^{-1} \frac{\partial \Sigma_{\theta}}{\partial \theta} \right) - \mathbf{r}^T \left[\frac{\partial \Sigma_{\theta}}{\partial \theta} \right] \mathbf{r} \right) \quad (\text{A.4})$$

A.2 Derivadas Parciais da Função de Covariância

A matriz de covariância é

$$\Sigma_{\theta}(\mathbf{h}) = \begin{pmatrix} \sigma_1^2 M(\mathbf{h} | \nu_1, a) & \\ \rho \sigma_2 \sigma_1 M\left(\mathbf{h} \mid \frac{\nu_1 + \nu_2}{2}, a\right) & \sigma_2^2 M(\mathbf{h} | \nu_2, a) \end{pmatrix}$$

Assim, as derivadas parciais de $\Sigma_{\theta}(\mathbf{h})$ com relação às componentes de $\boldsymbol{\theta}$ são:

- Derivada de Σ_{θ} com respeito a σ_1^2 :

$$\frac{\partial \Sigma_{\theta}(\mathbf{h})}{\partial \sigma_1^2} = \begin{pmatrix} M(\mathbf{h} | \nu_1, a) & \\ \frac{\rho \sigma_2}{2 \sigma_1} M\left(\mathbf{h} \mid \frac{\nu_1 + \nu_2}{2}, a\right) & \mathbf{0}_{N \times N} \end{pmatrix}$$

- Derivada de Σ_{θ} com respeito a σ_2^2 :

$$\frac{\partial \Sigma_{\theta}(\mathbf{h})}{\partial \sigma_2^2} = \begin{pmatrix} \mathbf{0}_{N \times N} & \\ \frac{\rho \sigma_1}{2 \sigma_2} M\left(\mathbf{h} \mid \frac{\nu_1 + \nu_2}{2}, a\right) & M(\mathbf{h} | \nu_2, a) \end{pmatrix}$$

- Derivada de Σ_{θ} com respeito a ρ :

$$\frac{\partial \Sigma_{\theta}(\mathbf{h})}{\partial \rho} = \begin{pmatrix} \mathbf{0}_{N \times N} & \\ \sigma_1 \sigma_2 M\left(\mathbf{h} \mid \frac{\nu_1 + \nu_2}{2}, a\right) & \mathbf{0}_{N \times N} \end{pmatrix}$$

- Derivada de Σ_{θ} com respeito a τ_1^2 :

$$\frac{\partial \Sigma_{\theta}(\mathbf{h})}{\partial \tau_1^2} = \begin{pmatrix} \mathbf{I} & \mathbf{0}_{N \times N} \\ \mathbf{0}_{N \times N} & \mathbf{0}_{N \times N} \end{pmatrix}$$

- Derivada de Σ_{θ} com respeito a τ_2^2 :

$$\frac{\partial \Sigma_{\theta}(\mathbf{h})}{\partial \tau_2^2} = \begin{pmatrix} \mathbf{0}_{N \times N} & \mathbf{0}_{N \times N} \\ \mathbf{0}_{N \times N} & \mathbf{I} \end{pmatrix}$$

- Derivada de Σ_{θ} com respeito a a :

$$\frac{\partial \Sigma_{\theta}(\mathbf{h})}{\partial a} = \begin{pmatrix} \sigma_1^2 M'(\mathbf{h}|\nu_1, a) & \\ \rho \sigma_1 \sigma_2 M'(\mathbf{h} | \frac{\nu_1 + \nu_2}{2}, a) & \\ & \sigma_2^2 M'(\mathbf{h}|\nu_2, a) \end{pmatrix}$$

Note que M' é a derivada de $M(\mathbf{h}|\nu_k, a)$ com respeito a a para $k = 1, 2, 3$, em que $\nu_3 = (\nu_1 + \nu_2)/2$. Isto é, para $\nu > 0$ em geral,

$$\frac{\partial M(\mathbf{h}|\nu, a)}{\partial a} = \frac{2^{1-\nu} \|\mathbf{h}\|^\nu}{\Gamma(\nu)} \left[\nu a^{\nu-1} K_\nu(a\|\mathbf{h}\|) + a^\nu \frac{\partial K_\nu(a\|\mathbf{h}\|)}{\partial a} \right]$$

e como (ABRAMOWITZ; STEGUN, 1972, p.376)

$$\frac{\partial K_\nu(a\|\mathbf{h}\|)}{\partial a} = \|\mathbf{h}\| \left[\frac{\nu}{a\|\mathbf{h}\|} K_\nu(a\|\mathbf{h}\|) - K_{\nu+1}(a\|\mathbf{h}\|) \right]$$

então

$$\frac{\partial M(\mathbf{h}|\nu, a)}{\partial a} = \frac{2^{1-\nu} \|\mathbf{h}\|^\nu}{\Gamma(\nu)} \left[\nu a^{\nu-1} K_\nu(a\|\mathbf{h}\|) + a^\nu \|\mathbf{h}\| \left(\frac{\nu}{a\|\mathbf{h}\|} K_\nu(a\|\mathbf{h}\|) - K_{\nu+1}(a\|\mathbf{h}\|) \right) \right].$$

Ao simplificar a última equação, obtem-se que

$$M(\mathbf{h}|\nu, a) = \frac{2^{1-\nu} \|\mathbf{h}\|^\nu a^{\nu-1}}{\Gamma(\nu)} \left[2\nu K_\nu(a\|\mathbf{h}\|) - a\|\mathbf{h}\| K_{\nu+1}(a\|\mathbf{h}\|) \right].$$

APÊNDICE B – Avaliação estável do gradiente para o Modelo de Covariância Matérn Bivariada

B.1 Inversa

O objetivo é obter a inversa de

$$\Sigma_{\theta}(\mathbf{h}) = \begin{pmatrix} C_{11}(\mathbf{h}) & C_{12}(\mathbf{h}) \\ C_{21}(\mathbf{h}) & C_{22}(\mathbf{h}) \end{pmatrix} \quad (\text{B.1})$$

então ao definirmos

$$\Sigma_{\theta}^{-1}(\mathbf{h}) = \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix}$$

é possível obter que

$$\begin{aligned} C_{11}^*(\mathbf{h}) &= \left[C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \right]^{-1} \\ C_{12}^*(\mathbf{h}) &= - \left[C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \right]^{-1} C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) &= -C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \left[C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \right]^{-1} \\ C_{22}^*(\mathbf{h}) &= C_{22}^{-1}(\mathbf{h}) + C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \left[C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \right]^{-1} C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h}). \end{aligned} \quad (\text{B.2})$$

Se também definirmos

$$\begin{aligned} V_1 &:= C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \\ V_2 &:= C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \end{aligned} \quad (\text{B.3})$$

podemos obter uma forma simplificada para o sistema (B.2), já que como $C_{11}, C_{12}, C_{21}, C_{22}$

são simétricas

$$\begin{aligned}
 C_{11}^*(\mathbf{h}) &= V_2^{-1} \\
 C_{12}^*(\mathbf{h}) &= -V_2^{-1}V_1^T \\
 C_{21}^*(\mathbf{h}) &= -V_1V_2^{-1} \\
 C_{22}^*(\mathbf{h}) &= C_{22}^{-1}(\mathbf{h}) + V_1V_2^{-1}V_1^T
 \end{aligned} \tag{B.4}$$

pois $V_1 = C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h})$ fornece $V_1^T = C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})$. Note ainda que $C_{12}^*(\mathbf{h}) = C_{21}^{*T}(\mathbf{h})$, pois

$$\begin{aligned}
 V_2^T &= C_{11}(\mathbf{h}) - (C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{12}(\mathbf{h}))^T \\
 &= C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})(C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h}))^T \\
 &= C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{12}(\mathbf{h}) \\
 &= V_2,
 \end{aligned}$$

e

$$\begin{aligned}
 C_{12}^{*T}(\mathbf{h}) &= -(V_2^{-1}V_1^T)^T \\
 &= V_1^T(V_2^{-1})^T \\
 &= V_1^T(V_2^{-1}) \\
 &= C_{21}^*(\mathbf{h}).
 \end{aligned}$$

B.2 Logaritmo do Determinante

Se $\Sigma_\theta(\mathbf{h})$ é como dada em (B.1), então

$$\det(\Sigma_\theta(\mathbf{h})) = |\Sigma_\theta(\mathbf{h})| = |C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h})||C_{22}(\mathbf{h})|$$

Portanto, com o logaritmo fica

$$\log(|\Sigma_\theta(\mathbf{h})|) = \log |C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h})| + \log |C_{22}(\mathbf{h})| \tag{B.5}$$

B.3 Traço e Forma Quadrática

Para avaliar $\mathbf{r}^T \frac{\partial \Sigma_\theta}{\partial \theta} \mathbf{r}$ e $\text{tr} \left(\Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta} \right)$ a estrutura de blocos também pode nos ajudar a deixar o algoritmo mais estável e rápido. Sendo a inversa de Σ_θ definida como

$$\Sigma_\theta^{-1}(\mathbf{h}) = \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix}, \tag{B.6}$$

se

$$\mathbf{r} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \mathbf{y} = \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} C_{11}^*(\mathbf{h})\mathbf{y}_1 + C_{12}^*(\mathbf{h})\mathbf{y}_2 \\ C_{21}^*(\mathbf{h})\mathbf{y}_1 + C_{22}^*(\mathbf{h})\mathbf{y}_2 \end{pmatrix}$$

em que \mathbf{y}_1 e \mathbf{y}_2 são, respectivamente, os vetores resposta associados às primeira e segunda variáveis multivariadas, temos que

$$\begin{aligned} \mathbf{y}^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\mathbf{h}) \mathbf{y} &= \mathbf{y}^T \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix} \mathbf{y} \\ &= (y_1, y_2)^T \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= (y_1, y_2)^T \begin{pmatrix} C_{11}^*(\mathbf{h})y_1 + C_{12}^*(\mathbf{h})y_2 \\ C_{21}^*(\mathbf{h})y_1 + C_{22}^*(\mathbf{h})y_2 \end{pmatrix} \\ &= y_1^T C_{11}^*(\mathbf{h})y_1 + y_1^T C_{12}^*(\mathbf{h})y_2 + y_2^T C_{21}^*(\mathbf{h})y_1 + y_2^T C_{22}^*(\mathbf{h})y_2 \\ &= y_1^T C_{11}^*(\mathbf{h})y_1 + 2y_1^T C_{12}^*(\mathbf{h})y_2 + y_2^T C_{22}^*(\mathbf{h})y_2 \end{aligned} \tag{B.7}$$

pois como $C_{12}^{*T}(\mathbf{h}) = C_{21}^*(\mathbf{h})$ obtemos que

$$\begin{aligned} y_1^T C_{12}^*(\mathbf{h})y_2 &= tr(y_1^T C_{12}^*(\mathbf{h})y_2) = tr(y_1^T C_{12}^*(\mathbf{h})y_2)^T = tr((C_{12}^*(\mathbf{h})y_2)^T y_1) = \\ &= tr(y_2^T C_{12}^{*T}(\mathbf{h})y_1) = y_2^T C_{21}^*(\mathbf{h})y_1. \end{aligned}$$

Também é possível escrever o resultado anterior em termos de (B.7), o que fornece

$$y_1^T V_2^{-1} y_1 - 2y_1^T V_2^{-1} V_1^T y_2 + y_2^T C_{22}^{-1} y_2 - y_2^T V_1 V_2^{-1} V_1^T y_2$$

Para terminar é preciso obter as derivadas em relação a cada componente de $\boldsymbol{\theta}$.

B.3.1 Caso em que $\theta_k = \sigma_1^2$

Traço:

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\mathbf{h}) \frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\theta}}}{\partial \sigma_1^2} &= \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix} \begin{pmatrix} M(\mathbf{h}|\nu_1, a) \\ \frac{\rho\sigma_2}{2\sigma_1} M(\mathbf{h}|\nu_3, a) \quad \mathbf{0}_{N \times N} \end{pmatrix} \\ &= \begin{pmatrix} C_{11}^*(\mathbf{h})M(\mathbf{h}|\nu_1, a) + C_{12}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1} M(\mathbf{h}|\nu_3, a) & C_{11}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1} M(\mathbf{h}|\nu_3, a) \\ C_{21}^*(\mathbf{h})M(\mathbf{h}|\nu_1, a) + C_{22}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1} M(\mathbf{h}|\nu_3, a) & C_{21}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1} M(\mathbf{h}|\nu_3, a) \end{pmatrix} \end{aligned}$$

Portanto como $C_{12}^*(\mathbf{h}) = C_{21}^{*T}(\mathbf{h})$

$$\begin{aligned}
 tr\left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\mathbf{h})\frac{\partial\boldsymbol{\Sigma}_{\boldsymbol{\theta}}}{\partial\sigma_1^2}\right) &= tr\left(C_{11}^*(\mathbf{h})M(\mathbf{h}|\nu_1, a)\right) + 2tr\left(C_{12}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1}M(\mathbf{h}|\nu_3, a)\right) \\
 &= tr\left(C_{11}^*(\mathbf{h})M(\mathbf{h}|\nu_1, a)\right) + \frac{\rho\sigma_2}{\sigma_1}tr\left(C_{12}^*(\mathbf{h})M(\mathbf{h}|\nu_3, a)\right) \\
 &= \frac{1}{\sigma_1^2}tr\left(C_{11}^*(\mathbf{h})C_{11}(\mathbf{h})\right) + \frac{1}{\sigma_1^2}tr\left(C_{12}^*(\mathbf{h})C_{12}(\mathbf{h})\right) \\
 &= \frac{1}{\sigma_1^2}\left[tr\left(C_{11}^*(\mathbf{h})C_{11}(\mathbf{h})\right) + tr\left(C_{12}^*(\mathbf{h})C_{12}(\mathbf{h})\right)\right]
 \end{aligned} \tag{B.8}$$

Forma Quadrática:

$$\begin{aligned}
 y^T\left[\frac{\partial\Sigma(\mathbf{h})}{\partial\theta_k}\right]y &= \begin{pmatrix} y_1^T & y_2^T \end{pmatrix} \begin{pmatrix} M(\mathbf{h}|\nu_1, a) & \\ \rho\frac{\sigma_2}{2\sigma_1}M(\mathbf{h}|\nu_3, a) & \mathbf{0}_{N\times N} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\
 &= \begin{pmatrix} y_1^T & y_2^T \end{pmatrix} \begin{pmatrix} M(\mathbf{h}|\nu_1, a)y_1 + \rho\frac{\sigma_2}{2\sigma_1}M(\mathbf{h}|\nu_3, a)y_2 \\ \rho\frac{\sigma_2}{2\sigma_1}M(\mathbf{h}|\nu_3, a)y_1 \end{pmatrix} \\
 &= y_1^T M(\mathbf{h}|\nu_1, a)y_1 + y_1^T \rho\frac{\sigma_2}{2\sigma_1}M(\mathbf{h}|\nu_3, a)y_2 + y_2^T \rho\frac{\sigma_2}{2\sigma_1}M(\mathbf{h}|\nu_3, a)y_1 \\
 &= y_1^T M(\mathbf{h}|\nu_1, a)y_1 + \rho\frac{\sigma_2}{\sigma_1}y_1^T M(\mathbf{h}|\nu_3, a)y_2 \\
 &= \frac{1}{\sigma_1^2}y_1^T \left[C_{11}(\mathbf{h}|\nu_1, a)y_1 + C_{12}(\mathbf{h}|\nu_3, a)y_2 \right]
 \end{aligned} \tag{B.9}$$

B.3.2 Caso em que $\theta_k = \sigma_2^2$

Traço:

$$\begin{aligned} \Sigma_{\theta}^{-1}(\mathbf{h}) \frac{\partial \Sigma_{\theta}}{\partial \sigma_2^2} &= \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix} \begin{pmatrix} \mathbf{0}_{N \times N} & \\ \frac{\rho \sigma_1}{2\sigma_2} M(\mathbf{h}|\nu_3, a) & M(\mathbf{h}|\nu_2, a) \end{pmatrix} \\ &= \begin{pmatrix} C_{12}^*(\mathbf{h}) \frac{\rho \sigma_1}{2\sigma_2} M(\mathbf{h}|\nu_3, a) & C_{11}^*(\mathbf{h}) \frac{\rho \sigma_1}{2\sigma_2} M(\mathbf{h}|\nu_3, a) + C_{12}^*(\mathbf{h}) M(\mathbf{h}|\nu_2, a) \\ C_{22}^*(\mathbf{h}) \frac{\rho \sigma_1}{2\sigma_2} M(\mathbf{h}|\nu_3, a) & C_{21}^*(\mathbf{h}) \frac{\rho \sigma_1}{2\sigma_2} M(\mathbf{h}|\nu_3, a) + C_{22}^*(\mathbf{h}) M(\mathbf{h}|\nu_2, a) \end{pmatrix} \end{aligned}$$

Portanto como $C_{12}^*(\mathbf{h}) = C_{21}^{*T}(\mathbf{h})$

$$\begin{aligned} tr \left(\Sigma_{\theta}^{-1}(\mathbf{h}) \frac{\partial \Sigma_{\theta}}{\partial \sigma_2^2} \right) &= tr \left(C_{22}^*(\mathbf{h}) M(\mathbf{h}|\nu_2, a) \right) + 2tr \left(C_{12}^*(\mathbf{h}) \frac{\rho \sigma_1}{2\sigma_2} M(\mathbf{h}|\nu_3, a) \right) \\ &= tr \left(C_{22}^*(\mathbf{h}) M(\mathbf{h}|\nu_2, a) \right) + \frac{\rho \sigma_1}{\sigma_2} tr \left(C_{12}^*(\mathbf{h}) M(\mathbf{h}|\nu_3, a) \right) \\ &= \frac{1}{\sigma_2^2} tr \left(C_{22}^*(\mathbf{h}) C_{22}(\mathbf{h}) \right) + \frac{1}{\sigma_2^2} tr \left(C_{12}^*(\mathbf{h}) C_{12}(\mathbf{h}) \right) \\ &= \frac{1}{\sigma_2^2} \left[tr \left(C_{22}^*(\mathbf{h}) C_{22}(\mathbf{h}) \right) + tr \left(C_{12}^*(\mathbf{h}) C_{12}(\mathbf{h}) \right) \right] \end{aligned} \tag{B.10}$$

Forma Quadrática:

$$\begin{aligned} y^T \left[\frac{\partial \Sigma(\mathbf{h})}{\partial \theta_k} \right] y &= \begin{pmatrix} y_1^T & y_2^T \end{pmatrix} \begin{pmatrix} \mathbf{0}_{N \times N} & \\ \rho \frac{\sigma_1}{2\sigma_2} M(\mathbf{h}|\nu_3, a) & M(\mathbf{h}|\nu_2, a) \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= \begin{pmatrix} y_1^T & y_2^T \end{pmatrix} \begin{pmatrix} \rho \frac{\sigma_1}{2\sigma_2} M(\mathbf{h}|\nu_3, a) y_2 \\ \rho \frac{\sigma_1}{2\sigma_2} M(\mathbf{h}|\nu_3, a) y_1 + M(\mathbf{h}|\nu_2, a) y_2 \end{pmatrix} \\ &= y_2^T M(\mathbf{h}|\nu_1, a) y_2 + y_1^T \rho \frac{\sigma_1}{2\sigma_2} M(\mathbf{h}|\nu_3, a) y_2 + y_2^T \rho \frac{\sigma_1}{2\sigma_2} M(\mathbf{h}|\nu_3, a) y_1 \\ &= y_2^T M(\mathbf{h}|\nu_1, a) y_2 + \rho \frac{\sigma_1}{\sigma_2} y_1^T M(\mathbf{h}|\nu_3, a) y_2 \\ &= \frac{1}{\sigma_2^2} \left[y_2^T C_{22}(\mathbf{h}|\nu_1, a) + y_1^T C_{12}(\mathbf{h}|\nu_3, a) \right] y_2 \end{aligned} \tag{B.11}$$

B.3.3 Caso em que $\theta_k = \rho$

Traço:

$$\begin{aligned}\Sigma_{\theta}^{-1}(\mathbf{h})\frac{\partial\Sigma_{\theta}}{\partial\rho} &= \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix} \begin{pmatrix} \mathbf{0}_{N\times N} & \\ \sigma_1\sigma_2M(\mathbf{h}|\nu_3, a) & \mathbf{0}_{N\times N} \end{pmatrix} \\ &= \begin{pmatrix} C_{12}^*(\mathbf{h})\sigma_1\sigma_2M(\mathbf{h}|\nu_3, a) & C_{11}^*(\mathbf{h})\sigma_1\sigma_2M(\mathbf{h}|\nu_3, a) \\ C_{22}^*(\mathbf{h})\sigma_1\sigma_2M(\mathbf{h}|\nu_3, a) & C_{21}^*(\mathbf{h})\sigma_1\sigma_2M(\mathbf{h}|\nu_3, a) \end{pmatrix}\end{aligned}$$

Portanto como $C_{12}^*(\mathbf{h}) = C_{21}^{*T}(\mathbf{h})$

$$tr\left(\Sigma_{\theta}^{-1}(\mathbf{h})\frac{\partial\Sigma_{\theta}}{\partial\rho}\right) = 2\sigma_1\sigma_2tr\left(C_{12}^*(\mathbf{h})M(\mathbf{h}|\nu_3, a)\right) \quad (\text{B.12})$$

Forma Quadrática:

$$\begin{aligned}y^T\left[\frac{\partial\Sigma(\mathbf{h})}{\partial\theta_k}\right]y &= \begin{pmatrix} y_1^T & y_2^T \end{pmatrix} \begin{pmatrix} \mathbf{0}_{N\times N} & \\ \sigma_1\sigma_2M(\mathbf{h}|\nu_3, a) & \mathbf{0}_{N\times N} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= \begin{pmatrix} y_1^T & y_2^T \end{pmatrix} \begin{pmatrix} \sigma_1\sigma_2M(\mathbf{h}|\nu_3, a)y_2 \\ \sigma_1\sigma_2M(\mathbf{h}|\nu_3, a)y_1 \end{pmatrix} \\ &= y_1^T\sigma_1\sigma_2M(\mathbf{h}|\nu_3, a)y_2 + y_2^T\sigma_1\sigma_2M(\mathbf{h}|\nu_3, a)y_1 \\ &= 2\sigma_1\sigma_2y_1^TM(\mathbf{h}|\nu_3, a)y_2\end{aligned} \quad (\text{B.13})$$

B.3.4 Caso em que $\theta_k = a$
Traço:

$$\begin{aligned} \Sigma_{\theta}^{-1}(\mathbf{h}) \frac{\partial \Sigma_{\theta}}{\partial a} &= \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix} \begin{pmatrix} \sigma_1^2 M'(\mathbf{h}|\nu_1, a) & \rho\sigma_1\sigma_2 M'(\mathbf{h}|\nu_3, a) \\ \rho\sigma_1\sigma_2 M'(\mathbf{h}|\nu_3, a) & \sigma_2^2 M'(\mathbf{h}|\nu_2, a) \end{pmatrix} \\ &= \begin{pmatrix} C_{11}^*(\mathbf{h})\sigma_1^2 M'(\mathbf{h}|\nu_1, a) + C_{12}^*(\mathbf{h})\rho\sigma_1\sigma_2 M'(\mathbf{h}|\nu_3, a) \\ C_{21}^*(\mathbf{h})\sigma_1^2 M'(\mathbf{h}|\nu_1, a) + C_{22}^*(\mathbf{h})\rho\sigma_1\sigma_2 M'(\mathbf{h}|\nu_3, a) \\ C_{11}^*(\mathbf{h})\rho\sigma_1\sigma_2 M'(\mathbf{h}|\nu_3, a) + C_{12}^*(\mathbf{h})\sigma_2^2 M'(\mathbf{h}|\nu_2, a) \\ C_{21}^*(\mathbf{h})\rho\sigma_1\sigma_2 M'(\mathbf{h}|\nu_3, a) + C_{22}^*(\mathbf{h})\sigma_2^2 M'(\mathbf{h}|\nu_2, a) \end{pmatrix} \end{aligned}$$

Portanto como $C_{12}^*(\mathbf{h}) = C_{21}^{*T}(\mathbf{h})$

$$\begin{aligned} tr\left(\Sigma_{\theta}^{-1}(\mathbf{h}) \frac{\partial \Sigma_{\theta}}{\partial a}\right) &= \sigma_1^2 tr\left(C_{11}^*(\mathbf{h}) M'(\mathbf{h}|\nu_1, a)\right) + \\ &\quad \sigma_2^2 tr\left(C_{22}^*(\mathbf{h}) M'(\mathbf{h}|\nu_2, a)\right) + \\ &\quad 2\rho\sigma_1\sigma_2 tr\left(C_{12}^*(\mathbf{h}) M'(\mathbf{h}|\nu_3, a)\right) \end{aligned} \quad (\text{B.14})$$

onde (ABRAMOWITZ; STEGUN, 1972)

$$M'(\mathbf{h}|\nu, a) = \frac{\partial M(\mathbf{h}|\nu, a)}{\partial a} = \frac{\left(\frac{2}{a}\right)^{1-\nu} d^{\nu}}{\Gamma(\nu)} \left[2\nu K_{\nu}(ad) - ad K_{\nu+1}(ad) \right] \quad (\text{B.15})$$

Forma Quadrática:

$$\begin{aligned} y^T \left[\frac{\partial \Sigma(\mathbf{h})}{\partial \theta_k} \right] y &= \begin{pmatrix} y_1^T & y_2^T \end{pmatrix} \begin{pmatrix} \sigma_1^2 M'(\mathbf{h}|\nu_1, a) & \\ \rho\sigma_1\sigma_2 M'(\mathbf{h}|\nu_3, a) & \sigma_2^2 M'(\mathbf{h}|\nu_2, a) \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= \begin{pmatrix} y_1^T & y_2^T \end{pmatrix} \begin{pmatrix} \sigma_1^2 M'(\mathbf{h}|\nu_1, a) y_1 + \rho\sigma_1\sigma_2 M'(\mathbf{h}|\nu_3, a) y_2 \\ \rho\sigma_1\sigma_2 M'(\mathbf{h}|\nu_3, a) y_1 + \sigma_2^2 M'(\mathbf{h}|\nu_2, a) y_2 \end{pmatrix} \\ &= y_1^T \sigma_1^2 M'(\mathbf{h}|\nu_1, a) y_1 + y_1^T \rho\sigma_1\sigma_2 M'(\mathbf{h}|\nu_3, a) y_2 + \\ &\quad y_2^T \rho\sigma_1\sigma_2 M'(\mathbf{h}|\nu_3, a) y_1 + y_2^T \sigma_2^2 M'(\mathbf{h}|\nu_2, a) y_2 \\ &= \sigma_1^2 y_1^T M'(\mathbf{h}|\nu_1, a) y_1 + \sigma_2^2 y_2^T M'(\mathbf{h}|\nu_2, a) y_2 + 2\rho\sigma_1\sigma_2 y_1^T M'(\mathbf{h}|\nu_3, a) y_2 \end{aligned} \quad (\text{B.16})$$

B.3.5 Caso em que $\theta_k = \tau_1^2$

Traço:

$$\begin{aligned} \text{tr} \left(\Sigma_{\theta}^{-1}(\mathbf{h}) \frac{\partial \Sigma_{\theta}}{\partial \tau_1^2} \right) &= \text{tr} \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix} \begin{pmatrix} I & O \\ O & O \end{pmatrix} = \text{tr} \begin{pmatrix} C_{11}^*(\mathbf{h}) & O \\ C_{21}^*(\mathbf{h}) & O \end{pmatrix} \\ &= \text{tr}(C_{11}^*(\mathbf{h})) \end{aligned}$$

Forma Quadrática:

$$y^T \left[\frac{\partial \Sigma(\mathbf{h})}{\partial \theta_k} \right] y = \begin{pmatrix} y_1^T & y_2^T \end{pmatrix} \begin{pmatrix} I & O \\ O & O \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = y_1^T y_1 \quad (\text{B.17})$$

 B.3.6 Caso em que $\theta_k = \tau_2^2$

Traço:

$$\begin{aligned} \text{tr} \left(\Sigma_{\theta}^{-1}(\mathbf{h}) \frac{\partial \Sigma_{\theta}}{\partial \tau_2^2} \right) &= \text{tr} \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix} \begin{pmatrix} O & O \\ O & I \end{pmatrix} = \text{tr} \begin{pmatrix} O & C_{12}^*(\mathbf{h}) \\ O & C_{22}^*(\mathbf{h}) \end{pmatrix} \\ &= \text{tr}(C_{22}^*(\mathbf{h})) \end{aligned}$$

Forma Quadrática:

$$y^T \left[\frac{\partial \Sigma(\mathbf{h})}{\partial \theta_k} \right] y = \begin{pmatrix} y_1^T & y_2^T \end{pmatrix} \begin{pmatrix} O & O \\ O & I \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = y_2^T y_2 \quad (\text{B.18})$$

APÊNDICE C – Informação de Fisher

Seja $\text{vec}(\mathbf{Y}_{n \times 2}) = \mathbf{Z}_{N \times 1} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Considere $\boldsymbol{\Sigma}$ parametrizada por $\boldsymbol{\theta} = (\sigma_1^2, \sigma_2^2, a, \rho)$. A log-verossimilhança $l(\boldsymbol{\theta})$ é

$$l(\boldsymbol{\theta}) \propto -\frac{1}{2} \log(\det(\boldsymbol{\Sigma}_{\boldsymbol{\theta}})) - \frac{1}{2} (\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\mathbf{z} - \boldsymbol{\mu}).$$

Assuma, sem perda de generalidade, que $\boldsymbol{\mu} = \mathbf{0}$, de forma que a derivada de $l(\boldsymbol{\theta})$ em relação a qualquer parâmetro θ de $\boldsymbol{\theta}$ é

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \theta} = -\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\theta}}}{\partial \theta} \right) + \frac{1}{2} \mathbf{z}^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \left[\frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\theta}}}{\partial \theta} \right] \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \mathbf{z}.$$

A matriz de Informação de Fisher $\mathcal{I}(\boldsymbol{\theta})$ é, portanto,

$$\mathcal{I}(\boldsymbol{\theta}) = \mathbb{E} \left(\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \right).$$

Se denotarmos $\dot{l}_k(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_k}$ então

$$\mathcal{I}(\boldsymbol{\theta}) = \mathbb{E} \begin{pmatrix} \dot{l}_1^2(\boldsymbol{\theta}) & \cdots & \dot{l}_1(\boldsymbol{\theta}) \dot{l}_4(\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \dot{l}_4(\boldsymbol{\theta}) \dot{l}_1(\boldsymbol{\theta}) & \cdots & \dot{l}_4^2(\boldsymbol{\theta}) \end{pmatrix}.$$

C.1 Resultado Preliminares

Aqui são apresentados alguns resultados básicos para calcular a esperança envolvendo a matriz de Informação de Fisher. As demonstrações estão adaptadas de (SEBER; LEE, 2012, pp.9–11) e (SEARLE, 1971, pp.54–57), respectivamente.

Resultado 1. Esperança de Forma Quadrática Seja \mathbf{Z} um vetor $N \times 1$ de variáveis aleatórias e \mathbf{A} uma matriz $N \times N$ simétrica. Se $\mathbb{E}[\mathbf{Z}] = \boldsymbol{\mu}$ e $\text{Var}(\mathbf{Z}) = \boldsymbol{\Sigma}$, então

$$\mathbb{E}[\mathbf{Z}^T \mathbf{A} \mathbf{Z}] = \text{tr}(\mathbf{A} \boldsymbol{\Sigma}) + \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu}$$

e se \mathbf{Z} tiver distribuição $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, então

$$\text{Var}[\mathbf{Z}^T \mathbf{A} \mathbf{Z}] = 2 \text{tr}(\mathbf{A} \boldsymbol{\Sigma}) + 4 \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\Sigma} \mathbf{A} \boldsymbol{\mu}$$

Demonstração. Como $\mathbf{Z}^T \mathbf{A} \mathbf{Z}$ é uma forma quadrática, então $\mathbb{E}[\mathbf{Z}^T \mathbf{A} \mathbf{Z}]$ é um número, e igual ao seu traço. Pela linearidade e propriedade cíclica do traço, temos

$$\begin{aligned} \mathbb{E}[\mathbf{Z}^T \mathbf{A} \mathbf{Z}] &= \text{tr} \left\{ \mathbb{E}[\mathbf{Z}^T \mathbf{A} \mathbf{Z}] \right\} \\ &= \mathbb{E} \left\{ \text{tr}[\mathbf{A} \mathbf{Z} \mathbf{Z}^T] \right\} \\ &= \text{tr} \left\{ \mathbf{A} \mathbb{E}[\mathbf{Z} \mathbf{Z}^T] \right\} \\ &= \text{tr} \left\{ \mathbf{A} \boldsymbol{\Sigma} \right\} + \text{tr} \left\{ \mathbf{A} \boldsymbol{\mu} \boldsymbol{\mu}^T \right\}. \end{aligned}$$

Para determinar $\text{Var}[\mathbf{Z}^T \mathbf{A} \mathbf{Z}]$, assumindo $\mathbf{Z} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, temos que a função geradora de momentos de $\mathbf{Z}^T \mathbf{A} \mathbf{Z}$ para $|t|$ pequeno (isto é, tal que $\boldsymbol{\Sigma}^{-1} - 2t\mathbf{A}$ é positiva definida), é

$$\begin{aligned} \mathcal{M}(t) &= \mathbb{E}(e^{t\mathbf{Z}^T \mathbf{A} \mathbf{Z}}) = \int_{-\infty}^{\infty} \frac{e^{t\mathbf{z}^T \mathbf{A} \mathbf{z}}}{(2\pi)^{\frac{N}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})} d\mathbf{z} \\ &= \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \frac{1}{2}\boldsymbol{\mu}^{*T} (\boldsymbol{\Sigma}^{-1} - 2t\mathbf{A}) \boldsymbol{\mu}^*}}{(2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu}^*)^T [\boldsymbol{\Sigma}^{-1} - 2t\mathbf{A}](\mathbf{z}-\boldsymbol{\mu}^*)} d\mathbf{z} \\ &= \frac{1}{|\boldsymbol{\Sigma}^{-1} - 2t\mathbf{A}|^{\frac{1}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \frac{1}{2}\boldsymbol{\mu}^{*T} (\boldsymbol{\Sigma}^{-1} - 2t\mathbf{A}) \boldsymbol{\mu}^*} \\ &= \frac{1}{|\mathbf{I}_{N \times N} - 2t\mathbf{A}\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}\boldsymbol{\mu}^T [\mathbf{I} - (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1}] \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}} \end{aligned}$$

onde $\boldsymbol{\mu}^* = [\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma}]^{-1} \boldsymbol{\mu}$. Note que $\mathcal{M}(0) = 1$. Usando a propriedade da função geradora de momentos

$$\mathbb{E}[(\mathbf{Z}^T \mathbf{A} \mathbf{Z})^2] = \left. \frac{d^2 \mathcal{M}(t)}{dt^2} \right|_{t=0}$$

temos que

$$\begin{aligned} \frac{d\mathcal{M}(t)}{dt} &= -\frac{1}{2} \frac{1}{|\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma}|^{\frac{3}{2}}} e^{-\frac{1}{2}\boldsymbol{\mu}^T [\mathbf{I} - (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1}] \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}} \left(\frac{d}{dt} |\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma}| \right) \\ &\quad + \frac{1}{2} \frac{1}{|\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}\boldsymbol{\mu}^T [\mathbf{I} - (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1}] \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}} \left(\frac{d}{dt} \boldsymbol{\mu}^T (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right) \\ &= -\frac{1}{2} \mathcal{M}(t) \left(\frac{1}{|\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma}|} \frac{d}{dt} |\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma}| - \frac{d}{dt} \boldsymbol{\mu}^T (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right) \\ &= \frac{1}{2} \mathcal{M}(t) \left(2\text{tr} \left\{ (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \mathbf{A}\boldsymbol{\Sigma} \right\} + 2\boldsymbol{\mu}^T (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \mathbf{A}\boldsymbol{\Sigma} (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right) \\ &= \mathcal{M}(t) \left(\text{tr} \left\{ (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \mathbf{A}\boldsymbol{\Sigma} \right\} + \boldsymbol{\mu}^T (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \mathbf{A}\boldsymbol{\Sigma} (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right), \\ \frac{d^2 \mathcal{M}(t)}{dt^2} &= \mathcal{M}(t) \left(\text{tr} \left\{ (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \mathbf{A}\boldsymbol{\Sigma} \right\} + \boldsymbol{\mu}^T (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \mathbf{A}\boldsymbol{\Sigma} (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right)^2 \\ &\quad + \mathcal{M}(t) \frac{d}{dt} \text{tr} \left\{ (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \mathbf{A}\boldsymbol{\Sigma} \right\} \\ &\quad + \mathcal{M}(t) \frac{d}{dt} \boldsymbol{\mu}^T (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \mathbf{A}\boldsymbol{\Sigma} (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}. \end{aligned}$$

Usando a regra da cadeia de cálculo matricial $\frac{dg(\mathbf{U}(t))}{dt} = \text{tr} \left(\frac{dg(\mathbf{U}(t))}{d\mathbf{U}} \frac{d\mathbf{U}(t)}{dt} \right)$, temos

$$\frac{d}{dt} \text{tr} \left\{ (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \mathbf{A}\boldsymbol{\Sigma} \right\} = 2\text{tr} \left((\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \mathbf{A}\boldsymbol{\Sigma} (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \mathbf{A}\boldsymbol{\Sigma} \right),$$

e

$$\begin{aligned} \frac{d}{dt} \boldsymbol{\mu}^T (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \mathbf{A}\boldsymbol{\Sigma} (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} = \\ 4\boldsymbol{\mu}^T (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \mathbf{A}\boldsymbol{\Sigma} (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \mathbf{A}\boldsymbol{\Sigma} (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}. \end{aligned}$$

Portanto

$$\mathbb{E}[(\mathbf{Z}^T \mathbf{A} \mathbf{Z})^2] = \left(\text{tr}\{\mathbf{A}\boldsymbol{\Sigma}\} + \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} \right)^2 + 2\text{tr}\{\mathbf{A}\boldsymbol{\Sigma} \mathbf{A} \boldsymbol{\Sigma}\} + 4\boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\Sigma} \mathbf{A} \boldsymbol{\mu}.$$

□

Resultado 2. *Covariância de Formas Quadráticas* Seja \mathbf{Z} um vetor $N \times 1$ de variáveis aleatórias e $\mathbf{A}_1, \mathbf{A}_2$ matrizes $N \times N$ simétricas. Se $\mathbb{E}[\mathbf{Z}] = \boldsymbol{\mu}$ e $\text{Var}(\mathbf{Z}) = \boldsymbol{\Sigma}$, então

$$\text{Cov}[\mathbf{Z}^T \mathbf{A}_1 \mathbf{Z}, \mathbf{Z}^T \mathbf{A}_2 \mathbf{Z}] = 2\text{tr}\left(\mathbf{A}_1 \boldsymbol{\Sigma} \mathbf{A}_2 \boldsymbol{\Sigma}\right) + 4\boldsymbol{\mu}^T \mathbf{A}_1 \boldsymbol{\Sigma} \mathbf{A}_2 \boldsymbol{\mu}$$

Demonstração. O **Resultado 1** vale para $\boldsymbol{\Sigma}$ de posto incompleto, substituindo as restrições de existência da função geradora de momentos baseadas em matrizes positivas definidas por positivas semi-definidas onde necessário (SEARLE, 1971, p.66). Se definirmos o vetor $\tilde{\mathbf{Z}}_{4N \times 1} = (\mathbf{Z}^T, \mathbf{Z}^T, \mathbf{Z}^T, \mathbf{Z}^T)^T$, então

$$\mathbb{E}(\tilde{\mathbf{Z}}) = \mathbf{1}_{4 \times 1} \otimes \boldsymbol{\mu}, \quad \text{Var}(\tilde{\mathbf{Z}}) = (\mathbf{1}_{4 \times 1} \mathbf{1}_{1 \times 4}^T) \otimes \boldsymbol{\Sigma}$$

e

$$\mathbf{A} = \frac{1}{2} \begin{pmatrix} \mathbf{0}_{N \times N} & \mathbf{A}_1 & \mathbf{0}_{N \times N} & \mathbf{0}_{N \times N} \\ \mathbf{A}_1 & \mathbf{0}_{N \times N} & \mathbf{0}_{N \times N} & \mathbf{0}_{N \times N} \\ \mathbf{0}_{N \times N} & \mathbf{0}_{N \times N} & \mathbf{0}_{N \times N} & \mathbf{A}_2 \\ \mathbf{0}_{N \times N} & \mathbf{0}_{N \times N} & \mathbf{A}_2 & \mathbf{0}_{N \times N} \end{pmatrix},$$

temos que $\tilde{\mathbf{Z}}^T \mathbf{A} \tilde{\mathbf{Z}} = \mathbf{Z}^T \mathbf{A}_1 \mathbf{Z} + \mathbf{Z}^T \mathbf{A}_2 \mathbf{Z}$. Naturalmente, $\text{Var}[\tilde{\mathbf{Z}}^T \mathbf{A} \tilde{\mathbf{Z}}] = \text{Var}[\mathbf{Z}^T \mathbf{A}_1 \mathbf{Z}] + \text{Var}[\mathbf{Z}^T \mathbf{A}_2 \mathbf{Z}] + 2\text{Cov}[\mathbf{Z}^T \mathbf{A}_1 \mathbf{Z}, \mathbf{Z}^T \mathbf{A}_2 \mathbf{Z}]$ o que implica, ao reorganizar os termos, no resultado desejado.

□

C.2 Resultado Geral

Assuma, sem perda de generalidade, que $\boldsymbol{\mu} = \mathbf{0}$. Sendo $\dot{l}_i(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_i}$ e $\dot{l}_2(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_j}$, onde $i \neq j$, temos que

$$\begin{aligned}
 \dot{l}_i(\boldsymbol{\theta})\dot{l}_j(\boldsymbol{\theta}) &= \frac{1}{4} \left\{ \left[\text{tr}(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}) - \mathbf{z}^T (\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1}) \mathbf{z} \right] \left[\text{tr}(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j}) - \mathbf{z}^T (\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \boldsymbol{\Sigma}^{-1}) \mathbf{z} \right] \right\} \\
 &= \frac{1}{4} \left[\text{tr}(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}) \text{tr}(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j}) - \text{tr}(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}) \mathbf{z}^T (\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \boldsymbol{\Sigma}^{-1}) \mathbf{z} \right. \\
 &\quad \left. - \text{tr}(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j}) \mathbf{z}^T (\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1}) \mathbf{z} + \mathbf{z}^T (\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1}) \mathbf{z} \mathbf{z}^T (\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \boldsymbol{\Sigma}^{-1}) \mathbf{z} \right]
 \end{aligned}$$

Agora pelo resultado 1 temos que

$$\begin{aligned}
 E \left[\text{tr}(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}) \mathbf{z}^T (\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \boldsymbol{\Sigma}^{-1}) \mathbf{z} \right] &= E \left[\text{tr}(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}) \mathbf{z}^T \boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \pm \frac{1}{2} \frac{\partial \boldsymbol{\Sigma}^T}{\partial \theta_j} \right) \boldsymbol{\Sigma}^{-1} \mathbf{z} \right] \\
 &= \frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}) E \left[\mathbf{z}^T \boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} + \frac{\partial \boldsymbol{\Sigma}^T}{\partial \theta_j} \right) \boldsymbol{\Sigma}^{-1} \mathbf{z} \right] \\
 &\quad + \frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}) E \left[\mathbf{z}^T \boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} - \frac{\partial \boldsymbol{\Sigma}^T}{\partial \theta_j} \right) \boldsymbol{\Sigma}^{-1} \mathbf{z} \right] \\
 &= \frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \right) \text{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} + \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}^T}{\partial \theta_j} \right)
 \end{aligned}$$

já que $E \left[\mathbf{z}^T \boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} - \frac{\partial \boldsymbol{\Sigma}^T}{\partial \theta_j} \right) \boldsymbol{\Sigma}^{-1} \mathbf{z} \right] = 0$, pois a esperança de uma forma quadrática com matriz antissimétrica é zero e $\left(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} - \frac{\partial \boldsymbol{\Sigma}^T}{\partial \theta_j} \right)$ é anti-simétrica por construção. De forma análoga,

$$E \left[\text{tr}(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j}) \mathbf{z}^T (\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1}) \mathbf{z} \right] = \frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} + \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}^T}{\partial \theta_i} \right) \text{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \right)$$

o que implica que

$$\begin{aligned}
 E[\dot{l}_i(\boldsymbol{\theta})\dot{l}_j(\boldsymbol{\theta})] &= \frac{1}{4} \left[\text{tr}(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}) \text{tr}(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j}) - \right. \\
 &\quad \frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} + \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}^T}{\partial \theta_i} \right) \text{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \right) - \\
 &\quad \left. \frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} + \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}^T}{\partial \theta_j} \right) \text{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \right) + \right. \\
 &\quad \left. E \left(\mathbf{z}^T (\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1}) \mathbf{z} \mathbf{z}^T (\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \boldsymbol{\Sigma}^{-1}) \mathbf{z} \right) \right]
 \end{aligned}$$

Por outro lado de acordo com o resultado 2 e já que $\boldsymbol{\mu} = \mathbf{0}$, temos que

$$\begin{aligned}
 & E \left[\mathbf{z}^T (\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1}) \mathbf{z} \mathbf{z}^T (\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \boldsymbol{\Sigma}^{-1}) \mathbf{z} \right] = \\
 & Cov \left[\mathbf{z}^T (\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1}) \mathbf{z}, \mathbf{z}^T (\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \boldsymbol{\Sigma}^{-1}) \mathbf{z} \right] = \\
 & \frac{1}{4} Cov \left[\mathbf{z}^T \left(\boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} + \frac{\partial \boldsymbol{\Sigma}^T}{\partial \theta_i} \right) \boldsymbol{\Sigma}^{-1} \right) \mathbf{z}, \mathbf{z}^T \left(\boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} + \frac{\partial \boldsymbol{\Sigma}^T}{\partial \theta_j} \right) \boldsymbol{\Sigma}^{-1} \right) \mathbf{z} \right] = \\
 & \frac{2}{4} tr \left(\boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} + \frac{\partial \boldsymbol{\Sigma}^T}{\partial \theta_i} \right) \boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} + \frac{\partial \boldsymbol{\Sigma}^T}{\partial \theta_j} \right) \right)
 \end{aligned}$$

Assim

$$\begin{aligned}
 E[\dot{l}_i(\boldsymbol{\theta}) \dot{l}_j(\boldsymbol{\theta})] &= \frac{1}{4} \left[tr(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}) tr(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j}) - \right. \\
 & \quad \frac{1}{2} tr \left(\boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} + \frac{\partial \boldsymbol{\Sigma}^T}{\partial \theta_i} \right) \right) tr \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \right) - \\
 & \quad \frac{1}{2} tr \left(\boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} + \frac{\partial \boldsymbol{\Sigma}^T}{\partial \theta_j} \right) \right) tr \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \right) + \\
 & \quad \left. \frac{1}{2} tr \left(\boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} + \frac{\partial \boldsymbol{\Sigma}^T}{\partial \theta_i} \right) \boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} + \frac{\partial \boldsymbol{\Sigma}^T}{\partial \theta_j} \right) \right) \right] \tag{C.1}
 \end{aligned}$$

Para o caso de simetria, C.1 torna-se

$$E[\dot{l}_i(\boldsymbol{\theta}) \dot{l}_j(\boldsymbol{\theta})] = \frac{1}{4} \left[tr \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \right) - tr(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}) tr(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j}) \right] \tag{C.2}$$

C.3 Informação de Fisher para σ_1^2 e σ_2^2

De acordo com os resultados do apêndice B os elementos da diagonal principal de $\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\sigma_1^2} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\sigma_2^2}$ são

$$\begin{pmatrix} C_{11}^*(\mathbf{h}) M(\mathbf{h}|\nu_1, a) + C_{12}^*(\mathbf{h}) \frac{\rho \sigma_2}{2 \sigma_1} M(\mathbf{h}|\nu_3, a) & C_{11}^*(\mathbf{h}) \frac{\rho \sigma_2}{2 \sigma_1} M(\mathbf{h}|\nu_3, a) \\ C_{21}^*(\mathbf{h}) M(\mathbf{h}|\nu_1, a) + C_{22}^*(\mathbf{h}) \frac{\rho \sigma_2}{2 \sigma_1} M(\mathbf{h}|\nu_3, a) & C_{21}^*(\mathbf{h}) \frac{\rho \sigma_2}{2 \sigma_1} M(\mathbf{h}|\nu_3, a) \end{pmatrix}$$

$$\begin{pmatrix} C_{12}^*(\mathbf{h}) \frac{\rho \sigma_1}{2 \sigma_2} M(\mathbf{h}|\nu_3, a) & C_{11}^*(\mathbf{h}) \frac{\rho \sigma_1}{2 \sigma_2} M(\mathbf{h}|\nu_3, a) + C_{12}^*(\mathbf{h}) M(\mathbf{h}|\nu_2, a) \\ C_{22}^*(\mathbf{h}) \frac{\rho \sigma_1}{2 \sigma_2} M(\mathbf{h}|\nu_3, a) & C_{21}^*(\mathbf{h}) \frac{\rho \sigma_1}{2 \sigma_2} M(\mathbf{h}|\nu_3, a) + C_{22}^*(\mathbf{h}) M(\mathbf{h}|\nu_2, a) \end{pmatrix}$$

$$\left[C_{11}^*(\mathbf{h})M(\mathbf{h}|\nu_1, a) + C_{12}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1}M(\mathbf{h}|\nu_3, a) \right] C_{12}^*(\mathbf{h})\frac{\rho\sigma_1}{2\sigma_2}M(\mathbf{h}|\nu_3, a) +$$

$$C_{11}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1}M(\mathbf{h}|\nu_3, a)C_{22}^*(\mathbf{h})\frac{\rho\sigma_1}{2\sigma_2}M(\mathbf{h}|\nu_3, a)$$

e

$$\left[C_{21}^*(\mathbf{h})M(\mathbf{h}|\nu_1, a) + C_{22}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1}M(\mathbf{h}|\nu_3, a) \right] * \\ \left[C_{11}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1}M(\mathbf{h}|\nu_3, a) + C_{12}^*(\mathbf{h})\frac{\rho\sigma_1}{2\sigma_2}M(\mathbf{h}|\nu_2, a) \right] + \\ C_{21}^*(\mathbf{h})\frac{\rho\sigma_1}{2\sigma_2}M(\mathbf{h}|\nu_3, a) + C_{22}^*(\mathbf{h})M(\mathbf{h}|\nu_2, a)$$

Ou seja, a esperança em \mathbf{Z} é duas vezes a soma dos traços desses dois elementos.

Ou seja:

$$2 * \left\{ \left[C_{11}^*(\mathbf{h})M(\mathbf{h}|\nu_1, a) + C_{12}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1}M(\mathbf{h}|\nu_3, a) \right] C_{12}^*(\mathbf{h})\frac{\rho\sigma_1}{2\sigma_2}M(\mathbf{h}|\nu_3, a) + \right. \\ \left. \frac{\rho^2}{4}C_{11}^*(\mathbf{h})M(\mathbf{h}|\nu_3, a)C_{22}^*(\mathbf{h})M(\mathbf{h}|\nu_3, a) + \right. \\ \left[C_{21}^*(\mathbf{h})M(\mathbf{h}|\nu_1, a) + C_{22}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1}M(\mathbf{h}|\nu_3, a) \right] * \\ \left[C_{11}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1}M(\mathbf{h}|\nu_3, a) + C_{12}^*(\mathbf{h})\frac{\rho\sigma_1}{2\sigma_2}M(\mathbf{h}|\nu_2, a) \right] + \\ \left. C_{21}^*(\mathbf{h})\frac{\rho\sigma_1}{2\sigma_2}M(\mathbf{h}|\nu_3, a) + C_{22}^*(\mathbf{h})M(\mathbf{h}|\nu_2, a) \right\}$$

Assim

$$\begin{aligned}
 E\left\{i_1(\boldsymbol{\theta})i_2(\boldsymbol{\theta})\right\} &= \frac{1}{4}\left\{\frac{-1}{\sigma_1^2\sigma_2^2}\left[tr\left(C_{11}^*(\mathbf{h})C_{11}(\mathbf{h})\right)+tr\left(C_{12}^*(\mathbf{h})C_{12}(\mathbf{h})\right)\right]^* \right. \\
 &\quad \left. \left[tr\left(C_{22}^*(\mathbf{h})C_{22}(\mathbf{h})\right)+tr\left(C_{12}^*(\mathbf{h})C_{12}(\mathbf{h})\right)\right] \right. \\
 &\quad + \\
 &\quad \left. 2tr\left(\left[C_{11}^*(\mathbf{h})M(\mathbf{h}|\nu_1,a)+C_{12}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1}M(\mathbf{h}|\nu_3,a)\right]C_{12}^*(\mathbf{h})\frac{\rho\sigma_1}{2\sigma_2}M(\mathbf{h}|\nu_3,a)+ \right. \right. \\
 &\quad \left. \left. \frac{\rho^2}{4}C_{11}^*(\mathbf{h})M(\mathbf{h}|\nu_3,a)C_{22}^*(\mathbf{h})M(\mathbf{h}|\nu_3,a)\right)\right) + \\
 &\quad \left. 2tr\left(\left[C_{21}^*(\mathbf{h})M(\mathbf{h}|\nu_1,a)+C_{22}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1}M(\mathbf{h}|\nu_3,a)\right]^* \right. \right. \\
 &\quad \left. \left. \left[C_{11}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1}M(\mathbf{h}|\nu_3,a)+C_{12}^*(\mathbf{h})\frac{\rho\sigma_1}{2\sigma_2}M(\mathbf{h}|\nu_2,a)\right]+ \right. \right. \\
 &\quad \left. \left. C_{21}^*(\mathbf{h})\frac{\rho\sigma_1}{2\sigma_2}M(\mathbf{h}|\nu_3,a)+C_{22}^*(\mathbf{h})M(\mathbf{h}|\nu_2,a)\right)\right)\left\} \tag{C.3}
 \end{aligned}$$

C.4 Informação de Fisher para σ_1^2 e ρ

De acordo com os resultados do apêndice B os elementos da diagonal principal de $\Sigma^{-1}\frac{\partial\Sigma}{\sigma_1^2}\Sigma^{-1}\frac{\partial\Sigma}{a}$ são

$$\begin{aligned}
 &\left[C_{11}^*M(\mathbf{h}|\nu_1,a)+\frac{\rho\sigma_2}{2\sigma_1}C_{12}^*M(\mathbf{h}|\nu_3,a)\right]\sigma_1\sigma_2C_{12}^*M(\mathbf{h}|\nu_3,a)+ \\
 &\quad \frac{\rho\sigma_2}{2\sigma_1}C_{11}^*M(\mathbf{h}|\nu_3,a)\sigma_2\sigma_1C_{22}^*M(\mathbf{h}|\nu_3,a)
 \end{aligned}$$

e

$$\begin{aligned}
 &\left[C_{21}^*M(\mathbf{h}|\nu_1,a)+\frac{\rho\sigma_2}{2\sigma_1}C_{22}^*M(\mathbf{h}|\nu_3,a)\right]\sigma_1\sigma_2C_{11}^*M(\mathbf{h}|\nu_3,a)+ \\
 &\quad \frac{\rho\sigma_2}{2\sigma_1}C_{21}^*M(\mathbf{h}|\nu_3,a)\sigma_2\sigma_1C_{21}^*M(\mathbf{h}|\nu_3,a)
 \end{aligned}$$

Portanto

$$\begin{aligned}
 E\left\{\dot{l}_1(\boldsymbol{\theta})\dot{l}_4(\boldsymbol{\theta})\right\} &= \frac{1}{4} \left\{ \frac{-1}{\sigma_1^2} \left[\text{tr} \left(C_{11}^*(\mathbf{h})C_{11}(\mathbf{h}) \right) + \text{tr} \left(C_{12}^*(\mathbf{h})C_{12}(\mathbf{h}) \right) \right] * \right. \\
 &\quad \left[2\sigma_1\sigma_2 \text{tr} \left(C_{12}^*(\mathbf{h})M(\mathbf{h}|\nu_3, a) \right) \right] + \\
 &\quad 2\text{tr} \left(\left[C_{11}^*M(\mathbf{h}|\nu_1, a) + \frac{\rho\sigma_2}{2\sigma_1}C_{12}^*M(\mathbf{h}|\nu_3, a) \right] \sigma_1\sigma_2 C_{12}^*M(\mathbf{h}|\nu_3, a) + \right. \\
 &\quad \left. \frac{\rho\sigma_2}{2\sigma_1}C_{11}^*M(\mathbf{h}|\nu_3, a)\sigma_2\sigma_1 C_{22}^*M(\mathbf{h}|\nu_3, a) \right) + \\
 &\quad 2\text{tr} \left(\left[C_{21}^*M(\mathbf{h}|\nu_1, a) + \frac{\rho\sigma_2}{2\sigma_1}C_{22}^*M(\mathbf{h}|\nu_3, a) \right] \sigma_1\sigma_2 C_{11}^*M(\mathbf{h}|\nu_3, a) + \right. \\
 &\quad \left. \left. \frac{\rho\sigma_2}{2\sigma_1}C_{21}^*M(\mathbf{h}|\nu_3, a)\sigma_2\sigma_1 C_{21}^*M(\mathbf{h}|\nu_3, a) \right) \right\} \tag{C.4}
 \end{aligned}$$

C.5 Informação de Fisher para σ_1^2 e a

De acordo com os resultados do apêndice B os elementos da diagonal principal de $\Sigma^{-1} \frac{\partial \Sigma}{\sigma_1^2} \Sigma^{-1} \frac{\partial \Sigma}{a}$ são

$$\begin{aligned}
 &\left(C_{11}^*M(\mathbf{h}|\nu_1, a) + \frac{\rho\sigma_2}{2\sigma_1}C_{12}^*M(\mathbf{h}|\nu_3, a) \right) \left(C_{11}^*\sigma_1^2 M'(\mathbf{h}|\nu_1, a) + \rho\sigma_2\sigma_1 C_{12}^*M'(\mathbf{h}|\nu_3, a) \right) + \\
 &\frac{\rho\sigma_2}{2\sigma_1}C_{11}^*M(\mathbf{h}|\nu_3, a) \left[C_{21}^*\sigma_1^2 M'(\mathbf{h}|\nu_1, a) + \rho\sigma_2\sigma_1 C_{22}^*M'(\mathbf{h}|\nu_3, a) \right]
 \end{aligned}$$

e

$$\begin{aligned}
 &\left(C_{21}^*M(\mathbf{h}|\nu_1, a) + \frac{\rho\sigma_2}{2\sigma_1}C_{22}^*M(\mathbf{h}|\nu_3, a) \right) \left(C_{11}^*\rho\sigma_2\sigma_1 M'(\mathbf{h}|\nu_3, a) + \sigma_2^2 C_{12}^*M'(\mathbf{h}|\nu_2, a) \right) + \\
 &\frac{\rho\sigma_2}{2\sigma_1}C_{21}^*M(\mathbf{h}|\nu_3, a) \left[C_{21}^*\rho\sigma_2\sigma_1 M'(\mathbf{h}|\nu_3, a) + \sigma_2^2 C_{22}^*M'(\mathbf{h}|\nu_2, a) \right]
 \end{aligned}$$

Portanto

$$\begin{aligned}
 E\left\{\dot{l}_1(\boldsymbol{\theta})\dot{l}_3(\boldsymbol{\theta})\right\} = & \frac{1}{4} \left\{ \frac{-1}{\sigma_1^2} \left[tr \left(C_{11}^*(\mathbf{h})C_{11}(\mathbf{h}) \right) + tr \left(C_{12}^*(\mathbf{h})C_{12}(\mathbf{h}) \right) \right] * \right. \\
 & \left[\sigma_1^2 tr \left(C_{11}^*(\mathbf{h})M'(\mathbf{h}|\nu_1, a) \right) + \sigma_2^2 tr \left(C_{22}^*(\mathbf{h})M'(\mathbf{h}|\nu_2, a) \right) + \right. \\
 & \left. \left. 2\rho\sigma_1\sigma_2 tr \left(C_{12}^*(\mathbf{h})M'(\mathbf{h}|\nu_3, a) \right) \right] \right. \\
 & + 2 * tr \left[\left(C_{11}^*M(\mathbf{h}|\nu_1, a) + \frac{\rho\sigma_2}{2\sigma_1}C_{12}^*M(\mathbf{h}|\nu_3, a) \right) \right. \\
 & \left. \left(C_{11}^*\sigma_1^2M'(\mathbf{h}|\nu_1, a) + \rho\sigma_2\sigma_1C_{12}^*M'(\mathbf{h}|\nu_3, a) \right) + \right. \\
 & \left. \left. \frac{\rho\sigma_2}{2\sigma_1}C_{11}^*M(\mathbf{h}|\nu_3, a) \left(C_{21}^*\sigma_1^2M'(\mathbf{h}|\nu_1, a) + \rho\sigma_2\sigma_1C_{22}^*M'(\mathbf{h}|\nu_3, a) \right) \right] + \right. \\
 & 2 * tr \left[\left(C_{21}^*M(\mathbf{h}|\nu_1, a) + \frac{\rho\sigma_2}{2\sigma_1}C_{22}^*M(\mathbf{h}|\nu_3, a) \right) \right. \\
 & \left. \left(C_{11}^*\rho\sigma_2\sigma_1M'(\mathbf{h}|\nu_3, a) + \sigma_2^2C_{12}^*M'(\mathbf{h}|\nu_2, a) \right) + \right. \\
 & \left. \left. \frac{\rho\sigma_2}{2\sigma_1}C_{21}^*M(\mathbf{h}|\nu_3, a) \left[C_{21}^*\rho\sigma_2\sigma_1M'(\mathbf{h}|\nu_3, a) + \sigma_2^2C_{22}^*M'(\mathbf{h}|\nu_2, a) \right] \right] \right\} \quad (C.5)
 \end{aligned}$$

C.6 Informação de Fisher para σ_2^2 e a

De acordo com os resultados do apêndice B os elementos da diagonal principal de $\Sigma^{-1}\frac{\partial\Sigma}{\sigma_2^2}\Sigma^{-1}\frac{\partial\Sigma}{a}$ são

$$\begin{aligned}
 & \frac{\rho\sigma_1}{2\sigma_2}C_{12}^*M(\mathbf{h}|\nu_3, a) \left[C_{11}^*\sigma_1^2M'(\mathbf{h}|\nu_1, a) + \rho\sigma_2\sigma_1C_{12}^*M'(\mathbf{h}|\nu_3, a) \right] + \\
 & \left(\frac{\rho\sigma_1}{2\sigma_2}C_{11}^*M(\mathbf{h}|\nu_3, a) + C_{12}^*M(\mathbf{h}|\nu_2, a) \right) \left(C_{21}^*\sigma_1^2M'(\mathbf{h}|\nu_1, a) + \rho\sigma_2\sigma_1C_{22}^*M'(\mathbf{h}|\nu_3, a) \right)
 \end{aligned}$$

e

$$\frac{\rho\sigma_1}{2\sigma_2}C_{22}^*M(\mathbf{h}|\nu_3, a) \left[C_{11}^*\rho\sigma_2\sigma_1M'(\mathbf{h}|\nu_3, a) + \sigma_2^2C_{12}^*M'(\mathbf{h}|\nu_2, a) \right] +$$

$$\left(\frac{\rho\sigma_1}{2\sigma_2}C_{21}^*M(\mathbf{h}|\nu_3, a) + C_{22}^*M(\mathbf{h}|\nu_2, a) \right) \left(C_{21}^*\rho\sigma_2\sigma_1M'(\mathbf{h}|\nu_3, a) + \sigma_2^2C_{22}^*M'(\mathbf{h}|\nu_3, a) \right)$$

Portanto

$$E\{i_2(\boldsymbol{\theta})i_3(\boldsymbol{\theta})\} = \frac{1}{4} \left\{ \frac{-1}{\sigma_2^2} \left[tr \left(C_{22}^*(\mathbf{h})C_{22}(\mathbf{h}) \right) + tr \left(C_{12}^*(\mathbf{h})C_{12}(\mathbf{h}) \right) \right] * \right.$$

$$\left[\sigma_1^2 tr \left(C_{11}^*(\mathbf{h})M'(\mathbf{h}|\nu_1, a) \right) + \sigma_2^2 tr \left(C_{22}^*(\mathbf{h})M'(\mathbf{h}|\nu_2, a) \right) + \right.$$

$$\left. \left. 2\rho\sigma_1\sigma_2 tr \left(C_{12}^*(\mathbf{h})M'(\mathbf{h}|\nu_3, a) \right) \right] + \right.$$

$$2 * tr \left[\frac{\rho\sigma_1}{2\sigma_2}C_{12}^*M(\mathbf{h}|\nu_3, a) \left(C_{11}^*\sigma_1^2M'(\mathbf{h}|\nu_1, a) + \rho\sigma_2\sigma_1C_{12}^*M'(\mathbf{h}|\nu_3, a) \right) + \right.$$

$$\left(\frac{\rho\sigma_1}{2\sigma_2}C_{11}^*M(\mathbf{h}|\nu_3, a) + C_{12}^*M(\mathbf{h}|\nu_2, a) \right) * \left.$$

$$\left(C_{21}^*\sigma_1^2M'(\mathbf{h}|\nu_1, a) + \rho\sigma_2\sigma_1C_{22}^*M'(\mathbf{h}|\nu_3, a) \right) \right] +$$

$$2 * tr \left[\frac{\rho\sigma_1}{2\sigma_2}C_{22}^*M(\mathbf{h}|\nu_3, a) \left(C_{11}^*\rho\sigma_2\sigma_1M'(\mathbf{h}|\nu_3, a) + \sigma_2^2C_{12}^*M'(\mathbf{h}|\nu_2, a) \right) + \right.$$

$$\left(\frac{\rho\sigma_1}{2\sigma_2}C_{21}^*M(\mathbf{h}|\nu_3, a) + C_{22}^*M(\mathbf{h}|\nu_2, a) \right) * \left.$$

$$\left(C_{21}^*\rho\sigma_2\sigma_1M'(\mathbf{h}|\nu_3, a) + \sigma_2^2C_{22}^*M'(\mathbf{h}|\nu_3, a) \right) \right]$$

$$\left. \right\}$$

(C.6)

C.7 Informação de Fisher para σ_2^2 e ρ

De acordo com os resultados do apêndice B os elementos da diagonal principal de $\Sigma^{-1}\frac{\partial\Sigma}{\sigma_2^2}\Sigma^{-1}\frac{\partial\Sigma}{a}$ são

$$\frac{\rho\sigma_1}{2\sigma_2}C_{12}^*M(\mathbf{h}|\nu_3, a)C_{12}^*\sigma_1\sigma_2\sigma_1^2M'(\mathbf{h}|\nu_3, a) + \sigma_1\sigma_2C_{11}^*M(\mathbf{h}|\nu_3, a)C_{22}^*M'(\mathbf{h}|\nu_3, a)$$

e

$$\begin{aligned} & \sigma_1 \sigma_2 C_{22}^* M(\mathbf{h}|\nu_3, a) \left[\frac{\rho \sigma_1}{2 \sigma_2} C_{11}^* M(\mathbf{h}|\nu_3, a) + C_{12}^* M(\mathbf{h}|\nu_2, a) \right] + \\ & \sigma_1 \sigma_2 C_{21}^* M(\mathbf{h}|\nu_3, a) \left[\frac{\rho \sigma_1}{2 \sigma_2} C_{21}^* M(\mathbf{h}|\nu_3, a) + C_{22}^* M(\mathbf{h}|\nu_2, a) \right] \end{aligned}$$

Portanto

$$\begin{aligned} E \left\{ \dot{l}_2(\boldsymbol{\theta}) \dot{l}_3(\boldsymbol{\theta}) \right\} &= \frac{1}{4} \left\{ 2 * tr \left(\left[\frac{\rho \sigma_1}{2 \sigma_2} C_{12}^* M(\mathbf{h}|\nu_3, a) C_{12}^* \sigma_1 \sigma_2 \sigma_1^2 M(\mathbf{h}|\nu_3, a) + \right. \right. \\ & \left. \left. \sigma_1 \sigma_2 C_{11}^* M(\mathbf{h}|\nu_3, a) C_{22}^* M(\mathbf{h}|\nu_3, a) \right] * \right. \\ & \left. \left[\sigma_1 \sigma_2 C_{22}^* M(\mathbf{h}|\nu_3, a) \left[\frac{\rho \sigma_1}{2 \sigma_2} C_{11}^* M(\mathbf{h}|\nu_3, a) + C_{12}^* M(\mathbf{h}|\nu_2, a) \right] \right] \right) + \\ & 2 * tr \left(\sigma_1 \sigma_2 C_{21}^* M(\mathbf{h}|\nu_3, a) \left[\frac{\rho \sigma_1}{2 \sigma_2} C_{21}^* M(\mathbf{h}|\nu_3, a) + C_{22}^* M(\mathbf{h}|\nu_2, a) \right] \right) - \\ & \left. tr \left(\frac{1}{\sigma_2^2} \left[y_2^T C_{22}(\mathbf{h}|\nu_1, a) + y_1^T C_{12}(\mathbf{h}|\nu_3, a) \right] y_2 \right) * tr \left(2 \sigma_1 \sigma_2 y_1^T M(\mathbf{h}|\nu_3, a) y_2 \right) \right\} \end{aligned} \tag{C.7}$$

C.8 Informação de Fisher para a e ρ

De acordo com os resultados do apêndice B os elementos da diagonal principal de $\Sigma^{-1} \frac{\partial \Sigma}{\sigma_2^2} \Sigma^{-1} \frac{\partial \Sigma}{a}$ são

$$\begin{aligned} & \sigma_1 \sigma_2 C_{12}^* M(\mathbf{h}|\nu_3, a) \left[\sigma_1^2 C_{11}^* M'(\mathbf{h}|\nu_1, a) + \rho \sigma_1 \sigma_2 C_{12}^* M(\mathbf{h}|\nu_3, a) \right] + \\ & \sigma_1 \sigma_2 C_{11}^* M(\mathbf{h}|\nu_3, a) \left[\sigma_1^2 C_{21}^* M'(\mathbf{h}|\nu_1, a) + \rho \sigma_1 \sigma_2 C_{22}^* M(\mathbf{h}|\nu_3, a) \right] \end{aligned}$$

e

$$\begin{aligned} & \sigma_1 \sigma_2 C_{22}^* M(\mathbf{h}|\nu_3, a) \left[\rho \sigma_1 2 \sigma_2 C_{11}^* M'(\mathbf{h}|\nu_1, a) + \sigma_2^2 C_{12}^* M(\mathbf{h}|\nu_3, a) \right] + \\ & \sigma_1 \sigma_2 C_{21}^* M(\mathbf{h}|\nu_3, a) \left[\rho \sigma_1 2 \sigma_2 C_{21}^* M'(\mathbf{h}|\nu_1, a) + \sigma_2^2 C_{12}^* M(\mathbf{h}|\nu_2, a) \right] \end{aligned}$$

Portanto

$$\begin{aligned} E \left\{ \dot{l}_2(\boldsymbol{\theta}) \dot{l}_3(\boldsymbol{\theta}) \right\} = & \frac{1}{4} \left\{ 2 * tr \left(\sigma_1 \sigma_2 C_{12}^* M(\mathbf{h}|\nu_3, a) \left[\sigma_1^2 C_{11}^* M'(\mathbf{h}|\nu_1, a) + \rho \sigma_1 \sigma_2 C_{12}^* M(\mathbf{h}|\nu_3, a) \right] + \right. \right. \\ & \left. \left. \sigma_1 \sigma_2 C_{11}^* M(\mathbf{h}|\nu_3, a) \left[\sigma_1^2 C_{21}^* M'(\mathbf{h}|\nu_1, a) + \rho \sigma_1 \sigma_2 C_{22}^* M(\mathbf{h}|\nu_3, a) \right] \right) + \right. \\ & 2 * tr \left(\sigma_1 \sigma_2 C_{22}^* M(\mathbf{h}|\nu_3, a) \left[\rho \sigma_1 2 \sigma_2 C_{11}^* M'(\mathbf{h}|\nu_1, a) + \sigma_2^2 C_{12}^* M(\mathbf{h}|\nu_3, a) \right] + \right. \\ & \left. \left. \sigma_1 \sigma_2 C_{21}^* M(\mathbf{h}|\nu_3, a) \left[\rho \sigma_1 2 \sigma_2 C_{21}^* M'(\mathbf{h}|\nu_1, a) + \sigma_2^2 C_{12}^* M(\mathbf{h}|\nu_2, a) \right] \right) - \right. \\ & \left. tr \left(\frac{1}{\sigma_2^2} \left[y_2^T C_{22}(\mathbf{h}|\nu_1, a) + y_1^T C_{12}(\mathbf{h}|\nu_3, a) \right] y_2 \right) * tr \left(2 \sigma_1 \sigma_2 y_1^T M(\mathbf{h}|\nu_3, a) y_2 \right) \right\} \end{aligned} \tag{C.8}$$

APÊNDICE D – Simulação de Processos Gaussianos Espaciais com Estrutura Matérn Bivariada

Simular um processo gaussiano com função de covariância Matérn Bivariada pode ser feito pela biblioteca `RandomFields` (SCHLATHER et al., 2015) no pacote estatístico R (R Core Team, 2020). Nele é preciso primeiro criar uma objeto que represente a estrutura de covariância desejada, por meio da função “`RMbiwm`”. Entretanto a parametrização da função de covariância do pacote difere um pouco daquela que foi adotada nesta discussão, como mostram as equações abaixo. Já o código logo após mostra como simular o processo exibido na Figura 12, que reflete a estrutura ilustrada em 11.

$$C_{ii}(\mathbf{h}) = \sigma_i^2 M(\mathbf{h}|a, \nu_i)$$

$$\underbrace{C_{ij}(\mathbf{h}) = \rho \sigma_i \sigma_j M(\mathbf{h}|a, \nu_i), i \neq j}_{\text{Versão Adotada no Texto}}$$

$$\underbrace{C_{ij}(\mathbf{h}) = c_{ij} M(\mathbf{h}|1/s_{ij}, \nu_i)}_{\text{Versão do RandomFields}}$$

```

1 library(RandomFields)
2 x <- y <- seq(-10, 10, 0.5)
3
4 modelo <- RMbiwm(nudiag=c(0.5, 1.5), nured=1, rhored=0.8,
5                   cdiag=c(1, 3), s=c(1, 1, 1))
6 biwm_sim <- RFsimulate(modelo, x, y)

```

Na rotina acima temos que o argumento

- `nudiag` representa o vetor (ν_1, ν_2) ;
- `cdiag` representa o vetor (σ_1^2, σ_2^2) ;
- `s` representa o vetor $(1/a_1, 1/a_2, 1/a_3)$. No caso do modelo bivariado reduzido `s` torna-se $(1/a, 1/a, 1/a)$;
- `nured` é dado por $\nu_{12} = \nu_{21} = 0.5(\nu_{11} + \nu_{22}) * \nu_{red}$ e então para que tenhamos o modelo reduzido é preciso que $\nu_{red} = 1$;

- rho representa ρ e precisa respeitar a restrição $|\rho| \leq \frac{(\nu_1\nu_2)^2}{(\nu_1 + \nu_2)/2}$

As simulações desta seção foram feitas por meio do pacote RandomFields (SCHLATHER et al., 2015), versão 3.3.8. Para a simulação foi criada uma malha regular de que vai de -10 até 10 . Nela, temos 201 pontos em cada coordenada, o que leva a um total de 40401 localizações. É importante ressaltar que todos os modelos aqui configuram funções de covariância válidas.

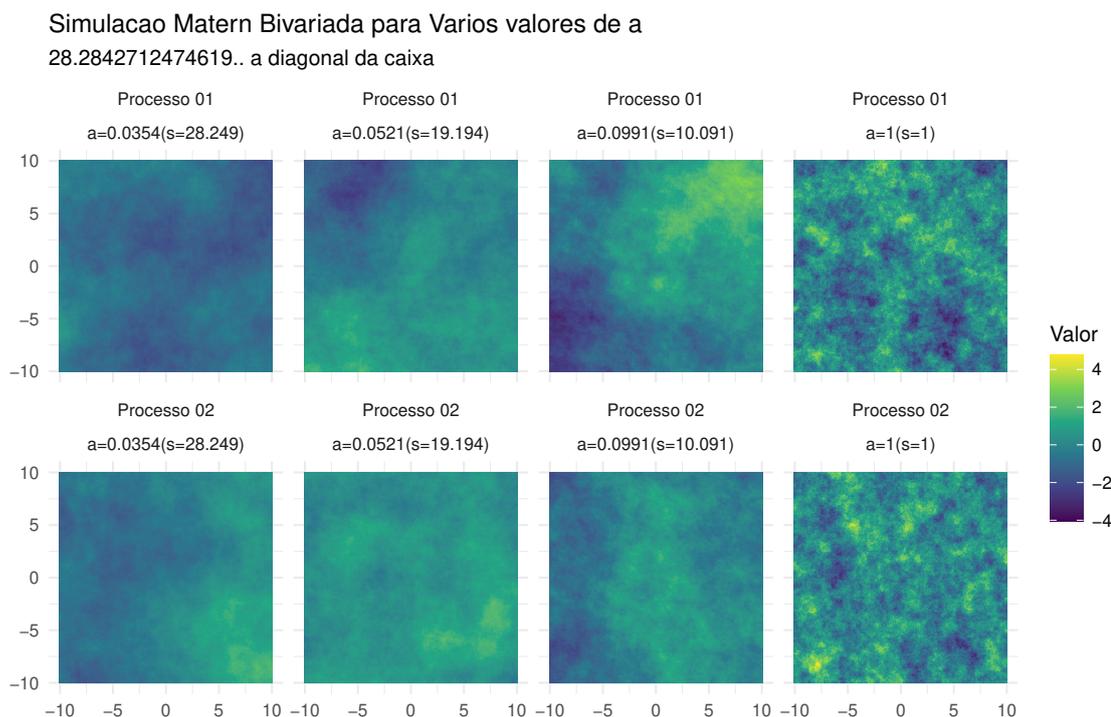


Figura 29 – Processo gaussiano simulado com $\sigma_1^2 = 1, \sigma_2^2 = 1, \rho = 0.2, \nu_1 = 0.5, \nu_2 = 0.5$ e $a = 1.0000, 0.0991, 0.0521, 0.0354$ o que levam a, respectivamente $s = 1.0000, 10.0948, 19.1895, 28.2843$, onde $s = 28.2843$ é a diagonal da caixa em que o processo foi simulado.

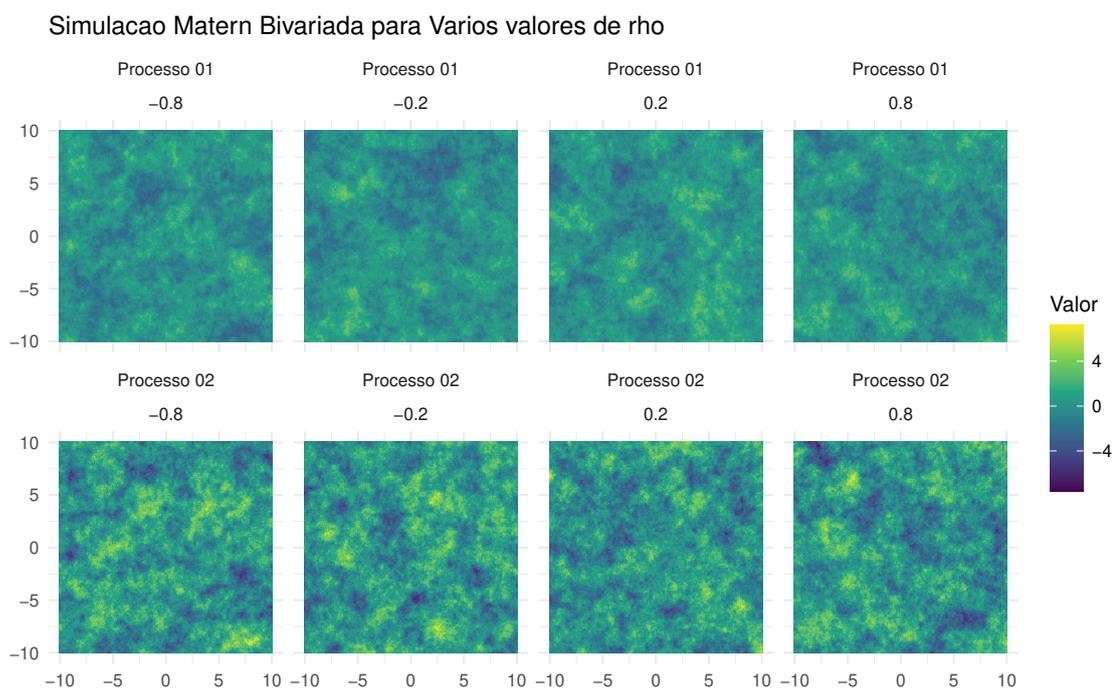


Figura 30 – Processo gaussiano simulado com $\sigma_1^2 = 1, \sigma_2^2 = 1, a = 1, \nu_1 = 0.5, \nu_2 = 0.5$ e $\rho = -0.8, -0.2, 0.2, 0.8$

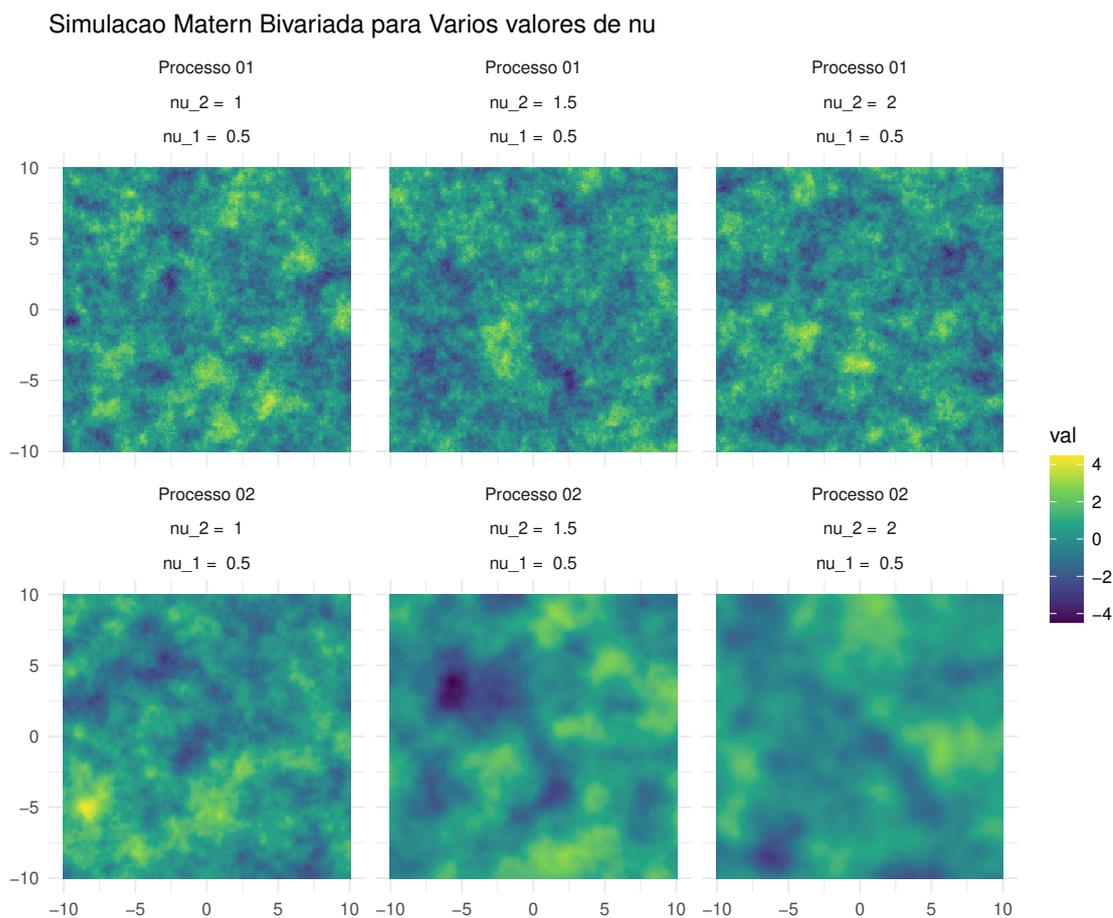


Figura 31 – Processo gaussiano simulado com $\sigma_1^2 = 1, \sigma_2^2 = 1, \rho = 0.2, a = 1, \nu_1 = 0.5$, e $\nu_2 = 1, 1.5, 2.0$

APÊNDICE E – Sistema Brasileiro de Classificação de Solos

De acordo com a página da EMBRAPA-Solos (EMBRAPA, 2020) o SiBCS é “Um sistema taxonômico de solos [...] que tem como finalidade classificar todos os solos existentes no Brasil.” O diagrama ternário da Figura 32 é a ilustração da atual classificação de solos, obtido pelo pacote `soiltext` (MOEYS, 2018), através da função SiBCS, e a Tabela 5 exibe a tradução de abreviações. É importante citar que as classificações média-arenosa e arenosa-média estão de acordo com o SiBCS, sendo que a primeira compreende a classe textural areia, enquanto que a segunda compreende a classe textural franco-arenosa (mais fina), com mais de 520gkg^{-1} de areia.

Tabela 5 – Abreviações do tipo de solo segundo o sistema SiBCS (EMBRAPA, 2020)

Abreviação	Classificação
MA	Muito Argilosa
A	Argilosa
S	Siltosa
MeAr	Média Arenosa
MeA	Argilosa
MeS	Média Siltosa
ArMe	Arenosa Média
MAr	Muito Arenosa

Triângulo textural: SiBCS 2013 (Embrapa)

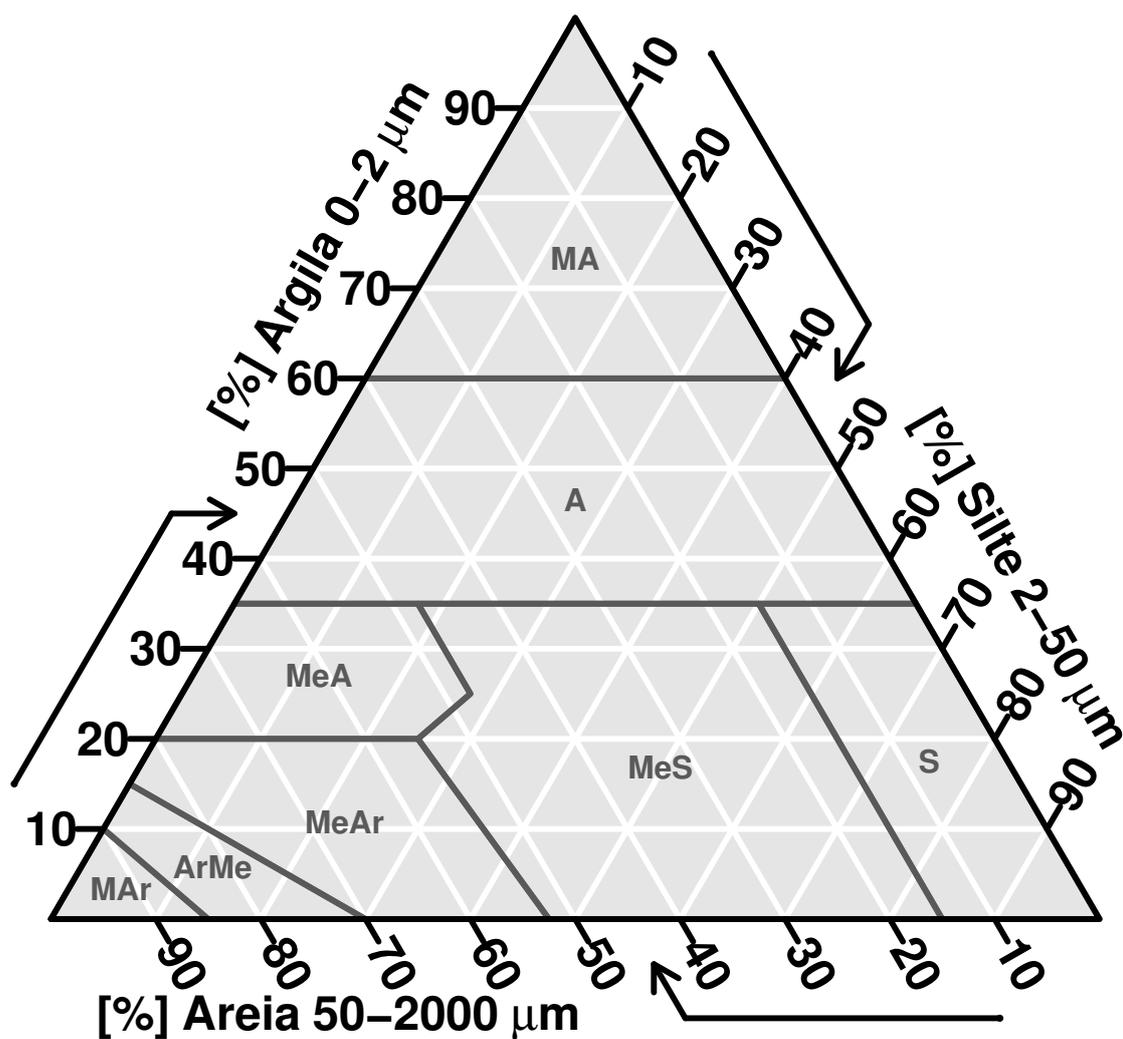


Figura 32 – Diagrama ternário representativo da classificação de solos da SiBCS. As setas indicam a direção de início das proporções (0% → 100%) para facilitar orientação.

APÊNDICE F – Resultados de Krigagem para dados CTB0809

Tabela 6 – Quantidades observadas e previstas de areia, junto com o respectivo erro padrão de predição. O maior erro padrão é de 0.24, enquanto que o menor é de 0.12

Id	Areia Observada	Areia Predita	Erro de Predição
1	0.16	0.11	0.14
2	0.10	0.15	0.18
3	0.50	0.19	0.22
4	0.15	0.12	0.15
5	0.05	0.10	0.13
6	0.38	0.13	0.16
7	0.21	0.13	0.16
8	0.11	0.22	0.24
9	0.09	0.17	0.20
10	0.10	0.18	0.21
11	0.14	0.11	0.15
12	0.08	0.09	0.12
13	0.11	0.10	0.13
14	0.06	0.14	0.17
15	0.15	0.16	0.18

Tabela 7 – Quantidades observadas e previstas de argila, junto com o respectivo erro padrão de predição. O maior erro padrão é de 0.37, enquanto que o menor é de 0.24

Id	Argila Observada	Argila Predita	Erro de Predição
1	0.56	0.72	0.28
2	0.80	0.71	0.28
3	0.44	0.69	0.29
4	0.49	0.77	0.24
5	0.85	0.77	0.24
6	0.44	0.68	0.30
7	0.71	0.71	0.29
8	0.85	0.64	0.31
9	0.56	0.59	0.34
10	0.83	0.66	0.31
11	0.74	0.76	0.25
12	0.85	0.55	0.37
13	0.78	0.77	0.25
14	0.88	0.63	0.33
15	0.69	0.60	0.33

Tabela 8 – Quantidades observadas e preditas de argila, junto com o respectivo erro padrão de predição. O maior erro padrão é de 0.36, enquanto que o menor é de 0.15.

Id	Silte Observada	Silte Predita	Erro de Predição
1	0.28	0.17	0.22
2	0.10	0.14	0.19
3	0.06	0.12	0.16
4	0.36	0.11	0.15
5	0.10	0.13	0.18
6	0.18	0.18	0.23
7	0.08	0.17	0.22
8	0.04	0.14	0.18
9	0.35	0.24	0.28
10	0.07	0.16	0.21
11	0.12	0.13	0.17
12	0.07	0.35	0.36
13	0.11	0.13	0.18
14	0.06	0.22	0.27
15	0.16	0.24	0.28

APÊNDICE G – Resultados dos K Vizinhos Mais Próximos

Como forma de comparar o desempenho da metodologia proposta neste texto com outras técnicas de modelagem populares, estimou-se também o desempenho do classificador K vizinhos mais próximos (HASTIE; TIBSHIRANI; FRIEDMAN, 2009) para classificações de solo descritas no apêndice E e encontradas em CTB0562. O procedimento aqui utilizado para estimar o número de K foi usar a Validação Cruzada (HASTIE; TIBSHIRANI; FRIEDMAN, 2009) “deixe um de fora” (*leave one out*) no conjunto de dados CTB0562 e avaliar seu desempenho no conjunto de dados CTB0809. O motivo de se criar um classificador sobre as etiquetas obtidas da classificação SIBCS, ao invés de um regressor sobre as composições, é devido à hipótese de que o classificador terá um desempenho melhor do que o regressor. A Tabela 9 a seguir mostra quais foram as acurácias obtidas a partir da validação cruzada. A maior acurácia encontrada foi de 59% o que forneceu como melhor classificador encontrado aquele que com $K = 5$. A matriz de confusão (HASTIE; TIBSHIRANI; FRIEDMAN, 2009) gerada a partir da previsão do classificador nos dados de CTB0809 e as categorias observadas está na Tabela 10. Nota-se que seu desempenho é muito pouco superior à krigagem pois o algoritmo KNN só acertou uma categoria a mais que a krigagem.

Tabela 9 – Taxas de acurácias nos dados CTB0809, para cada algoritmo KNN ajustado por validação cruzada “deixe um de fora” nos dados CTB0562. O melhor resultado foi obtido com $K = 5$, com acurácia de 59.1%.

K	Acurácia
1	54.80
2	52.70
3	51.60
4	55.90
5	59.10
6	54.80
7	57.00
8	52.70
9	49.50
10	51.60

Tabela 10 – Classificador criados com base em 5 vizinhos mais próximos contra classificação do solo observada. As linhas representam as categorias observadas em CTB0809, enquanto que as colunas representam as categorias previstas pelo algoritmo.

	A	MA	MeA	MeA/MeAr	MeAr	MeS
A	1	4	0	0	0	0
MA	0	10	0	0	0	0

Por fim as Tabelas 11 a 20 a seguir exibem as categorias previstas nos dados CTB0809, junto as probabilidades de pertencer à categoria prevista. Nota-se que nenhuma das categorias foi prevista com probabilidade igual a $1/8 = 0.125$, ou seja, não ocorreram empates na classificação.

Tabela 11 – Categorias observadas junto com classificações e probabilidades de classe previstas pelo algoritmo KNN com 1 vizinho mais próximo.

	Observado	Predito	Probabilidade
1	A	MA	1.00
2	MA	MA	1.00
3	A	MA	1.00
4	A	MA	1.00
5	MA	MA	1.00
6	A	A	1.00
7	MA	MA	1.00
8	MA	MA	1.00
9	A	MA	1.00
10	MA	MA	1.00
11	MA	MA	1.00
12	MA	A	1.00
13	MA	MA	1.00
14	MA	MA	1.00
15	MA	MeS	1.00

Tabela 12 – Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 2 vizinhos mais próximos.

	Observado	Predito	Probabilidade
1	A	MA	1.00
2	MA	MA	1.00
3	A	MA	1.00
4	A	MA	1.00
5	MA	MA	1.00
6	A	MA	0.50
7	MA	MA	1.00
8	MA	MA	0.50
9	A	A	0.50
10	MA	MA	1.00
11	MA	MA	1.00
12	MA	MA	0.50
13	MA	MA	1.00
14	MA	MA	1.00
15	MA	MA	0.50

Tabela 13 – Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 3 vizinhos mais próximos.

	Observado	Predito	Probabilidade
1	A	MA	1.00
2	MA	MA	1.00
3	A	MA	1.00
4	A	MA	1.00
5	MA	MA	1.00
6	A	MA	0.67
7	MA	MA	1.00
8	MA	A	0.33
9	A	A	0.67
10	MA	MA	1.00
11	MA	MA	1.00
12	MA	A	0.67
13	MA	MA	1.00
14	MA	MA	0.67
15	MA	MeS	0.67

Tabela 14 – Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 4 vizinhos mais próximos.

	Observado	Predito	Probabilidade
1	A	MA	1.00
2	MA	MA	1.00
3	A	MA	1.00
4	A	MA	1.00
5	MA	MA	1.00
6	A	MA	0.75
7	MA	MA	1.00
8	MA	MA	0.50
9	A	A	0.50
10	MA	MA	0.75
11	MA	MA	0.75
12	MA	MA	0.50
13	MA	MA	1.00
14	MA	MA	0.75
15	MA	MeS	0.50

Tabela 15 – Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 5 vizinhos mais próximos.

	Observado	Predito	Probabilidade
1	A	MA	1.00
2	MA	MA	0.80
3	A	MA	0.80
4	A	MA	0.80
5	MA	MA	0.80
6	A	MA	0.60
7	MA	MA	0.80
8	MA	MA	0.60
9	A	A	0.60
10	MA	MA	0.60
11	MA	MA	0.80
12	MA	MA	0.60
13	MA	MA	1.00
14	MA	MA	0.80
15	MA	MeS	0.40

Tabela 16 – Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 6 vizinhos mais próximos.

	Observado	Predito	Probabilidade
1	A	MA	0.83
2	MA	MA	0.83
3	A	MA	0.67
4	A	MA	0.83
5	MA	MA	0.67
6	A	MA	0.57
7	MA	MA	0.83
8	MA	MA	0.50
9	A	A	0.67
10	MA	MA	0.67
11	MA	MA	0.83
12	MA	MA	0.67
13	MA	MA	0.83
14	MA	MA	0.67
15	MA	MA	0.50

Tabela 17 – Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 7 vizinhos mais próximos.

	Observado	Predito	Probabilidade
1	A	MA	0.86
2	MA	MA	0.71
3	A	MA	0.57
4	A	MA	0.71
5	MA	MA	0.71
6	A	MA	0.57
7	MA	MA	0.86
8	MA	MA	0.43
9	A	A	0.71
10	MA	MA	0.57
11	MA	MA	0.86
12	MA	MA	0.57
13	MA	MA	0.86
14	MA	MA	0.75
15	MA	MA	0.43

Tabela 18 – Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 8 vizinhos mais próximos.

	Observado	Predito	Probabilidade
1	A	MA	0.88
2	MA	MA	0.75
3	A	MA	0.50
4	A	MA	0.62
5	MA	MA	0.62
6	A	MA	0.50
7	MA	MA	0.75
8	MA	MA	0.38
9	A	A	0.62
10	MA	A	0.50
11	MA	MA	0.75
12	MA	MA	0.62
13	MA	MA	0.88
14	MA	MA	0.75
15	MA	MA	0.50

Tabela 19 – Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 9 vizinhos mais próximos.

	Observado	Predito	Probabilidade
1	A	MA	0.78
2	MA	MA	0.67
3	A	MA	0.56
4	A	MA	0.67
5	MA	MA	0.67
6	A	MA	0.44
7	MA	MA	0.78
8	MA	MA	0.44
9	A	A	0.56
10	MA	MA	0.40
11	MA	MA	0.78
12	MA	MA	0.67
13	MA	MA	0.89
14	MA	MA	0.78
15	MA	MA	0.56

Tabela 20 – Categorias observadas junto com classificações e probabilidades de classe preditas pelo algoritmo KNN com 10 vizinhos mais próximos.

	Observado	Predito	Probabilidade
1	A	MA	0.64
2	MA	MA	0.60
3	A	MA	0.60
4	A	MA	0.60
5	MA	MA	0.60
6	A	MA	0.40
7	MA	MA	0.70
8	MA	A	0.40
9	A	A	0.50
10	MA	A	0.40
11	MA	MA	0.70
12	MA	MA	0.70
13	MA	MA	0.80
14	MA	MA	0.70
15	MA	MA	0.60

APÊNDICE H – Validação Cruzada Tipo “Deixe-um-de-fora” na Co-Krigagem em CTB0562

É possível realizar a Co-Krigagem com validação cruzada do tipo ‘deixe-um-de-fora’ para comparar seu desempenho com o classificador de “vizinhos-mais-próximos”. Para isto seguiu-se o procedimento descrito em (HASTIE; TIBSHIRANI; FRIEDMAN, 2009), de forma a consecutivamente remover uma única amostra de CTB0562 e, com o restante das observações, estimar os parâmetros pelo método descrito na Seção 3.2 e também realizar a predição daquela composição deixada de fora. Este procedimento fornece 93 predições relativas às 93 composições extraídas para a aplicação em CTB0562. Ao prosseguir desta maneira obtêm-se as classificações das predições de acordo com a Tabela 32, o que fornece a matriz de confusão (HASTIE; TIBSHIRANI; FRIEDMAN, 2009), dispostas na Tabela 21, resultante da comparação entre as etiquetas preditas pelo método de validação cruzada e aquelas observadas em CTB0562 estão. Nota-se que a acurácia é de aproximadamente 61,29%.

Tabela 21 – Validação Cruzada Tipo “Deixe-um-de-fora” na Co-Krigagem em CTB0562

	A	MA	MeA	MeA/MeAr	MeAr	MeS
A	12	7	1	1	4	6
MA	12	45	0	0	1	3
MeA	1	0	0	0	0	0