

Caroline Ferreira da Cruz Gomes

Modelagem de covariância espacial bivariada  
baseada em aproximações separáveis

Niterói - RJ, Brasil

16 de dezembro de 2022

**Caroline Ferreira da Cruz Gomes**

**Modelagem de covariância espacial  
bivariada baseada em aproximações  
separáveis**

**Trabalho de Conclusão de Curso**

Monografia apresentada para obtenção do grau de Bacharel em  
Estatística pela Universidade Federal Fluminense.

Orientador(a): Prof. Dr. Rafael Santos Erbisti

Niterói - RJ, Brasil

16 de dezembro de 2022

**Caroline Ferreira da Cruz Gomes**

**Modelagem de covariância espacial bivariada  
baseada em aproximações separáveis**

Monografia de Projeto Final de Graduação sob o título “*Modelagem de covariância espacial bivariada baseada em aproximações separáveis*”, defendida por Caroline Ferreira da Cruz Gomes e aprovada em 16 de dezembro de 2022, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

---

**Prof. Dr. Rafael Santos Erbisti**  
Departamento de Estatística – UFF

---

**Prof. Dra. Patrícia Lusié Velozo da Costa**  
Departamento de Estatística – UFF

---

**Prof. Dra. Viviana das Graças Ribeiro Lobo**  
Departamento de Métodos Estatísticos – IM – UFRJ

Niterói, 16 de dezembro de 2022

Ficha catalográfica automática - SDC/BIME  
Gerada com informações fornecidas pelo autor

G633m Gomes, Caroline Ferreira da Cruz  
Modelagem de covariância espacial bivariada baseada em aproximações separáveis / Caroline Ferreira da Cruz Gomes. - 2022.  
63 f.

Orientador: Rafael Santos Erbisti.  
Trabalho de Conclusão de Curso (graduação)-Universidade Federal Fluminense, Instituto de Matemática e Estatística, Niterói, 2022.

1. Estatística espacial. 2. Modelos não separáveis. 3. Aproximações separáveis. 4. Inferência Bayesiana. 5. Produção intelectual. I. Erbisti, Rafael Santos, orientador. II. Universidade Federal Fluminense. Instituto de Matemática e Estatística. III. Título.

CDD - XXX

# Resumo

Os recursos computacionais se mostram cada vez mais necessários na área da Estatística e para diversas ciências, porém lidar com dados de alta dimensionalidade ainda é um desafio. Em estatística espacial, na área da geoestatística, as análises de dados multivariados espaciais necessitam da especificação de uma função de covariância cruzada, a qual define a dependência entre as variáveis do vetor resposta e as localizações do domínio espacial. Entretanto, o custo computacional para realizar o procedimento de inferência e predição pode ser proibitivo. Como consequência, o uso de modelos complexos é inviável. Neste trabalho, será considerado o modelo de covariância não separável para dados multivariados, apresentado em [Porcu, Bevilacqua e Genton \(2016\)](#), e a forma de aproximação da matriz de covariância cheia a partir de duas matrizes separáveis de menor dimensão, apresentada em [Genton \(2007\)](#) e [Erbisti, Fonseca e Alves \(2018\)](#). O método de aproximação é aplicado apenas no cálculo da verossimilhança, mantendo a interpretação do modelo original. Um estudo de simulação foi conduzido para comparar a performance inferencial dos modelos aproximados, bem como o ganho em eficiência computacional. Diferentemente de [Erbisti, Fonseca e Alves \(2018\)](#), o objetivo é investigar o uso dessas aproximações separáveis para a matriz de covariância obtida a partir de funções que permitem diferentes alcances espaciais para cada componente do vetor resposta. O procedimento de estimação dos modelos foi feito sob o enfoque bayesiano e implementado no *software* R. Por fim, desenvolveu-se um pacote no R para o cálculo da função de verossimilhança utilizando as aproximações separáveis.

Palavras-chave: Estatística espacial. Modelos lineares. Modelos separáveis. Modelos não separáveis. Aproximações separáveis. Inferência Bayesiana.

# Agradecimentos

Para todos aqueles que de alguma forma me incentivaram a buscar conhecimento, um muito obrigada!

Agradeço aos meus pais, Lilian Cruz e Darlys Gomes, e a minha tia, Darlene Queiroz, por me inspirarem desde pequena a questionar, entender e construir ideias. O exemplo que recebi de vocês foi, e continua a ser, fundamental para o meu crescimento pessoal e intelectual.

Agradeço a Markson Arguello por toda paciência e amor, por acreditar em mim quando eu me questionava se seria capaz de absorver e aplicar os conhecimentos passados durante a graduação em Estatística.

Agradeço aos meus amigos, Ingrid Marrocos, Nicolly Amaral, Vitor Rodrigues, Igor Sampaio e Patrícia Mello. Vocês tornaram este desafio que é a graduação muito mais leve. Aprendi e sigo aprendendo com cada um de vocês.

E a Rafael Erbisti, meu orientador, obrigada por tanta paciência e dedicação. Sou grata por cada reunião que tivemos e que permitiu a aprender não somente o conteúdo discutido, e também a como apresentá-lo de forma mais acessível. Sua didática inspira!

Obrigada também à Thamires Louzada Marques por toda paciência e por toda ajuda no desenvolvimento deste trabalho, um trabalho que começou como um projeto de iniciação científica. Todas as reuniões que tivemos contribuíram demais para a realização deste trabalho, um grande muito obrigada!

Um muito obrigada aos docentes e funcionários do curso de Estatística na Universidade Federal Fluminense por contribuírem com a minha formação.

# Conteúdo

## Lista de Figuras

## Lista de Tabelas

<b>1 Introdução</b>	p. 11
1.1 Objetivos	p. 14
1.2 Organização	p. 14
<b>2 Materiais e Métodos</b>	p. 15
2.1 Modelo espacial univariado	p. 15
2.2 Modelo espacial multivariado	p. 17
2.2.1 Modelos separáveis	p. 19
2.2.2 Modelos não separáveis	p. 21
2.2.3 Aproximações separáveis	p. 22
2.2.4 Erro de aproximação	p. 23
2.3 Inferência bayesiana	p. 24
2.3.1 Simulações estocásticas feitas por Monte Carlo via Cadeias de Markov (MCMC)	p. 25
2.3.2 Algoritmo amostrador de Gibbs	p. 26
2.3.3 Método Metropolis-Hastings	p. 27
2.3.4 Distribuições condicionais completas	p. 28
<b>3 Análise dos resultados</b>	p. 31

3.1	Comportamento do erro de aproximação conforme aumento de não se-	
	parabilidade . . . . .	p. 31
3.2	Definição de cenários com diferentes alcances espaciais . . . . .	p. 32
3.3	Desempenho das aproximações separáveis no cálculo da função de veros-	
	similhança . . . . .	p. 34
3.4	Curvas de nível da verossimilhança . . . . .	p. 36
3.5	Estimação dos modelos . . . . .	p. 39
3.6	Pacote desenvolvido . . . . .	p. 46
4	Conclusões	p. 51
	Referências	p. 53
	Apêndice 1 – Curvas de correlação espacial	p. 55
	Apêndice 2 – Cadeias geradas pelo método MCMC	p. 58



# Lista de Figuras

1	Comportamento dos erros padronizados segundo grau de não separabilidade medido a partir do incremento nos alcances espaciais. . . . .	p. 32
2	Comportamento da função de correlação para os cenários descritos na Tabela 1 . . . . .	p. 33
3	Redução de custo computacional(em percentual) no cálculo da função de verossimilhança em diferentes cenários de separabilidade usando aproximações separáveis. . . . .	p. 35
4	Curvas de contorno no cenário 1. Linhas pretas: matriz de covariância cheia. Linhas cinzas: matriz de covariância aproximada. Linhas tracejadas: o verdadeiro valor de cada parâmetro. . . . .	p. 36
5	Curvas de contorno no cenário 2. Linhas pretas: matriz de covariância cheia. Linhas cinzas: matriz de covariância aproximada. Linhas tracejadas: o verdadeiro valor de cada parâmetro. . . . .	p. 37
6	Curvas de contorno no cenário 3. Linhas pretas: matriz de covariância cheia. Linhas cinzas: matriz de covariância aproximada. Linhas tracejadas: o verdadeiro valor de cada parâmetro. . . . .	p. 38
7	Curvas de contorno no cenário 4. Linhas pretas: matriz de covariância cheia. Linhas cinzas: matriz de covariância aproximada. Linhas tracejadas: o verdadeiro valor de cada parâmetro. . . . .	p. 39
8	Curvas das densidades das distribuições <i>a posteriori</i> dos parâmetros geradas com a estrutura não separável, a estrutura aproximada e as respectivas distribuições <i>a priori</i> no cenário 1 . . . . .	p. 40
9	Curvas das densidades das distribuições <i>a posteriori</i> dos parâmetros geradas com a estrutura não separável, a estrutura aproximada e as respectivas distribuições <i>a priori</i> no cenário 2 . . . . .	p. 42

10	Curvas das densidades das distribuições <i>a posteriori</i> dos parâmetros geradas com a estrutura não separável, a estrutura aproximada e as respectivas distribuições <i>a priori</i> no cenário 3	p. 44
11	Curvas das densidades das distribuições <i>a posteriori</i> dos parâmetros geradas com a estrutura não separável, a estrutura aproximada e as respectivas distribuições <i>a priori</i> no cenário 4	p. 45
12	Descrição do pacote <code>cov.approx</code> no <i>software</i> R e suas respectivas funções, variáveis e autores	p. 46
13	Função <code>dist.calc()</code> do pacote <code>cov.approx</code> no <i>software</i> R e suas respectivas variáveis e autores	p. 47
14	Função <code>dmvnorm.chol()</code> do pacote <code>cov.approx</code> no <i>software</i> R e suas respectivas variáveis e autores	p. 48
15	Função <code>likelihood.approx()</code> do pacote <code>cov.approx</code> no <i>software</i> R e suas respectivas variáveis e autores	p. 49
16	Função <code>sub.matrix()</code> do pacote <code>cov.approx</code> no <i>software</i> R e suas respectivas variáveis e autores	p. 50
17	Comportamento das funções de correlação Cauchy e Exponencial para o cenário 1	p. 55
18	Comportamento das funções de correlação Cauchy e Exponencial para o cenário 2	p. 56
19	Comportamento das funções de correlação Cauchy e Exponencial para o cenário 3	p. 56
20	Comportamento das funções de correlação Cauchy e Exponencial para o cenário 4	p. 57
21	Cadeias geradas pelo método MCMC para cada parâmetro no cenário 1 para a estrutura cheia	p. 58
22	Cadeias geradas pelo método MCMC para cada parâmetro no cenário 1 para a estrutura aproximada	p. 59
23	Cadeias geradas pelo método MCMC para cada parâmetro no cenário 1 para a estrutura cheia	p. 59

24	Cadeias geradas pelo método MCMC para cada parâmetro no cenário 1	
	para a estrutura aproximada . . . . .	p. 60
25	Cadeias geradas pelo método MCMC para cada parâmetro no cenário 1	
	para a estrutura cheia . . . . .	p. 60
26	Cadeias geradas pelo método MCMC para cada parâmetro no cenário 1	
	para a estrutura aproximada . . . . .	p. 61
27	Cadeias geradas pelo método MCMC para cada parâmetro no cenário 1	
	para a estrutura cheia . . . . .	p. 61
28	Cadeias geradas pelo método MCMC para cada parâmetro no cenário 1	
	para a estrutura aproximada . . . . .	p. 62

# Lista de Tabelas

1	Possíveis Cenários operativos. . . . .	p. 32
2	Tempo necessário (em segundos) para calcular a função de verossimi- lhança da matriz de covariância cheia e na de estrutura aproximada. (Intel(R) Core(TM) i5-8250U, 1.60GHz, 8GB RAM). . . . .	p. 34
3	Resumo da distribuição a <i>posteriori</i> para os parâmetros do cenário 1, considerando a estrutura de modelos não separável e a estrutura aproxi- mada . . . . .	p. 41
4	Resumo da distribuição a <i>posteriori</i> para os parâmetros do cenário 2, considerando a estrutura de modelos não separável e a estrutura aproxi- mada . . . . .	p. 43
5	Resumo da distribuição a <i>posteriori</i> para os parâmetros do cenário 3, considerando a estrutura de modelos não separável e a estrutura aproxi- mada . . . . .	p. 44
6	Resumo da distribuição a <i>posteriori</i> para os parâmetros do cenário 4, considerando a estrutura de modelos não separável e a estrutura aproxi- mada . . . . .	p. 45

# 1 Introdução

Com o avanço da obtenção e dados georreferenciados, análises e modelos estatísticos podem agregar mais informações e trazer melhores resultados ao considerar o arranjo espacial dos dados. Algumas áreas como a agricultura, astronomia, epidemiologia, ciência ambiental e a ciência do clima atualmente se beneficiam do uso da estatística espacial. Na área da ciência ambiental, [Mousavifazl, Alizadh e Ghahraman \(2013\)](#) apresentaram um modelo estatístico espacial capaz de prever a qualidade da água utilizando dados georreferenciados e variáveis como os níveis de nitrato, sódio e da condutividade elétrica presente nos lençóis freáticos na cidade de Mashhad, no Irã. O estudo de [Mousavifazl, Alizadh e Ghahraman \(2013\)](#) verificou uma forte correlação espacial entre as variáveis, tornando assim mais favorável a aplicação da estatística espacial em relação aos métodos clássicos da estatística.

Dados espaciais podem ser classificados de acordo com sua natureza. Há um consenso na área de estatística espacial proposto por [Cressie \(1993\)](#) que decompõe a estatística espacial em três grandes áreas: dados de área, geoestatística e processos pontuais. Este trabalho utiliza dados georreferenciados, logo ao assumir que o espaço geográfico contém infinitos pontos, a geoestatística irá trabalhar com uma realização parcial desse processo, onde as amostras irão conter localizações fixas ao longo de uma região de interesse, possivelmente em diferentes instantes do tempo de forma que os pontos que não foram observados podem ser obtidos através de técnicas como a interpolação.

Ao analisar dados desse tipo, espera-se que as medidas feitas em localizações próximas entre si sejam altamente correlacionadas, enquanto que para localizações separadas por grandes distâncias ocorra um comportamento mais independente. De fato, a distância entre as localizações é um conceito muito presente e de grande relevância na estatística espacial.

Na outra decomposição, intitulada dados de área, as observações são provenientes de uma coleção numerável de localizações no espaço, como por exemplo a coleta de dados

de satélites, onde cada *pixel* das imagens obtidas serão tratadas como dados espaciais. A outra área de estatística espacial, denominada processos pontuais, possui como variável mais relevante a localização da ocorrência de um dado evento, cujo objetivo não será necessariamente estudar uma variável ao longo de uma região no espaço e sim a região em si, segundo [Cressie \(1993\)](#).

A aplicação de modelos espaciais tem crescido substancialmente em diversas áreas. Frequentemente, observa-se mais de uma variável, ou seja, há um vetor de respostas em diversas localizações do espaço. De forma geral, dada a localização geográfica, é possível obter informações sobre diferentes variáveis, por exemplo, estações de monitoramento do ar podem medir níveis de diversos poluentes, temperatura, umidade do ar, velocidade do vento, etc. Nesse sentido, o desafio está não apenas em entender o comportamento dessas variáveis ao longo do espaço, mas também de considerar a associação entre elas. Com efeito, avaliar de forma conjunta variáveis que são correlacionadas entre si promove resultados melhores tanto em termos de ajuste quanto em predição do que ao se trabalhar com modelos independentes. Quando duas ou mais variáveis estão sendo analisadas de forma conjunta, as análises e modelos desenvolvidos são multivariados.

Em modelos espaciais multivariados para dados georreferenciados, a função de covariância cruzada é responsável por determinar e descrever a dependência espacial entre componentes do vetor resposta ao longo do espaço. As funções de covariância devem descrever de forma conjunta as correlações espaciais para serem consideradas adequadas, como descrito em [Schmidt e Sansó \(2006\)](#). Há, na literatura, diversas abordagens capazes de modelar dados espaciais multivariados. [Banerjee, Carlin e Gelfand \(2004\)](#) apresentam alguns métodos de modelagem para esse tipo de dados. Um deles é o modelo separável.

Ao assumir separabilidade na estrutura de covariância, se impõe às variáveis do vetor resposta o mesmo padrão de comportamento no espaço, ou seja, variáveis diferentes acabam tendo a mesma estrutura de dependência no espaço. De certa forma, essa característica é uma restrição no procedimento de modelagem, uma vez que é possível variáveis como, por exemplo, temperatura e monóxido de carbono apresentarem comportamentos diferentes ao longo do espaço. Por outro lado, assumir separabilidade faz com que a matriz de covariância possa ser escrita a partir do produto de Kronecker de duas matrizes de menor dimensão: uma com a dependência das observações no espaço e outra com a dependência entre as variáveis. Modelos separáveis são muito vantajosos em termos computacionais por ser possível separar a matriz de covariância em duas matrizes de menor dimensão, porém são limitados por utilizarem estruturas de dependência espacial iguais,

o que, frequentemente, não retrata a realidade.

Na literatura diversos autores buscaram relaxar a suposição de separabilidade das funções de covariância cruzada, para assim suprir suas limitações e tornar o processo mais realista. Historicamente, [Goulard e Voltz \(1992\)](#) e [Wackernagel \(2003\)](#) foram pioneiros em modelos lineares que buscavam decompor o processo espacial multivariado em conjuntos de componentes espaciais que não são correlacionadas, de forma que a função de covariância cruzada pode ser construída com diferentes alcances espaciais para cada componente do vetor resposta. Trabalhar com alcances espaciais distintos para cada variável significa que o modelo permite estruturas espaciais diferentes para cada componente do vetor resposta, diferentemente do modelo separável.

A construção de modelos de estrutura não separável pode ser feita a partir de misturas de funções separáveis. Por exemplo, [Ma \(2002\)](#) e [Ma \(2003\)](#) mostraram ser possível obter modelos em convoluções de funções puramente temporais e puramente espaciais, tendo o espaço-tempo por abordagem. Essa convolução ou mistura de funções traz como benefício modelos de covariância não separáveis em uma variedade suficiente. Outros exemplos de classes de funções de covariância cruzada baseadas em misturas podem ser encontrados em [Porcu e Zastavnyi \(2011\)](#), [Fonseca e Steel \(2011\)](#) e [Erbisti, Fonseca e Alves \(2018\)](#). Mais recentemente [Porcu, Bevilacqua e Genton \(2016\)](#) mostraram ser possível escolher intuitivamente os parâmetros que definem as funções de covariância e covariância cruzada através de uma família de covariância paramétrica flexível que pode ser utilizada em análise de dados espaciais multivariados na esfera.

O modelo espacial não separável apresenta, em muitos aspectos, melhor desempenho em comparação ao modelo separável. Entretanto, a matriz de covariância em modelos espaciais multivariados pode ter alta dimensionalidade, uma vez que sua dimensão depende da quantidade de localizações e de variáveis observadas. Dessa forma, fazer inferência sobre os parâmetros desconhecidos pode ser um procedimento não trivial em termos computacionais. Se um processo gaussiano é assumido, por exemplo, calcular a inversa e o determinante dessa matriz de covariância pode ser proibitivo.

Uma possível solução para este problema é baseada em aproximações separáveis da matriz de covariância, decompondo-a em duas matrizes separáveis de menor dimensão, como proposto em [Genton \(2007\)](#) para o caso espaço-temporal e [Erbisti, Fonseca e Alves \(2018\)](#), para o caso multivariado espacial. A solução por aproximações pode ser aplicada à função de verossimilhança e assim, o procedimento de inferência será feito de forma mais rápida, sem perder a interpretação e flexibilidade do modelo multivariado não separável.

Este trabalho foi desenvolvido também como um projeto de iniciação científica, e financiado pela Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

## 1.1 Objetivos

A proposta deste trabalho é avaliar e quantificar o ganho computacional proporcionado pelas aproximações separáveis da matriz de covariância cheia ao utilizar a função de covariância não separável. Diferente do proposto em [Erbisti, Fonseca e Alves \(2018\)](#), este trabalho utilizará funções de covariância cruzada com alcances espaciais diferentes, propostas em [Porcu, Bevilacqua e Genton \(2016\)](#). Os objetivos específicos são:

- avaliar e comparar o tempo computacional para o cálculo da função de verossimilhança dos modelos não separáveis completo e aproximado utilizando a função de correlação Cauchyna estrutura da função de covariância;
- avaliar a função de verossimilhança completa e aproximada a partir das curvas de nível;
- comparar os modelos completo e aproximado em termos de ajuste;
- construir um pacote no *software* R para o cálculo da função de verossimilhança utilizando as aproximações separáveis.

## 1.2 Organização

Neste trabalho, o Capítulo [2](#) apresenta os métodos a serem usados ao longo de duas seções. Na Seção [2.1](#) é apresentado o modelo espacial para uma única variável. Já a Seção [2.2](#) traz os modelos espaciais utilizados quando há mais uma variável a ser estudada. Por fim, o Capítulo [2](#) se encerra na Seção [2.3](#) onde são apresentados alguns conceitos de inferência bayesiana e de simulações estocásticas feitas por Monte Carlo via Cadeias de Markov (MCMC). O Capítulo [3](#) traz os cenários propostos para as análises realizadas, assim como a estimação do modelo e o pacote desenvolvido no *software* R. O Capítulo [4](#) exhibe a conclusão e algumas discussões sobre este trabalho.



## 2 Materiais e Métodos

Este capítulo se propõe a apresentar as técnicas a serem usadas para as análises feitas neste trabalho, as quais se baseiam em modelos espaciais multivariados separáveis e não separáveis e também em aproximações separáveis da matriz de covariância. A Seção 2.1 apresenta o caso mais simples de modelo, o modelo univariado onde apenas uma variável é analisada.

A Seção 2.2 apresenta o modelo multivariado, onde neste trabalho será apresentado o caso bivariado, nesta seção as Subseções 2.2.1, 2.2.2, 2.2.3, 2.2.4 apresentam, respectivamente, os modelos separáveis, os modelos não separáveis, as aproximações separáveis, e por fim no erro de aproximação. A Seção 2.3 traz os métodos de inferência utilizados, seguindo o ponto de vista bayesiano.

### 2.1 Modelo espacial univariado

Um campo aleatório é adotado quando se deseja descrever, analisar e gerar previsões sobre um evento  $Y$  de interesse com coordenadas dadas pelo vetor  $\mathbf{s}$ , cuja região espacial contém  $n$  localizações, e pode ser descrito como  $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ ,  $D \subset \mathbb{R}^d$ , a considerar que  $d \geq 1$ , sempre será uma variável dimensional positiva. Usualmente, extrai-se dos pontos georreferenciados as informações de latitude e longitude, gerando uma análise com  $d = 2$  dimensões. Cressie (1993) define a ligação entre os dados georreferenciados e um campo aleatório (ou processo estocástico), como:

$$\mathbf{Y} = [Y(\mathbf{s}_1), Y(\mathbf{s}_2), Y(\mathbf{s}_3), \dots, Y(\mathbf{s}_n)]^T. \quad (2.1)$$

As distribuições finito-dimensionais que caracterizam os pontos georreferenciados  $\mathbf{s}_i$  com  $i = 1, \dots, n$ ,  $\mathbf{s}_i \in D$ , descrevem um campo aleatório. Para o caso em que essas distribuições finito-dimensionais seguem distribuição Normal, trata-se de um campo aleatório Gaussiano.

**Definição 2.1** Um **campo aleatório gaussiano** é um campo aleatório onde todas as distribuições finito dimensionais, por exemplo,  $F(\mathbf{s}_1, \dots, \mathbf{s}_n)$  são distribuições normais multivariadas, para quaisquer escolhas de  $n$  e localizações  $\mathbf{s}_1, \dots, \mathbf{s}_n$ .

Lidar com dados espaciais traz a necessidade de detalhar dois conceitos importantes: o primeiro diz respeito à **estacionariedade** do processo, definido em Schmidt e Sansó (2006) como sendo útil para simplificar o procedimento de inferência em processos estocásticos e exige que as funções de covariância e média não se alterem sob translações, onde a distribuição de probabilidade de dois ou mais eventos ocorrem de forma simultânea, incondicional e permanece a mesma conforme o deslocamento do tempo.

**Definição 2.2** Um campo aleatório é **estritamente estacionário** se para qualquer coleção finita de estações  $\mathbf{s}_1, \dots, \mathbf{s}_n$  e qualquer  $\mathbf{u} \in G$ , a distribuição conjunta de  $(Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))$  é a mesma do que a de  $(Y(\mathbf{s}_1 + \mathbf{u}), \dots, Y(\mathbf{s}_n + \mathbf{u}))$ .

Outro conceito relevante definido em Schmidt e Sansó (2006) é o de isotropia, traz a ideia de que a dependência entre as localidades, representadas pelas coordenadas  $\mathbf{s}$  e  $\mathbf{s}'$  é dada apenas pela distância euclidiana entre elas definida por  $\mathbf{h} = \mathbf{s} - \mathbf{s}'$  de forma que  $h = \|\mathbf{h}\| = \|\mathbf{s} - \mathbf{s}'\|$ .

Este trabalho irá se basear em um campo aleatório gaussiano e, conforme apresentado em Schmidt e Sansó (2006), para que um processo gaussiano seja especificado por completo, é necessário especificar apenas a média e a função de covariância. Assumindo um processo estacionário e isotrópico, tem-se

$$E(Y(\mathbf{s})) = \mu(\mathbf{s}) = \mu, \quad \forall \mathbf{s} \in D. \quad (2.2)$$

e a função de covariância

$$\text{Cov}(Y(\mathbf{s}), Y(\mathbf{s}')) = C(\|\mathbf{s} - \mathbf{s}'\|) = C(h) \quad \forall \mathbf{s}, \mathbf{s}' \in D. \quad (2.3)$$

Nesse sentido, é possível definir o modelo espacial univariado,

$$\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (2.4)$$

sendo  $\boldsymbol{\mu}$  a média do processo representado como um vetor que tem como dimensões as  $n$  localizações, e também tem-se que  $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ , e  $\mathbf{X}$  como a matriz de variáveis explicativas e  $\boldsymbol{\beta}$  o vetor de parâmetros associados às regressoras. Adota-se, neste caso,  $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$ ,  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$  o erro aleatório.

Cada elemento da matriz de covariância  $\boldsymbol{\Sigma}$  será definido pela função de covariância,

dada por  $\Sigma = \sigma^2 \rho(h, \phi)$ , com  $\sigma^2$  sendo a variância do processo espacial e  $\rho(h, \phi)$  uma função de correlação espacial válida.

Na literatura, há diversas funções de correlação espacial, dentre elas, apresentam-se abaixo duas funções paramétricas:

1. Função de correlação Cauchy

$$\rho(h, \phi) = \left(1 + \frac{h}{\phi}\right)^{-\alpha}, \quad (2.5)$$

onde  $h$  é a distância euclidiana,  $\phi$  é o parâmetro de alcance espacial que mede a distância no espaço na qual ainda há dependência espacial,  $\phi > 0$ , e  $\alpha$  é um parâmetro de suavização, com  $\alpha > 0$ .

2. Função exponencial

$$\rho(h, \phi) = \exp\left\{\frac{-h}{\phi}\right\}, \quad (2.6)$$

onde  $h$  é a distância euclidiana e  $\phi$  é o parâmetro de alcance espacial.

Neste trabalho será utilizada a função de covariância espacial Cauchy, uma vez que a classe Cauchy possui um decaimento mais lento e permite longos alcances, isto é, em casos em que as variáveis têm uma dependência espacial que se estica muito ao longo do espaço a Cauchy vai conseguir se adaptar melhor do que a função Exponencial.

Note que a escolha não é baseada em uma restrição e sim nas características de longo alcance da Cauchy, que são amplamente estudadas neste trabalho, mas não existe impedimento para usar a função de covariância Exponencial. Para fins de comparação, as figuras presentes no Capítulo [I](#) do Apêndice traz as curvas de correlação com a função Cauchy e com a Exponencial.

## 2.2 Modelo espacial multivariado

No caso multivariado também será assumido um processo gaussiano, como visto na Equação [\(2.1\)](#), porém, diferentemente do caso univariado, tem-se  $Y_k$  com  $k = 1, \dots, p$  variáveis que serão avaliadas de forma simultânea na mesma localização  $\mathbf{s} \in D$  e para este trabalho, será adotado  $D \subset \mathfrak{R}^2$ . Desta forma, agora,  $\mathbf{Y}(\mathbf{s})$  possui dimensão  $1 \times p$ , sendo  $\mathbf{s}$  a localização no espaço  $D$ . O desafio desta análise é compreender a dependência entre as variáveis mensuradas em cada localização no espaço e entre as localizações. Assim, se  $\mathbf{Y}$  segue um processo gaussiano, para especificá-lo por completo é necessário definir as

funções de média e covariância cruzada. A função de covariância cruzada é responsável por descrever a dependência entre os componentes do vetor resposta entre si e ao longo do espaço. Entretanto, essas funções de covariância cruzada não são simples de serem definidas, uma vez que para qualquer localização, a matriz de covariância gerada deve ser positiva definida (BANERJEE; CARLIN; GELFAND, 2004).

Seguindo Wackernagel (2003), as funções de covariância cruzada  $C_{kj}(h)$ , do conjunto de  $p$  variáveis aleatórias  $Y_k(\mathbf{s})$ ,  $\mathbf{s} \in D$ , considerando um processo estacionário e isotrópico, podem ser definidas da seguinte forma: Se

$$E(Y_k(\mathbf{s})) = m_k, \quad k = 1, \dots, p. \quad (2.7)$$

então, a estrutura de covariância cruzada é dada por

$$E[(Y_k(\mathbf{s}) - m_k(\mathbf{s}))(Y_j(\mathbf{s}') - m_j(\mathbf{s}'))] = C_{kj}(h), \quad k, j = 1, \dots, p, \quad \forall \mathbf{s}, \mathbf{s}' \in D \text{ e } h = \|\mathbf{s} - \mathbf{s}'\|. \quad (2.7)$$

onde a média de cada variável  $Y_k(\mathbf{s})$ , em cada localização do domínio  $\mathbf{s}$ , é igual à constante  $m_k$ .

Com o objetivo de trazer a dependência entre as observações medidas nas  $n$  localizações para cada uma das duas variáveis a serem estudadas, constrói-se uma matriz de covariância  $\Sigma$  com três blocos: o primeiro bloco traz a dependência espacial para a primeira variável, o segundo bloco a dependência espacial para a segunda variável e um terceiro bloco, na diagonal da matriz, onde a primeira e a segunda variável são analisadas de forma conjunta.

Este trabalho não irá definir quais são as duas variáveis a serem usadas, uma vez que o intuito é que os resultados obtidos sejam válidos para quaisquer variáveis que possuem dependência espacial.

Cada elemento da matriz de covariância  $\Sigma$  será definido pela função de covariância, dada por  $\Sigma = \sigma^2 \rho(h, \phi)$ , com  $\sigma^2$  sendo a variância do processo espacial e  $\rho(h, \phi)$  uma função de correlação espacial válida.

A função de covariância cruzada é apresentada para qualquer  $p \geq 2$ , no entanto, neste trabalho, será avaliado o caso de  $p = 2$  variáveis resposta. Para o caso particular bivariado, ou seja, assumindo que  $\mathbf{Y}(\mathbf{s}) = [Y_1(\mathbf{s}), Y_2(\mathbf{s})]'$ , a matriz de covariância cheia

resultante segue o seguinte formato

$$\boldsymbol{\Sigma} = \begin{bmatrix} C_{11}(h) & C_{12}(h) \\ C_{21}(h) & C_{22}(h) \end{bmatrix}. \quad (2.8)$$

Na literatura, há diversas formas paramétricas para  $C_{kj}(h)$ . Na Subseção 2.2.1 são apresentadas as estruturas de covariância mais simples, denominadas de funções de covariância cruzada separáveis.

### 2.2.1 Modelos separáveis

Para dados multivariados, pode-se denotar  $\mathbf{Y}(\mathbf{s})$  como sendo o vetor composto, por exemplo, por duas variáveis como temperatura e umidade na localização  $\mathbf{s}$ , sendo  $\{\mathbf{Y}(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^p; \mathbf{Y} \in \mathbb{R}^p\}$  e a covariância pode ser definida pela seguinte função:

$$C_{kj}(h) = a_{kj}\rho(h, \phi). \quad (2.9)$$

sendo  $a_{kj}$  um elemento da matriz  $\mathbf{A}$  e esta com dimensão  $p \times p$ , positiva definida, que traz a dependência entre as variáveis. A considerar  $\mathbf{Y}_1$  e  $\mathbf{Y}_2$  como os vetores dos dados da primeira e segunda variável, respectivamente, e assume-se que  $a_{11}$  representa a variância de  $\mathbf{Y}_1$ ,  $a_{12}$  representa  $\text{Cov}(\mathbf{Y}_1, \mathbf{Y}_2)$ ,  $a_{21}$  representa  $\text{Cov}(\mathbf{Y}_2, \mathbf{Y}_1)$  e  $a_{22}$  representa a variância de  $\mathbf{Y}_2$ ; e  $\rho(h, \phi)$  é um função de correlação espacial válida, como as descritas na Seção 2.1. Neste caso, teremos os elementos da matriz  $\mathbf{R}_{kj}$  formados por  $\rho(h_{kj}, \phi)$ , com dimensão  $n \times n$ , sendo  $\mathbf{R}$  a matriz de dependência puramente espacial.

Neste trabalho, utilizaremos funções de correlação espacial da família Cauchy, tal escolha não é baseada em uma restrição e sim nas características de longo alcance da Cauchy, ou seja, não há impedimento para usar a função de covariância Exponencial, por exemplo. Para o modelo bivariado, as funções de covariância separáveis são dadas por:

$$C_{11}(h) = a_{11} \left(1 + \frac{h}{\phi}\right)^{-1}, \quad (2.10)$$

$$C_{22}(h) = a_{22} \left(1 + \frac{h}{\phi}\right)^{-1}, \quad (2.11)$$

$$C_{12}(h) = a_{12} \left(1 + \frac{h}{\phi}\right)^{-1}, \quad (2.12)$$

onde  $\phi$  é o alcance espacial, responsável por trazer a distância até o ponto em que as correlações das variáveis ao longo do espaço não é desprezível. Por exemplo, suponha que

tenham sido coletados os dados de temperatura e monóxido de carbono em Copacabana, Ipanema e Campo Grande, bairros da cidade do Rio de Janeiro. A correlação entre as variáveis medidas em Copacabana e Ipanema tende a ser muito maior por serem bairros bem próximos um do outro. Já a correlação entre Copacabana e Campo Grande tende a diminuir, por terem uma distância maior. Com efeito, o parâmetro que é responsável por trazer a distância até o ponto em que a correlação espacial não é desprezível é o alcance espacial.

Neste modelo, é possível construir a matriz de covariância cheia  $\Sigma$ , de dimensão  $np \times np$ , da seguinte forma

$$\Sigma = \mathbf{A} \otimes \mathbf{R}. \quad (2.13)$$

onde  $\otimes$  representa o produto de Kronecker. A matriz  $\mathbf{A}$  traz somente a dependência entre as variáveis no espaço e possui dimensão  $p \times p$ , sendo  $p$  o número de variáveis, e a matriz  $\mathbf{R}$  traz somente a dependência entre as localizações no espaço e possui dimensão  $n \times n$ . Este modelo apresenta vantagens computacionais interessantes no procedimento de inferência, pois

$$|\Sigma| = |\mathbf{A} \otimes \mathbf{R}| = |\mathbf{A}|^n |\mathbf{R}|^p. \quad (2.14)$$

e

$$\Sigma^{-1} = \mathbf{A}^{-1} \otimes \mathbf{R}^{-1}. \quad (2.15)$$

Demonstrações e maiores detalhes sobre o produto de Kronecker apresentados nas Equações (2.13), (2.14), (2.15) podem ser vistos em Steeb (1997).

Assim sendo, ao invés de ser preciso calcular os determinantes e as inversas de uma matriz de dimensão  $np \times np$ , basta calcular de matrizes de menor dimensão, uma  $p \times p$  e outra de dimensão  $n \times n$ , de forma que essa redução da dimensão é muito vantajosa em diminuir o tempo de processamento computacional. Entretanto, as funções separáveis apresentam algumas limitações.

Uma das limitações mais importantes dos modelos separáveis está associada aos alcances espaciais. Como mencionado anteriormente, o alcance espacial mede a distância na qual ainda há correlação/dependência no espaço. Desta forma, pensando em modelagem espacial conjunta de variáveis aleatórias, não é razoável supor que estas apresentem o mesmo comportamento de dependência ao longo do espaço. Observe que nas Equações (2.10), (2.11) e (2.12) a função de correlação espacial é a mesma, independente da variável avaliada, isto é, o alcance espacial, definido como  $\phi$ , é igual para todas as variáveis, impondo o mesmo comportamento no espaço.

### 2.2.2 Modelos não separáveis

O modelo espacial de estrutura não separável traz como sua maior vantagem permitir que diferentes alcances espaciais sejam adotados no modelo e, portanto, é mais próxima à realidade dos dados espaciais.

Diversos autores buscaram relaxar as limitações dos modelos separáveis e utilizaram modelos não separáveis para ter mais flexibilidade. Para este trabalho a estrutura não separável de covariância utilizada será a apresentada em [Porcu, Bevilacqua e Genton \(2016\)](#), sendo

$$C_{kj}(h) = \sigma_k \sigma_j \rho_{kj} \left(1 + \frac{h}{\phi_{kj}}\right)^{-1}. \quad (2.16)$$

onde  $h$  é a distância euclidiana,  $\sigma_k$  e  $\sigma_j$  são os desvios-padrão das variáveis  $\mathbf{Y}_k$  e  $\mathbf{Y}_j$ ,  $k, j = 1, \dots, p$ ,  $\rho_{kj}$  é a correlação entre as variáveis  $\mathbf{Y}_k$  e  $\mathbf{Y}_j$ . Para a função na Equação [\(2.16\)](#) ser válida, ou seja, gerar uma matriz de covariância positiva definida, é necessário impor as seguintes restrições, válidas para função de correlação Cauchy, conforme apresentado em [Porcu, Bevilacqua e Genton \(2016\)](#):

$$\rho_{kj}^2 \leq (\phi_{kk} + \phi_{jj}) / (\phi_{kj})^2, \text{ com } k, j = 1, 2, k \neq j \quad (2.17)$$

$$\text{e } \phi_{kj} = (\phi_{kk} + \phi_{jj}) / 2 \quad (2.18)$$

Observe que as estruturas a serem discutidas em diante permitem alcances espaciais diferentes para variáveis distintas.

No caso bivariado, essa função pode ser escrita da seguinte forma

$$C_{11}(h) = \sigma_1^2 \left(1 + \frac{h}{\phi_{11}}\right)^{-1}, \quad (2.19)$$

$$C_{22}(h) = \sigma_2^2 \left(1 + \frac{h}{\phi_{22}}\right)^{-1}, \quad (2.20)$$

$$C_{12}(h) = \sigma_1 \sigma_2 \rho_{12} \left(1 + \frac{h}{\phi_{12}}\right)^{-1}. \quad (2.21)$$

O modelo não separável considera que os alcances espaciais podem ser diferentes, como apresentado em [Porcu, Bevilacqua e Genton \(2016\)](#), portanto terá como restrição  $\rho_{12}^2 \leq (\phi_{11} + \phi_{22}) / (\phi_{12})^2$ , onde  $\phi_{12} = (\phi_{11} + \phi_{22}) / 2$ .

Apesar dos modelos não separáveis serem mais flexíveis do que os modelos separáveis, a matriz de covariância cheia  $\Sigma$ , de dimensão  $np \times np$ , não pode ser escrita da mesma forma que no modelo separável. Nesse caso, no procedimento de inferência, será necessário inverter e calcular o determinante de uma matriz de dimensão  $np \times np$ , trazendo um custo de processamento computacional muito elevado.

Uma proposta para que os modelos espaciais aumentem a sua capacidade de lidar com diferentes alcances e tenham um menor custo de processamento, é utilizar aproximações separáveis.

### 2.2.3 Aproximações separáveis

Com o objetivo de unir os ganhos do modelo separável em termos computacionais com os ganhos do modelo não separável em termos de variedade de alcances espaciais permitidos, faz-se o uso de aproximações separáveis.

As aproximações separáveis são aplicadas às matrizes de covariância que são não separáveis, de forma que as propriedades de tais matrizes sejam mantidas, como apresentado em [Genton \(2007\)](#) e [Erbisti, Fonseca e Alves \(2018\)](#). [Genton \(2007\)](#) também especifica que tal aproximação será feita pelo produto de Kronecker para uma matriz de covariância cheia a partir da norma de Frobenius.

Antes de apresentar a solução por aproximações separáveis, alguns conceitos devem ser definidos.

1. **Operador  $vec(\cdot)$ :** o operador  $vec(\cdot)$  transforma uma matriz  $\mathbf{A}$  de dimensão  $n_1 \times n_2$  em um vetor  $vec(\mathbf{A}) \in \mathfrak{R}^{n_1 n_2}$ , empilhando as colunas uma em cima da outra, como por exemplo:

$$\mathbf{A} = [\mathbf{a}_1 \mathbf{a}_2 \dots \mathbf{a}_n], \text{ então } vec(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{bmatrix} \quad (2.22)$$

Mais detalhes sobre operador  $vec(\cdot)$  podem ser encontrados em [Steeb \(1997\)](#).

2. **Norma de Frobenius:** tem-se que uma matriz  $\mathbf{A}$  de dimensão  $p \times p$  com elemento  $a_{kj}$ ,  $k, j = 1, \dots, p$  possui para  $\mathbf{A}(\|\mathbf{A}\|_F)$  a seguinte norma de Frobenius:

$$\|\mathbf{A}\|_F = \left( \sum_{k=1}^p \sum_{j=1}^p a_{kj}^2 \right)^{1/2}. \quad (2.23)$$



3. **Proximidade do produto de Kronecker para uma matriz de covariância cheia:** seja a matriz de covariância cheia  $\Sigma$  de dimensão  $np \times np$ . O desafio é encontrar duas matrizes  $\mathbf{A} \in \mathfrak{R}^{p \times p}$  e  $\mathbf{R} \in \mathfrak{R}^{n \times n}$  tal que  $\|\Sigma - \mathbf{A} \otimes \mathbf{R}\|_F$  seja minimizada.
4. **Decomposição em Valores Singulares :** [Golub e Loan \(1996\)](#) apresentam um teorema que aponta que se  $\mathbf{B} \in \mathfrak{R}^{n_2 \times p_2}$  é uma matriz real, então há duas matrizes ortogonais  $\mathbf{U} = [u_1, \dots, u_{n_2}] \in \mathfrak{R}^{n_2 \times n_2}$  e  $\mathbf{V} = [v_1, \dots, v_{p_2}] \in \mathfrak{R}^{p_2 \times p_2}$ , tais que

$$\mathbf{U}^T \mathbf{B} \mathbf{V} = \text{diag}(w_1, \dots, w_r) \in \mathfrak{R}^{n_2 \times p_2}, \quad r = \text{posto}(\mathbf{B}) = \min\{n_2, p_2\} \quad (2.24)$$

onde  $w_1 \geq w_2 \geq w_r \geq 0$ .

Baseado em [Genton \(2007\)](#) e [Erbisti, Fonseca e Alves \(2018\)](#), a proximidade do produto de Kronecker para uma matriz de covariância cheia é feita através da Decomposição em Valor Singulares de uma versão permutada da matriz de covariância cheia  $\Sigma$  de dimensão  $np \times np$ , que será construída a partir da função de covariância não separável apresentada na Equação [\(2.16\)](#).

O primeiro passo é rearranjar  $\Sigma$  em outra matriz  $\mathfrak{S}(\Sigma) \in \mathfrak{R}^{n_2 \times p_2}$ , de tal forma que a soma dos quadrados  $\|\Sigma - \mathbf{R} \otimes \mathbf{A}\|_F$  seja a mesma de  $\|\mathfrak{S}(\Sigma) - \text{vec}(\mathbf{R}) \otimes \text{vec}(\mathbf{A})^T\|_F$ . Para tal, [Genton \(2007\)](#) e [Erbisti, Fonseca e Alves \(2018\)](#) apresentam  $\|\Sigma - \mathbf{R} \otimes \mathbf{A}\|_F = \|\mathfrak{S}(\Sigma) - \text{vec}(\mathbf{R}) \otimes \text{vec}(\mathbf{A})^T\|_F$  e também  $\|\Sigma\|_F = \|\mathfrak{S}(\Sigma)\|_F$ .

Uma vez rearranjada a matriz de covariância cheia, o problema da aproximação se reduz ao cálculo do posto da matriz retangular  $\mathfrak{S}(\Sigma) \in \mathfrak{R}^{n_2 \times p_2}$ , solução que também pode ser encontrada em [Golub e Loan \(1996\)](#). Com base na Decomposição de Valores Singulares, a solução é dada por

$$\text{vec}(\mathbf{R}) = \sqrt{w_1} \mathbf{u}_1 \quad \text{vec}(\mathbf{A}) = \sqrt{w_1} \mathbf{v}_1. \quad (2.25)$$

Sendo  $\mathbf{v}_1$  a primeira coluna da matriz  $\mathbf{V} \in \mathfrak{R}^{p_2 \times p_2}$  e  $\mathbf{u}_1$  a primeira coluna da matriz  $\mathbf{U} \in \mathfrak{R}^{n_2 \times n_2}$  e  $w_1$  o primeiro autovalor de  $\mathbf{W}$ , o vetor que contém os valores singulares da matriz permutada  $\mathfrak{S}(\Sigma)$  de tamanho  $\min\{n^2, p^2\}$  e ordenado de forma decrescente .

## 2.2.4 Erro de aproximação

Aproximações separáveis estão sujeitas a erros. A fim de medir se a aproximação é adequada, [Genton \(2007\)](#) e [Erbisti, Fonseca e Alves \(2018\)](#) definem  $k_\Sigma(\mathbf{A}, \mathbf{R})$  como o erro de aproximação por separabilidade da matriz  $k_\Sigma$  aproximada pelo produto de Kronecker de duas matrizes  $\mathbf{A}$  e  $\mathbf{R}$ .

E o erro da aproximação pode ser calculado por

$$k_{\Sigma}(\mathbf{A}, \mathbf{R}) = \frac{\|\Sigma - \mathbf{A} \otimes \mathbf{R}\|_F}{\|\Sigma\|_F}. \quad (2.26)$$

Como apresentado em [Erbisti, Fonseca e Alves \(2018\)](#), caso  $\Sigma$  seja separável,  $k_{\Sigma}(\mathbf{A}, \mathbf{R}) = 0$ . Este erro varia, então, de 0 a  $\sqrt{1 - \frac{1}{r}}$ , onde  $r = \min\{n^2, p^2\}$ . Como solução, utiliza-se o erro da aproximação padronizado

$$k_{\Sigma}^*(\mathbf{A}, \mathbf{R}) = \frac{k_{\Sigma}(\mathbf{A}, \mathbf{R})}{\sqrt{1 - \frac{1}{r}}}. \quad (2.27)$$

O erro aproximado torna mais fácil de conhecer qual o valor máximo do ajuste, pois  $k_{\Sigma}^*(\mathbf{A}, \mathbf{R})$  irá variar de 0 a 1.

## 2.3 Inferência bayesiana

Neste trabalho, as estimativas para os parâmetros do modelo proposto serão feitas sob a abordagem bayesiana. Os métodos a serem discutidos nesta seção trazem os conceitos básicos de inferência bayesiana aplicados à geoestatística, de acordo com [Schmidt e Sansó \(2006\)](#) e [Erbisti, Fonseca e Alves \(2018\)](#) e também sobre inferência bayesiana sem aplicações em geoestatística por [DeGROOT e Schervish \(2002\)](#).

A inferência bayesiana traz a ideia de se combinar informações já conhecidas a respeito de um vetor de parâmetros  $\boldsymbol{\theta}$  com os dados observados  $\mathbf{Y}(\mathbf{s})$  através da função de verossimilhança. Tal procedimento busca trazer informações atualizadas sobre os parâmetros a serem estimados e com a  $p(\boldsymbol{\theta})$ , a distribuição a *priori*, e  $l(\boldsymbol{\theta}, \mathbf{y})$ , a função de verossimilhança, é possível inferir sobre  $\boldsymbol{\theta}$  a partir de  $p(\boldsymbol{\theta}|\mathbf{y})$  também conhecida como distribuição a *posteriori*. Portanto, seguindo o Teorema de Bayes, tem-se que

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}. \quad (2.28)$$

Mais detalhes sobre o Teorema de Bayes podem ser vistos em [DeGROOT e Schervish \(2002\)](#).

Assim, a considerar o caso bivariado, o vetor  $\mathbf{Y}(\mathbf{s})$  traz os dados das duas variáveis observadas nas localizações  $\mathbf{s}_1, \dots, \mathbf{s}_n \in D$ , com vetor de médias  $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ , onde  $\mathbf{X}$  é a matriz que contém os dados das variáveis explicativas latitude e longitude, e  $\boldsymbol{\beta}$  o vetor que contém o intercepto e o coeficiente angular de cada variável, e tem  $\Sigma_{np \times np}$  como a estrutura de covariância. Os dados são representados por meio de um processo gaussiano,

$\mathbf{Y} \sim N_{2n}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$  e a função de verossimilhança pode ser descrita como:

$$L(\boldsymbol{\theta}; \mathbf{y}) = \frac{1}{\sqrt{2\pi}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \quad (2.29)$$

Em inferência bayesiana, a distribuição a *priori* é feita com o conhecimento prévio que se possui dos dados, onde o pesquisador irá supor uma distribuição para  $\boldsymbol{\theta}$ .

Para os modelos separáveis apresentados nas Equações (2.10), (2.11) e (2.12), o vetor paramétrico  $\boldsymbol{\theta}$  será composto pelas quantidades desconhecidas a serem estimadas, denotadas por  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi, \mathbf{A})$ , sendo  $\boldsymbol{\beta} = (\beta_{10}, \beta_{11}, \beta_{12}, \dots, \beta_{1q}, \beta_{20}, \beta_{21}, \beta_{22}, \dots, \beta_{2q}, \dots, \beta_{pq})$ , onde tem-se  $p$  variáveis e  $q = 1, 2, \dots, 2(p+1)$  covariâncias, incluindo o intercepto. Os alcances espaciais são representados pelo escalar  $\phi$ ; e  $\mathbf{A}$ , matriz de elementos  $a_{kj}$  traz as dependências entre as variáveis. Com efeito, para o caso separável obtém-se a seguinte distribuição a *priori*, assumindo independência dos parâmetros

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\beta})p(\phi)p(\mathbf{A}). \quad (2.30)$$

Em modelos não separáveis,  $\boldsymbol{\theta}$  terá diferentes parâmetros de tal forma que  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi_{11}, \phi_{22}, \sigma_1, \sigma_2, \rho_{12})$ . Os alcances espaciais  $\phi_{11}$  e  $\phi_{22}$ , e  $\sigma_1, \sigma_2$  como as variâncias do processo espacial e  $\rho_{12}$  como a correlação entre as variáveis  $\mathbf{Y}_1$  e  $\mathbf{Y}_2$ . Desta forma, é definida a distribuição a *priori* para o caso não separável, assumindo independência dos parâmetros

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\beta})p(\phi_{11})p(\phi_{22})p(\sigma_1)p(\sigma_2)p(\rho_{12}). \quad (2.31)$$

Com o conhecimento prévio sobre as variáveis, atribui-se uma distribuição a *priori* para cada parâmetro, sendo  $p = 2$  e  $\boldsymbol{\beta} \sim N_{2(q+1)}(\mathbf{b}, \mathbf{B})$ ,  $q = 2$ , e seja  $med(h)$  a mediana das distâncias espaciais,  $\phi_{kj} \sim Gama(a_{kj}, a_{kj} \times med(h))$ ,  $k, j = 1, 2$ , onde, por simplicidade, será utilizado  $a_{kj}^* = a_{kj} \times med(h)$ . A demais distribuições a *priori* são  $\sigma_{kj} \sim Gama(c_{kj}, d_{kj})$ ,  $k, j = 1, 2$  e  $\rho_{12} \sim Unif(-(\frac{\phi_{11} + \phi_{22}}{\phi_{12}^2}), (\frac{\phi_{11} + \phi_{22}}{\phi_{12}^2}))$ .

### 2.3.1 Simulações estocásticas feitas por Monte Carlo via Cadeias de Markov (MCMC)

O segundo passo é determinar a distribuição a *posteriori*, que será obtida através de métodos computacionais, onde, caso não tenham a sua forma analítica conhecida, terão suas amostras geradas através de simulações estocásticas de Monte Carlo via Cadeias de Markov. Tais simulações têm por objetivo trazer uma amostra com as propriedades

probabilísticas da distribuição a *posteriori* de forma que o pesquisador poderá atualizar as informações provenientes da distribuição a *priori*.

As formas de se gerar as amostras da distribuição a *posteriori* de cada parâmetro irá variar de acordo com a forma analítica da distribuição condicional completa a *posteriori* de cada parâmetro. Para o caso em que a distribuição a *posteriori* possua forma analítica fechada, isto é, quando tiver o núcleo de sua função equivalente ao núcleo da função de densidade de uma distribuição de probabilidade conhecida, será usado o Algoritmo amostrador de Gibbs, proposto por [Geman e Geman \(1984\)](#). Nesse método será usado a própria forma analítica fechada da distribuição a *priori* e maiores detalhes de seu algoritmo podem ser vistos na Subseção [2.3.2](#).

Para os casos em que a distribuição a *posteriori* de um parâmetro não tenha forma analítica fechada, será utilizado simulações estocásticas via MCMC utilizando Método de Metropolis-Hastings, inicialmente proposto por [Metropolis et al. \(1953\)](#) para distribuições simétricas e posteriormente generalizado por [Hastings \(1970\)](#). Este método é apresentado na Subseção [2.3.3](#).

Maiores detalhes sobre as simulações de Monte Carlo via Cadeias de Markov (MCMC) podem ser vistas em [Paulino, Turkman e Murteira \(2003\)](#).

### 2.3.2 Algoritmo amostrador de Gibbs

Um dos métodos mais simples de fazer simulações estocásticas para obter uma amostra da distribuição a *posteriori* é pelo Algoritmo amostrador de Gibbs. Com tal método é possível utilizar a distribuição condicional completa a *posteriori* para gerar as amostras. De forma generalizada, a considerar como  $\Theta$  o vetor paramétrico que contém os parâmetros de interesse  $\Theta = (\theta_1, \theta_2, \dots, \theta_g)$  e os valores iniciais  $\Theta^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_g^{(1)})$ , e tendo  $M$  como o número de iterações da cadeia e sendo  $\mathbf{y}$  o vetor de dados observados, tem-se que  $p(\theta_i | \boldsymbol{\theta}_{-i}, \mathbf{y})$ ,  $i = 1, \dots, g$  são as condicionais completas. Com efeito, é possível obter o seguinte Algoritmo de Gibbs:

1. Para  $j$  de 1 até  $M - 1$ , faça:
2. Amostre  $\theta_1^{(j+1)}$  de  $p(\theta_1 | \theta_2^{(j)}, \dots, \theta_g^{(j)}, \mathbf{y})$ ;
3. Amostre  $\theta_2^{(j+1)}$  de  $p(\theta_2 | \theta_1^{(j+1)}, \theta_3^{(j)}, \dots, \theta_g^{(j)}, \mathbf{y})$ ;

Com isso, será gerada uma amostra considerando os valores iniciais posteriormente definidos,  $(\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_g^{(1)})$ ,  $(\theta_1^{(2)}, \theta_2^{(2)}, \dots, \theta_g^{(2)})$ , ...,  $(\theta_1^{(M)}, \theta_2^{(M)}, \dots, \theta_g^{(M)})$  como uma sequência

dependente parâmetros.

No Algoritmo de Gibbs são aplicadas as técnicas de *burn-in* (aquecimento) e *thinning* (espaçamento). A técnica de *burn-in* retira os valores iniciais da cadeia até que os dados da amostra sejam próximos à distribuição proposta e a técnica do espaçamento suaviza a autocorrelação entre os valores gerados.

### 2.3.3 Método Metropolis-Hastings

Em um caso em que não é possível obter uma forma analítica fechada da distribuição *a posteriori*, entra a proposta do método de Metropolis-Hastings para gerar tais amostras via MCMC.

Este método tem como conceitos básicos propor uma distribuição de probabilidade e gerar valores aleatórios a partir dela que poderão ser aceitos ou não.

Ao assumir o mesmo vetor paramétrico de interesse  $\Theta = (\theta_1, \theta_2, \dots, \theta_g)$  e tendo  $M$  como o número de iterações da cadeia e sendo  $\mathbf{y}$  os dados observados, uma vez que é de interesse gerar amostras da distribuição  $p(\theta_1, \theta_2, \dots, \theta_g | \mathbf{y})$  e as distribuições condicionais completas  $p(\theta_i | \theta_{-i}, \mathbf{y})$ ,  $i = 1, \dots, g$ , são impossíveis de se amostrar, para cada  $i$  há uma distribuição proposta  $q_i(\cdot)$  em que pode se gerar amostras aleatórias.

Com efeito, tem-se o algoritmo do método de Método Metropolis-Hastings:

1. Inicialize  $\Theta^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_g^{(1)})$
2. Para  $j$  de 1 até  $M - 1$ , faça:
  - (a) Para  $i$  de 1 até  $g$ , faça:
    - i. Gerar um valor aleatório  $\theta_1^*$  da distribuição proposta

$$\theta_1^* \sim q_i(\cdot | \theta_i^{j-1}) \quad (2.32)$$

- ii. Aceitar  $\theta^*$  proposto com probabilidade

$$\alpha = \min \left\{ 1, \frac{p(\theta^* | \theta_{-i}^{(a)}, \mathbf{y}) q(\theta_i^{(j-1)} | \theta^*)}{p(\theta_i^{(j-1)} | \theta_{-i}^{(a)}, \mathbf{y}) q(\theta^* | \theta_i^{(j-1)})} \right\} \quad (2.33)$$

$$\text{onde } \theta_{-i}^{(a)} = (\theta_1^{(j)}, \dots, \theta_{i-1}^{(j)}, \theta_{i+1}^{(j-1)}, \dots, \theta_g^{(j-1)})$$

3. Avaliar a convergência da Cadeia de Markov

O método apresentado possui algumas ressalvas, uma vez que utiliza uma distribuição de probabilidade a ser proposta pelo pesquisador. É possível que a cadeia em suas iterações iniciais não represente bem a distribuição escolhida e que, ao longo da cadeia, os valores gerados sejam bastante autocorrelacionados. Para sanar esses dois pontos, são aplicadas as técnicas de *burn-in* (aquecimento) e *thinning* (espaçamento). A técnica de *burn-in* retira os valores iniciais da cadeia até que os dados da amostra sejam próximos à distribuição proposta. Por outro lado, a técnica do espaçamento irá suavizar a autocorrelação entre os valores.

A etapa descrita na Equação (2.33) será utilizada na construção da taxa de aceitação da Cadeia. Mais detalhes sobre a taxa de aceitação podem ser vistos em Paulino, Turkman e Murteira (2003).

### 2.3.4 Distribuições condicionais completas

Definidos os métodos a serem utilizados na Seção 2.3 e nas Subseções 2.3.2 e 2.3.3, o cálculo da distribuição conjunta *posteriori* é dado por:

$$\begin{aligned}
p(\boldsymbol{\beta}, \phi_{11}, \phi_{22}, \sigma_1, \sigma_2, \rho_{12} | \mathbf{y}) &\propto L(\boldsymbol{\beta}, \sigma_1, \sigma_2, \phi_{11}, \phi_{22}, \rho_{12}; \mathbf{y}) \times \\
& p(\boldsymbol{\beta}) p(\phi_{11}) p(\phi_{22}) p(\sigma_1) p(\sigma_2) p(\rho_{12}) \\
&\propto \left[ (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \right] \times \\
& \left[ (2\pi)^{-(q+1)/2} |\mathbf{B}|^{-1/2} \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \mathbf{b})^T \mathbf{B}^{-1} (\boldsymbol{\beta} - \mathbf{b}) \right\} \right] \times \\
& \left[ \frac{a_{11}}{\Gamma(a_{11}^*)} (\phi_{11})^{a_{11}-1} \exp \left\{ -\frac{\phi_{11}}{a_{11}^*} \right\} \right] \times \\
& \left[ \frac{a_{22}}{\Gamma(a_{22}^*)} (\phi_{22})^{a_{22}-1} \exp \left\{ -\frac{\phi_{22}}{a_{22}^*} \right\} \right] \times \\
& \left[ \frac{d_{11}}{\Gamma(c_{11})} (\sigma_1)^{-c_{11}-1} \exp \left\{ -\frac{d_{11}}{\sigma_1} \right\} \right] \times \\
& \left[ \frac{d_{22}}{\Gamma(c_{22})} (\sigma_2)^{-c_{22}-1} \exp \left\{ -\frac{d_{22}}{\sigma_2} \right\} \right] \times \\
& \left[ \left( \frac{\phi_{11} + \phi_{22}}{(\phi_{12})^2} \right)^{1/2} \right]. \tag{2.34}
\end{aligned}$$

Conforme apresentado na Seção 2.3, os parâmetros que não apresentarem uma forma analítica fechada no cálculo de suas distribuições condicionais completas terão sua dis-

tribuição a *posteriori* estimada pelo método de Metropolis-Hastings, os que possuírem forma analítica fechada serão gerados pelo Algoritmo de Gibbs. Com efeito, o processo geral é feito pelo método do Amostrador de Gibbs com passos de Metropolis-Hastings.

Considerando que  $\boldsymbol{\theta}^- = (\phi_{11}, \phi_{22}, \sigma_1, \sigma_2, \rho_{12})$  é o vetor com os parâmetros de interesse que não contém o vetor  $\boldsymbol{\beta}$  e que  $\boldsymbol{\beta} \sim N_{2(q+1)}(\mathbf{b}, \mathbf{B})$ , tem-se que

$$p(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\theta}^-) \propto \exp\left\{ \frac{-1}{2} \left[ (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + (\boldsymbol{\beta} - \mathbf{B})^T \boldsymbol{\Sigma} (\boldsymbol{\beta} - \mathbf{B}) \right] \right\}. \quad (2.35)$$

Supondo que  $\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\theta}^- \sim N_{2n}(b_0^*, b_1^*)$ , é esperado que

$$p(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\theta}^-) \propto \exp\left\{ \frac{-1}{2} \left[ (\boldsymbol{\beta} - b_0^*)^T (b_1^*)^{-1} (\boldsymbol{\beta} - b_0^*) \right] \right\}. \quad (2.36)$$

Pelas propriedades de Álgebra Linear, supondo  $\mathbf{A}$  e  $\mathbf{B}$  duas matrizes reais,  $(\mathbf{A} - \mathbf{B})^T (\mathbf{A} - \mathbf{B}) = \mathbf{A}^T \mathbf{A} - \mathbf{A}^T \mathbf{B} - \mathbf{B}^T \mathbf{A} + \mathbf{B}^T \mathbf{B}$ , pode ser afirmar que  $\mathbf{A}^T \mathbf{B} = \mathbf{B}^T \mathbf{A}$ , então  $(\mathbf{A} - \mathbf{B})^T (\mathbf{A} - \mathbf{B}) = \mathbf{A}^T \mathbf{A} - 2\mathbf{A}^T \mathbf{B} + \mathbf{B}^T \mathbf{B}$ , com efeito é possível reescrever a Equação (2.36) de forma que, considerando apenas as expressões que dependem do parâmetro  $\boldsymbol{\beta}$  e obtendo a seguinte expressão

$$p(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\theta}^-) \propto \exp\left\{ \frac{-1}{2} \left[ \boldsymbol{\beta}^T (b_1^*)^{-1} \boldsymbol{\beta} - 2\boldsymbol{\beta}^T (b_1^*)^{-1} (b_0^*) \right] \right\}. \quad (2.37)$$

E assim, utilizando a Equação (2.35) e o resultado esperado dado pela Equação (2.36), tem-se que

$$p(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\theta}^-) \propto \exp\left\{ \frac{-1}{2} \left[ \boldsymbol{\beta}^T (\mathbf{X}^T \boldsymbol{\Sigma} \mathbf{X}) \boldsymbol{\beta} - 2\boldsymbol{\beta}^T (\mathbf{Y} \mathbf{X}^T \boldsymbol{\Sigma}^{-1}) + \boldsymbol{\beta}^T (b_1^{-1}) \boldsymbol{\beta} \right] \right\} \quad (2.38)$$

Logo, a distribuição condicional completa de  $\boldsymbol{\beta}$  é dada por

$$\begin{aligned} \boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\theta}^- &\sim N_{2(q+1)}(\mathbf{b}_0^*, \mathbf{b}_1^*), \\ \text{sendo } \mathbf{b}_1^* &= [\mathbf{X}^t \boldsymbol{\Sigma}^{-1} \mathbf{X} + \mathbf{B}^{-1}]^{-1} \text{ e } \mathbf{b}_0^* = [\mathbf{X}^t \boldsymbol{\Sigma}^{-1} \mathbf{X} + \mathbf{B}^{-1} \mathbf{B}] \mathbf{b}_1^*. \end{aligned} \quad (2.38)$$

Uma vez que a distribuição de  $\boldsymbol{\beta}$  possui forma analítica fechada, as amostras da distribuição condicional completa a *posteriori* foram geradas pelo amostrador de Gibbs.

Para o parâmetro  $\phi_{11}$ , a considerar a *priori* é dada por  $\phi_{11} \sim \text{Gama}(a_{11}, a_{11} \times \text{med}(h))$ , com  $\text{med}(h)$  sendo a mediana das distâncias espaciais,  $k, j = 1, 2$ . Considerando, por

simplicação, que  $a_{11}^* = a_{11} \times med(h)$ , a distribuição condicional completa é dada por

$$p(\phi_{11} | \mathbf{y}, \boldsymbol{\beta}, \phi_{22}, \sigma_1, \sigma_2, \rho_{12}) \propto \left[ |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ \frac{-1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \right] \times \quad (2.39)$$

$$\left[ (\phi_{11})^{a_{11}-1} \exp \left\{ \frac{-\phi_{11}}{a_{11}^*} \right\} \right] \quad (2.40)$$

Note que uma vez que  $\boldsymbol{\Sigma}$  é construída em blocos como mostra a Equação (2.8) e uma vez que o parâmetro  $\phi_{11}$  não pode ser isolado, conclui-se que não há forma analítica fechada. A distribuição condicional completa de  $\phi_{22}$  é equivalente a do parâmetro  $\phi_{11}$ . Assim, os alcances espaciais foram amostrados a partir do método de Metropolis-Hastings.

A variância da variável 1 representada por  $\sigma_1$  tem como *priori*  $\sigma_1 \sim Gama(c_{kj}, d_{kj})$ ,  $k, j = 1, 2$  e a sua distribuição condicional completa pode ser descrita por

$$p(\sigma_1 | \mathbf{y}, \boldsymbol{\beta}, \phi_{11}, \phi_{22}, \sigma_2, \rho_{12}) \propto \left[ |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ \frac{-1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \right] \times \quad (2.41)$$

$$\left[ (\sigma_1)^{-c_{11}-1} \exp \left\{ \frac{-d_{11}}{\sigma_1} \right\} \right] \quad (2.42)$$

Com isso, obtém-se o mesmo resultado dos alcances espaciais  $\phi_{11}$  e  $\phi_{22}$ , como  $\sigma_1$  está contido em  $\boldsymbol{\Sigma}$  e não é possível de ser isolado, não possui forma analítica fechada e assim a distribuição a condicional completa a *posteriori* de  $\sigma_1$  será gerada a partir do método de Metropolis-Hastings, e o equivalente para a variância da variável 2,  $\sigma_2$ .

O parâmetro  $\rho_{12}$  tem sua distribuição condicional completa da seguinte forma:

$$p(\rho_{12} | \mathbf{y}, \boldsymbol{\beta}, \phi_{11}, \phi_{22}, \sigma_2, \rho_{12}) \propto \left[ |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ \frac{-1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \right] \quad (2.43)$$

Note que  $\rho_{12}$  está contido em  $\boldsymbol{\Sigma}$  e não pode ser isolado. Desta forma,  $\rho_{12}$  não possuirá forma analítica fechada e a amostra a posteriori desse parâmetro será gerada pelo método de Metropolis-Hastings.



## 3 Análise dos resultados

Este capítulo se propõe a apresentar os resultados da qualidade das aproximações separáveis, a exploração do comportamento de cenários com diferentes alcances espaciais e o resultado do procedimento de inferência sob o ponto vista bayesiano, utilizando de dados simulados. Neste capítulo também será apresentado o pacote desenvolvido para o *software* R com as funções utilizadas para gerar as análises.

### 3.1 Comportamento do erro de aproximação conforme aumento de não separabilidade

A fim de compreender o comportamento do erro de aproximação padronizado ao lidar com diferentes graus de separabilidade, tais erros foram obtidos como visto na Equação (2.26), utilizando a função de correlação Cauchy (Porcu, Bevilacqua e Genton (2016)), conforme descrito pelas Equações (2.19), (2.20), (2.21) e com o parâmetro de suavização  $\alpha$  fixado em  $\alpha = 1$ . Para tal, fixou-se  $\phi_{11} = \phi_{22} = 0,1$  e foi feito um incremento de 0,6 com saltos de 0,01 em  $\phi_{22}$ . De fato, iniciou-se a análise considerando alcances iguais e a cada incremento no alcance espacial  $\phi_{22}$  foi se obtendo cenários não separáveis.

A Figura 1 apresenta o comportamento dos erros padronizados. Nota-se que o erro de aproximação cresce conforme aumenta a diferença dos alcances espaciais, porém não chega a altos valores, uma vez que varia de 0 a 1 por se tratar do erro padronizado, e obteve valor máximo de aproximadamente 8%.

O crescimento do erro visto na Figura 1 conforme a diferença entre os alcances espaciais aumenta é esperado, pois quanto maior for o grau de não separabilidade do modelo, maior será o erro padronizado da aproximação separável. É possível também constatar que o erro de aproximação não mostra tendências de crescimento.

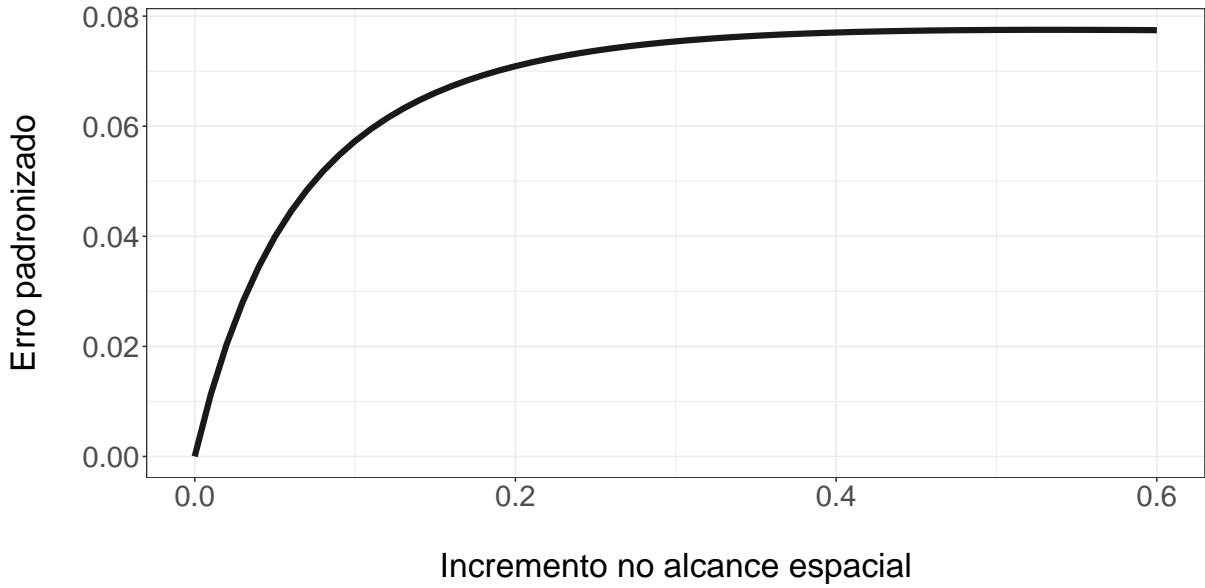


Figura 1: Comportamento dos erros padronizados segundo grau de não separabilidade medido a partir do incremento nos alcances espaciais.

## 3.2 Definição de cenários com diferentes alcances espaciais

Este trabalho trabalha com diferentes cenários operativos, onde cada um possui uma determinada diferença entre os alcances espaciais. Essa definição de cenários é importante para avaliar diferentes estruturas de não separabilidade. O caso mais simples é dado pelo cenário 1, cuja estrutura é separável e todos os alcances espaciais  $\phi_{k,j}$ ,  $k, j = 1, 2$  são idênticos. Nos cenários 2, 3 e 4, tem-se o caso não separável onde a diferença entre  $\phi_{11}$  e  $\phi_{22}$  aumenta ao longo dos cenários, isto é, o grau de não separabilidade irá aumentar a cada cenário, aumentando assim o desafio de utilizar as aproximações separáveis.

Tabela 1: Possíveis Cenários operativos.

Cenário 1	Cenário 2	Cenário 3	Cenário 4
$\phi_{11} = 0,10$	$\phi_{11} = 0,10$	$\phi_{11} = 0,10$	$\phi_{11} = 0,10$
$\phi_{22} = 0,10$	$\phi_{22} = 0,20$	$\phi_{22} = 0,50$	$\phi_{22} = 0,70$
$\phi_{12} = 0,10$	$\phi_{12} = 0,15$	$\phi_{12} = 0,30$	$\phi_{12} = 0,35$

Conforme apresentado em [Erbisti, Fonseca e Alves \(2018\)](#), variáveis com diferentes alcances espaciais promovem diferentes correlações ao longo do espaço. A correlação linear entre as variáveis  $\rho(h, \phi)$  tende a diminuir com o aumento da distância entre elas. Para melhor compreensão do comportamento da função de correlação espacial de cada cenário apresentado na Tabela [1](#), a Figura [2](#) apresenta as funções teóricas de correlação espacial

dos cenários 1, 2, 3 e 4. Para este estudo, os parâmetros  $\phi_{11}$  e  $\phi_{22}$  foram fixados de acordo com o respectivo cenário e os demais parâmetros foram fixados em  $\rho_{12} = 0,4$ ,  $\sigma_1 = 1$ ,  $\sigma_2 = 2$  e, por simplicidade, o vetor de médias  $\boldsymbol{\mu} = \mathbf{0}$ .

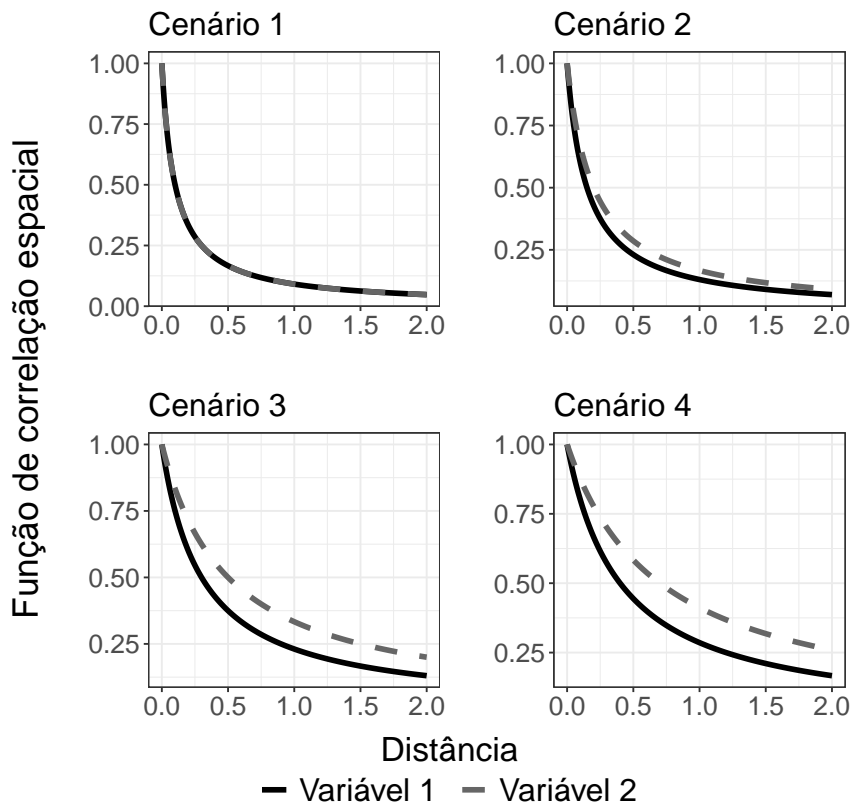


Figura 2: Comportamento da função de correlação para os cenários descritos na Tabela 1

Nota-se pela Figura 2 que com alcances espaciais idênticos, como é caso do cenário 1 descrito na Tabela 1, as curvas de correlação espacial das variáveis 1 de 2 estão sobrepostas, ou seja, seria um caso em que as duas variáveis possuem o mesmo comportamento ao longo do espaço.

Para os cenários 2, 3 e 4 onde os alcances espaciais possuem comportamentos diferentes no espaço, a Figura 2 mostra que a função de correlação das variáveis  $Y_1$  e  $Y_2$  já não estão mais sobrepostas e ficam cada vez mais distantes conforme o comportamento das variáveis se torna cada vez mais distinto ao longo do espaço.

### 3.3 Desempenho das aproximações separáveis no cálculo da função de verossimilhança

Para compreender como trabalhar com as aproximações reduz o custo computacional, foi analisado o tempo necessário para calcular a função de verossimilhança utilizando a matriz de covariância cheia, isto é, de estrutura não separável e a matriz de covariância aproximada. Os dados da Tabela 2 foram gerados para o caso bivariado seguindo o modelo  $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$ . Uma vez que um processo gaussiano é assumido,  $\mathbf{Y} \sim N_{2n}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , sendo  $\boldsymbol{\Sigma}$  construída através da função de covariância definida nas Equações (2.19), (2.20) e (2.21), considerando cada cenário operativo e  $\sigma_1 = 1$ ,  $\sigma_2 = 1,5$  e  $\rho_{12} = 0,4$ . Adotou-se  $n = 50, 100, 200, 500, 800$  e  $1000$  localizações espaciais geradas aleatoriamente em um quadrado  $[0, 1] \times [0, 1]$ . Para este estudo, foram feitas 200 réplicas do vetor resposta.

Tabela 2: Tempo necessário (em segundos) para calcular a função de verossimilhança da matriz de covariância cheia e na de estrutura aproximada. (Intel(R) Core(TM) i5-8250U, 1.60GHz, 8GB RAM).

Localizações	Cenário 1		Cenário 2		Cenário 3		Cenário 4	
	Cheia	Aprox.	Cheia	Aprox.	Cheia	Aprox.	Cheia	Aprox.
50	0,19	0,14	0,19	0,17	0,16	0,14	0,28	0,16
100	0,75	0,41	0,62	0,35	0,62	0,33	0,74	0,27
200	4,30	1,25	4,23	1,53	4,25	1,17	3,63	1,06
500	50,36	9,65	63,78	17,15	95,50	18,22	44,89	6,83
800	370,67	62,30	382,61	63,63	383,48	70,87	577,58	69,68
1000	764,07	110,52	778,06	114,14	774,63	112,30	337,11	45,75

Nota-se pela Tabela 2 que a matriz de covariância aproximada reduz o custo computacional ao se calcular a função de verossimilhança. É possível notar que para todos os cenários e em todas as localizações, o tempo para calcular a função de verossimilhança é menor para a matriz de covariância aproximada.

A redução obtida se dá pelo aumento das  $n$  localizações que consequentemente aumentam a dimensão  $np \times np$  da matriz de covariância  $\boldsymbol{\Sigma}$ , ou seja, não é esperado obter uma redução no tempo do cálculo da função de verossimilhança com diferentes cenários, e sim quando há uma maior quantidade  $n$  localizações. Desta forma, quanto maior é a dimensão de  $\boldsymbol{\Sigma}$ , maior será a redução do tempo computacional para o cálculo da verossimilhança.

A Figura 3 apresenta a redução em percentual obtida no cálculo da função de verossimilhança usando a matriz aproximada, a fim de proporcionar maior entendimento sobre os resultados previamente discutidos na Tabela 2.

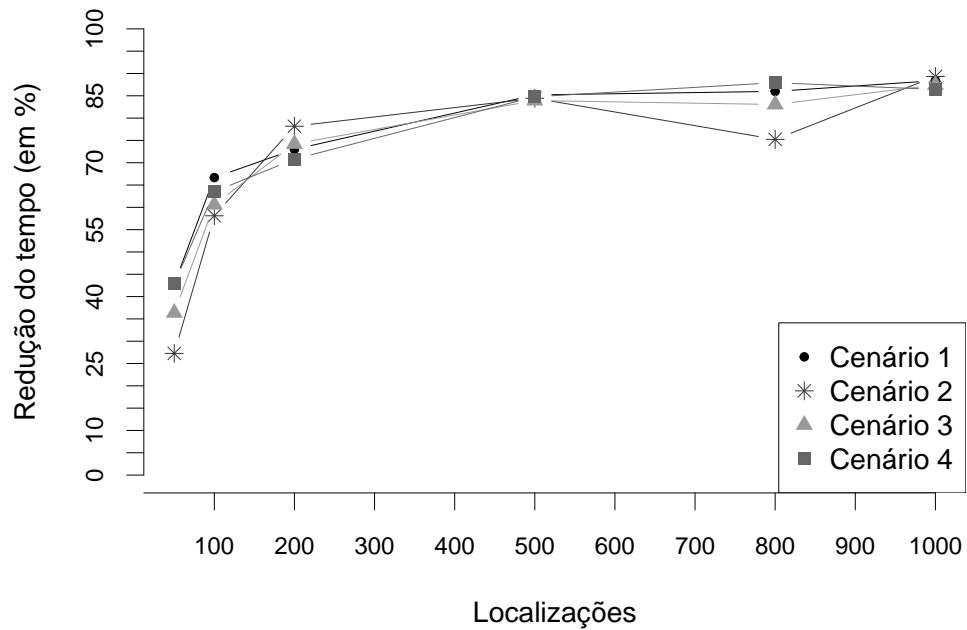


Figura 3: Redução de custo computacional(em percentual) no cálculo da função de verossimilhança em diferentes cenários de separabilidade usando aproximações separáveis.

Nota-se que o tempo para calcular a função de verossimilhança decresce conforme a dimensão da matriz de covariância aumenta. Também é possível notar que a redução do tempo é bem próxima para os diferentes cenários, sendo praticamente a mesma para os quatro cenários no caso em que se tem 1000 localizações, atingindo aproximadamente 85% de redução do tempo, resultado que se mantém dentro do esperado uma vez que a dimensão da matriz de covariância  $\Sigma$  é a responsável pelo custo computacional e não o grau de separabilidade de cada cenário.

### 3.4 Curvas de nível da verossimilhança

Esta análise foi desenvolvida com o objetivo de verificar a semelhança entre as funções de verossimilhança geradas com a matriz de covariância cheia e com a matriz de covariância aproximada, e também o quanto que tais densidades se aproximam do verdadeiro valor de cada parâmetro, o que é possível uma vez este estudo trabalha com dados simulados.

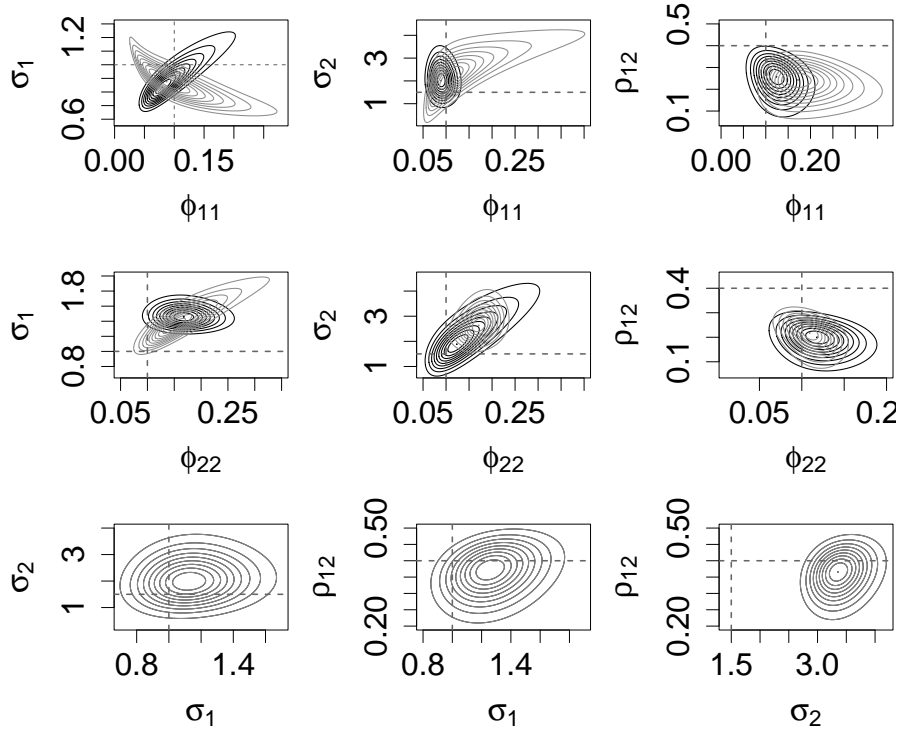


Figura 4: Curvas de contorno no cenário 1. Linhas pretas: matriz de covariância cheia. Linhas cinzas: matriz de covariância aproximada. Linhas tracejadas: o verdadeiro valor de cada parâmetro.

A construção desta análise foi feita para os parâmetros  $\theta = (\phi_{11}, \phi_{22}, \sigma_1, \sigma_2, \rho_{12})$ , onde, por simplicidade, o vetor de médias  $\mu = \mathbf{0}$ , os parâmetros  $\phi_{11}$  e  $\phi_{22}$  foram fixados de acordo com o cenário operativo e  $\sigma_1 = 1$ ,  $\sigma_2 = 1,5$  e  $\rho_{12} = 0,4$ . Foi utilizada apenas uma réplica e 250 localizações em um quadrado  $[0, 1] \times [0, 1]$ , onde cada dado é gerado pelo modelo  $\mathbf{Y} = \mu + \epsilon$  e por se tratar de um processo gaussiano a função de verossimilhança será a denotada na Equação (2.29), com a matriz de covariância  $\Sigma$  definida na Equação (2.8), onde os blocos são as funções mencionadas nas Equações (2.19), (2.20) e (2.21). Também foi considerado o domínio de cada parâmetro a restrição na função de covariância cruzada, de forma que  $\rho_{12}^2 \leq (\phi_{11} + \phi_{22})/(\phi_{12})^2$ , e  $\phi_{12} = (\phi_{11} + \phi_{22})/2$ .

Por se tratar de uma análise bidimensional onde apenas dois parâmetros irão variar e

os demais irão se manter fixos.

A Figura 4 apresenta as curvas de contorno, onde os parâmetros variam em pares e pode-se notar que para o cenário 1 em alguns pares as curvas são muito coincidentes e em alguns casos até idênticas, algo esperado por ser tratar de um cenário de estrutura separável. Nesse caso, como o modelo verdadeiro é o modelo separável, o erro de aproximação é zero pois a aproximação será exatamente o modelo separável.

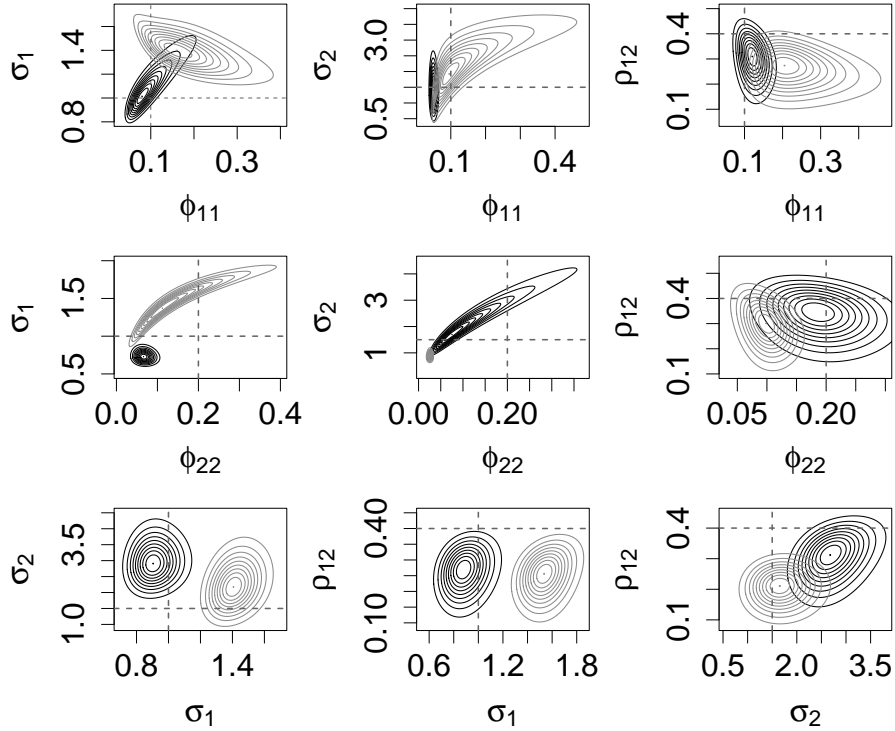


Figura 5: Curvas de contorno no cenário 2. Linhas pretas: matriz de covariância cheia. Linhas cinzas: matriz de covariância aproximada. Linhas tracejadas: o verdadeiro valor de cada parâmetro.

A Figura 5 apresenta as curvas de contorno da função de verossimilhança para o cenário 2, onde é possível notar que as curvas mais distantes entre si e já não há mais curvas sobrepostas como os casos da Figura 4, uma vez que os alcances espaciais agora são diferentes. Ainda sim é possível ver que as curvas se aproximam do verdadeiro valor do parâmetro.

Em um cenário mais não separável como o cenário 3, a Figura 6 apresenta curvas de contorno da função de verossimilhança com a matriz de covariância em sua estrutura não separável mais distante da estrutura aproximada.

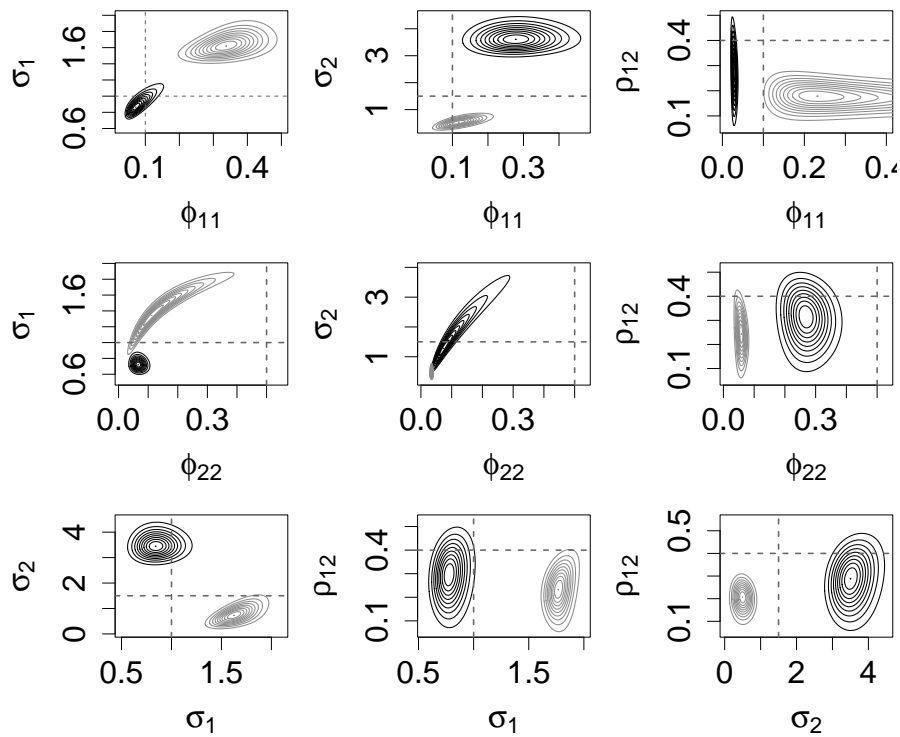


Figura 6: Curvas de contorno no cenário 3. Linhas pretas: matriz de covariância cheia. Linhas cinzas: matriz de covariância aproximada. Linhas tracejadas: o verdadeiro valor de cada parâmetro.

O caso mais não separável abordado neste trabalho, o cenário 4, apresentado pela Figura 7 exibe curvas de contorno com uma maior variação dos valores para alguns parâmetros e maior distância entre si.

De forma geral, essa análise gráfica é limitada. As análises são bidimensionais e podem não retratar adequadamente a relação conjunta verdadeira entre os parâmetros. A análise mais adequada para verificar a qualidade da aproximação em termos de informação dos parâmetros da matriz de covariância é considerar e comparar o ajuste dos modelos cheio e aproximado.



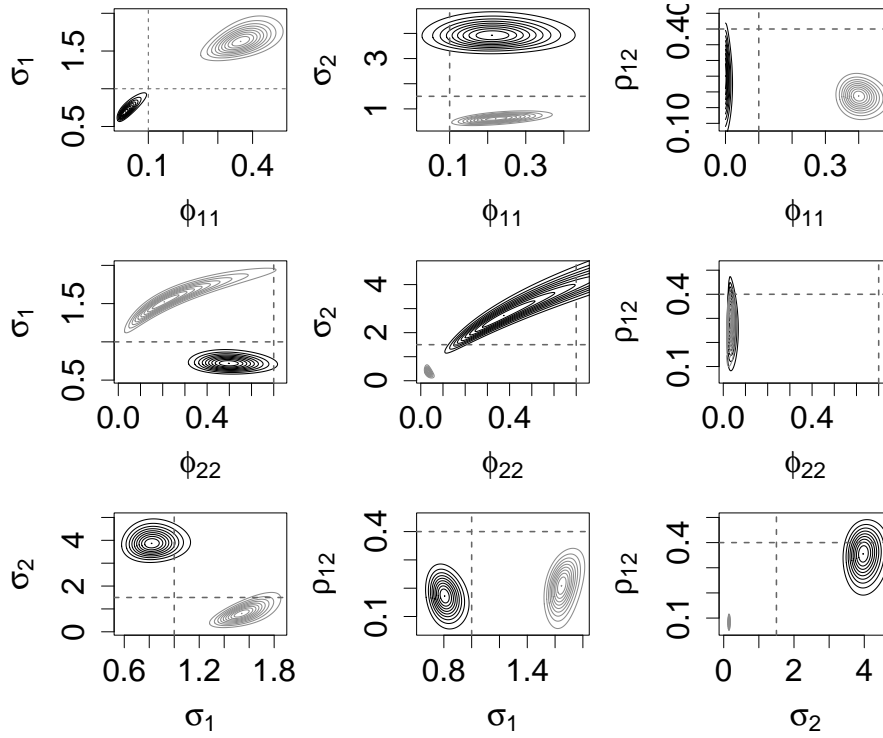


Figura 7: Curvas de contorno no cenário 4. Linhas pretas: matriz de covariância cheia. Linhas cinzas: matriz de covariância aproximada. Linhas tracejadas: o verdadeiro valor de cada parâmetro.

### 3.5 Estimação dos modelos

Nesta seção serão apresentados os resultados obtidos através das simulações estocásticas de Monte Carlo via Cadeias de Markov (MCMC para os casos em que a distribuição condicional completa *a posteriori* não possui forma conhecida e o Algoritmo de Gibbs para quando a distribuição condicional completa é conhecida.

Os dados utilizados na estimação dos modelos são os mesmos gerados na simulação da Seção 3.4. Considere a *priori*  $\beta \sim N_{2(q+1)}(\mathbf{b}, \mathbf{B})$ ,  $q = 0, 1, 2$ ,  $\phi_{kj} \sim Gama(a_{kj}, a_{kj} \times med(h))$ ,  $k, j = 1, 2$ , onde  $med(h)$  é a mediana das distâncias entre as localizações,  $\sigma_{kj} \sim Gama(c_{kj}, d_{kj})$ ,  $k, j = 1, 2$ , como visto na Seção 2.3.

Apresentando os resultados do procedimento de inferência feito sob a ótica bayesiana, a distribuição *a priori* foi definida de forma vaga a fim de que contemplasse diversos valores, a considerar:  $\beta \sim N_6(0; 1000I_6)$ , incluindo o intercepto,  $\phi_{11} \sim Gama(0, 1; 0, 0511)$ ,  $\phi_{22} \sim Gama(0, 1; 0, 0511)$ ,  $\sigma_1 \sim Gama(1; 0, 4)$ ,  $\sigma_2 \sim Gama(1; 0, 4)$ , e, por simplicidade,  $\rho_{12}$  não foi estimado.

Este trabalho buscou obter resultados sobre a distribuição *a posteriori* ao utilizar a

matriz de covariância  $\Sigma$  em sua estrutura não separável e na estrutura aproximada, percorrendo os quatro cenários previamente definidos na Seção 3.2. Para obter as amostras das distribuições *a posteriori*, foram utilizados os métodos de MCMC, com 100.000 iterações, retirando o *burn-in* de 50.000 iterações e com saltos de 50 iterações, resultando em uma amostra de tamanho 1.000 para cada elemento do vetor paramétrico.

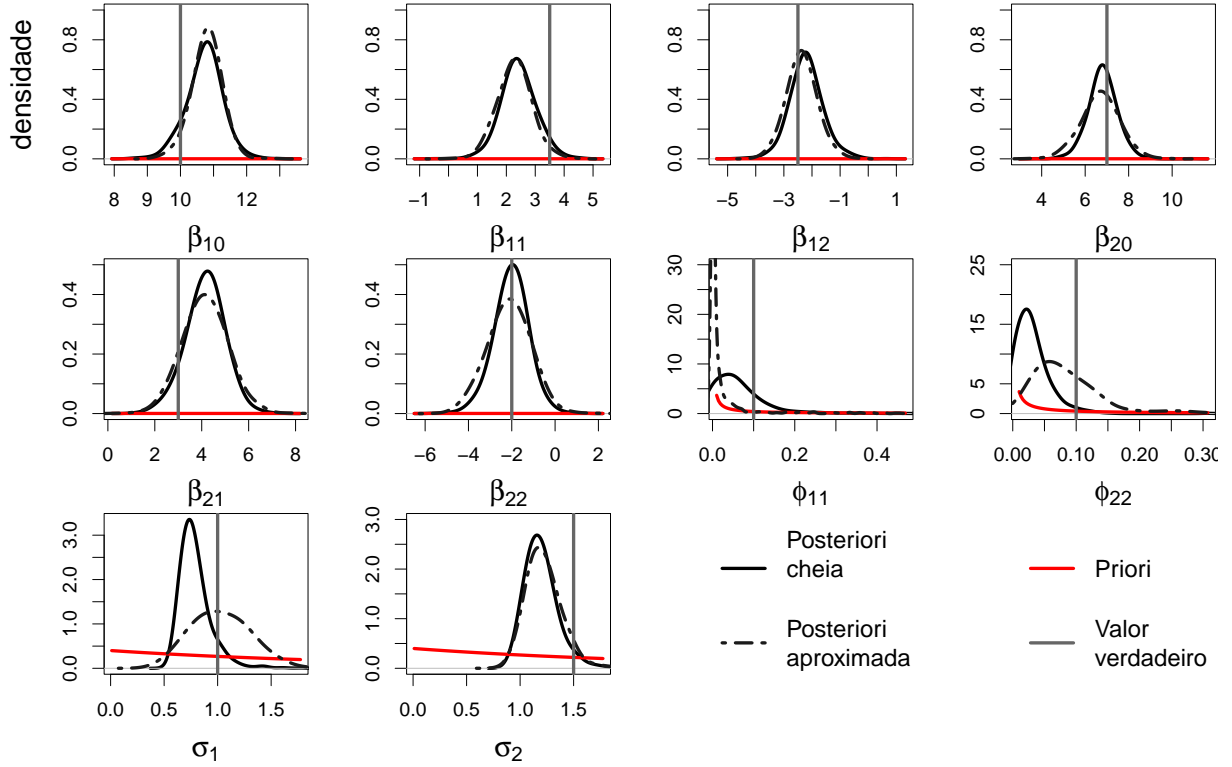


Figura 8: Curvas das densidades das distribuições *a posteriori* dos parâmetros geradas com a estrutura não separável, a estrutura aproximada e as respectivas distribuições *a priori* no cenário 1

A Figura 8 mostra as densidades das distribuições *a priori*, as distribuições *a posteriori* obtidas via MCMC, e o valor verdadeiro de cada parâmetro no cenário 1, feito que é possível devido a este trabalho utilizar dados simulados. Nota-se o ganho de informação no uso das distribuições *a posteriori* de estrutura não separável e de estrutura aproximada, além de que essas duas distribuições *a posteriori* se mostram bem próximas entre si no cenário 1, cenário este que assume que as duas variáveis terão o mesmo comportamento ao longo do espaço. MU

A Tabela 3 mostra a estimativa intervalar através do intervalo de credibilidade e a estimativa pontual pela média *a posteriori* para os parâmetros no cenário 1, a considerar a matriz de covariância sendo construída pela estrutura não separável e a estrutura aproximada. Pode-se observar que o intervalo de credibilidade contém o verdadeiro valor do parâmetro para todos os casos, com exceção do parâmetro  $\beta_{11}$  na estrutura aproximada.

Tabela 3: Resumo da distribuição a *posteriori* para os parâmetros do cenário 1, considerando a estrutura de modelos não separável e a estrutura aproximada

Parâmetro	Estrutura Não Separável			Estrutura Aproximada			Valor verdadeiro
	Média	2,5%	97,5%	Média	2,5%	97,5%	
$\beta_{01}$	10,728	9,614	11,767	10,794	9,804	11,573	10
$\beta_{11}$	2,422	1,242	3,655	2,281	1,191	3,441	3,5
$\beta_{12}$	-2,206	-3,38	-0,892	-2,366	-3,441	-1,275	-2,5
$\beta_{02}$	6,81	5,575	8,114	6,696	4,965	8,346	7
$\beta_{21}$	4,153	2,388	5,678	4,14	2,365	6,125	3
$\beta_{22}$	-2,03	-3,672	-0,517	-2,123	-4,028	-0,366	-2
$\phi_{11}$	0,062	1.35e-18	0,244	0,032	7.41e-09	0,357	0,1
$\phi_{22}$	0,031	8.17e-08	0,096	0,085	0,021	0,258	0,1
$\sigma_1$	0,79	0,603	1,149	1,027	0,549	1,558	1
$\sigma_2$	1,195	0,955	1,547	1,22	0,954	1,569	1,5

É possível também constatar que as médias de ambas as estruturas são próximas entre si e também do verdadeiro valor do parâmetro.

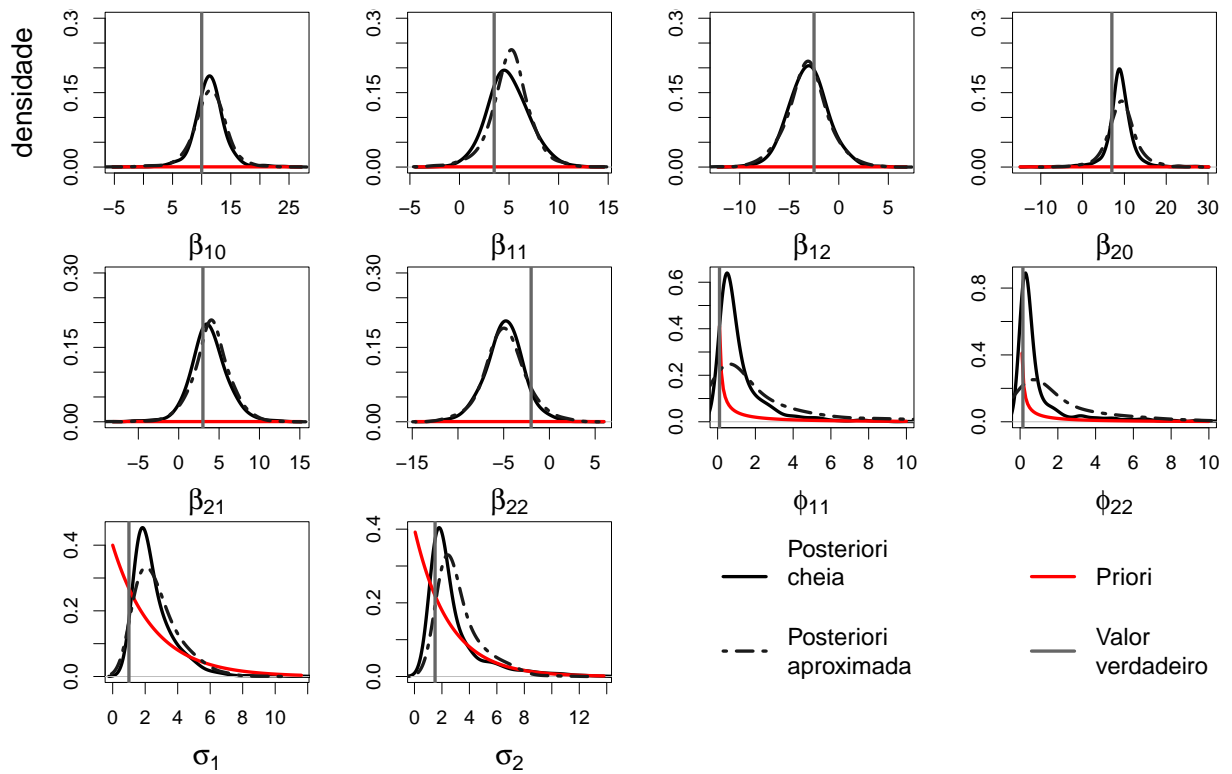


Figura 9: Curvas das densidades das distribuições *a posteriori* dos parâmetros geradas com a estrutura não separável, a estrutura aproximada e as respectivas distribuições *a priori* no cenário 2

Avaliando o cenário 2, é possível notar que as densidades das amostras da distribuição *a posteriori* estão bastante similares entre si e favorecem uma atualização satisfatória dos dados da distribuição *a priori*, com ambas se mantendo próximas ao verdadeiro valor de cada parâmetro.

Tabela 4: Resumo da distribuição a *posteriori* para os parâmetros do cenário 2, considerando a estrutura de modelos não separável e a estrutura aproximada

Parâmetro	Estrutura Não Separável			Estrutura Aproximada			Valor verdadeiro
	Média	2,5%	97,5%	Média	2,5%	97,5%	
$\beta_{01}$	11,386	5,862	16,296	11,426	5,306	17,315	10
$\beta_{11}$	4,944	1,215	9,247	5,124	0,482	9,033	3,5
$\beta_{12}$	-3,071	-6,811	0,941	-3,13	-7,338	1,306	-2,5
$\beta_{02}$	9,018	2,297	15,503	9,325	1,822	17,2	7
$\beta_{21}$	3,757	-0,349	8,442	4,068	-0,5	8,635	3
$\beta_{22}$	-4,912	-9,024	-0,791	-4,838	-9,591	-0,058	-2
$\phi_{11}$	1,249	0,177	5,525	2,731	0,099	14,514	0,1
$\phi_{22}$	1,165	0,048	7,623	2,698	0,154	14,832	0,2
$\sigma_1$	2,523	1,258	5,821	2,732	1,054	5,847	1
$\sigma_2$	2,828	1,147	9,077	3,237	1,41	7,274	1,5

Ao analisar a estimativa pontual pela média *posteriori* e o intervalo de credibilidade, apresentados na Tabela 4, observa-se que em alguns casos as médias da estrutura não separável e da estrutura aproximada são mais distintas entre si, em comparação às estimativas de ambas as estruturas do cenário 1. O verdadeiro valor do parâmetro está contido no intervalo de credibilidade de todos os parâmetros, exceto para  $\phi_{11}$  na estrutura não separável e para o parâmetro  $\sigma_1$  as duas estruturas não contêm o verdadeiro valor do parâmetro.

A Figura 10 exibe um comportamento da densidade da amostra da distribuição a *posteriori* na estrutura aproximada mais distinto da densidade da *posteriori* na estrutura não separável. No entanto, em todos os casos, a distribuição a *posteriori* das duas estruturas se mantém bem próxima ao verdadeiro valor dos parâmetros apresentados.

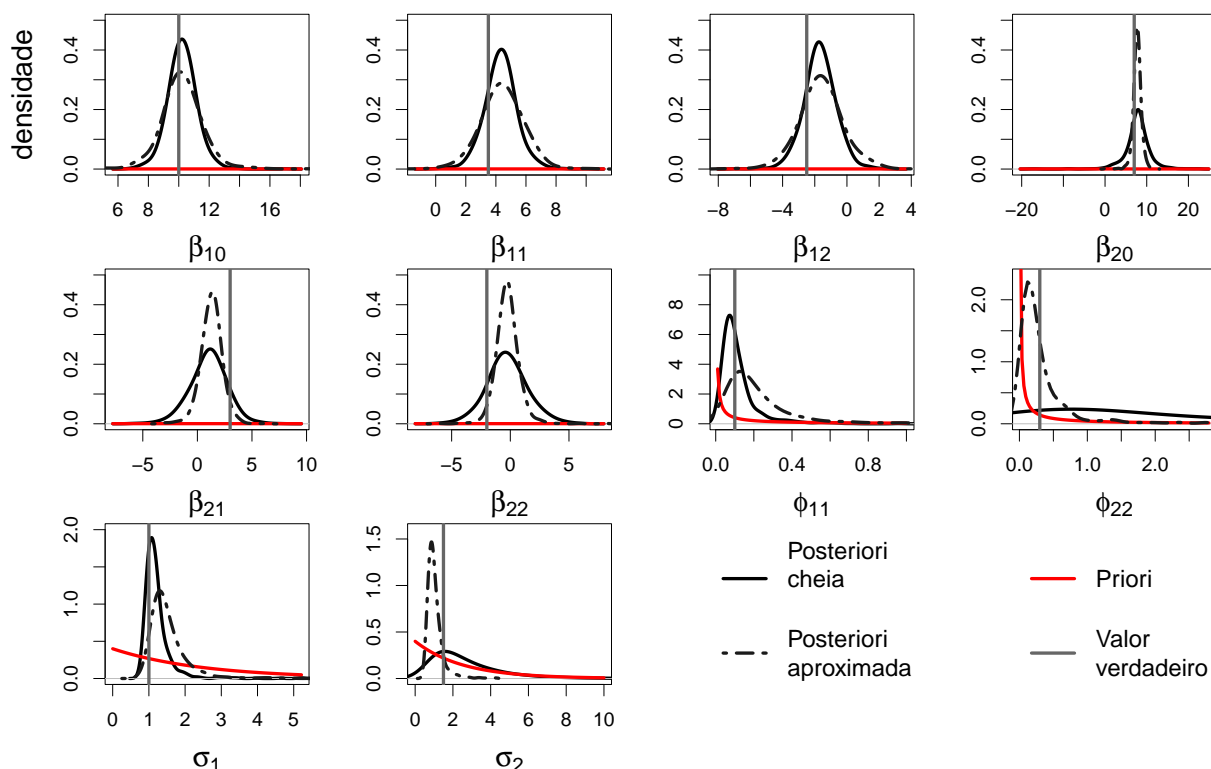


Figura 10: Curvas das densidades das distribuições *a posteriori* dos parâmetros geradas com a estrutura não separável, a estrutura aproximada e as respectivas distribuições *a priori* no cenário 3

Tabela 5: Resumo da distribuição *a posteriori* para os parâmetros do cenário 3, considerando a estrutura de modelos não separável e a estrutura aproximada

Parâmetro	Estrutura Não Separável			Estrutura Aproximada			Valor verdadeiro
	Média	2,5%	97,5%	Média	2,5%	97,5%	
$\beta_{01}$	10,231	8,381	12,126	10,187	7,402	12,934	10
$\beta_{11}$	4,325	2,27	6,446	4,421	1,709	7,028	3,5
$\beta_{12}$	-1,678	-3,597	0,169	-1,601	-4,253	1,303	-2,5
$\beta_{02}$	7,757	1,634	13,58	7,745	5,729	9,767	7
$\beta_{21}$	1,024	-2,342	4,066	1,285	-0,617	3,175	3
$\beta_{22}$	-0,258	-3,639	3,294	-0,248	-2,012	1,67	-2
$\phi_{11}$	0,134	0,027	0,483	0,279	0,052	1,124	0,1
$\phi_{22}$	2,936	0,109	16,777	0,408	0,025	2,342	0,5
$\sigma_1$	1,209	0,852	2,056	1,536	0,962	2,95	1
$\sigma_2$	2,607	0,812	7,624	1,034	0,6	2,166	1,5

A fim de avaliar o comportamento da média *a posteriori* e o intervalo de credibilidade para o cenário 3, a Tabela 5 mostra que o valor real do parâmetro está contido em todos os intervalos de credibilidade para ambas as estruturas e as médias *a posteriori* são similares antes da primeira casa decimal, com exceção para os parâmetros  $\phi_{22}$  e  $\sigma_2$ .

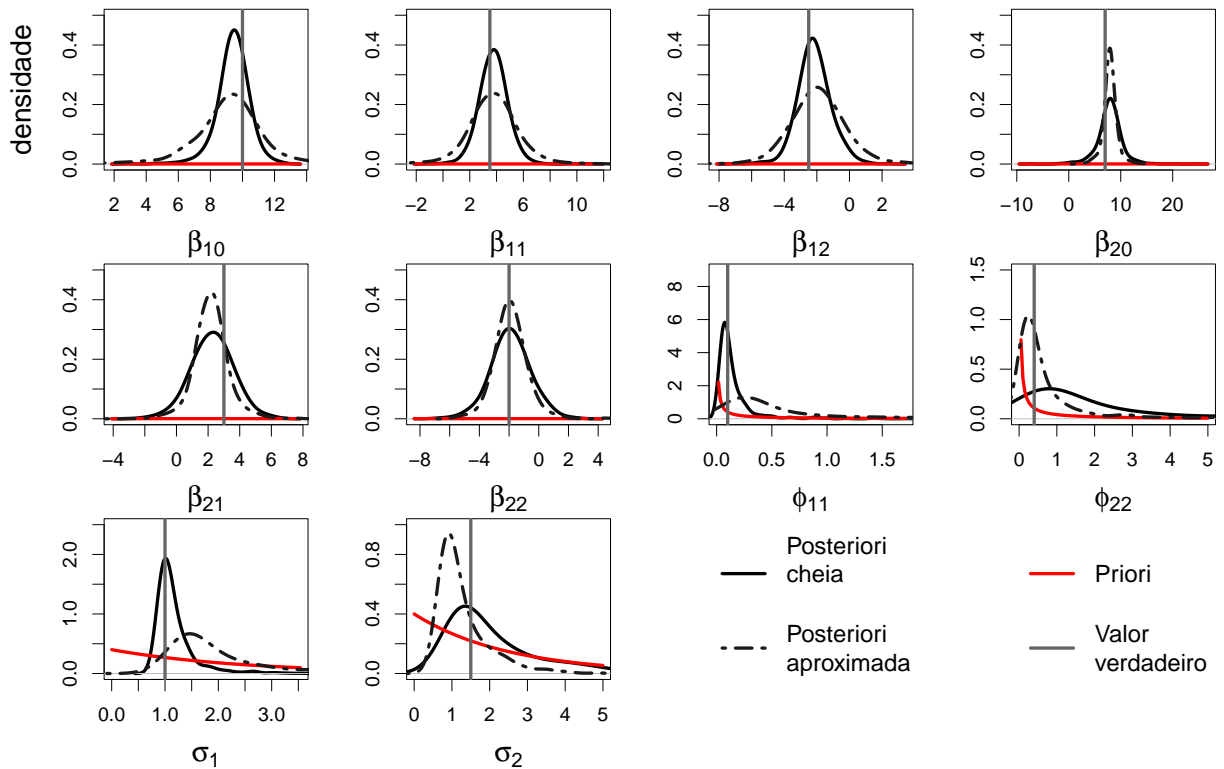


Figura 11: Curvas das densidades das distribuições *a posteriori* dos parâmetros geradas com a estrutura não separável, a estrutura aproximada e as respectivas distribuições *a priori* no cenário 4

Tabela 6: Resumo da distribuição *a posteriori* para os parâmetros do cenário 4, considerando a estrutura de modelos não separável e a estrutura aproximada

Parâmetro	Estrutura Não Separável			Estrutura Aproximada			Valor verdadeiro
	Média	2,5%	97,5%	Média	2,5%	97,5%	
$\beta_{01}$	9,421	7,267	11,277	9,192	4,067	13,702	10
$\beta_{11}$	3,771	1,725	5,765	3,768	0,233	7,3	3,5
$\beta_{12}$	-2,152	-3,986	-0,122	-1,935	-4,927	1,33	-2,5
$\beta_{02}$	7,85	2,502	12,69	7,934	4,879	10,885	7
$\beta_{21}$	2,262	-0,368	4,802	2,175	-0,03	4,529	3
$\beta_{22}$	-1,894	-4,642	0,796	-1,902	-4,018	0,427	-2
$\phi_{11}$	0,147	0,02	0,663	0,789	0,099	4,543	0,1
$\phi_{22}$	2,661	0,206	13,543	0,936	0,057	5,939	0,7
$\sigma_1$	1,165	0,81	2,224	2,005	1,045	4,813	1
$\sigma_2$	2,109	0,82	5,663	1,269	0,604	3,163	1,5

Adentrando o cenário com a maior diferença entre os alcances espaciais, a Figura 11 apresenta as densidades das distribuições *a posteriori* bem próximas do valor verdadeiro e um pouco mais distantes entre si, em comparação aos demais cenários.

Para o cenário 4, a Tabela 6 mostra que as médias das duas estruturas estão mais distantes entre si e que o intervalo de credibilidade contém o valor real do parâmetro em

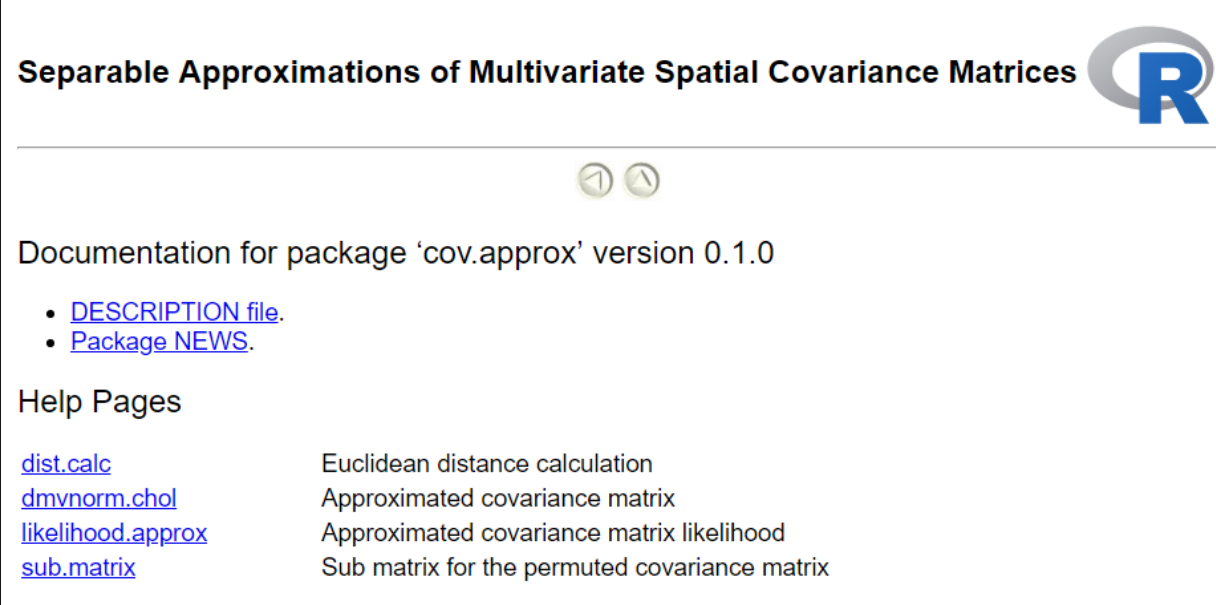
todos os casos, exceto para o parâmetro  $\sigma_1$  na estrutura aproximada, com uma pequena diferença entre o valor real.


De forma geral, independente do cenário operativo avaliado, o ajuste dos parâmetros do modelo aproximado foi próximo em comparação ao modelo não separável, indicando que as aproximações em termos de ajuste são satisfatórias.

## 3.6 Pacote desenvolvido

No decorrer deste trabalho, um pacote no *software* R foi desenvolvido com todas as funções usadas para calcular a função de verossimilhança com o uso das aproximações separáveis. O pacote foi denominado `cov.approx` e foi submetido ao CRAN (*The Comprehensive R Archive Network*) e segue aguardando aprovação para ser devidamente publicado e estar disponível de forma gratuita.

No pacote `cov.approx` foram construídas quatro funções, são elas: `dist.calc()`, `dmvnorm.chol()`, `likelihood.approx()` e `sub.matrix()`, como mostra a Figura 12.



**Separable Approximations of Multivariate Spatial Covariance Matrices** 

Documentation for package 'cov.approx' version 0.1.0

- [DESCRIPTION file](#).
- [Package NEWS](#).

Help Pages

<a href="#">dist.calc</a>	Euclidean distance calculation
<a href="#">dmvnorm.chol</a>	Approximated covariance matrix
<a href="#">likelihood.approx</a>	Approximated covariance matrix likelihood
<a href="#">sub.matrix</a>	Sub matrix for the permuted covariance matrix

Figura 12: Descrição do pacote `cov.approx` no *software* R e suas respectivas funções, variáveis e autores

A função `dist.calc()` é responsável por calcular a distância euclidiana entre as localizações, como mostra a Figura 13. Caso o usuário tenha problemas ao utilizar a função, basta consultar as informações contidas na Figura 13 ou, caso seja uma situação em que o tipo da variável não é adequado, como se for usada uma variável do tipo `char` em uma variável que deveria ser `numeric`, a função irá retornar o erro específico.



dist.calc {cov.approx}	R Documentation
<b>Euclidean distance calculation</b>	
<b>Description</b>	
Computes the <u>Euclidean distance</u> between two points. Given by the expression $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$	
<b>Usage</b>	
<code>dist.calc(x1, y1, x2, y2)</code>	
<b>Arguments</b>	
x1 numeric; the x coordinate of the first point.	
y1 numeric; the y coordinate of the first point.	
x2 numeric; the x coordinate of the second point.	
y2 numeric; the y coordinate of the second point.	
<b>Value</b>	
The Euclidean distance between two points.	
<b>Author(s)</b>	
Caroline Ferreira da Cruz Gomes <carolinefcg@id.uff.br>	
Rafael Santos Erbisti <rebisti@id.uff.br>	
Thamires Louzada Marques <thamireslouzada@id.uff.br>	

Figura 13: Função `dist.calc()` do pacote `cov.approx` no *software* R e suas respectivas variáveis e autores

dmvnorm.chol {cov.approx}
R Documentation

## Approximated covariance matrix

### Description

Computes the approximated covariance matrix  $\Sigma$  of dimension  $np \times np$  (where  $n$  stand for the number of locations and  $p$  for the number of variables) by calculating the Kronecker product of matrices  $A$  and  $R$ .

### Usage

```
dmvnorm.chol(
  y_vector,
  R,
  A,
  mu_vector = rep(0, length = ncol(y_vector)),
  log = FALSE
)
```

### Arguments

`y_vector` vector or matrix of quantiles. If `y_vector` is a matrix, each row is taken to be a quantile.

`R` matrix; the correlations matrix of dimension  $n \times n$ .

`A` matrix; it should be positive definite of dimension  $p \times p$ .

`mu_vector` mean vector; default is `rep(0, length = ncol(x))`.

`log` logical; indicates if the logarithm scale should be applied (log-normal distribution). The default is `FALSE`.

### Value

The approximated covariance matrix.

### Author(s)

Caroline Ferreira da Cruz Gomes <carolinefcg@id.uff.br>  
 Rafael Santos Erbisti <rebisti@id.uff.br>  
 Thamires Louzada Marques <thamireslouzada@id.uff.br>

Figura 14: Função `dmvnorm.chol()` do pacote `cov.approx` no *software* R e suas respectivas variáveis e autores

Com o objetivo de calcular a matriz de covariância  $\Sigma$  utilizando as aproximações separáveis, a função `dmvnorm.chol()` é uma subfunção da função `likelihood.aprox()`, apresentada pela Figura 14. A função `dmvnorm.chol()` irá calcular o produto de Kronecker de duas matrizes  $\mathbf{A}$  e  $\mathbf{R}$ , que serão calculadas previamente dentro da função `likelihood.aprox()`. Assim como as demais funções do pacote `cov.approx`, caso as variáveis de entrada da função `dmvnorm.chol()` não sejam adequadas devido à sua natureza, a função retornará o erro para que o usuário possa corrigi-lo.

likelihood.aprox {cov.approx}
R Documentation

## Approximated covariance matrix likelihood

### Description

Computes the likelihood of the approximated covariance matrix.

### Usage

```
likelihood.aprox(y_vector, mu_vector, mat_full, log = FALSE)
```

### Arguments

`y_vector` numeric; the response variable array with the estimated values.

`mu_vector` numeric; the mean values array. It should be an array composed by real numbers with the mean values. The default is

`mat_full` matrix; the full covariance matrix, composed by real numbers. Its dimension should be  $2n \times 2n$ , where "n" is the number of locations.

`log` logical; it indicates if the logarithm scale should be applied (log-normal distribution). The default is `FALSE`.

### Value

The approximated covariance matrix likelihood.

### Author(s)

Caroline Ferreira da Cruz Gomes <carolinefcg@id.uff.br>  
 Rafael Santos Erbisti <rerbisti@id.uff.br>  
 Thamires Louzada Marques <thamireslouzada@id.uff.br>

Figura 15: Função `likelihood.aprox()` do pacote `cov.approx` no *software* R e suas respectivas variáveis e autores

A Figura 15 mostra a função `likelihood.aprox()`, função amplamente utilizada nos resultados contidos neste trabalho, uma vez que é ela que irá calcular a função de verossimilhança utilizando a matriz de covariância de estrutura aproximada. O algoritmo

presente na função para obter as matrizes  $\mathbf{A}$  e  $\mathbf{R}$  é similar ao apresentado na Subseção [2.2.3](#). Nessa função também é possível especificar se a função de verossimilhança será a normal ou a log-normal através do parâmetro da função `log`. Esta função também irá averiguar se as entradas são adequadas para obter os resultados corretos.

sub.matrix {cov.approx}
R Documentation

## Sub matrix for the permuted covariance matrix

### Description

Subsets a matrix in partitions based on the chosen amount of rows and columns. Click [here](#) for more details.

### Usage

```
sub.matrix(mat, r, c)
```

### Arguments

`mat` matrix; the matrix to be subset.  
`r` numeric; the number of rows of each subset matrix.  
`c` numeric; the number of columns of each subset matrix.

### Value

The sub matrix.

### Author(s)

Caroline Ferreira da Cruz Gomes <carolinefcg@id.uff.br>  
Rafael Santos Erbisti <rerbisti@id.uff.br>  
Thamires Louzada Marques <thamireslouzada@id.uff.br>

Figura 16: Função `sub.matrix()` do pacote `cov.approx` no *software* R e suas respectivas variáveis e autores

A Figura [16](#) exhibe a função `sub.matrix()`, que tem por objetivo obter submatrizes de uma dada matriz com base em um determinado número de linhas e colunas que, por definição, uma submatriz de dimensão  $q \times w$  extraída de uma matriz  $n \times p$ , tem por restrição que  $q \leq n$  e  $w \leq p$ . Esta restrição também está implementada na função, assim como a verificação de adequação das variáveis de entrada.

Vale ressaltar que o pacote é capaz de realizar aproximações para qualquer  $p \geq 2$ . Neste trabalho as análises foram restritas apenas aos modelos bivariados.

## 4 Conclusões

Os objetivos deste trabalho de conclusão de curso e também projeto de iniciação científica financiado pela Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) consistem em averiguar o quanto as aproximações separáveis podem reduzir o custo computacional no cálculo da função de verossimilhança em comparação ao modelo cheio, e fazer inferência sobre os dados utilizando as duas estruturas e também de permitir que estas funções possam ser utilizadas de forma livre e gratuita para usuários que utilizam o *software* R.

A escolha por utilizar modelos não separáveis foi dada com o principal objetivo de permitir que dados mais próximos à realidade sejam mais explorados, uma vez que a estrutura não separável torna possível que as variáveis possuam comportamentos distintos no espaço, proporcionando análises mais realísticas.

O início deste trabalho foi dado na avaliação dos erros de aproximação padronizados, a fim de que fosse conhecido qual o comportamento de tal erro e se seria possível construir os cenários presentes na Tabela 2 e uma vez que tal erro demonstrou baixo valor (atingindo valor máximo de 8% de erro, como apresentado na Seção 2.2.4), concluiu-se que é viável trabalhar com a diferença 0,6 entre os alcances espaciais e que o erro de aproximação padronizado não apresentou padrões de crescimento caso a diferença entre os alcances espaciais fosse ainda maior. Observou-se também que o comportamento da função de correlação entre as duas variáveis simuladas tiveram resultados cada vez mais distintos conforme o comportamento delas ao longo do espaço também se tornou mais diferente.

O propósito deste trabalho foi lidar com dados multivariados, mais especificamente do caso bivariado. Nele, foi apresentada a função de covariância cruzada para o caso não separável definida nas Equações (2.19), (2.20) e (2.21), de forma que a matriz de covariância foi construída em blocos como visto na Equação (2.8). Ao longo de todo trabalho, essa matriz de covariância gerada por funções não separáveis foi aproximada, e assim foi possível reduzir consideravelmente o tempo de processamento com o uso de

tais aproximações. Foi possível notar também que tal redução de custo computacional foi dada devido à dimensionalidade da matriz de covariância cruzada, que varia de de acordo com o número de localizações e variáveis utilizadas.

Os cenários construídos permitiram que a função `likelihood.approx()` desenvolvida para o pacote `cov.approx` alcançasse resultados satisfatórios nos casos em que a estrutura é separável e também quando a estrutura não é separável.

A análise bidimensional feita pelas curvas de contorno do cálculo da função de verossimilhança mostrou que as duas estruturas (cheia e aproximada) se distanciaram entre si e do valor real do parâmetro conforme os cenários se tornaram cada vez mais não separáveis. Entende-se que a limitação de variar apenas dois parâmetros por vez e usar apenas uma réplica possa ter levados a resultados não tão favoráveis em cenários que possuem um alcances espaciais tão distintos. Os resultados da estimação sob a ótica da inferência bayesiana foram satisfatórios para a matriz de covariância aproximada e se mostram próximos aos valores verdadeiros.

Para trabalhos futuros, é interessante construir funções para dados multivariados com maior dimensão superior ao caso bivariado, assim como explorar outras famílias de funções de covariância.

Além disso, espera-se realizar ajuste de modelos com mais réplicas para que seja possível comparar os resultados para amostras distintas. É de interesse também generalizar as análises para o caso de  $p \geq 3$ .

Este trabalho, seus resultados e o pacote submetidos ao CRAN são essenciais para a otimização do tempo computacional necessário ao se trabalhar com dados espaciais bivariados.

# Referências

- BANERJEE, S.; CARLIN, B. P.; GELFAND, A. E. *Hierarchical Modeling and Analysis for Spatial Data. Monographs on Statistics and Applied Probability*. 1st. ed. [S.l.]: Chapman & Hall/CRC, 2004.
- CRESSIE, N. *Statistics for Spatial Data*. [S.l.]: New York: Wiley, 1993.
- DEGROOT, M. H.; SCHERVISH, M. J. *Probability and statistics*. 3. ed. [S.l.]: Boston: Addison-Wesley Pub, 2002.
- ERBISTI, R. S.; FONSECA, T. C. O.; ALVES, M. B. *Covariance Modeling for Multivariate Spatial Processes Based on Separable Approximations*. [S.l.]: Springer Proceedings in Mathematics Statistics, 2018.
- FONSECA, T. C. O.; STEEL, M. F. J. *A General Class of Nonseparable Space-time Covariance Models*. 22. ed. [S.l.]: Environmetrics, 2011. 2,224-242 p.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *EEE Transactions on Pattern Analysis and Machine Intelligence*, v. 6, n. 6, 1984.
- GENTON, M. *Separable Approximations Of Space-time Covariance Matrices*. 18 681–695. ed. [S.l.]: Environmetrics, 2007.
- GOLUB, G. H.; LOAN, C. F. V. *Matrix Computations*. [S.l.]: The Johns Hopkins University Press, 1996.
- GOULARD; VOLTZ. *Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix*. 24(3):269-286. ed. [S.l.]: Journal of the International Association for Mathematical Geology, 1992.
- HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, v. 57, n. 1, 1970.
- MA, C. *Spatio-temporal covariance functions generated by mixtures*. 34. ed. [S.l.]: Mathematical Geology, 2002. 965–975 p.
- MA, C. *Spatio-temporal stationary covariance models. Journal of Multivariate analysis*. 86. ed. [S.l.]: Journal of Multivariate Analysis, 2003. 97–107 p.
- METROPOLIS, N. et al. Equation of state calculations by fast computing machines. *THE JOURNAL OF CHEMICAL PHYSICS*, v. 21, n. 6, 1953.
- MOUSAVIFAZL, H.; ALIZADH, A.; GHAHRAMAN, B. Application of geostatistical methods for determining nitrate concentrations in groundwater (case study of mashhad plain, iran). *International Journal of Agriculture and Crop Sciences*, 2013.

- PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. *Estatística Bayesiana*. [S.l.]: Fundação Calouste Gulbenkian, 2003.
- PORCU, E.; BEVILACQUA, M.; GENTON, M. G. *Spatio-Temporal Covariance and Cross Covariance Functions of the Great Circle Distance on a Sphere*. 111, 514, 888–898. ed. [S.l.]: Journal of the American Statistical Association, 2016.
- PORCU, E.; ZASTAVNYI, V. *Characterization theorems for some classes of covariance functions associated to vector valued random fields*. 102. ed. [S.l.]: Journal of Multivariate Analysis, 2011. 1293–1301 p.
- SCHMIDT, A. M.; SANSÓ, B. *Modelagem Bayesiana da Estrutura de Covariância de Processos Espaciais e Espaço Temporais*. 1. ed. [S.l.]: 17th SINAPE, 2006.
- STEEB, W. H. *Matrix Calculus and Kronecker Product with Applications and C++ Programs*. [S.l.]: World Scientific: Singapore, 1997.
- WACKERNAGEL, H. *Multivariate Geostatistics: An Introduction with Applications*. 111, 514, 888–898. ed. [S.l.]: Springer-Verlag, Berlin, 2003.



## APÊNDICE 1 – Curvas de correlação espacial

Como mencionado na Seção 2.1, este trabalho irá adotar a função de correlação Cauchy descrita em Porcu, Bevilacqua e Genton (2016), uma vez que tal função de correlação espacial possui um decaimento mais lento e permite longos alcances, isto é, em casos em que as variáveis têm uma dependência espacial que se estica muito ao longo do espaço. As Figuras 2 apresenta as funções de correlação espacial utilizando a Cauchy e a Exponencial em cada um dos cenários operativos descritos na Tabela 1. Para este estudo, os parâmetros  $\phi_{11}$  e  $\phi_{22}$  foram fixados de acordo com o respectivo cenário e os demais parâmetros foram fixados em  $\rho_{12} = 0,4$ ,  $\sigma_1 = 1$ ,  $\sigma_2 = 2$  e, por simplicidade, o vetor de médias  $\boldsymbol{\mu} = \mathbf{0}$ .

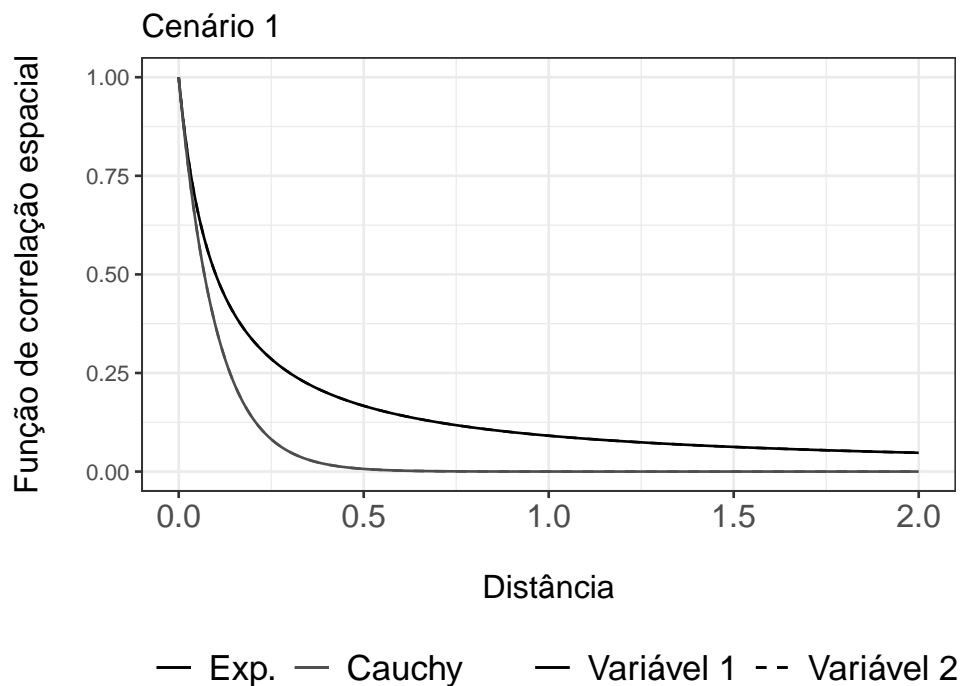


Figura 17: Comportamento das funções de correlação Cauchy e Exponencial para o cenário 1

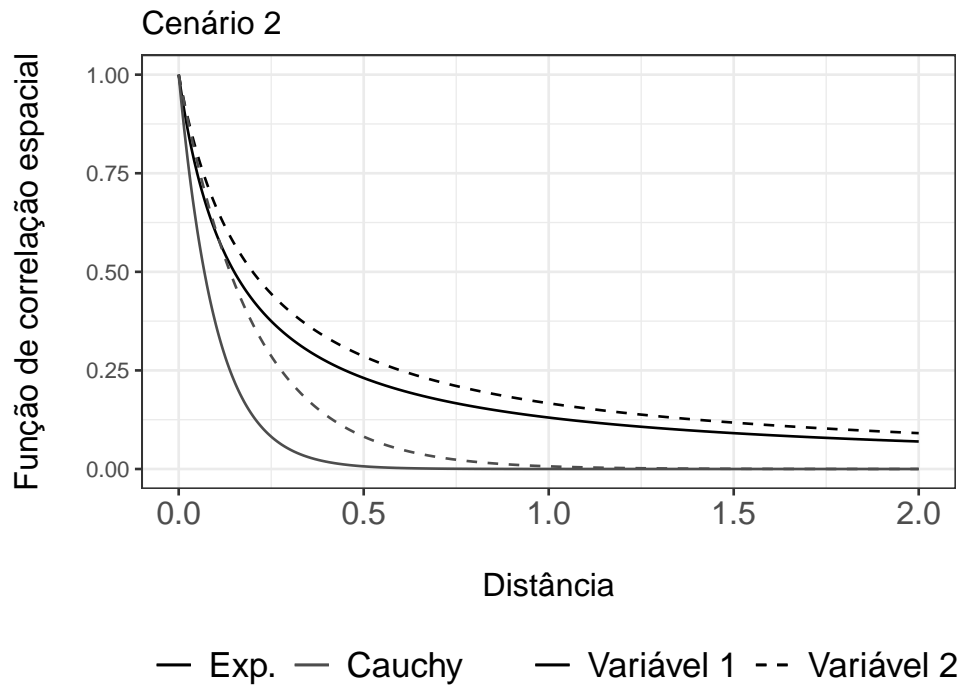


Figura 18: Comportamento das funções de correlação Cauchy e Exponencial para o cenário 2

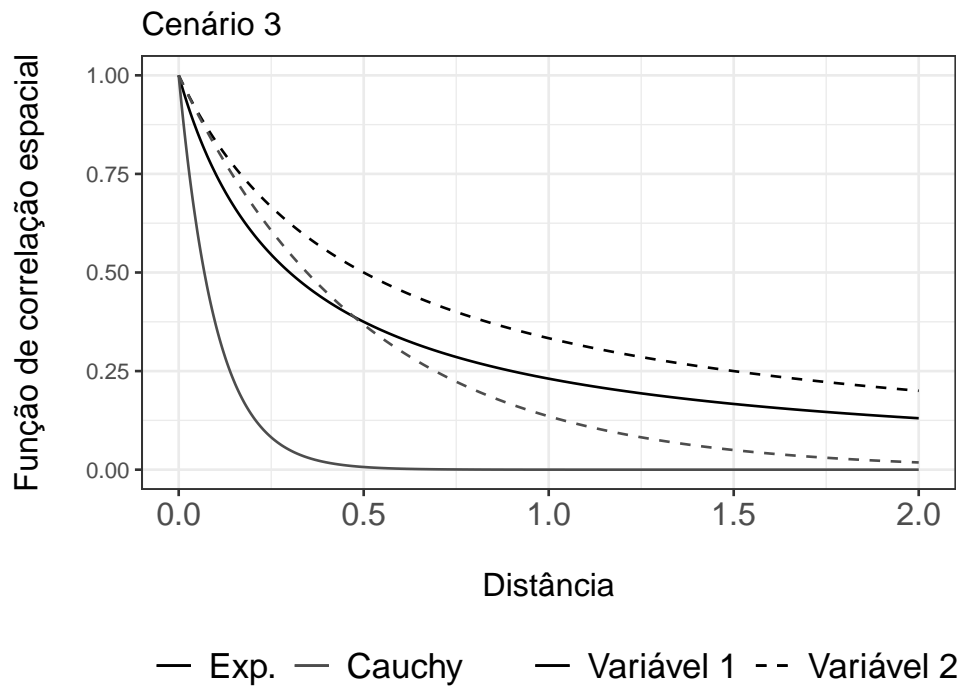


Figura 19: Comportamento das funções de correlação Cauchy e Exponencial para o cenário 3

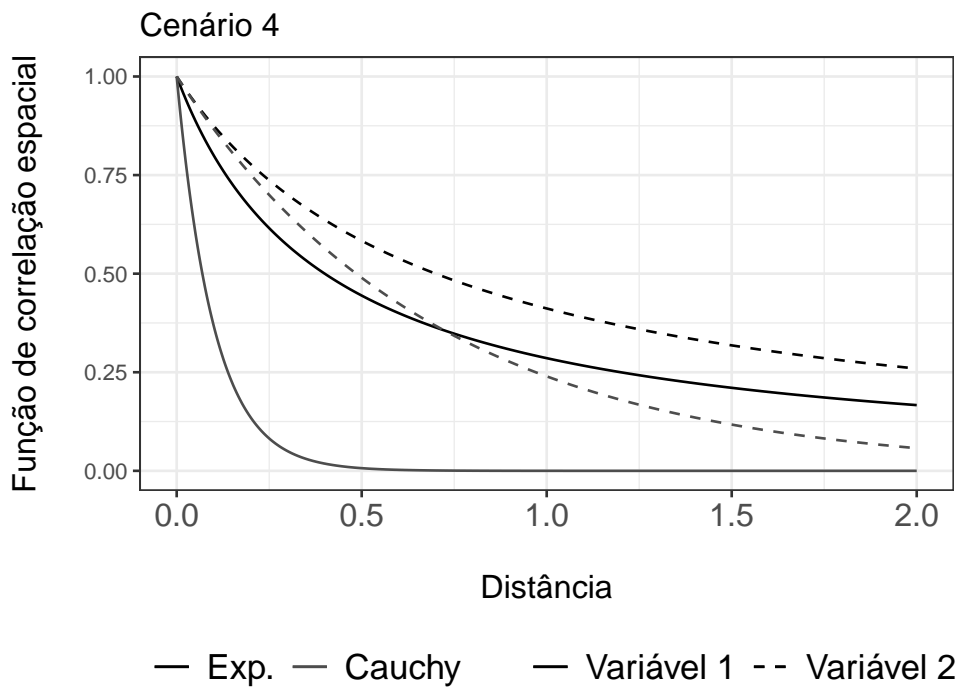


Figura 20: Comportamento das funções de correlação Cauchy e Exponencial para o cenário 4

## APÊNDICE 2 – Cadeias geradas pelo método MCMC

Este apêndice exibe as cadeias de Markov geradas pelo método MCMC, conforme mencionado na Seção 2.3, para cada cenário definido na Seção 3.2 e utilizando a matriz de covariância em sua estrutura não separável e na sua estrutura aproximada.

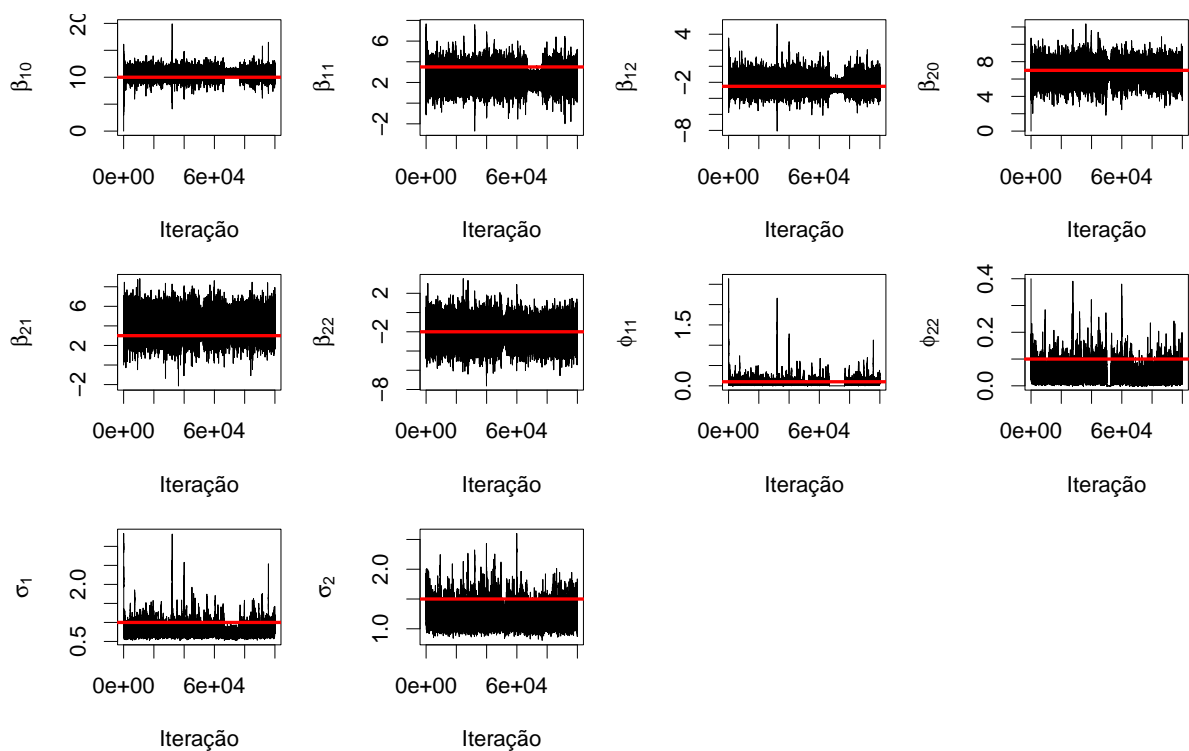


Figura 21: Cadeias geradas pelo método MCMC para cada parâmetro no cenário 1 para a estrutura cheia

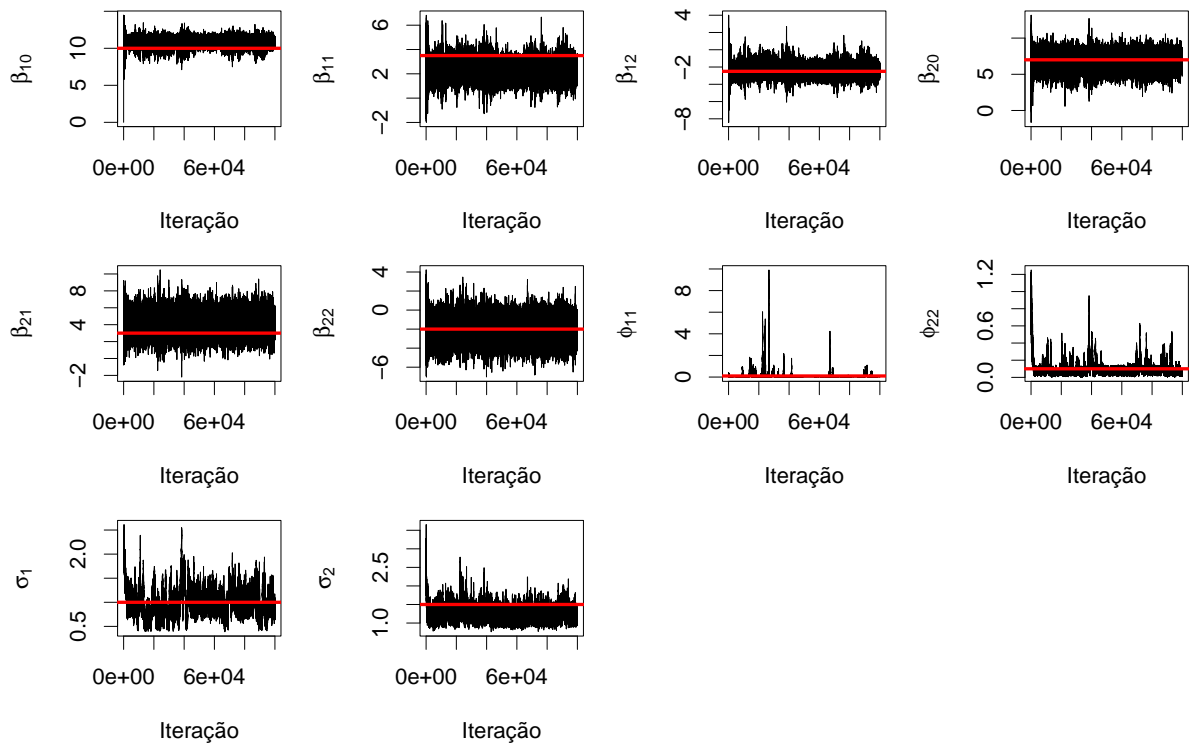


Figura 22: Cadeias geradas pelo método MCMC para cada parâmetro no cenário 1 para a estrutura aproximada

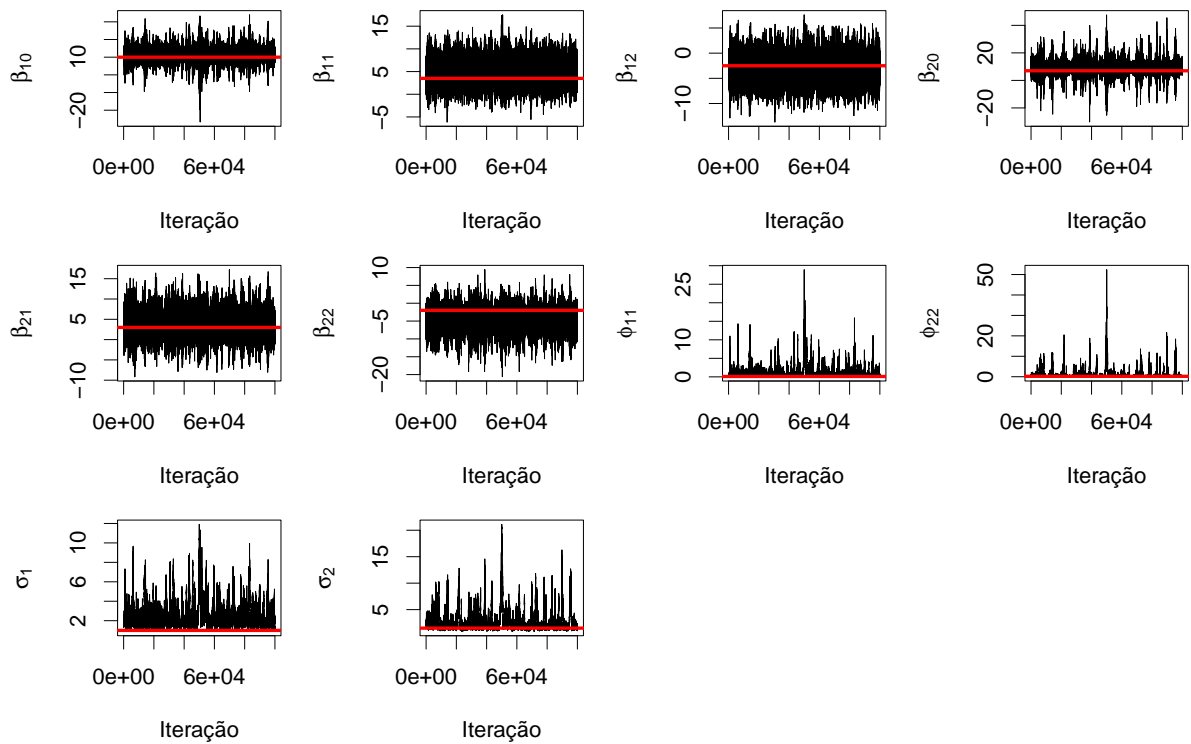


Figura 23: Cadeias geradas pelo método MCMC para cada parâmetro no cenário 1 para a estrutura cheia

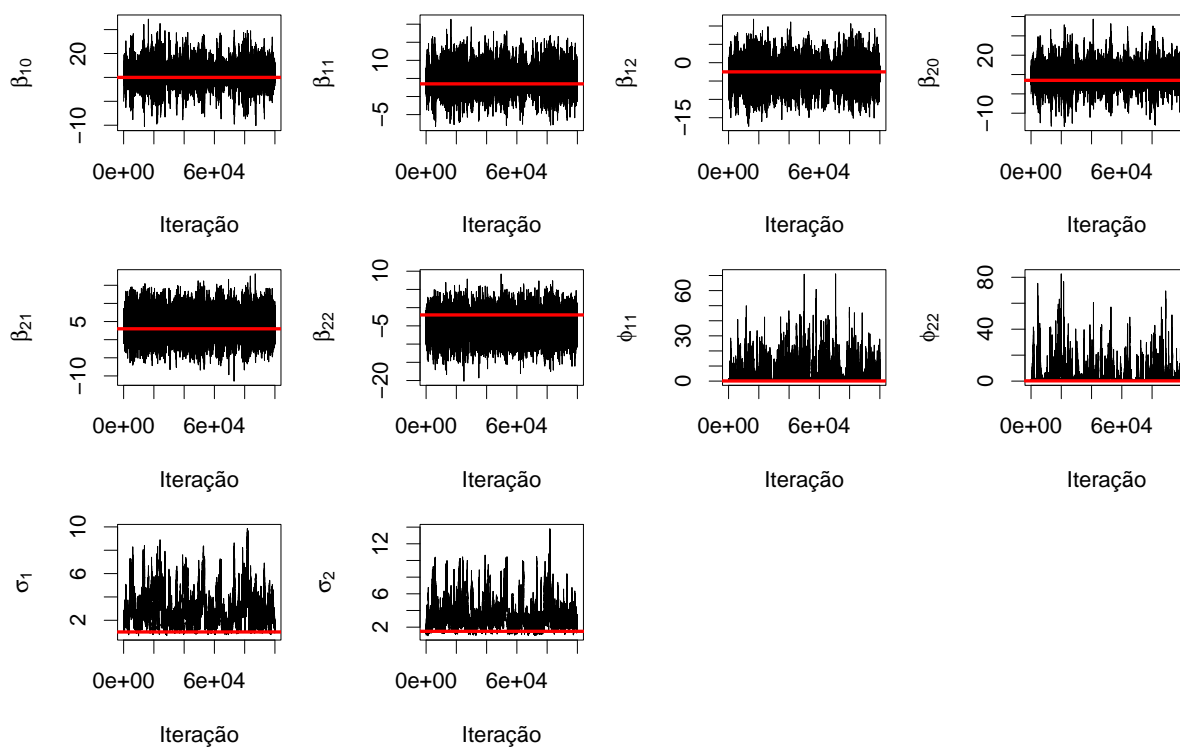


Figura 24: Cadeias geradas pelo método MCMC para cada parâmetro no cenário 1 para a estrutura aproximada

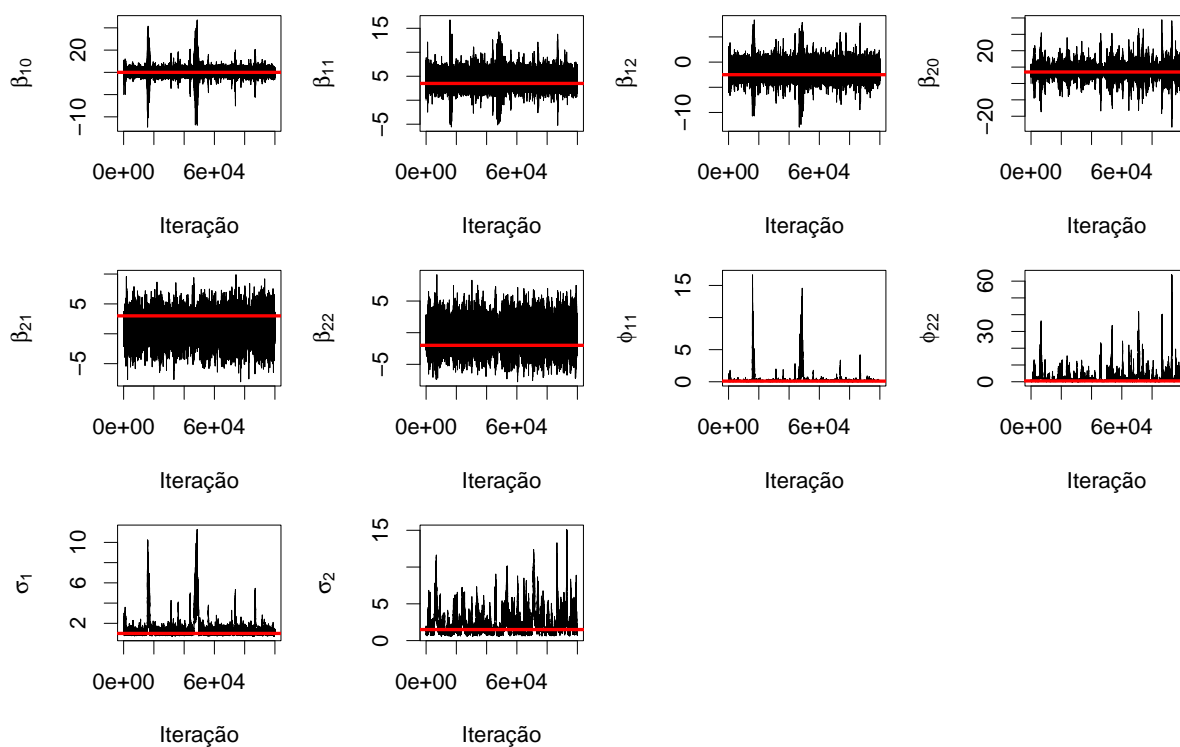


Figura 25: Cadeias geradas pelo método MCMC para cada parâmetro no cenário 1 para a estrutura cheia

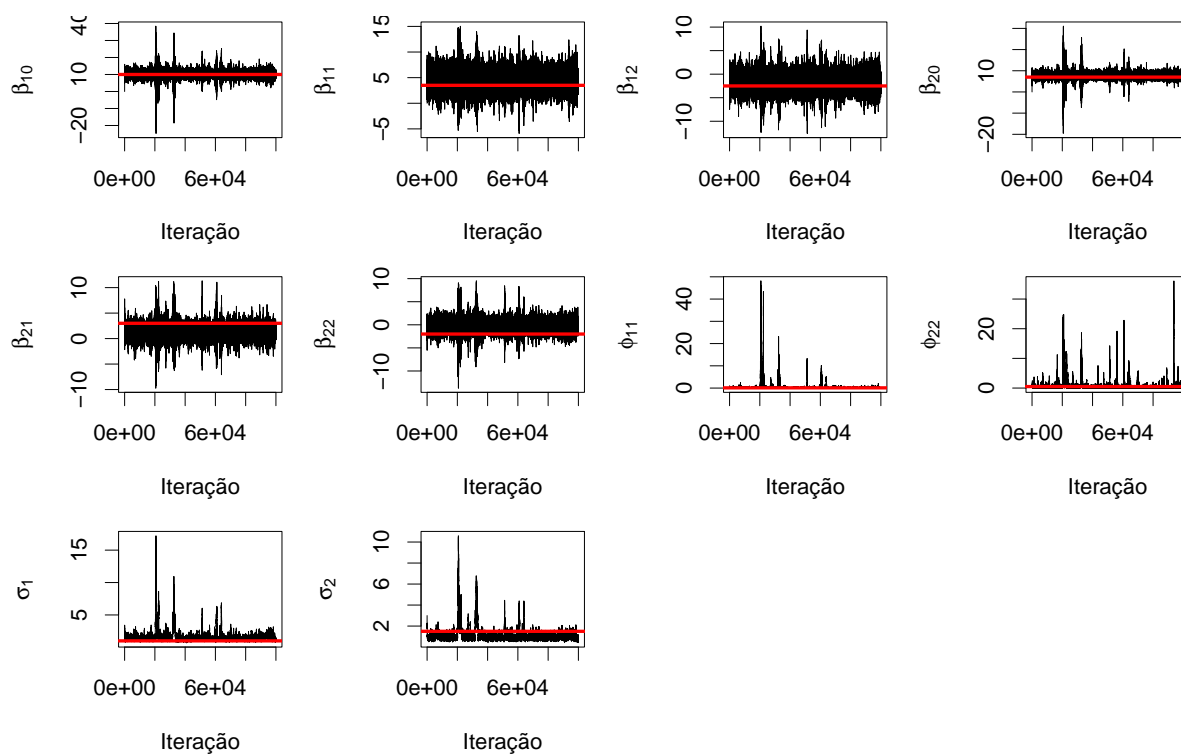


Figura 26: Cadeias geradas pelo método MCMC para cada parâmetro no cenário 1 para a estrutura aproximada

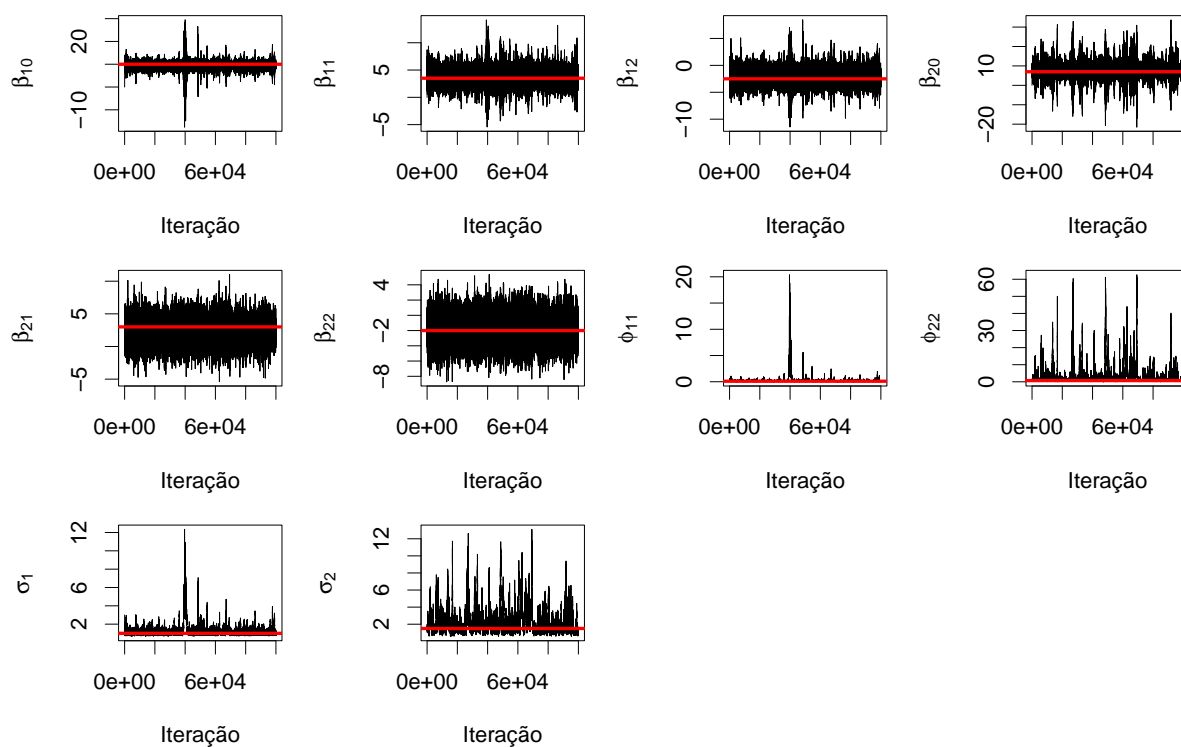


Figura 27: Cadeias geradas pelo método MCMC para cada parâmetro no cenário 1 para a estrutura cheia

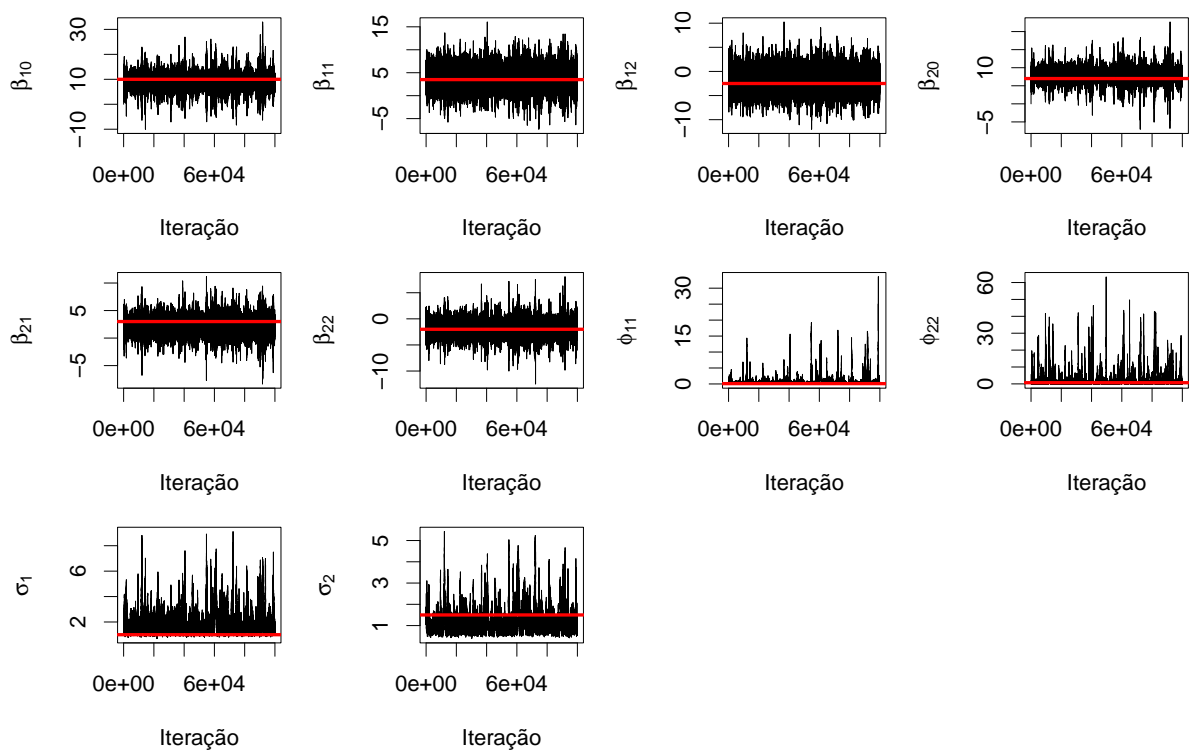


Figura 28: Cadeias geradas pelo método MCMC para cada parâmetro no cenário 1 para a estrutura aproximada