



RAFAEL AGOSTINHO FERREIRA

**PROCESSO DE COX MARCADO MODULADO POR
PROCESSOS GAUSSIANOS PARA CONFIGURAÇÕES
PONTUAIS UNIDIMENSIONAIS**

**LAVRAS – MG
2022**

RAFAEL AGOSTINHO FERREIRA

**PROCESSO DE COX MARCADO MODULADO POR PROCESSOS GAUSSIANOS
PARA CONFIGURAÇÕES PONTUAIS UNIDIMENSIONAIS**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Prof. DSc. João Domingos Scalon
Orientador

Prof. DSc. Deive Ciro de Oliveira
Coorientador

**LAVRAS – MG
2022**

**Ficha Catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Ferreira, Rafael Agostinho

Processo de Cox marcado modulado por processos Gaussianos para configurações pontuais unidimensionais / Rafael Agostinho Ferreira. – Lavras : UFLA, 2022.

142 p. : il.

Orientador: Prof. DSc. João Domingos Scalon.

Coorientador: Prof. DSc. Deive Ciro de Oliveira.

Tese (doutorado)- Universidade Federal de Lavras, 2022.

Bibliografia.

1. Inferência variacional Bayesiana. 2. Modelagem em processos pontuais. 3. Processos Gaussianos esparsos multivariados. I. Scalon, João Domingos. II. Oliveira, Deive Ciro de. III. Título.

RAFAEL AGOSTINHO FERREIRA

**PROCESSO DE COX MARCADO MODULADO POR PROCESSOS GAUSSIANOS
PARA CONFIGURAÇÕES PONTUAIS UNIDIMENSIONAIS**

**GAUSSIAN PROCESSES MODULATED MARKED COX PROCESS FOR
ONE-DIMENSIONAL POINT PATTERNS**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 28 de Janeiro de 2022.

Profa. DSc. Evelise Roman Corbalan Gois Freire	UFLA
Prof. DSc. Julio Silvio de Sousa Bueno Filho	UFLA
Prof. DSc. Marcelo Silva de Oliveira	UFLA
Prof. DSc. Denismar Alves Nogueira	UNIFAL-MG

Prof. DSc. João Domingos Scalon
Orientador

Prof. DSc. Deive Ciro de Oliveira
Coorientador

**LAVRAS – MG
2022**

*Aos meus amados avós que partiram durante o
desenvolvimento dessa pesquisa e deixaram uma saudade imensa:
Albertina dos Reis Ferreira, Pedro Agostinho e Vicentina Procópio Agostinho.
Dedico!*

AGRADECIMENTOS

À Deus pelo dom da vida, por me dar saúde e forças para continuar a trilhar meus sonhos, me mostrando em todos momentos que estava comigo nessa caminhada. Ao Senhor toda a minha gratidão!

Ao meu pai Paulo Roberto Ferreira e minha mãe Márcia Aparecida Agostinho Ferreira por me encorajarem, acreditarem no meu potencial e sempre orarem por mim. Amo vocês!

Às minhas irmãs Caroline e Veridiane, pelos conselhos, incentivos e por sempre estarem presentes. À minha prima, irmã do coração e melhor amiga Camila, por estar ao meu lado em todos os momentos. E ao meu cunhado Gustavo, por todo o apoio e amizade.

À minha amada sobrinha Ana Júlia, por trazer luz e alegria ao meu coração.

À toda minha família de Lavras: Antônio Henrique, Flávio, Isah, Kelly, Lauro, Neander, Silvana, Thaís, Thiago, Vanésia e, em especial, minha querida tia Ilma, por me receberem e me acolherem com tanto carinho e amor. Nada será o suficiente para retribuir tudo o que fizeram por mim.

Aos meus colegas de doutorado: Carlos, Cláudio, Cristian, Eleanderson, Ernandes, Luciano, Rodnei, Rodrigo e Vania por toda amizade durante esses anos. Em especial, gostaria de agradecer à minha amiga Patricia por todo companheirismo e apoio nas horas difíceis.

À todos os colegas e amigos de Pós-Graduação, pela troca de conhecimentos e pela convivência que tivemos.

Aos meus orientadores Dr. João Domingos Scalon e Dr. Deive Ciro de Oliveira por terem aceitado me orientar. Com excelência, paciência e muita sabedoria, contribuíram para minha formação e para que minha pesquisa pudesse ser concluída.

Ao Grupo de Pesquisa em Estatística Espacial (GPS) pela troca de conhecimentos.

Aos professores do Programa de Pós-graduação em Estatística e Experimentação Agropecuária que tiveram um papel fundamental no meu aprendizado.

Aos servidores do Departamento de Estatística (DES), em especial a secretária Nádia por toda disposição em ajudar e a sanar minhas dúvidas, sempre com muita simpatia.

À Universidade Federal de Lavras (UFLA) e ao Departamento de Estatística (DES) por me darem a oportunidade de realizar meu doutorado.

Aos membros da banca examinadora por todas as contribuições feitas em minha tese.

À todos os parentes e amigos não citados anteriormente, mas não menos importantes, que sempre torcem pelo meu sucesso.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

*Esta é minha ordem: Seja forte e corajoso!
Não tenha medo nem desanime, pois o Senhor,
seu Deus, estará com você por onde você andar.
Josué 1:9*

RESUMO

A teoria de processos pontuais é uma área da Estatística bastante importante para descrever o comportamento de um determinado fenômeno aleatório cuja a realização resulta em um conjunto de pontos distribuídos de forma aleatória que representam ocorrências de natureza pontual. Esses pontos, quando indexados nos reais unidimensionais, podem representar o momento exato de ocorrência. No entanto, o processo pode estar definido em qualquer conjunto de indexação, seja ele o tempo ou não. Uma das maneiras de se estudar a realização do processo pontual é através da função de intensidade, que descreve uma taxa média de ocorrências. Diversos modelos para descrever o comportamento da intensidade de um processo pontual foram propostos na literatura, incluindo a recente contribuição de Lloyd et al. (2015), baseado na classe de processos de Cox na qual a função de intensidade é descrita em função de um processo estocástico Gaussiano. A abordagem de Lloyd et al. (2015) se baseia em um método de estimação variacional com a inclusão de um método esparsos, o que permite que o modelo consiga lidar com uma grande quantidade de observações. Além disso, informações adicionais associadas às ocorrências do processo pontual podem ser incorporadas ao modelo, sendo essas informações denominadas por marcas. Desse modo, o objetivo da presente tese foi propor um esquema de modelagem para descrever a intensidade de processos pontuais marcados, na qual a marca é uma variável de natureza qualitativa, composta por duas categorias. A proposta se tratou de uma extensão do modelo de Lloyd et al. (2015), na qual a função de intensidade marcada, baseada em duas categorias, foi modelada em função de um processo Gaussiano esparsos bivariado. Seguindo a proposta de Lloyd et al. (2015), o processo de estimação se baseou no método variacional Bayesiano, o que permitiu que a função de intensidade pudesse ser estimada para qualquer ponto pertencente ao conjunto de indexação. Como forma de exemplificar a proposta dessa tese, uma aplicação foi feita a partir de um conjunto de dados reais baseado em ocorrências de acidentes em rodovias federais brasileiras. O modelo proposto se mostrou promissor, sugerindo que outras extensões possam ser feitas a fim de que o modelo consiga descrever um conjunto muito maior de fenômenos estocásticos de natureza pontual.

Palavras-chave: Inferência Bayesiana variacional. Modelagem em processos pontuais. Processos Gaussianos esparsos multivariados.

ABSTRACT

The theory of point processes is a very important Statistics area to describe the behavior of a certain random phenomenon whose realization results in a set of random points that represent occurrences of a point nature. These points, when indexed by the onedimensional set, they can represent the exact moment of occurrence. However, it can be defined in any indexing set, whether it is time or not. One of the ways to study a point process is through the intensity function, which describes an average rate of occurrences. It have been proposed several models to describe the behavior of the intensity of a point process in the literature, including the recent contribution of Lloyd et al. (2015), based on the Cox processes' class in which the intensity function is described as a function of a stochastic Gaussian process . Lloyd et al. (2015) approach is based on a variational estimation method with the inclusion of a sparse method, which allows the model to handle a large number of observations. In addition, additional information associated with the occurrences of the point process can be incorporated into the model, which is called by marks. Thus, this thesis aimed to propose a modeling scheme to describe the intensity of a marked point processes, in which the mark is a qualitative variable, with two categories. The proposal was an extension of the Lloyd et al. (2015) model, in which the marked intensity function, based on two categories, was modeled as a function of a sparse bivariate Gaussian process. Following Lloyd et al. (2015), the estimation process was based on the Bayesian variational method, which allowed that the intensity function could be estimated for any point in the index set. As a way of exemplifying the proposal of this thesis, it was made an application from a set of real data based on the occurrence of accidents on Brazilian federal highways. The proposed model proved to be promising, suggesting that other extensions can be made so that the model can describe a much larger set of stochastic phenomena of a point nature.

Keywords: Variational Bayesian inference. Point process modeling. Multivariate sparse Gaussian process.

LISTA DE FIGURAS

<p>Figura 2.1 – Realizações de um processo estocástico \mathcal{X} associado a cada elemento do espaço amostral Ω, considerando um conjunto de indexação unidimensional (a) e bidimensional (b).</p>	18
<p>Figura 2.2 – Representação gráfica do comportamento de um processo estocástico \mathcal{X}, quanto à natureza de sua variável aleatória e sua indexação. Tanto as variáveis aleatórias quanto seu conjunto de indexação podem ser de natureza discreta ou contínua, apresentando quatro possíveis cenários. Ambos os exemplos são considerados os pontos de indexação do processo unidimensionais.</p>	20
<p>Figura 2.3 – Representação gráfica do comportamento de um processo estocástico \mathcal{X}, quanto à natureza de sua variável aleatória e sua indexação. Tanto as variáveis aleatórias quanto seu conjunto de indexação podem ser de natureza discreta ou contínua, apresentando quatro possíveis cenários. Ambos os exemplos são considerados os pontos de indexação pertencentes à uma subregião de \mathbb{R}^2, representando processos espaciais.</p>	21
<p>Figura 2.4 – Representação de uma configuração pontual unidimensional.</p>	28
<p>Figura 2.5 – Processo de contagem para uma configuração pontual na reta real, em que a medida $\mathbb{N}(a, b]$ estabelece o número de eventos contidos no intervalo $(a, b]$. No exemplo ilustrativo, a medida \mathbb{N} contabiliza 3 eventos ocorridos no intervalo semi-aberto $(a, b]$.</p>	28
<p>Figura 2.6 – Exemplo de uma função <i>step</i> para uma configuração de pontos no tempo, os quais estão contidos no \mathbb{R}^+. De forma análoga a uma função de distribuição, $\mathbb{N}(t)$ quantifica uma contagem acumulativa de eventos até o instante de tempo t. Por exemplo, 4 eventos ocorreram até o instante de tempo t_i.</p>	29
<p>Figura 2.7 – Exemplo de uma configuração pontual considerando uma sequência de pontos. Neste caso, apesar dos eventos serem gerados de forma aleatória no tempo, uma relação de ordem é estabelecida, de tal forma que $t_1 < t_2 < t_3 < \dots$.</p>	30

Figura 2.8 – Exemplo uma configuração pontual considerando uma sequência de intervalos entre os tempos de ocorrências. Neste caso, o estudo do processo se baseia nas diferenças, que é uma variável aleatória contínua e que assume apenas valores positivos.	30
Figura 2.9 – Ilustração do método variacional aplicado à inferência Bayesiana, na qual a densidade a posteriori de interesse (A) é aproximada pela densidade variacional (B) usando a divergência de Kullback-Leibler como critério de otimização (C). Uma vez obtida uma densidade variacional ótima (D), tem-se que $q^*(\theta) \approx p(\theta \mathbf{x})$ e a divergência de Kullback-Leibler é próximo de zero, isto é, $KL(q(\theta) p(\theta \mathbf{x})) \approx 0$	44
Figura 3.1 – Geolocalização de trecho da rodovia Fernão Dias (em vermelho), localizado na região sul do estado de Minas Gerais, no qual será estimada a intensidade das ocorrências de acidentes.	57
Figura 3.2 – Exemplificação da criação do conjunto de indexação. Distância percorrida (em verde) a partir do ponto inicial até a localização do evento.	58
Figura 3.3 – Base de dados de acidentes por ocorrência registrados no período de janeiro de 2017 a setembro de 2021 em um trecho da rodovia federal BR-381, considerando apenas os registros cuja a causa tenha sido “Velocidade Incompatível” e em condições meteorológicas de “chuva”, “garoa / chuvisco” e “nevoeiro / neblina”.	59
Figura 3.4 – Exemplificação da atribuição do tipo de marca (veículo de pequeno/ grande porte) associado ao evento (veículo(s) associado(s) à ocorrência do acidente).	60
Figura 3.5 – Base de dados de acidentes utilizada na segunda análise, criada a partir da junção das duas bases disponibilizadas pela Polícia Rodoviária Federal, na qual o evento é rotulado de acordo com o tipo de veículo envolvido: <i>Pequeno porte</i> ou <i>Grande porte</i>	61
Figura 5.1 – Histograma com a frequência absoluta do número de ocorrências de acidentes por quilômetro em um trecho da rodovia Fernão Dias (BR-381) entre os municípios de Extrema e Betim. Os traços em vermelho, I , representam as posições exatas (distância em quilômetros) de cada ocorrência.	79

Figura 5.2 – Gráfico da estimativa da função de intensidade dos registros de acidentes ao longo do trecho da BR-381 (entre Extrema e Betim), durante janeiro de 2017 à setembro de 2021, considerando apenas os acidentes cuja a causa principal tenha sido por “velocidade incompatível” e tenham ocorrido em situações de chuva, chuvisco, garoa ou granizo. A faixa na cor cinza em torno da estimativa de intensidade mostra o intervalo de credibilidade com 95% de probabilidade.	80
Figura 5.3 – Mapa de calor da projeção da função de intensidade estimada no trecho da rodovia federal BR-381 (entre Extrema e Betim). Áreas mais quentes (mais avermelhadas) representam áreas de maior intensidade.	82
Figura 5.4 – Histograma com a frequência absoluta do número de ocorrências de acidentes envolvendo veículos de pequeno porte (A) e grande porte (B) por quilômetro em um trecho da rodovia Fernão Dias (BR-381) entre os municípios de Extrema e Betim. Os traços em vermelho, I, representam as posições exatas (distância em quilômetros) de cada ocorrência.	84
Figura 5.5 – Estimativas das funções de intensidade dos registros de acidentes em trecho da BR-381, entre os municípios de Extrema e Betim, durante janeiro de 2017 à setembro de 2021, considerando os acidentes envolvendo veículos de pequeno porte (A) e grande porte (B), cuja a causa principal tenha sido por “velocidade incompatível” ocorridos em situações de chuva, chuvisco, garoa ou granizo. A faixa na cor cinza em torno das estimativas descrevem os intervalos de credibilidade com 95% de probabilidade.	86
Figura 5.6 – Projeção das funções de intensidades estimadas para as ocorrências de acidentes no trecho da rodovia federal BR-381, entre Extrema e Betim, considerando o porte do(s) veículo(s) envolvido(s), sendo (A) a intensidade de acidentes para veículos de pequeno porte e (B) para veículos de grande porte.	88

LISTA DE SÍMBOLOS

\mathbb{R}^d	Conjunto dos números reais no espaço d -dimensional
\mathcal{S}	Configuração pontual no espaço d -dimensional
\mathcal{S}_m	Configuração pontual marcado no espaço d -dimensional
$\mathcal{S}_m^{(j)}$	j -ésima replicação de um processo pontual marcado no espaço d -dimensional
N	Número de eventos observados em uma configuração pontual
\mathcal{K}	Número de categorias de uma marca de um processo pontual
\mathcal{T}	Domínio (ou espaço) dos eventos
\mathcal{M}	Domínio (ou espaço) das marcas
R	Número de replicações (amostras) de um processo pontual
ξ	Função que mensura o intervalo de tempo aleatório entre dois eventos pontuais
\mathcal{H}	Historia de um processo pontual.
$\lambda(\cdot \mathcal{H})$	Função de intensidade condicional
$\lambda(\cdot)$	Função de intensidade incondicional
$f(\cdot)$	Processo Gaussiano latente
$\mathbf{f}(\cdot)$	Vetor de processos Gaussianos latentes (Processo Gaussiano multivariado)
$f_i(\cdot)$	Processo Gaussiano latente associado à i -ésima marca
\mathcal{Z}	Conjunto de pontos de indução
M	Número de pontos de indução
$\mathbf{u} = f(\mathcal{Z})$	Processo Gaussiano aplicado nos M pontos de indução
\mathbf{u}_i	Processo Gaussiano da i -ésima marca aplicado nos M pontos de indução
$\bar{\mathbf{1}}$	Vetor unitário.
\bar{u}_i	Média a priori associada ao i -ésimo vetor de variáveis de indução Gaussianas, \mathbf{u}_i .
$\boldsymbol{\mu}$	Vetor funções de médias a priori/variacionais de $\mathbf{f} \mathbf{u}$.
$\tilde{\boldsymbol{\mu}}$	Vetor funções de médias variacionais de \mathbf{f} .
$\boldsymbol{\Sigma}$	Matriz de funções de covariâncias cruzadas a priori/variacionais de $\mathbf{f} \mathbf{u}$.
$\tilde{\boldsymbol{\Sigma}}$	Matriz de funções de covariâncias cruzadas variacionais de \mathbf{f} .
$\tilde{\mu}_i(s)$	Função de média para o i -ésimo processo Gaussiano.
$\tilde{\Sigma}_{ij}(s, s')$	Função de covariância cruzado entre processos Gaussianos i e j , para $i \neq j$.
$\tilde{\Sigma}_{ii}(s, s')$	Função de covariância do i -ésimo processo Gaussiano.
$\tilde{\sigma}_i^2(s)$	Função de variância do i -ésimo processo Gaussiano.
ρ_{ij}	Coefficiente de correlação entre o i -ésimo e j -ésimo processo Gaussiano.
γ_i	Desvio padrão a priori para o i -ésimo processo Gaussiano
$c(\cdot, \cdot; \alpha)$	Função de correlação baseado na distância entre pontos de um processo Gaussiano, controlado pelo parâmetro de escala, α .
$\mathbf{K}_{ij}^{(s, s')}$	Função de covariância cruzada entre o i -ésimo e j -ésimo processo Gaussiano, avaliados nos pontos de indexação $(s, s') \in \mathcal{T}$
$\mathbf{K}^{(s, s')}$	Matriz particionada contendo as funções de covariâncias $\mathbf{K}_{ij}^{(s, s')}$ a priori.

$\mathbf{K}_{ij}^{(z,z)}$	Matriz de covariâncias cruzadas entre o i -ésimo e j -ésimo vetor de variáveis de indução, \mathbf{u}_i e \mathbf{u}_j , avaliados nos pontos $z \in \mathcal{Z}$ pré-definidos.
$\mathbf{K}^{(z,z)}$	Matriz particionada contendo todas as matrizes $\mathbf{K}_{ij}^{(z,z)}$.
$\mathbf{K}_{ij}^{(s,z)}$	Função matriz de covariância cruzada entre o i -ésimo processo Gaussiano e o j -ésimo vetor de variáveis de indução, \mathbf{u}_j , avaliados nos pontos $s \in \mathcal{T}$ e $z \in \mathcal{Z}$, respectivamente.
$\mathbf{K}^{(s,z)}$	Matriz particionada contendo todas as matrizes funções de covariâncias $\mathbf{K}_{ij}^{(s,z)}$.
\mathbf{m}_i	Vetor de médias variacionais associado ao i -ésimo vetor de variáveis de indução, \mathbf{u}_i
\mathbf{m}	Vetor particionado contendo todos os vetores de médias variacionais \mathbf{m}_i .
\mathbf{S}_{ij}	Matriz de covariâncias cruzadas variacionais entre os i -ésimo e o j -ésimo vetor de variáveis de indução, \mathbf{u}_i e \mathbf{u}_j .
\mathbf{S}	Matriz particionada contendo todas as matrizes de covariâncias \mathbf{S}_{ij} .
Ψ	Matriz de dimensão $M \times M$ obtidas pela integração do produto das funções matrizes $K(z, s)$ e $K(s, z')$ em relação a $s \in \mathcal{T}$, para o modelo de Lloyd et al. (2015).
$KL(q \parallel p)$	Divergência de Kullback-Leibler entre duas funções (densidades) de probabilidade.
\mathcal{L}	Limite inferior do logaritmo da função de evidência (ou verossimilhança marginal).

SUMÁRIO

1	INTRODUÇÃO	13
2	REFERENCIAL TEÓRICO	16
2.1	Processos estocásticos	16
2.1.1	Processos estocásticos Gaussianos	24
2.1.2	Processos pontuais	26
2.1.2.1	Processos pontuais com marcas	32
2.2	Modelos para processos pontuais	34
2.2.1	Processos de Poisson	34
2.2.2	Processos de Cox	36
2.3	Inferência Bayesiana	38
2.3.1	Métodos de Monte Carlo via Cadeias de Markov	40
2.3.2	Inferência Bayesiana variacional	42
2.4	Modelo proposto por Lloyd et al. (2015)	47
2.5	Modelo proposto por Liang, Carlin e Gelfand (2009)	54
3	MATERIAIS E MÉTODOS	57
3.1	Conjunto de dados reais	57
3.2	Recursos computacionais	61
4	RESULTADOS TEÓRICOS	62
4.1	Modelo	62
4.2	Estimação	65
5	APLICAÇÃO EM DADOS DE ACIDENTES RODOVIÁRIOS	77
6	LIMITAÇÕES DO MODELO E FUTUROS TRABALHOS	89
7	CONCLUSÃO	91
	REFERÊNCIAS	92
	APENDICE A – Algoritmo Broyden–Fletcher–Goldfarb–Shanno (BFGS)	97
	APENDICE B – Alguns resultados importantes para o modelo de Lloyd et al. (2015)	101
	APENDICE C – Alguns resultados importantes para o modelo proposto	106

1 INTRODUÇÃO

Diversos tipos de fenômenos aleatórios ocorrem diariamente e, de alguma forma, podem afetar a vida humana, seja positivamente ou negativamente. Estas ocorrências podem se manifestar de diversas formas, apresentando inúmeros possíveis resultados, sendo alguns mais prováveis do que outros. A aleatoriedade presente nesses fenômenos é o principal objeto de estudo da Estatística. Conhecer como as coisas aleatórias se comportam é o que move a ciência Estatística, com o intuito de detectar possíveis padrões e prever futuras ocorrências.

Por isso, a Estatística é uma ciência de grande importância para os dias atuais. Sua principal ferramenta são dados observados a partir da realização de algum fenômeno aleatório em questão e seu principal objetivo é lidar, basicamente, com a coleta, análise e interpretação desses dados, a fim de fornecer informações relevantes para que se possa tomar decisões diante de alguma situação. Seus métodos e técnicas contribuem para que diversas áreas do conhecimento possam compreender o comportamento de determinados eventos presentes no cotidiano.

Dentre as diversas formas de análise de dados presentes na Estatística, a teoria de processos estocásticos se destaca como um dos principais métodos existentes. Apesar de ser um conceito bastante amplo, o termo “estocástico” geralmente está relacionado às ferramentas exploratórias e de modelagem que se baseiam no estudo da dependência entre variáveis aleatórias indexadas a um determinado conjunto. Em geral, esse conjunto de indexação pode representar a posição da variável aleatória no tempo (análise temporal), no espaço (análise espacial) ou em ambas (análise espaço-temporal).

Processos pontuais são mecanismos estocásticos que geram dados de natureza pontual, isto é, a informação primordial necessária para análise desse tipo de dado se encontra na posição exata da ocorrência do evento. Segundo Daley e Vere-Jones (2003), a teoria de processos pontuais teve seu início a partir de problemas estatísticos relacionado a demografia, seguros e tábuas de vida. No entanto, a teoria se desenvolveu, de fato, a partir de 1930. A princípio, seus métodos foram desenvolvidos para descrever ocorrências pontuais de eventos no domínio do tempo. A partir das contribuições feitas por diversos pesquisadores em relação ao contexto temporal, a teoria se expandiu para outros contextos, através da inclusão de marcas, extensões para problemas espaciais, espaço-temporais, entre outros.

No contexto temporal, diversos modelos para processos pontuais já foram propostos na literatura, entre eles destacam-se os *processos de Poisson*, *processos de Hawkes*, *processos self-correcting* e *processos de Cox*.

Uma sub-classe pertencente aos modelos de Cox, amplamente utilizada para descrever fenômenos desta natureza, são os *processos de Cox modulado por processos Gaussianos* ou, simplesmente, *processos de Cox Gaussianos*. Esse tipo de modelagem consiste em relacionar a função de intensidade de um processo pontual com um outro processo estocástico latente, cuja distribuição é Gaussiana. Essa relação pode se dar por meio de uma variedade de funções de ligação, o que permite uma maior flexibilização no processo de modelagem.

Uma proposta desenvolvida por Lloyd et al. (2015), que recebe o nome de *Variational Bayes for Point Processes (VBPP)*, consiste em modelar a função de intensidade por meio de um processo Gaussiano, utilizando um método inferencial Bayesiano variacional. Além disso, os autores incluíram uma abordagem esparsa sob o processo Gaussiano, a fim de contornar o problema de inversões de matrizes com alta dimensionalidade. Apesar de seu modelo apresentar bom desempenho, ele se limita apenas à uma abordagem envolvendo processos pontuais não marcados, isto é, o modelo considera apenas a informação da posição exata das ocorrências dos eventos, desprezando qualquer informação adicional que o evento pontual possa ter.

Assim, o objetivo dessa tese foi propor um modelo para a intensidade de processos pontuais, a partir da proposta de Lloyd et al. (2015), que utiliza um modelo de Cox modulado por processos Gaussianos para inferir sobre a intensidade de uma configuração pontual unidimensional. Especificamente, esta proposta consiste em estender o modelo de Lloyd et al. (2015) para processos pontuais marcados, utilizando as ideias apresentadas por Liang, Carlin e Gelfand (2009) para a construção de um modelo com marcas categóricas.

Com o objetivo exposto, chegou-se ao seguinte problema de pesquisa: Como modelar a função de intensidade de um processo de Cox marcado a partir de uma configuração pontual unidimensional considerando marcas de natureza qualitativa e assumindo que os atributos da marca podem ser correlacionados?

Devido a variedade de eventos que se comportam de maneira pontual, a presente proposta pode ser aplicada à diversas áreas do conhecimento, tais como a neurociência, epidemiologia, estudos sobre ocorrências de crimes, entre outros.

Desse modo, como forma de exemplificar a proposta metodológica dessa tese, objetivou-se um estudo aplicado à ocorrências de acidentes em rodovias federais brasileiras. Especifica-

mente para o conjunto de dados em questão, foram considerados os acidentes que ocorreram em um trecho da rodovia Fernão Dias (entre Extrema e Betim) no período de janeiro de 2017 à setembro de 2021. Primeiramente, uma estimativa da intensidade dessas ocorrências foram obtidas por meio do modelo de Lloyd et al. (2015). Posteriormente, os acidentes foram categorizados de acordo com o tipo de veículo envolvido no acidente (veículos de pequeno e grande porte), criando assim uma marca para os dados pontuais.

2 REFERENCIAL TEÓRICO

Nesse capítulo, serão introduzidos alguns tópicos importantes que darão suporte a construção de toda a teoria desenvolvida nesta tese. Primeiramente, serão abordadas as teorias de processos estocásticos e de processos Gaussianos. Após isso, será introduzida a teoria de processos pontuais e de algumas classes de modelos para sua função de intensidade. Posteriormente, métodos e técnicas de inferência Bayesiana serão discutidos visando apresentar a estimação variacional, método inferencial que será utilizado nesta tese. E, por fim, serão apresentados os modelos para a função de intensidade propostos por Lloyd et al. (2015) e de Liang, Carlin e Gelfand (2009), respectivamente, essenciais para a construção da tese.

2.1 Processos estocásticos

Em muitas situações, o interesse de um pesquisador pode estar não apenas no estudo do comportamento aleatório de um determinado fenômeno, mas de possíveis evoluções ou mudanças que este fenômeno venha a sofrer ao longo do tempo e/ou em diferentes regiões de um espaço de dimensão qualquer. A teoria de processos estocásticos, segundo Basu (2003), é a área da Estatística que se propõe a construir, matematicamente, os conceitos probabilísticos envolvendo tais fenômenos.

Considerando um conjunto de todas as possíveis realizações de um experimento aleatório, denominado por espaço amostral e denotado por $\Omega = \{\omega_1, \omega_2, \dots, \omega_i, \dots\}$, define-se uma variável aleatória X como sendo uma função que associa cada elemento de Ω a um valor pertencente ao conjunto dos reais \mathbb{R} , isto é, $X : \Omega \rightarrow \mathbb{R}$ (ALBUQUERQUE; FORTES; FINAMORE, 2008).

Para uma definição formal sobre processos estocásticos é necessário estabelecer um novo tipo de conjunto, que recebe o nome de *conjunto de indexação*. Denotado por $\mathcal{S} = \{s_1, s_2, \dots\}$, sendo $s_i \in \mathcal{T} \subset \mathbb{R}^d$ e $d \geq 1$, os elementos que compõem o conjunto de indexação podem, em geral, representar momentos no tempo, quando o conjunto é unidimensional ($d = 1$), ou coordenadas espaciais de algum fenômeno aleatório, quando a dimensão do conjunto é multidimensional ($d \geq 2$).

De acordo com Schabenberger e Gotway (2004) e Therrien (2012), um processo estocástico pode ser definido por meio de uma variável aleatória X mensurada em cada elemento

pertencente ao conjunto de indexação \mathcal{S} , formando assim uma família infinita de variáveis aleatórias, denotada por

$$\mathcal{X} = \{X(\mathbf{s}_1), X(\mathbf{s}_2), X(\mathbf{s}_3), \dots\}, \quad (2.1)$$

para $X(\mathbf{s}_i) : \Omega \rightarrow \mathbb{R}$, $\mathbf{s}_i \in \mathcal{T} \subset \mathbb{R}^d$, $i = 1, 2, 3, \dots$. De acordo com Albuquerque, Fortes e Finamore (2008), o processo estocástico pode ser visto como uma função de $\mathbf{s} \in \mathcal{T} \subset \mathbb{R}^d$ e de $\omega \in \Omega$.

Dado uma realização do experimento aleatório em questão, fixando $\omega \in \Omega$, a variável aleatória X se reduz em um único valor real, chamado de *realização da variável aleatória*, na qual é denotado por x . Considerando o conjunto de indexação \mathcal{S} , a realização da variável aleatória X em cada indexador $\mathbf{s}_i \in \mathcal{T}$ resultará em uma realização do processo estocástico, formando um conjunto infinito de valores reais x ,

$$\{x(\mathbf{s}_1, \omega), x(\mathbf{s}_2, \omega), x(\mathbf{s}_3, \omega), \dots\},$$

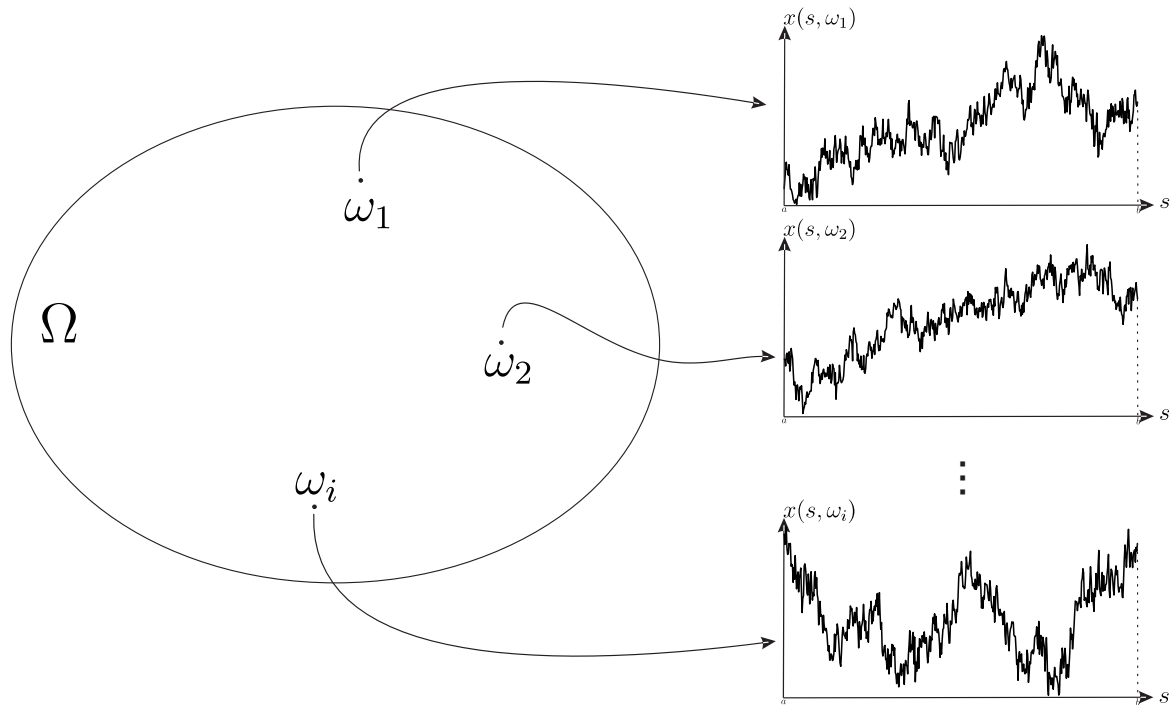
em que $x(\mathbf{s}_i, \omega)$ é a realização da variável aleatória X indexada ao ponto \mathbf{s}_i , para $i = 1, 2, \dots$, associada à um único elemento ω do espaço amostral Ω .

Na prática, a realização de uma variável aleatória de interesse é observada apenas em um conjunto finito de N pontos de indexação, formando assim um vetor de N observações de X , indexado por $\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N\}$, no qual é denotado por

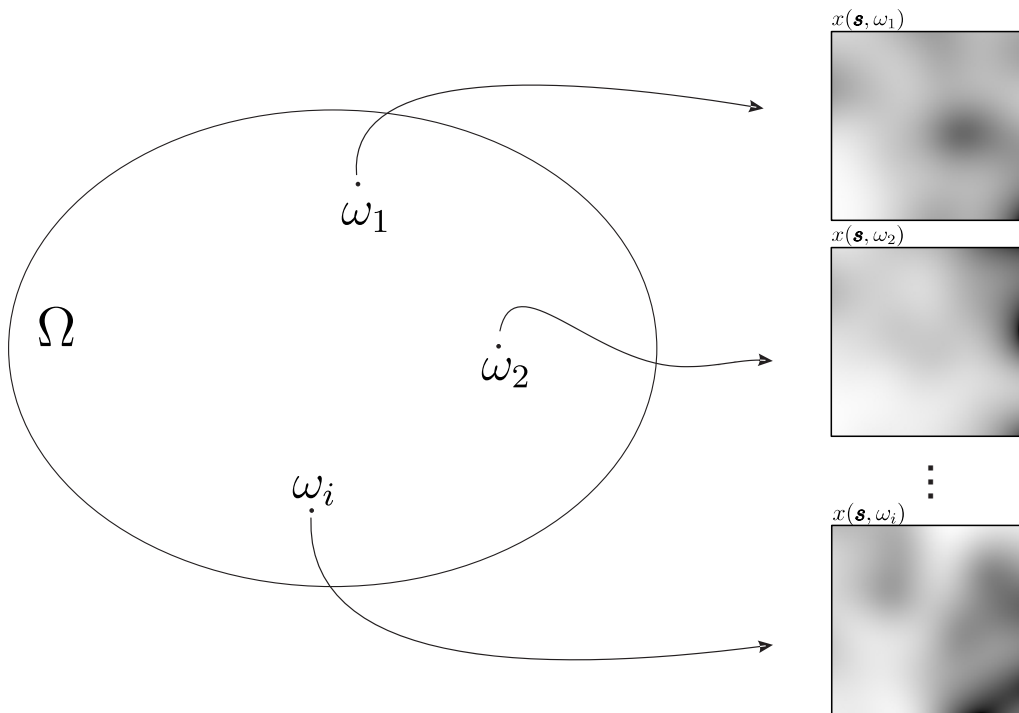
$$\{x(\mathbf{s}_1, \omega), x(\mathbf{s}_2, \omega), \dots, x(\mathbf{s}_N, \omega)\}. \quad (2.2)$$

Vale ressaltar que a realização do processo estocástico \mathcal{X} em (2.2) não corresponde à uma amostra de tamanho N , mas sim em uma única amostra de \mathcal{X} obtida em cada um dos N pontos de indexação observados, uma vez que a variável aleatória X está associada à um único elemento ω (SCHABENBERGER; GOTWAY, 2004). Além disso, a realização do processo representa uma observação incompleta, uma vez que a variável X não é obtida, na prática, para todos os pontos $\mathbf{s} \in \mathcal{T}$. Em outras palavras, a observação do processo \mathcal{X} é uma amostra de tamanho um obtida por meio de uma distribuição N -variada. A Figura 2.1 apresenta uma ilustração de realizações do processo estocástico \mathcal{X} considerando $\omega \in \Omega$ fixo e um conjunto de indexação unidimensional pertencentes à $\mathcal{T} = [a, b] \subset \mathbb{R}^1$ (Figura 2.1(a)) e um conjunto de indexação bidimensional ($d = 2$) pertencentes à $\mathcal{T} \subset \mathbb{R}^2$ (Figura 2.1(b)).

Figura 2.1 – Realizações de um processo estocástico \mathcal{X} associado a cada elemento do espaço amostral Ω , considerando um conjunto de indexação unidimensional (a) e bidimensional (b).



(a)



(b)

Fonte: Do autor (2021).

Segundo Albuquerque, Fortes e Finamore (2008), a partir da definição acima de um processo estocástico como sendo uma função de s e ω , pode-se ter diferentes formas de se tratar um determinado fenômeno aleatório. São elas:

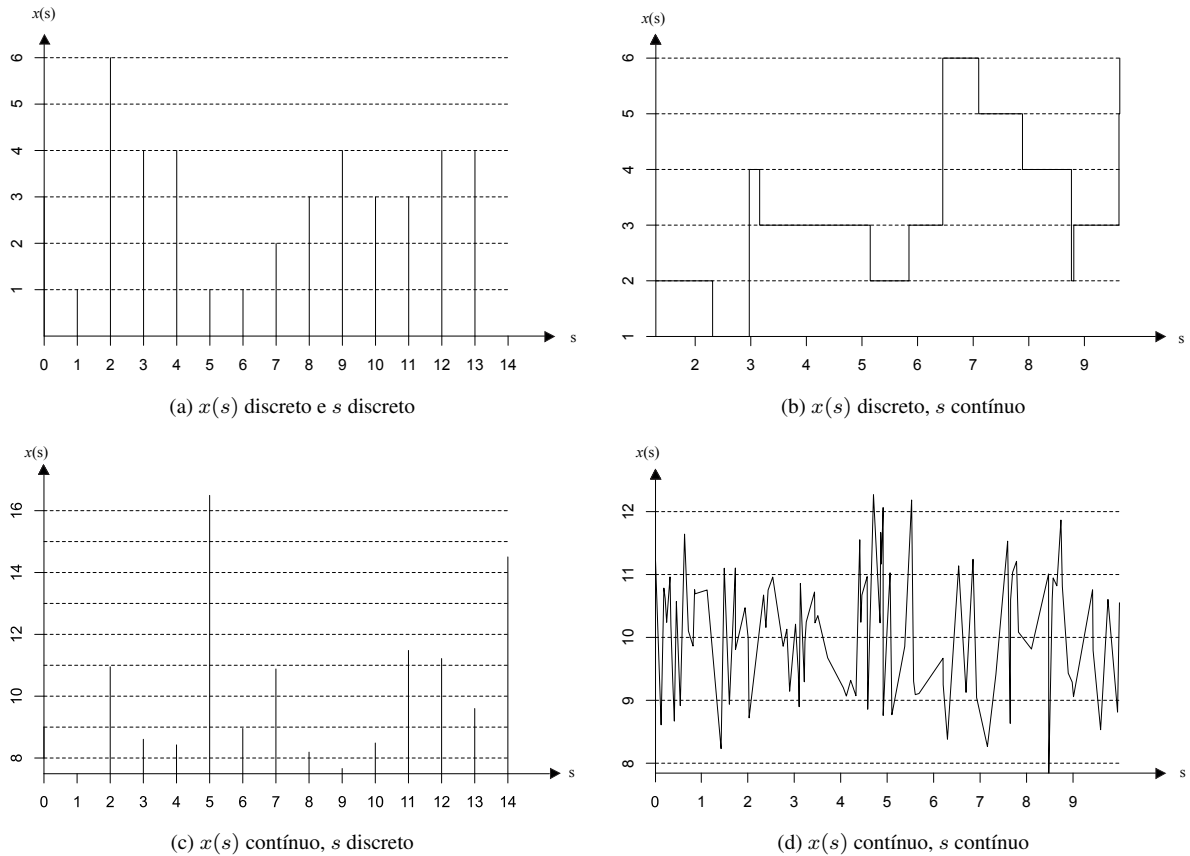
- a) um processo estocástico \mathcal{X} , considerando \mathcal{S} um conjunto infinito de pontos de indexação e ω variável;
- b) uma realização do processo \mathcal{X} , $\{x(\mathbf{s}_1), x(\mathbf{s}_2), \dots\}$, considerando \mathcal{S} um conjunto infinito e ω fixo;
- c) uma variável aleatória $X(\mathbf{s})$, fixando em um único ponto \mathbf{s} e considerando ω variável;
- d) um vetor aleatório de uma mesma variável X , $\{X(\mathbf{s}_1), X(\mathbf{s}_2), \dots, X(\mathbf{s}_N)\}$, para um conjunto finito de N pontos de indexação e ω variável;
- e) um valor real $x(\mathbf{s})$ como realização da variável aleatória $X(\mathbf{s})$, quando apenas um único ponto de indexação \mathbf{s} for considerado e ω for fixo.

Por questão de conveniência e simplicidade nas notações aqui apresentadas, será omitida das notações de processos estocásticos, bem como suas realizações, sua relação com os elementos do espaço amostral, Ω . Porém, é necessário reforçar que, apesar de não estar evidenciada, esta relação continuará presente implicitamente em toda a teoria apresentada nessa tese.

No que se refere ao conjunto de indexação \mathcal{S} e à natureza da variável aleatória associados ao processo estocástico, pode-se classificá-lo em quatro possíveis cenários. O primeiro é chamado de *processo estocástico discreto com indexação discreta* quando a variável aleatória X é discreta e \mathcal{S} apresenta um conjunto de valores discretos. O segundo cenário é chamado de *processo estocástico discreto com indexação contínua*, na qual a natureza da variável aleatória continua sendo discreta mas com um conjunto de pontos de indexação contínuo. Já o terceiro recebe o nome de *processo estocástico contínuo com indexação discreta*, quando a variável aleatória é contínua e com conjunto de indexação discreta. Por fim, o quarto cenário, denominado por *processo estocástico contínuo com indexação contínua*, se dá quando tanto a variável aleatória quanto o conjunto de indexação são de natureza contínua. A Figura 2.2 exemplifica uma realização do processo estocástico \mathcal{X} considerando suas quatro possíveis classificações, supondo um conjunto de pontos de indexação pertencentes ao intervalo $\mathcal{T} \subset \mathbb{R}^1$.

Já a Figura 2.3 mostram exemplos de uma realização de processos estocásticos cuja indexação está definida em uma subregião de um espaço bidimensional, considerando as quatro possíveis classificações que um processo pode apresentar. Esses exemplos são típicos de processos espaciais. A Figura 2.3 (A) representa uma realização de um processo estocástico contínuo com indexação contínua, comumente chamada de *processo de superfície contínua* ou,

Figura 2.2 – Representação gráfica do comportamento de um processo estocástico \mathcal{X} , quanto à natureza de sua variável aleatória e sua indexação. Tanto as variáveis aleatórias quanto seu conjunto de indexação podem ser de natureza discreta ou contínua, apresentando quatro possíveis cenários. Ambos os exemplos são considerados os pontos de indexação do processo unidimensionais.



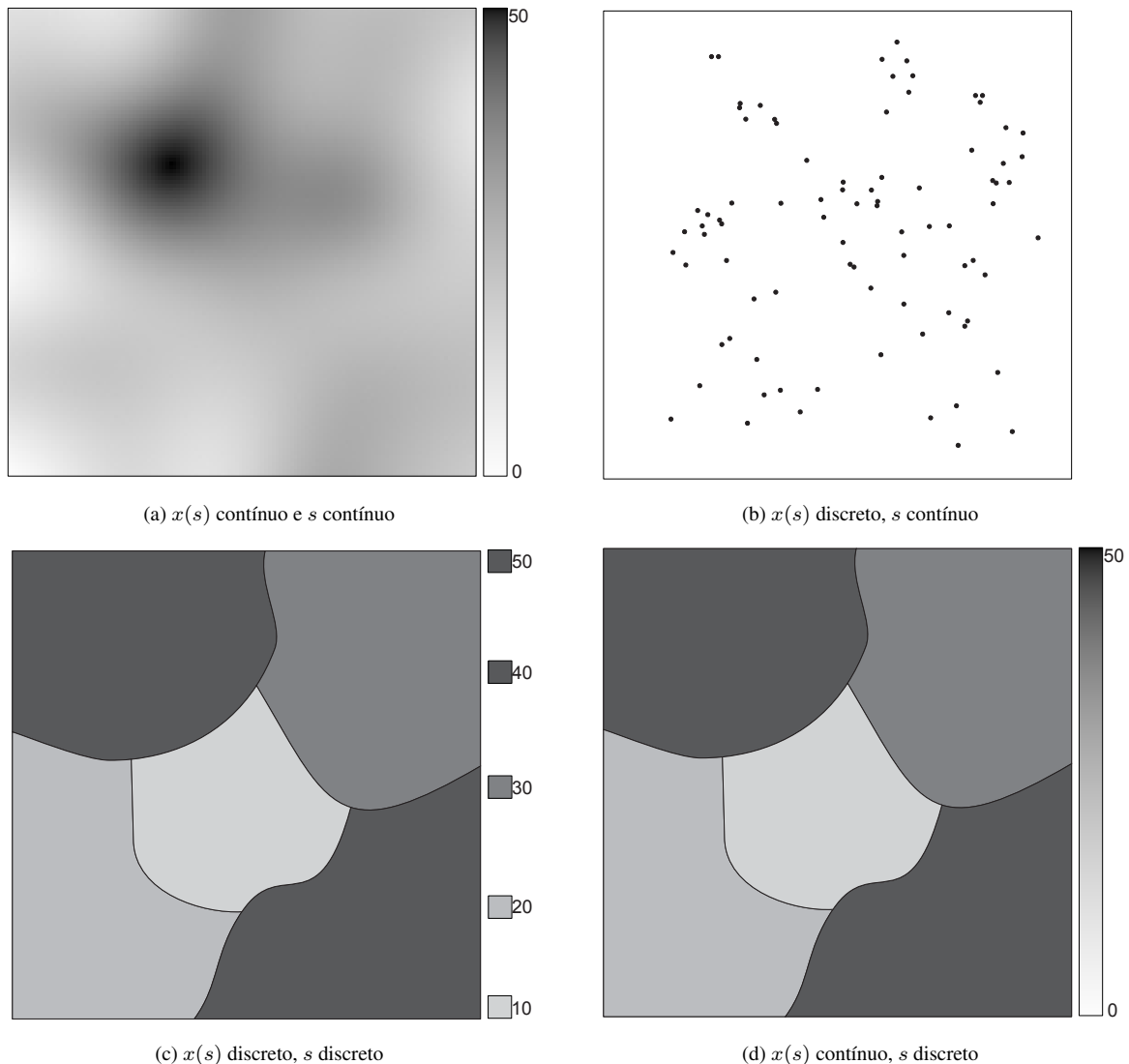
Fonte: Do autor (2021).

simplesmente, de *geoestatística*. Já a Figura 2.3 (B) exemplifica um processo estocástico em que a indexação é contínua e a variável aleatória associada é discreta (binária), indicando a ocorrência ou não ocorrência da variável. Esse exemplo é típico de um *processo estocástico pontual*, que será detalhado mais a frente. Por outro lado, as Figuras 2.3 (C) e 2.3 (D) são exemplos de uma realização de processos estocásticos com indexação discreta e variável aleatória discreta e contínua, respectivamente. Ambos os exemplos representam processos espaciais chamados de *dados em lattice*, ou apenas *dados de área*.

De acordo com Ripley (1981), a especificação de um processo estocástico é feita por meio da função de distribuição conjunta de dimensão finita, para qualquer subconjunto finito do processo. Isto é, se

$$F[x(\mathbf{s}_1), \dots, x(\mathbf{s}_N)] = p(X(\mathbf{s}_1) \leq x(\mathbf{s}_1), \dots, X(\mathbf{s}_N) \leq x(\mathbf{s}_N)) \quad (2.3)$$

Figura 2.3 – Representação gráfica do comportamento de um processo estocástico \mathcal{X} , quanto à natureza de sua variável aleatória e sua indexação. Tanto as variáveis aleatórias quanto seu conjunto de indexação podem ser de natureza discreta ou contínua, apresentando quatro possíveis cenários. Ambos os exemplos são considerados os pontos de indexação pertencentes à uma subregião de \mathbb{R}^2 , representando processos espaciais.



Fonte: Do Autor (2021).

for conhecida, para todo $N \geq 1$, então o processo estocástico estará completamente especificado (ALBUQUERQUE; FORTES; FINAMORE, 2008). Desse modo, é possível conhecer as distribuições marginais de cada variável aleatória associada ao processo, bem como qualquer distribuição bivariada ou, até mesmo, a distribuição conjunta de qualquer subconjunto de variáveis aleatórias de tamanho inferior à N (MORETTIN; TOLOI, 2006). Isso se deve ao fato de que, uma vez que o processo está especificado até a ordem N , é possível conhecer a função de distribuição (ou, de forma análoga, sua função densidade de probabilidade) para qualquer ordem menor que N por meio da integração (ALBUQUERQUE; FORTES; FINAMORE, 2008).

No entanto, trabalhar com a especificação completa do processo estocástico, como em (2.3), é extremamente complexa e em alguns casos impossível, uma vez que não é possível conhecer todas essas distribuições de dimensão finita, como indicam Albuquerque, Fortes e Finamore (2008) e Morettin e Toloi (2006). Uma alternativa que é utilizada na prática para estudar processos estocásticos é restringi-lo a uma especificação máxima de ordem $N = 2$, bem como estudar suas características baseada em momentos. O momento de um processo estocástico nada mais é do que o momento da variável aleatória X em qualquer ponto de indexação $\mathbf{s} \in \mathcal{T}$ (ALBUQUERQUE; FORTES; FINAMORE, 2008). Existem dois tipos de momentos: momento não central (ou momento centrado na origem) e momento central (ou momento centrado na média). Considerando um processo estocástico especificado até a ordem N , o r -ésimo momento produto não central para esse processo, tal que $r = r_1 + \dots + r_N$, é definido como

$$\begin{aligned} \mathbb{E}[X^{r_1}(\mathbf{s}_1) \dots X^{r_N}(\mathbf{s}_N)] &= \\ &= \int \dots \int x^{r_1}(\mathbf{s}_1) \dots x^{r_N}(\mathbf{s}_N) p(X(\mathbf{s}_1), \dots, X(\mathbf{s}_N)) \partial x(\mathbf{s}_1) \dots \partial x(\mathbf{s}_N), \end{aligned} \quad (2.4)$$

ao passo que, o r -ésimo momento produto central, para $r = r_1 + \dots + r_N$, é dado por

$$\begin{aligned} \mathbb{E}[(X(\mathbf{s}_1) - \mu(\mathbf{s}_1))^{r_1} \dots (X(\mathbf{s}_N) - \mu(\mathbf{s}_N))^{r_N}] &= \\ &= \int \dots \int (x(\mathbf{s}_1) - \mu(\mathbf{s}_1))^{r_1} \dots (x(\mathbf{s}_N) - \mu(\mathbf{s}_N))^{r_N} p(X(\mathbf{s}_1), \dots, X(\mathbf{s}_N)) \partial x(\mathbf{s}_1) \dots \partial x(\mathbf{s}_N), \end{aligned} \quad (2.5)$$

sendo $(x(\mathbf{s}_i) - \mu(\mathbf{s}_i))$ o desvio entre $x(\mathbf{s}_i)$ e sua respectiva média $\mu(\mathbf{s}_i)$ e $p(X(\mathbf{s}_1), \dots, X(\mathbf{s}_N))$ a função densidade de probabilidade de $X(\mathbf{s}_1), \dots, X(\mathbf{s}_N)$ obtida a partir de (2.3).

Considerando um processo estocástico especificado até a primeira ordem, com função densidade de probabilidade $p(X(\mathbf{s}))$, tem-se que o primeiro momento não central de X , pela expressão (2.4), é dado por

$$\mathbb{E}[X(\mathbf{s})] = \mu(\mathbf{s}) = \int x(\mathbf{s}) p(X(\mathbf{s})) dx(\mathbf{s}), \quad \mathbf{s} \in \mathcal{T}, \quad (2.6)$$

definido como a média da variável aleatória X associada ao ponto de indexação \mathbf{s} . Utilizando a expressão em (2.5) para o primeiro momento central, é possível verificar que ele é nulo, uma vez que $\mathbb{E}(X(\mathbf{s}) - \mu(\mathbf{s})) = \mu(\mathbf{s}) - \mu(\mathbf{s}) = 0$. Considerando ainda a expressão (2.5), tem-se que o segundo momento central de X é dado por

$$\mathbb{E}[(X(\mathbf{s}) - \mu(\mathbf{s}))^2] = \sigma^2(\mathbf{s}) = \int (x(\mathbf{s}) - \mu(\mathbf{s}))^2 p(X(\mathbf{s})) dx(\mathbf{s}), \quad \mathbf{s} \in \mathcal{T}, \quad (2.7)$$

na qual é definida como a função de variância de X em um ponto \mathbf{s} .

Agora, para um processo estocástico especificado até a ordem $N = 2$, com função densidade de probabilidade $p(X(\mathbf{s}_1), X(\mathbf{s}_2))$, tem-se que o segundo momento produto não central de $X(\mathbf{s}_1)$ e $X(\mathbf{s}_2)$, através de (2.4) e considerando $r_1 = r_2 = 1$ e $r = r_1 + r_2 = 2$, é dado por

$$\mathbb{E}[X(\mathbf{s}_1) X(\mathbf{s}_2)] = \int \int x(\mathbf{s}_1)x(\mathbf{s}_2) p(X(\mathbf{s}_1), X(\mathbf{s}_2)) \partial x(\mathbf{s}_1)\partial x(\mathbf{s}_2), \quad (2.8)$$

chamada de função de autocorrelação do processo estocástico, que correlaciona a variável aleatória X em dois pontos de indexação, $\mathbf{s}_1 \in \mathcal{T}$ e $\mathbf{s}_2 \in \mathcal{T}$, quaisquer. De forma análoga, considerando $r_1 = r_2 = 1$ e $r = r_1 + r_2 = 2$, o segundo momento produto central do processo é dado por

$$\begin{aligned} \mathbb{E}[(X(\mathbf{s}_1) - \mu(\mathbf{s}_1)) (X(\mathbf{s}_2) - \mu(\mathbf{s}_2))] &= \mathbb{E}[X(\mathbf{s}_1) X(\mathbf{s}_2)] - \mu(\mathbf{s}_1) \mu(\mathbf{s}_2) = \\ &= Cov[X(\mathbf{s}_1), X(\mathbf{s}_2)] = \\ &= \int \int (x(\mathbf{s}_1) - \mu(\mathbf{s}_1))(x(\mathbf{s}_2) - \mu(\mathbf{s}_2)) p(X(\mathbf{s}_1), X(\mathbf{s}_2)) \partial x(\mathbf{s}_1)\partial x(\mathbf{s}_2), \end{aligned} \quad (2.9)$$

sendo definido como a função de autocovariância do processo, mensurada em dois pontos de indexação quaisquer, $\mathbf{s}_1 \in \mathcal{T}$ e $\mathbf{s}_2 \in \mathcal{T}$. Note que a função de autocovariância (segundo momento central) depende da função de autocorrelação (segundo momento não central). Como um caso particular, supondo que $\mathbf{s}_1 = \mathbf{s}_2 = \mathbf{s}$, tem-se que a função de autocovariância se reduz à $\sigma^2(\mathbf{s})$, a variância do processo estocástico como na expressão em (2.7).

Para se fazer inferências sobre essas características relacionadas ao processo estocástico, tal como a média, a variância e a covariância, em geral, é necessário que múltiplas realizações do processo sejam obtidas para que as técnicas inferenciais estatísticas possam ser empregadas. No entanto, é muito comum em estudos envolvendo processos estocásticos que apenas uma única amostra seja obtida. Desse modo, os métodos clássicos de inferência baseada em múltiplas realizações de uma variável aleatória e amplamente utilizada na teoria estatística se tornam inviáveis (SCHABENBERGER; GOTWAY, 2004). Nesse caso, as chamadas *suposições de estacionariedade* são admitidas na teoria estocástica justamente para contornar essa falta de amostras observadas. De acordo com Albuquerque, Fortes e Finamore (2008), existem vários níveis de estacionariedade, que vão desde suposições mais fortes sob a estrutura probabilística do processo até às mais fracas.

O primeiro tipo de estacionariedade é a chamada *estacionariedade estrita*, também denominada de *estacionariedade forte*, que supõe que

$$F[x(\mathbf{s}_1), \dots, x(\mathbf{s}_N)] = F[x(\mathbf{s}_1 + \mathbf{h}), \dots, x(\mathbf{s}_N + \mathbf{h})], \quad (2.10)$$

isto é, a função de distribuição de dimensão finita do processo é invariante sob translações de comprimento \mathbf{h} . Em outras palavras, isso quer dizer que a distribuição do processo não se altera sob qualquer movimento feito nos pontos de indexação em $\mathcal{T} \subset \mathbb{R}^d$. No entanto, como o próprio nome já diz, essa suposição é muito forte, pois impõe restrições sobre toda a estrutura probabilística do processo estocástico e, portanto, poucos fenômenos atenderão a essas condições (SCHABENBERGER; GOTWAY, 2004).

Desse modo, uma suposição mais suave pode ser imposta ao processo, o que garante que mais fenômenos possam atendê-la de forma mais satisfatória. A *estacionariedade de segunda ordem* ou, simplesmente, *estacionariedade fraca* estabelece que

$$\mathbb{E}[X(\mathbf{s})] = \mu$$

e

$$Cov[X(\mathbf{s}_i), X(\mathbf{s}_j)] = Cov(\mathbf{h}),$$

sendo $\mathbf{h} = \|\mathbf{s}_i - \mathbf{s}_j\|$ a distância euclidiana entre os pontos \mathbf{s}_i e \mathbf{s}_j . Isso é, a estacionariedade de segunda ordem implica que a média do processo não se altera sob o conjunto de indexação e que a função de covariância do processo é dada apenas como função de \mathbf{h} . É possível verificar que as condições aqui estabelecidas são mais fracas, uma vez que a suposição de estacionariedade é feita sobre os momentos do processo estocástico e não sobre sua distribuição.

Para processos estocásticos indexados em um subconjunto de \mathbb{R}^d , para $d \geq 2$, um outro tipo de suposição é feito além da estacionariedade de segunda ordem. Quando o processo, além de ser invariante sob translação, também for invariante sob rotação (ou estacionário sob rotações), o processo estocástico é chamado de *isotrópico* (RIPLEY, 1981). Isso implica que a covariância do processo mensurado em dois pontos de indexação quaisquer irá depender unicamente da distância euclidiana que os separam, independente de sua localização e de sua direção em uma determinada região espacial.

2.1.1 Processos estocásticos Gaussianos

A partir dos conceitos básicos de processos estocásticos, diversos modelos estatísticos podem ser descritos dependendo da natureza da variável aleatória ou do conjunto de indexação associado à ela. Um caso particular de processos estocásticos, bastante importante para a compreensão dessa tese e amplamente utilizado na literatura, são os chamados *processos*

estocásticos Gaussianos. De acordo com Albuquerque, Fortes e Finamore (2008), os processos Gaussianos possuem bastante relevância na literatura devido à sua simplicidade teórica e a grande capacidade de explicar e modelar diversos fenômenos presentes na natureza. Particularmente para essa tese, o processo Gaussiano se torna de suma importância devido à sua relação com a classe de processos estocásticos de Cox, uma família de modelos amplamente utilizada para modelar configurações pontuais, teoria que será abordada posteriormente.

Segundo Dobrow (2016), um processo Gaussiano é definido por meio de uma variável aleatória Gaussiana, denotada por f , indexada por um conjunto de pontos $\{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \dots\} \in \mathcal{T} \subset \mathbb{R}^d$, sendo completamente especificado por uma função de média, denotada por $\mu(\mathbf{s})$, e função de covariância entre dois pontos $\mathbf{s} \in \mathcal{T}$ e $\mathbf{s}' \in \mathcal{T}$, denotada por $K(\mathbf{s}, \mathbf{s}')$. Ou seja,

$$f \sim \mathcal{PG}(\mu(\mathbf{s}), K(\mathbf{s}, \mathbf{s}')).$$

Além disso, considerando qualquer subconjunto de N pontos indexadores definido sob o processo, tem-se que o vetor de variáveis aleatórias $\mathbf{f} = [f(\mathbf{s}_1), f(\mathbf{s}_2), \dots, f(\mathbf{s}_N)]^\top$ segue uma distribuição Normal N -variada, com vetor de médias e matriz de covariâncias dados, respectivamente, por

$$\boldsymbol{\mu} = \begin{bmatrix} \mu(\mathbf{s}_1) \\ \mu(\mathbf{s}_2) \\ \vdots \\ \mu(\mathbf{s}_N) \end{bmatrix} \quad \text{e} \quad \mathbf{K} = \begin{bmatrix} K(\mathbf{s}_1, \mathbf{s}_1) & K(\mathbf{s}_1, \mathbf{s}_2) & \dots & K(\mathbf{s}_1, \mathbf{s}_N) \\ K(\mathbf{s}_2, \mathbf{s}_1) & K(\mathbf{s}_2, \mathbf{s}_2) & \dots & K(\mathbf{s}_2, \mathbf{s}_N) \\ \vdots & \vdots & \ddots & \vdots \\ K(\mathbf{s}_N, \mathbf{s}_1) & K(\mathbf{s}_N, \mathbf{s}_2) & \dots & K(\mathbf{s}_N, \mathbf{s}_N) \end{bmatrix},$$

e com função densidade de probabilidade conjunta

$$p(\mathbf{f}; \boldsymbol{\mu}, \mathbf{K}) = (2\pi)^{-N/2} |\mathbf{K}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{f} - \boldsymbol{\mu})^\top \mathbf{K}^{-1} (\mathbf{f} - \boldsymbol{\mu}) \right\}, \quad (2.11)$$

cujo os elementos da matriz \mathbf{K} podem ser calculados através de

$$K(\mathbf{s}_i, \mathbf{s}_j) = \gamma \times c(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta}), \quad (2.12)$$

sendo γ o parâmetro de variância do processo Gaussiano e $c(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta})$ uma função que correlaciona a variável Gaussiana em dois pontos distintos do conjunto de indexação, controlada pelo vetor de parâmetros $\boldsymbol{\theta}$ que parametriza a dependência estocástica existente entre as variáveis aleatórias.

Na literatura, foram propostos vários modelos para descrever a estrutura de correlação estocástica existente no processo Gaussiano, que também é chamada de *função kernel*. Para que uma determinada função de covariância e, conseqüentemente, de correlação seja considerada válida, ela deve ser caracterizada como uma função positiva semi-definida, de tal modo que a forma quadrática $\mathbf{a}^\top \mathbf{K} \mathbf{a}$ atenda a condição $\mathbf{a}^\top \mathbf{K} \mathbf{a} \geq 0$, para todos os vetores $\mathbf{a} \in \mathbb{R}^N$ (RASMUSSEN, 2006a).

Assumindo que o momento de segunda ordem do processo Gaussiano, $K(\mathbf{s}, \mathbf{s}')$, seja invariante sob translação e rotação, ou seja, estacionário de segunda ordem e isotrópico, alguns modelos para $c(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta})$ podem ser usados para descrevê-lo, de acordo com Rasmussen (2006a). A Tabela 2.1 apresenta alguns deles.

Tabela 2.1 – Exemplos de funções de correlação para processos Gaussianos e que são consideradas positiva semi-definidas.

Funções de correlação	Expressões para $c(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta})$	$\boldsymbol{\theta}$
Matérn	$\frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{\sqrt{2\nu}}{\alpha} \ \mathbf{s} - \mathbf{s}'\ \right)^\nu B_\nu \left(\frac{\sqrt{2\nu}}{\alpha} \ \mathbf{s} - \mathbf{s}'\ \right)$	$\boldsymbol{\theta} = \{\nu, \alpha\}$
Exponencial quadrático	$\exp \left\{ -\frac{(\mathbf{s} - \mathbf{s}')^2}{2\alpha^2} \right\}$	$\boldsymbol{\theta} = \{\alpha\}$
Exponencial	$\exp \left\{ -\frac{\ \mathbf{s} - \mathbf{s}'\ }{\alpha} \right\}$	$\boldsymbol{\theta} = \{\alpha\}$
β -exponencial	$\exp \left\{ -\left(\frac{\ \mathbf{s} - \mathbf{s}'\ }{\alpha} \right)^\beta \right\}$	$\boldsymbol{\theta} = \{\alpha, \beta\}$
Quártica racional	$\left(1 + \frac{(\mathbf{s} - \mathbf{s}')^2}{2\delta\alpha^2} \right)^\delta$	$\boldsymbol{\theta} = \{\alpha, \delta\}$

Dentre as várias funções de correlação mostradas na Tabela 2.1, destaca-se a função de correlação exponencial quadrática que, segundo Rasmussen (2006a), é uma das mais usadas em estudos desse tipo, e que será a função escolhida para o desenvolvimento dessa tese.

Como explicado anteriormente, diferentes modelos de processos estocásticos podem ser estabelecidos, considerando as diferentes especificações que, tanto a variável aleatória de interesse quanto o conjunto de indexação, podem assumir. A próxima subseção apresentará um novo modelo de processo estocástico, na qual o processo tem por objetivo descrever ocorrências de um determinado fenômeno aleatório, podendo ser usado para modelar uma série de aplicações, como ocorrências de doenças epidemiológicas, acidentes, terremotos, e muitos outros.

2.1.2 Processos pontuais

Um processo pontual, denotado por $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \dots\}$, é descrito como um fenômeno que gera pontos aleatoriamente sob um domínio $\mathcal{T} \subset \mathbb{R}^d$, para $d \geq 1$, a partir de um

modelo estocástico. A realização desse processo consiste na geração de um conjunto de pontos, denominados por eventos, de quantidade finita e distribuídos de forma irregular em \mathcal{T} (CRESSIE, 1993). O conjunto de eventos obtidos a partir da realização do processo pontual recebe várias denominações na literatura, tais como: configurações pontuais, padrão de pontos ou padrão pontual (SCHABENBERGER; GOTWAY, 2004). De acordo com Rasmussen (2018), a terminologia “ponto” é utilizada, pois o evento está associado a uma localização precisa de sua ocorrência, representando um ponto na linha temporal ou em uma região espacial.

Padrões pontuais distribuídos na reta real, $d = 1$, geralmente representam processos pontuais indexados no tempo, ou seja, é uma sequência de tempos aleatórios que representam o momento de ocorrência de um determinado evento, $\mathcal{S} = \{t_1, t_2, \dots, t_N\}$. Apesar da maior parte das aplicações envolvendo configurações pontuais unidimensionais estar associada ao domínio do tempo, podem haver situações em que os eventos estejam indexados em outro conjunto de indexação que não seja necessariamente o tempo. Com o objetivo de distinguir as as diferentes abordagens envolvendo processos pontuais, será utilizada a notação

$$\{t_1, t_2, \dots, t_N\}$$

para representar uma configuração pontual indexada no tempo,

$$\{s_1, s_2, \dots, s_N\}$$

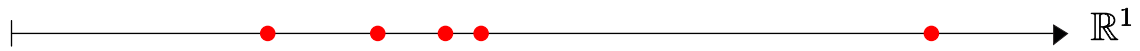
para descrever um padrão pontual indexado em qualquer outro subconjunto de \mathbb{R}^1 e que seja diferente do tempo, ao passo que

$$\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N\}$$

representará padrões de pontos indexados em qualquer subconjunto de \mathbb{R}^d , para $d \geq 2$. Quando os eventos de uma configuração pontual estão indexados em um plano, $d = 2$, ou em outra dimensão maior, $d \geq 3$, geralmente, eles representam localizações espaciais de ocorrências de um fenômeno (BADDELEY; BÁRÁNY; SCHNEIDER, 2007). Nessa tese, será dada uma ênfase maior para processos pontuais unidimensionais.

Na Figura 2.4 apresenta-se uma configuração pontual com indexação unidimensional.

Figura 2.4 – Representação de uma configuração pontual unidimensional.



Fonte: Do autor (2021).

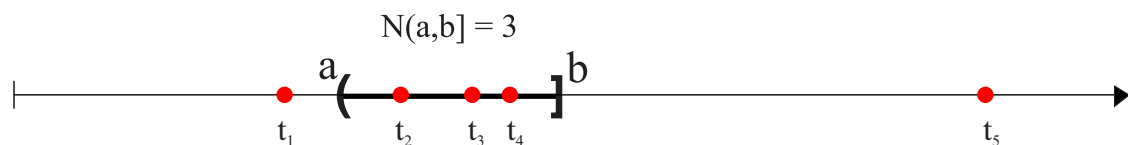
De acordo com Daley e Vere-Jones (2003), existem várias formas de lidar matematicamente com um processo pontual no domínio do tempo. Entre elas:

- a) **Medidas de contagem:** se um processo pontual é estudado por meio de uma contagem de eventos, um subconjunto (ou intervalo) $(a, b]$ da reta real deve ser estabelecido. Além disso, é especificada uma medida $\mathbb{N}(\cdot)$ que contabiliza o número de eventos ocorridos em $(a, b]$, definido como

$$\mathbb{N}(a, b) = \#\{t_i \in (a, b]\}, \quad \text{para } (a, b] \subset \mathbb{R}^1. \quad (2.13)$$

Assim, a medida $\mathbb{N}(\cdot)$, definida em (2.13), deve ser um valor inteiro e não negativo. Na Figura 2.5 apresenta-se um exemplo de uma contagem em uma configuração pontual na reta real;

Figura 2.5 – Processo de contagem para uma configuração pontual na reta real, em que a medida $\mathbb{N}(a, b]$ estabelece o número de eventos contidos no intervalo $(a, b]$. No exemplo ilustrativo, a medida \mathbb{N} contabiliza 3 eventos ocorridos no intervalo semi-aberto $(a, b]$.



Fonte: Do autor (2021).

- b) **Funções step:** considerando que a indexação do processo pontual ocorre no conjunto dos reais positivos, \mathbb{R}^+ , uma nova medida pode ser estabelecida. Considerando

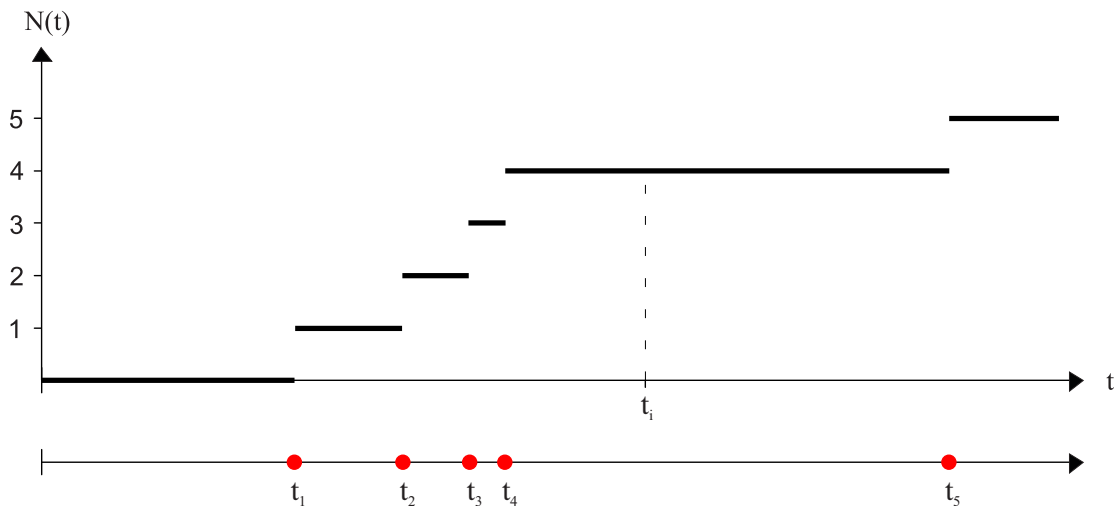
$$\mathbb{N}(t) = \mathbb{N}(0, t], \quad \text{para } 0 < t \leq \infty, \quad (2.14)$$

em que $\mathbb{N}(t)$ é uma medida que quantifica um número acumulado de ocorrências de eventos a partir de um ponto inicial (neste caso, o zero) até um instante de tempo t qualquer. Se, ao contrário do caso anterior, a indexação do processo pontual for em toda a reta real, a definição da função *step* em (2.14) pode ser reescrita como

$$N(t) = \begin{cases} N(0, t], & \text{se } t > 0 \\ 0, & \text{se } t = 0 \\ -N(t, 0], & \text{se } t < 0. \end{cases} \quad (2.15)$$

Com base nessas definições, a função $N(t)$ é uma função crescente, contínua à direita e que assume apenas valores inteiros. Seu comportamento pode ser visto na Figura 2.6;

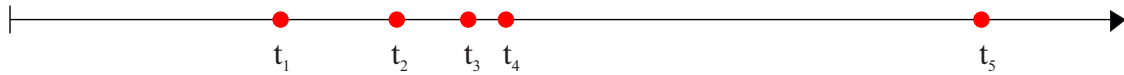
Figura 2.6 – Exemplo de uma função *step* para uma configuração de pontos no tempo, os quais estão contidos no \mathbb{R}^+ . De forma análoga a uma função de distribuição, $N(t)$ quantifica uma contagem acumulativa de eventos até o instante de tempo t . Por exemplo, 4 eventos ocorreram até o instante de tempo t_i .



Fonte: Do autor (2021).

- c) **Sequências de pontos:** esta é a forma mais direta e, no entanto, a menos usual para se trabalhar com um processo pontual na reta. Ao invés de se trabalhar com contagens acumuladas ou em sub-intervalos, a informação do tempo exato da ocorrência do evento é utilizada. No entanto, uma forte dependência entre os eventos é estabelecida, uma vez que $t_i < t_{i+1}$. Assim, uma estrutura de ordem é imposta entre os eventos, como mostra a Figura 2.7;

Figura 2.7 – Exemplo de uma configuração pontual considerando uma sequência de pontos. Neste caso, apesar dos eventos serem gerados de forma aleatória no tempo, uma relação de ordem é estabelecida, de tal forma que $t_1 < t_2 < t_3 < \dots$.



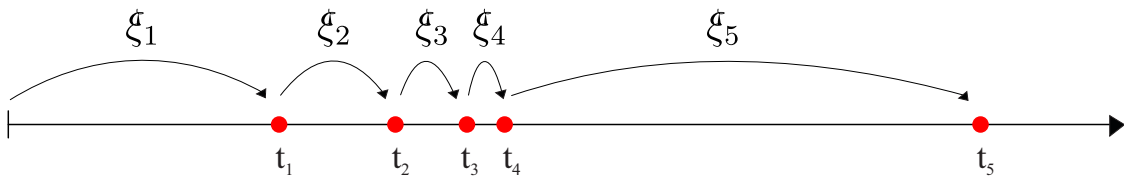
Fonte: Do autor (2021).

- d) **Sequências de intervalos:** por fim, uma última forma de se trabalhar com processos pontuais é considerar uma sequência de tempos intervalares entre eventos pontuais, definido como

$$\xi_i = t_i - t_{i-1}. \quad (2.16)$$

Como pode ser visto na Figura 2.8, ξ_i é uma variável aleatória que assume qualquer valor real positivo.

Figura 2.8 – Exemplo uma configuração pontual considerando uma sequência de intervalos entre os tempos de ocorrências. Neste caso, o estudo do processo se baseia nas diferenças, que é uma variável aleatória contínua e que assume apenas valores positivos.



Fonte: Do autor (2021).

De acordo com Rasmussen (2006b), uma característica importante associada a um processo pontual no tempo é o seu caráter evolutivo. Assim, os eventos ocorrem seguindo uma ordem temporal, que é algo natural e intrínseco a esse tipo de abordagem, sendo de suma importância para a sua modelagem. Dessa forma, a ocorrência de um evento no tempo atual pode depender do que ocorreu no passado. Segundo Baddeley, Bárány e Schneider (2007), essa característica evolutiva é o que diferencia a teoria de processos pontuais no domínio do tempo em relação a outros processos pontuais com conjunto de indexação diferente do tempo, tanto unidimensional quanto de múltiplas dimensões. Processos espaciais, por exemplo, não possuem essa característica, o que torna os seus métodos e técnicas diferentes do caso temporal.

Sendo assim, o conjunto composto por todos os tempos de ocorrências dos eventos antes de um instante de tempo t , é definido como a história do processo pontual, denotado por

$\mathcal{H}_t = \{t_1, t_2, \dots, t_{N-1}\}$, para $t_{N-1} < t$. Este conjunto é essencial para a caracterização de medidas que descrevem o processo pontual (RASMUSSEN, 2018).

Segundo Johnson (1996), um processo pontual no tempo é especificado pelas duas quantidades

$$p[\mathbb{N}[t, t + dt) = 1 \mid \mathcal{H}_t] = \lambda(t \mid \mathcal{H}_t) \times dt \quad (2.17)$$

e

$$p[\mathbb{N}[t, t + dt) > 1 \mid \mathcal{H}_t] = o(t, dt), \quad (2.18)$$

sendo que $\frac{o(t, dt)}{dt}$ tende à zero a medida que $dt \rightarrow 0$. Isto é, a equação (2.18) define que não mais do que um evento pode ocorrer em um intervalo infinitesimal após t .

Além disso, a equação em (2.17) descreve que a probabilidade de ocorrer um evento em um intervalo suficientemente pequeno em torno de t é igual ao produto do tamanho desse intervalo pela função $\lambda(t \mid \mathcal{H}_t)$. Essa função é conhecida por função de intensidade (ou função taxa) condicional do processo pontual que, segundo Daley e Vere-Jones (2003), é definida como

$$\lambda(t \mid \mathcal{H}_t) \approx \frac{\mathbb{E}[\mathbb{N}(dt) \mid \mathcal{H}_t]}{dt}, \quad (2.19)$$

o qual especifica o número esperado de eventos em uma pequena região em torno de t , dada a historia do processo até esse instante de tempo.

Apesar de uma configuração pontual conter N eventos observados, esse conjunto representa uma única realização do processo estocástico sob o tempo. Neste caso, hipóteses que simplificam a estrutura do processo se tornam essenciais. Uma suposição feita em relação ao processo pontual, é considerá-lo estacionário. Processos pontuais estacionários ocorrem se a sua distribuição não for alterada em decorrência de uma translação no tempo (COX; ISHAM, 1980). De acordo com Daley e Vere-Jones (2003), considerando r subconjuntos, $(a_1, b_1], (a_2, b_2], \dots, (a_r, b_r]$, contidos no conjunto dos números reais positivos, \mathbb{R}^1 , o processo pontual será estritamente estacionário se a distribuição conjunta de $\{\mathbb{N}(a_1 + t, b_1 + t], \mathbb{N}(a_2 + t, b_2 + t], \dots, \mathbb{N}(a_r + t, b_r + t]\}$ não depender de t , sendo $t \geq 0$.

De acordo com a definição, a suposição de estacionariedade é considerada uma suposição forte, pois leva em consideração a distribuição do processo como um todo. Assim como foi falado anteriormente para processos estocásticos em geral, suposições mais fracas podem ser estabelecidas, sendo baseadas apenas nas medidas de momentos do processo pontual. Neste

caso, supõe-se que o processo pontual será fracamente estacionário se a função de intensidade, que se associa ao momento de primeira ordem, for invariante sob translação no conjunto de indexação (COX; ISHAM, 1980).

2.1.2.1 Processos pontuais com marcas

Um estudo que leva em consideração apenas a distribuição pontual de eventos pode, em alguns casos, ser um fator limitante quando se quer obter informações mais precisas sob o fenômeno em questão. Segundo Daley e Vere-Jones (2003), em diversos modelos para processos pontuais, o momento (ou a localização referenciada, no caso espacial) da ocorrência do evento perde sua relevância primordial quando este carrega consigo uma ou mais informações adicionais. Tais informações são denominadas por *marcas*. Nesse caso, o processo pontual passa a ser apenas mais uma componente incorporada a um modelo mais complexo, o qual visa, não apenas o estudo da distribuição de eventos, mas as relações de dependência estocástica entre as marcas e evento/marca.

Dessa forma, define-se um processo pontual marcado como um conjunto aleatório de pares $\mathcal{S}_m = \{(s_1, A(s_1)), (s_2, A(s_2)), \dots\}$ definido em $\mathcal{T} \times \mathcal{M}$, sendo s_i denotado como o i -ésimo evento pertencente à uma subregião $\mathcal{T} \subseteq \mathbb{R}^d$ e $A(s_i)$ denota o valor (ou atributo) da marca associada ao ponto s_i e \mathcal{M} é o espaço das marcas no qual $A(s_i)$ está definida (CHIU et al., 2013). Mais especificamente para o caso temporal, uma configuração pontual marcada é uma sequência finita de eventos, $\mathcal{S}_m = \{(t_1, A(t_1)), (t_2, A(t_2)), \dots, (t_N, A(t_N))\}$, distribuídos sob os reais, $\mathcal{T} \subseteq \mathbb{R}^1$, que carrega consigo uma marca (DU et al., 2016). De acordo com Daley e Vere-Jones (2003), os eventos pontuais marcados estão definidos no produto dos espaços dos eventos \mathcal{T} e das marcas \mathcal{M} . Desse modo, se apenas o conjunto de informações sobre o tempo de ocorrência dos eventos for considerado, tal processo pode ser visto como um processo marginal de \mathcal{S}_m e, nesse contexto, recebe o nome de “processo pontual de base”.

Sobre a natureza das marcas, elas podem ser classificadas como categóricas ou quantitativas. Para o caso quantitativo, o espaço das marcas assume um conjunto de valores reais, que atribui às marcas associadas aos eventos de uma configuração pontual um valor real, isto é, $\mathcal{M} \subseteq \mathbb{R}^1$. Por outro lado, se as marcas forem categóricas, o espaço das marcas se reduz à um conjunto finito e enumerável, composto por \mathcal{K} categorias não numéricas, isto é, $\mathcal{M} = \{1, \dots, \mathcal{K}\}$. Nesse caso, a literatura denomina processo pontual com marcas categóricas, processo pontual multivariado e processos de multi-tipos (ou *multitypes*) como sinônimos.

De acordo com Rasmussen (2018), a distribuição das marcas associada a um evento no tempo t é especificada por meio de uma função de probabilidade (ou função densidade de probabilidade, para o caso quantitativo) condicional, $p(A(t) | t, \mathcal{H}_t)$, sendo que a história do processo agora depende tanto da referência temporal quanto das marcas passadas. Desse modo, o processo é especificado pela função de intensidade condicional marcada, denotada por

$$\lambda(t, A(t) | \mathcal{H}_t) = \lambda(t | \mathcal{H}_t) p(A(t) | t, \mathcal{H}_t), \quad (2.20)$$

em que $\lambda(t | \mathcal{H}_t)$ é a intensidade do processo de base, mas que também pode depender das marcas dos eventos passados.

Seguindo a definição para processos pontuais sem a existência das marcas, a função de intensidade condicional marcada é interpretada de forma análoga, incluindo apenas a informação da marca. Nesse caso, se a marca for categórica, a intensidade condicional descreve o número médio de eventos associados à k -ésima marca, $A(t) = k$, em um pequeno intervalo de tempo dt , sendo denotada por

$$\lambda(t, A(t) = k | \mathcal{H}_t) = \frac{\mathbb{E}[\mathbb{N}(dt \times k) | \mathcal{H}_t]}{dt}. \quad (2.21)$$

No entanto, se a marca for quantitativa, a intensidade em (2.20) é interpretada como o número esperado de eventos em um pequeno intervalo de tempo dt e em um pequeno intervalo de valores associados à marca, da , sendo denotado por

$$\lambda(t, A(t) | \mathcal{H}_t) = \frac{\mathbb{E}[\mathbb{N}(dt \times da) | \mathcal{H}_t]}{dt \times da}. \quad (2.22)$$

De acordo com Cox e Isham (1980), o objetivo principal na pesquisa sob a teoria de marcas se encontra no estudo da distribuição conjunta entre os pontos e as marcas, com um enfoque maior no estudo da existência de uma estrutura de correlação estocástica. Mais precisamente, três cenários possíveis podem acontecer:

- a) as marcas são variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) e independem do processo pontual;
- b) existe uma estrutura de dependência entre as marcas e os eventos de um processo pontual, sendo as marcas i.i.d. entre si;
- c) as marcas estão correlacionadas umas com as outras e também com os eventos.

Uma particularidade da teoria de processos pontuais marcados e que resulta em outra teoria é considerar as marcas de um processo pontual no tempo como sendo localizações referenciadas, em que o espaço das marcas é o \mathbb{R}^d . Neste caso, tem-se um processo pontual espaço temporal (RASMUSSEN, 2006b).

2.2 Modelos para processos pontuais

Nessa seção serão apresentados os processos de Poisson e de Cox, classes de modelos usuais para representar o mecanismo estocástico de um processo pontual. Até o presente momento, a teoria de processos pontuais foi, particularmente, apresentada considerando a indexação sob o ponto de vista temporal, devido à sua particularidade evolutiva do tempo. A partir de agora, os modelos serão apresentados para qualquer tipo de indexação unidimensional, incluindo o tempo. Desse modo, será usada a notação s ao invés de t para denotar a configuração pontual.

2.2.1 Processos de Poisson

Um processo de Poisson é a estrutura mais simples para a representação de uma configuração pontual, seja ela no domínio do tempo, no espaço ou em qualquer outro conjunto de indexação. De acordo com Kingman (1996), tal processo é utilizado para descrever o comportamento de fenômenos que se distribuem de forma completamente aleatória sobre o espaço de indexação. Considerando um processo pontual de indexação unidimensional e tomando r subconjuntos disjuntos, $(a_1, b_1]$, $(a_2, b_2]$, \dots , $(a_r, b_r]$, um processo de Poisson é completamente definido pela distribuição conjunta do número de elementos em cada subintervalo:

$$\begin{aligned} p[\mathbb{N}(a_1, b_1] = N_1, \dots, \mathbb{N}(a_r, b_r] = N_r] &= \prod_{i=1}^r p[\mathbb{N}(a_i, b_i] = N_i] \\ &= \prod_{i=1}^r \frac{\mathbb{M}(a_i, b_i]^{N_i}}{N_i!} \exp\{-\mathbb{M}(a_i, b_i]\}. \end{aligned} \quad (2.23)$$

A partir da distribuição conjunta em (2.23), duas propriedades podem ser obtidas:

- a) se $\mathbb{N}(a_i, b_i]$ denota o número de eventos aleatórios que ocorrem no intervalo arbitrário $(a_i, b_i] \subset \mathbb{R}^1$, então $\mathbb{N}(a_i, b_i] \sim \text{Poisson}(\mathbb{M}(a_i, b_i])$, sendo $\mathbb{M}(a_i, b_i] > 0$ o número médio de eventos no intervalo $b_i - a_i$;

b) se $(a_1, b_1]$, $(a_2, b_2]$, \dots , $(a_r, b_r]$ são subconjuntos disjuntos de \mathbb{R}^1 , então $\mathbb{N}(a_1, b_1]$, $\mathbb{N}(a_2, b_2]$, \dots , $\mathbb{N}(a_r, b_r]$ são variáveis aleatórias independentes.

Outra forma de escrever o comportamento médio do processo de Poisson, $\mathbb{M}(\cdot)$, é por meio da função de intensidade, definida em (2.19). Como esse processo possui a propriedade de independência entre as ocorrências de eventos, então a função de intensidade condicional para um processo pontual temporal se torna $\lambda(t | \mathcal{H}_t) = \lambda(t)$. Desse modo, o número médio de eventos no intervalo $(a_i, b_i]$ pode ser obtido por

$$\mathbb{M}(a_i, b_i] = \int_{a_i}^{b_i} \lambda(s) ds. \quad (2.24)$$

Segundo Kingman (1996), se a medida de intensidade, λ , for uma função que varia pelo conjunto indexador, então a média de eventos, definida em (2.24), para um pequeno intervalo $(a, b]$ em torno de s , pode ser aproximada por

$$\mathbb{M}(a, b] \approx \lambda(s) (b - a). \quad (2.25)$$

Nesse caso, o processo pontual recebe o nome de processo de Poisson *não homogêneo* (ou *não uniforme*), pois o comportamento da intensidade muda com s . Assim, $\lambda(s) (b - a)$ é interpretado como uma probabilidade aproximada de um evento ocorrer em um pequeno intervalo, de tamanho $b - a$. Essa probabilidade será alta se a intensidade neste intervalo também for, ou baixa, caso contrário.

Do mesmo modo, se a função de intensidade for considerada uma taxa constante por todo o conjunto de indexação, isto é, $\lambda(s) = \lambda$, o número médio de eventos em um intervalo arbitrário será exatamente igual à

$$\mathbb{M}(a, b] = \lambda (b - a). \quad (2.26)$$

Logo, o processo de Poisson é classificado como sendo *homogêneo* (ou *uniforme*). Esse tipo de processo coincide, também, com os chamados processos fracamente estacionários definidos anteriormente, pois o comportamento médio do processo não muda devido à uma translação de \mathcal{S} . Em outras palavras, considerando dois subconjuntos $(a_1, b_1]$ e $(a_2, b_2]$, sendo $(a_2, b_2] = (a_1 + s, b_1 + s]$, o número médio de eventos em cada um deles será o mesmo, uma vez que a intensidade é constante e o comprimento dos intervalos são os mesmos.

Dado que o processo de Poisson (homogêneo e não homogêneo) é considerado a classe de modelos mais simples em toda a teoria de processos pontuais, outros diversos tipos de classes

de modelos foram formuladas por meio de uma generalização em sua teoria. Dentre elas, se destaca a classe de processos de Poisson duplamente estocásticos, o qual será apresentada na próxima subseção.

2.2.2 Processos de Cox

Os processos de Poisson duplamente estocásticos representam uma classe importante de modelos para processos pontuais, o qual surgiu por meio de uma generalização dos processos de Poisson, a partir dos estudos de Lundberg (1940). Posteriormente, esta classe veio a ser denominada como processos de Cox, graças às contribuições feitas por Cox (1955).

Um processo de Cox pode ser definido como um processo pontual dirigido por um processo de intensidade, λ . Isto é, a função de intensidade que descreve o comportamento médio do processo pontual também é fruto de um mecanismo aleatório. Daí surgiu a terminologia de um processo “duplamente estocástico”. Sua relação com o processo de Poisson é que, dada uma realização do processo de intensidade, o processo pontual condicional segue um processo de Poisson não homogêneo com função de intensidade $\lambda(s)$. Segundo Adams, Murray e MacKay (2009), a motivação por trás desse tipo de abordagem surge a partir da necessidade de se modelar uma função de intensidade de um processo de Poisson que varia ao longo da indexação, mas que sua forma funcional não é conhecida. Desse modo, um novo processo estocástico é incorporado ao modelo a fim de modelar não parametricamente a variação da intensidade ao longo de sua indexação.

Segundo Kingman (1996), definir o processo de Cox é impor um conjunto de suposições sob a distribuição conjunta condicional das contagens em subintervalos, dado $\lambda(s)$. Desse modo, tomando r subconjuntos disjuntos, $(a_1, b_1]$, $(a_2, b_2]$, \dots , $(a_r, b_r]$, sua distribuição conjunta, condicionado a uma realização de $\lambda(s)$, é dada por:

$$\begin{aligned} p[\mathbb{N}(a_1, b_1] = N_1, \dots, \mathbb{N}(a_r, b_r] = N_r \mid \lambda(s)] &= \prod_{i=1}^r p[\mathbb{N}(a_i, b_i] = N_i \mid \lambda(s)] \\ &= \prod_{i=1}^r \frac{\mathbb{M}(a_i, b_i]^{N_i}}{N_i!} \exp\{-\mathbb{M}(a_i, b_i]\}, \end{aligned} \quad (2.27)$$

em que

$$\mathbb{M}(a_i, b_i] = \int_{a_i}^{b_i} \lambda(s) ds, \quad (2.28)$$

e $\lambda(s)$ é um processo estocástico contínuo indexado sobre \mathcal{T} e que assume valores positivos na reta real.

De acordo com Kingman (1996), o processo de intensidade pode assumir várias formas, o que torna o processo de Cox uma classe de modelos para processos pontuais bastante flexível. Uma forma de caracterizar o processo de intensidade é considerá-lo em função de um outro processo estocástico, denotado por $f(s)$. Considerando que $f(s)$ é um processo Gaussiano, o processo de Cox recebe a terminologia *processo de Cox modulado por um processo Gaussiano* ou, simplesmente, *processo de Cox Gaussiano*, que assume a forma

$$\lambda(s) = g(f(s)), \quad (2.29)$$

em que $f(s)$ também é um processo estocástico contínuo indexado em \mathcal{T} e que pode assumir qualquer valor real, sendo parametrizado por uma função de média e uma função de covariância entre dois pontos indexadores quaisquer; $g(\cdot)$ é uma função de ligação não negativa para garantir que as propriedades de $\lambda(s)$ sejam mantidas.

Na literatura, diversas propostas para a modelagem utilizando um processo de Cox Gaussiano foram feitas. Dentre elas, destacam-se

- a) *Processo de Cox log-Gaussiano*: desenvolvido por Møller, Syversveen e Waagepetersen (1998), em que seu processo de intensidade é modelado considerando

$$\lambda(s) = \exp \{f(s)\}, \quad (2.30)$$

sendo $g(\cdot) = \exp\{\cdot\}$.

Tal modelo é amplamente utilizado no contexto espacial, tendo diversas extensões feitas a partir da proposta de Møller, Syversveen e Waagepetersen (1998). Dentre elas, destacam-se os trabalhos de Brix e Diggle (2001) e Diggle, Rowlingson e Su (2005) para o contexto espaço-temporal; Kelsall e Wakefield (2002) com uma variação do modelo que permite um ajuste a dados de área; Diggle et al. (2013) adaptado para processos multivariados; entre outros;

- b) *Processo de Cox Gaussiano Sigmoidal*: proposto por Adams, Murray e MacKay (2009), considera a intensidade como

$$\lambda(s) = \lambda^* \sigma(f(s)), \quad (2.31)$$

em que λ^* é o valor máximo que a intensidade pode assumir e $\sigma(f(s)) = (1 + e^{-f(s)})^{-1}$ é uma função logística;

c) *Processo de Cox Gaussiano Quadrático*: tal processo foi primeiramente apresentado por Møller, Syversveen e Waagepetersen (1998) como um processo de Cox Qui-quadrado, pois este leva consigo algumas características da distribuição. No entanto, o modelo só foi explorado, de fato, por Lloyd et al. (2015), que considera o processo de intensidade como

$$\lambda(s) = f^2(s). \quad (2.32)$$

Outros modelos baseados no processo de Cox, mas que não levam em consideração o processo Gaussiano, foram desenvolvidos na literatura. Dentre eles, destacam-se os processos de Neyman-Scott (BARTLETT, 1964), processos de Cox *Shot Noise* (MØLLER, 2003), *Shot Noise* Generalizado (MØLLER; TORRISI, 2005), entre outros.

Na próxima seção, será exposta a teoria Bayesiana usada para a estimação de parâmetros. Esta seção se faz necessária, pois ela dará suporte para a compreensão dos modelos propostos por Lloyd et al. (2015) e Liang, Carlin e Gelfand (2009), que serão abordados posteriormente e se baseiam em processos de Cox Gaussiano com enfoque Bayesiano variacional.

2.3 Inferência Bayesiana

O objetivo da Estatística é conhecer e descrever o comportamento de determinados fenômenos aleatórios, tendo como ferramenta principal amostras obtidas a partir de suas ocorrências. Conhecer seu comportamento é também conhecer suas principais características. Em termos estatísticos, as características de um fenômeno são descritas por meio de um conjunto de vetor de parâmetros, $\theta = \{\theta_1, \dots, \theta_k\}$. Em geral, os valores desses parâmetros são quantidades desconhecidas e incertas e, portanto, métodos e técnicas de inferência estatística são utilizados a fim de estimá-los por meio de tais amostras.

De acordo com Bolstad (2004), a inferência sobre parâmetros pode ser realizada por meio de diversas óticas inferenciais, sendo os métodos frequentista (ou clássico) e bayesiano os principais. A abordagem frequentista se baseia exclusivamente na informação dos dados e um estimador para θ é construído a partir de uma distribuição amostral. Além disso, os parâmetros são considerados como quantidades fixas.

Nos métodos bayesianos, a incerteza envolvida em relação ao valor paramétrico real é tratada como algo aleatório. Desse modo, uma distribuição de probabilidade pode ser atribuída

ao parâmetro, associando um peso maior a valores que podem ser mais prováveis (GHOSH; DELAMPADY; SAMANTA, 2007).

A teoria Bayesiana teve seu desenvolvimento a partir dos estudos do reverendo Thomas Bayes descrito no artigo *An essay towards solving a problem in the doctrine of chances*, publicado em 1763, o qual recebeu o nome de teorema de Bayes em sua homenagem. Seu método começou a ser utilizado no século XIX, porém só em meados do século XX é que a teoria ganhou força e se consolidou entre os pesquisadores (BOLSTAD, 2004).

O método consiste em incorporar uma informação probabilística prévia dos valores paramétricos no processo de estimação, o qual é chamada de distribuição a priori de θ , descrita por $p(\theta)$. Esta priori, segundo Ghosh, Delampady e Samanta (2007), quantifica a incerteza sob θ antes dos dados serem observados.

A distribuição conjunta dos dados amostrais, $\mathbf{x} = \{x_1, \dots, x_n\}$, recebe o nome de função de verossimilhança, descrita por $p(\mathbf{x} | \theta)$, que, quando combinada com a distribuição a priori, $p(\theta)$, obtém-se uma distribuição a posteriori, por meio da relação estabelecida no teorema de Bayes:

$$p(\theta | \mathbf{x}) = \frac{p(\mathbf{x} | \theta) p(\theta)}{p(\mathbf{x})} = \frac{p(\mathbf{x} | \theta) p(\theta)}{\int_{\Theta} p(\mathbf{x} | \theta) p(\theta) d\theta}. \quad (2.33)$$

A distribuição a posteriori, descrita em (2.33), pode ser vista como uma quantificação da incerteza em relação aos parâmetros desconhecidos após a introdução da informação amostral no processo de estimação. Segundo Ghosh, Delampady e Samanta (2007), o aprendizado que se ganha sobre θ por meio da função de verossimilhança é o que fornece a transição da distribuição a priori para a posteriori. Assim, os dados amostrais representam uma componente chave que afeta o conhecimento prévio que se tem sob θ , quando $p(\mathbf{x} | \theta)$ é considerado como função de θ , para x_1, \dots, x_n fixos (GELMAN et al., 2014).

Uma forma equivalente à expressão em (2.33) desconsidera a componente presente no denominador, a distribuição marginal dos dados, também chamada de verossimilhança marginal (CONGDON, 2003) ou de função de evidência (BLEI; KUCUKELBIR; MCAULIFFE, 2017). Uma vez que $p(\mathbf{x})$ não depende do conjunto de parâmetros θ , pode-se considerá-la como uma simples constante normalizadora. Desse modo, a posteriori em (2.33) pode ser reescrita como

$$p(\theta | \mathbf{x}) \propto p(\mathbf{x} | \theta) p(\theta). \quad (2.34)$$

em que \propto representa proporcionalidade.

De acordo com Ghosh, Delampady e Samanta (2007), as informações associadas à posteriori de θ podem ser apresentadas de diversas formas, por meio de medidas descritivas associadas a sua distribuição, tais como a média a posteriori

$$\mathbb{E}(\theta | \mathbf{x}) = \int_{\Theta} \theta p(\theta | \mathbf{x}) d\theta, \quad (2.35)$$

seus quantis, como por exemplo, sua mediana a posteriori

$$\text{Med}(\theta | \mathbf{x}) = \arg_{\theta} \int_{-\infty}^{\theta} p(\theta | \mathbf{x}) d\theta = 0,5, \quad (2.36)$$

a variância a posteriori

$$\begin{aligned} \text{Var}(\theta | \mathbf{x}) &= \mathbb{E}[(\theta - \mathbb{E}(\theta | \mathbf{x}))^2 | \mathbf{x}] \\ &= \int_{\Theta} (\theta - \mathbb{E}(\theta | \mathbf{x}))^2 f(\theta | \mathbf{x}) d\theta, \end{aligned} \quad (2.37)$$

ou, então, o desvio padrão a posteriori, dado por $DP(\theta | \mathbf{x}) = \sqrt{\text{Var}(\theta | \mathbf{x})}$.

Entretanto, o cálculo de integrais como em (2.35), (2.36) e em (2.37) pode ser, em algumas situações, analiticamente intratável. Uma forma de contornar tal problema é utilizando métodos de amostragem Monte Carlo, amplamente utilizado em problemas como esse. Esses métodos possibilitam que estimativas dessas medidas descritivas possam ser tomadas, sendo que essas irão convergir para suas grandezas a posteriori à medida que o número de simulações Monte Carlo se torna cada vez maior (PRESS, 2003).

2.3.1 Métodos de Monte Carlo via Cadeias de Markov

Mais precisamente, quando a distribuição a posteriori não é analiticamente tratável, as amostras podem ser tomadas utilizando algoritmos Monte Carlo via Cadeias de Markov (ou apenas MCMC, usando a sigla em inglês para o termo *Markov Chain Monte Carlo*). Os métodos MCMC tem por finalidade gerar amostras Monte Carlo de uma distribuição alvo (nesse caso, a distribuição a posteriori) a partir de um mecanismo estocástico que é descrito por uma cadeia de Markov (PRESS, 2003). Isto é, esses métodos simulam uma cadeia de Markov cujo estado atual (nesse caso, a última observação gerada da distribuição a posteriori) depende do estado anterior (observação anterior gerada). De acordo com Ghosh, Delampady e Samanta (2007), a ideia dos métodos MCMC é que, para um dado número de simulações relativamente grande, a cadeia de Markov converge para uma distribuição estacionária, sendo esta distribuição

a densidade a posteriori exata. Dentre os diversos métodos existentes na literatura, os algoritmos de Metropolis-Hastings e amostrador de Gibbs são uns dos principais métodos MCMC.

O algoritmo de Metropolis-Hastings, proposto por Metropolis et al. (1953) e Hastings (1970), consiste em tomar amostras de uma função densidade de probabilidade que, neste caso, é a densidade a posteriori $p(\boldsymbol{\theta} \mid \mathbf{x})$. Para isso, um valor inicial, $\boldsymbol{\theta}^{(0)}$, é tomado a partir de uma distribuição qualquer. Dado que $i - 1$ valores amostrais foram simulados, um novo valor amostral, $\boldsymbol{\theta}'$, é obtido a partir de uma densidade de probabilidade arbitrária, $q(\boldsymbol{\theta}^{(i-1)}, \boldsymbol{\theta}')$, denominada de núcleo (ou densidade) de transição. Esse novo valor é considerado um candidato para ocupar a i -ésima amostra simulada e será aceito com probabilidade

$$\alpha(\boldsymbol{\theta}^{(i-1)}, \boldsymbol{\theta}') = \min \left\{ \frac{p(\boldsymbol{\theta}' \mid \mathbf{x})}{p(\boldsymbol{\theta}^{(i-1)} \mid \mathbf{x})} \frac{q(\boldsymbol{\theta}^{(i-1)}, \boldsymbol{\theta}')}{q(\boldsymbol{\theta}', \boldsymbol{\theta}^{(i-1)})}, 1 \right\}, \quad (2.38)$$

em que $\alpha(\boldsymbol{\theta}^{(i-1)}, \boldsymbol{\theta}')$ é denominada como probabilidade de aceitação. Assim, se este valor for aceito, a i -ésima amostra da simulação receberá o valor $\boldsymbol{\theta}'$, isto é, $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}'$. Caso contrário, a i -ésima amostra recebe o valor anterior, isto é, $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^{(i-1)}$. O processo é repetido até que as amostras geradas descrevam uma distribuição estacionária (ROBERT; CASELLA, 2009).

Já o amostrador de Gibbs, proposto por Geman e Geman (1984) e consolidado por Gelfand e Smith (1990) em análises Bayesianas, diferentemente do algoritmo de Metropolis-Hastings, gera amostras de Monte Carlo dos parâmetros de interesse considerando suas respectivas distribuições condicionais completas a posteriori com formas conhecidas.

A técnica de amostragem de Gibbs consiste em decompor o vetor de parâmetros $\boldsymbol{\theta}$ em k sub-vetores, isto é, $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_k]$. Atribuindo valores iniciais para o conjunto de parâmetros, denotados por $\boldsymbol{\theta}^{(0)} = [\boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_2^{(0)}, \dots, \boldsymbol{\theta}_k^{(0)}]$, uma nova amostra é tomada, de forma iterativa, a partir do seguinte procedimento:

obtem-se a amostra $\boldsymbol{\theta}_1^{(1)}$ através de $p(\boldsymbol{\theta}_1 \mid \mathbf{x}, \boldsymbol{\theta}_2^{(0)}, \boldsymbol{\theta}_3^{(0)}, \dots, \boldsymbol{\theta}_k^{(0)})$

obtem-se a amostra $\boldsymbol{\theta}_2^{(1)}$ através de $p(\boldsymbol{\theta}_2 \mid \mathbf{x}, \boldsymbol{\theta}_1^{(1)}, \boldsymbol{\theta}_3^{(0)}, \dots, \boldsymbol{\theta}_k^{(0)})$

⋮

obtem-se a amostra $\boldsymbol{\theta}_k^{(1)}$ através de $p(\boldsymbol{\theta}_k \mid \mathbf{x}, \boldsymbol{\theta}_1^{(1)}, \boldsymbol{\theta}_2^{(1)}, \dots, \boldsymbol{\theta}_{k-1}^{(1)})$

A primeira iteração é finalizada após as k amostras terem sido geradas, obtendo-se $\boldsymbol{\theta}^{(1)} = [\boldsymbol{\theta}_1^{(1)}, \boldsymbol{\theta}_2^{(1)}, \dots, \boldsymbol{\theta}_k^{(1)}]$. Assim, uma nova iteração é feita, $\boldsymbol{\theta}^{(2)}$, através das condicionais

completas, apresentadas anteriormente, e condicionada nas amostras geradas na iteração anterior.

Assim, de acordo com Paulino, Turkman e Murteira (2003) sucessivas amostras de θ serão obtidas por meio de uma cadeia markoviana com um núcleo de transição dado por

$$q(\theta^{(i-1)}, \theta^{(i)}) = \prod_{j=1}^k p(\theta_j^{(i)} | \mathbf{x}, \theta_1^{(i)}, \dots, \theta_{j-1}^{(i)}, \theta_{j+1}^{(i-1)}, \dots, \theta_k^{(i-1)}). \quad (2.39)$$

Supondo t iterações realizadas pelo amostrador de Gibbs, $\theta^{(t)} = [\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_k^{(t)}]$ irá convergir em distribuição para um vetor de variáveis aleatórias cuja distribuição conjunta é a própria distribuição a posteriori, $p(\theta_1, \dots, \theta_k | \mathbf{x})$, a medida que $t \rightarrow \infty$ (BERNARDO; SMITH, 1994). Da mesma forma, a t -ésima iteração para o i -ésimo sub-vetor $\theta_i^{(t)}$ também irá convergir em distribuição para um vetor de variáveis aleatórias cuja distribuição será a marginal a posteriori de θ_i , isto é, $p(\theta_i | \mathbf{x})$. Desse modo, a amostra gerada na t -ésima iteração, para t grande e considerando que houve convergência da cadeia de Markov, para $\theta_i^{(t)}$ será uma amostra aleatória gerada de $p(\theta_i | \mathbf{x})$.

No entanto, os algoritmos baseados em simulação MCMC podem demandar um tempo computacional elevado, principalmente para problemas de inferência com um grande número de parâmetros (BISHOP, 2011). Em outros casos, principalmente em situações na qual o conjunto de dados é muito grande ou os modelos são bastante complexos, os algoritmos MCMC se tornam inviáveis para solucionar problemas inferenciais (BLEI; KUCUKELBIR; MCAULIFFE, 2017). Neste caso, a inferência Bayesiana variacional se torna uma boa alternativa.

2.3.2 Inferência Bayesiana variacional

O método variacional, em sua essência, é um método de otimização usado para maximizar (ou minimizar) uma integral que envolve funções que são desconhecidas (RUSTAGI, 1976). De acordo com Rustagi (1976), o método surgiu no século XVII para solucionar problemas de engenharia a partir dos estudos de Newton, que buscava obter o formato do casco de um navio que assegurasse o mínimo de resistência da água. Com o passar dos anos, o método foi aprimorado e passou a ser usado em diferentes áreas do conhecimento e aplicações. Especificamente no contexto estatístico, a inferência Bayesiana atrelada ao uso variacional teve seu início na década de 1980 (ZHANG et al., 2019).

Segundo Blei, Kucukelbir e McAuliffe (2017), a inferência Bayesiana variacional é uma abordagem alternativa aos algoritmos MCMC para obter aproximações de densidades a poste-

riori difíceis de serem obtidas analiticamente pelo teorema de Bayes em (2.33), devido a integração presente na função de evidência $p(\mathbf{x})$.

Este esquema de aproximação consiste em propor uma família de distribuições densidade de probabilidade \mathcal{Q} associadas ao vetor de parâmetros desconhecidos do modelo, sendo cada $q(\boldsymbol{\theta}) \in \mathcal{Q}$ uma densidade candidata para representar a verdadeira posteriori de interesse (ZHANG et al., 2019). As densidades que compõem \mathcal{Q} são chamadas de *densidades variacionais*. De acordo com Blei, Kucukelbir e McAuliffe (2017), o objetivo da inferência variacional é obter a densidade variacional $q(\cdot)$ que melhor se aproxima do comportamento aleatório da real densidade a posteriori, $p(\boldsymbol{\theta} \mid \mathbf{x})$. Para isso, é necessário utilizar de alguma medida que consiga mensurar o grau de similaridade entre duas distribuições. Na literatura, diversas medidas foram propostas para esta finalidade, como a divergência- α , divergência- f e a discrepância de Stein, como apontado por Zhang et al. (2019). No entanto, a medida mais utilizada para esta finalidade é a *divergência de Kullback-Leibler* ou, simplesmente, divergência KL, devido à sua simplicidade teórica em relação as demais e por apresentar soluções fechadas para determinados modelos. Considerando duas funções densidades de probabilidade, $q(\boldsymbol{\theta})$ e $p(\boldsymbol{\theta} \mid \mathbf{x})$, a divergência KL é definida como

$$\begin{aligned} KL(q(\boldsymbol{\theta}) \parallel p(\boldsymbol{\theta} \mid \mathbf{x})) &= \int q(\boldsymbol{\theta}) \log \left(\frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta} \mid \mathbf{x})} \right) d\boldsymbol{\theta} \\ &= \mathbb{E}_q \left[\log \left(\frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta} \mid \mathbf{x})} \right) \right], \end{aligned} \quad (2.40)$$

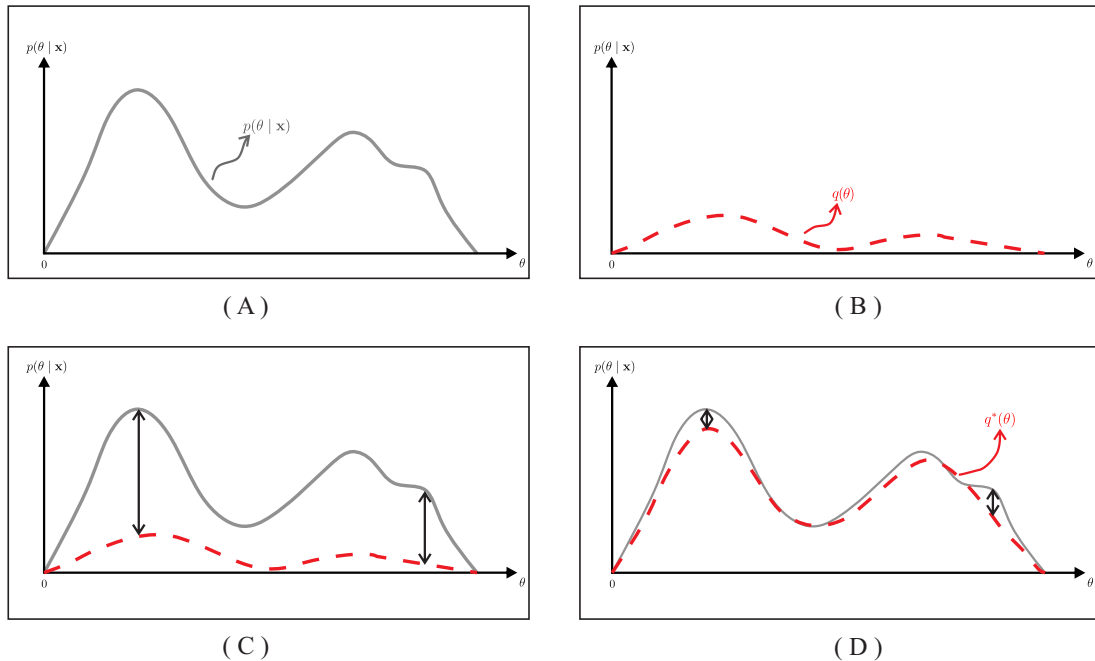
sendo $KL(q(\cdot) \parallel p(\cdot)) \geq 0$. Se a distribuição variacional for exatamente igual a distribuição a posteriori verdadeira, $q(\cdot) = p(\cdot)$, tem-se que $KL(q(\cdot) \parallel p(\cdot)) = 0$.

Sendo assim, o objetivo da inferência variacional é, simplesmente, minimizar a divergência KL entre a densidade variacional e a densidade a posteriori de interesse, a fim tornar o comportamento probabilístico de ambas o mais similar possível (BLEI; KUCUKELBIR; MCAULIFFE, 2017), como mostra a Figura 2.9. Em outras palavras, a inferência variacional se reduz a obter uma densidade variacional dentre todas aquelas pertencentes a \mathcal{Q} , tal que a divergência KL seja mínima:

$$q^*(\boldsymbol{\theta}) = \underset{q(\boldsymbol{\theta}) \in \mathcal{Q}}{\operatorname{argmin}} KL(q(\boldsymbol{\theta}) \parallel p(\boldsymbol{\theta} \mid \mathbf{x})). \quad (2.41)$$

Blei, Kucukelbir e McAuliffe (2017) ressaltam que a finalidade do uso do método variacional na inferência Bayesiana é a mesma dos algoritmos MCMC: obter a distribuição a posteriori de interesse. No entanto, o caminho que cada método segue é diferente. Enquanto

Figura 2.9 – Ilustração do método variacional aplicado à inferência Bayesiana, na qual a densidade a posteriori de interesse (A) é aproximada pela densidade variacional (B) usando a divergência de Kullback-Leibler como critério de otimização (C). Uma vez obtida uma densidade variacional ótima (D), tem-se que $q^*(\theta) \approx p(\theta | \mathbf{x})$ e a divergência de Kullback-Leibler é próximo de zero, isto é, $KL(q(\theta) \parallel p(\theta | \mathbf{x})) \approx 0$.



Fonte: Adaptado de Silveira (2018).

os métodos MCMC se baseiam em simular amostras Monte Carlo, os métodos variacionais se baseiam em otimização. Outra característica apresentada pelos autores é que a distribuição estacionária obtida pelos métodos MCMC resulta em uma amostra da verdadeira posteriori ao passo que, a densidade variacional otimizada não é exata, apenas uma aproximação da posteriori.

No entanto, otimizar a divergência KL em (2.40) se torna inviável, pois a expressão envolve a verdadeira distribuição a posteriori que inclui a função de evidência, difícil de ser calculada. Desenvolvendo a expressão em (2.40), tem-se que

$$\begin{aligned}
KL(q(\boldsymbol{\theta}) \parallel p(\boldsymbol{\theta} \mid \mathbf{x})) &= \mathbb{E}_q \left[\log \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta} \mid \mathbf{x})} \right] \\
&= \mathbb{E}_q [\log q(\boldsymbol{\theta})] - \mathbb{E}_q [\log p(\boldsymbol{\theta} \mid \mathbf{x})] \\
&= \mathbb{E}_q [\log q(\boldsymbol{\theta})] - \mathbb{E}_q \left[\log \left(\frac{p(\mathbf{x} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{x})} \right) \right] \\
&= \mathbb{E}_q [\log q(\boldsymbol{\theta})] - \mathbb{E}_q [\log p(\boldsymbol{\theta})] - \mathbb{E}_q [\log p(\mathbf{x} \mid \boldsymbol{\theta})] + \mathbb{E}_q [\log p(\mathbf{x})] \\
&= \mathbb{E}_q \left[\log \left(\frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta})} \right) \right] - \mathbb{E}_q [\log p(\mathbf{x} \mid \boldsymbol{\theta})] + \log p(\mathbf{x}) \\
&= KL(q(\boldsymbol{\theta}) \parallel p(\boldsymbol{\theta})) - \mathbb{E}_q [\log p(\mathbf{x} \mid \boldsymbol{\theta})] + \log p(\mathbf{x}) \\
&= -\mathcal{L} + \log p(\mathbf{x}).
\end{aligned} \tag{2.42}$$

Ou seja, a divergência KL entre a densidade variacional e a densidade a posteriori é fatorada em dois termos: logaritmo da função de evidência, $\log p(\mathbf{x})$, e o *limite inferior da log evidência*, denotado por \mathcal{L} . Este termo recebe esse nome devido a sua relação com a função de log-evidência, uma vez que

$$\begin{aligned}
KL(q(\cdot) \parallel p(\cdot)) &\geq 0 \\
\log p(\mathbf{x}) - \mathcal{L} &\geq 0 \\
\log p(\mathbf{x}) &\geq \mathcal{L}.
\end{aligned}$$

Ou seja, a quantidade \mathcal{L} é o menor valor que $\log p(\mathbf{x})$ pode assumir.

Uma vez que $KL(q(\cdot) \parallel p(\cdot))$ é uma medida difícil de ser minimizada por conta da presença do termo $\log p(\mathbf{x})$ na expressão, uma alternativa é otimizar o termo \mathcal{L} . Como \mathcal{L} é inversamente proporcional a divergência KL, minimizar $KL(q(\cdot) \parallel p(\cdot))$ é análogo a maximizar \mathcal{L} . Desse modo, a distribuição variacional ótima em (2.41) pode ser encontrada, de maneira análoga, obtendo uma densidade variacional, dentre todas aquelas pertencentes a família de densidades \mathcal{Q} , que maximize o termo \mathcal{L} . Isto é,

$$q^*(\boldsymbol{\theta}) = \operatorname{argmax}_{q(\boldsymbol{\theta}) \in \mathcal{Q}} \mathcal{L}, \tag{2.43}$$

sendo $\mathcal{L} = \mathbb{E}_q [\log p(\mathbf{x} \mid \boldsymbol{\theta})] - KL(q(\boldsymbol{\theta}) \parallel p(\boldsymbol{\theta}))$ uma quantidade que depende tanto dos hiperparâmetros que parametrizam a distribuição variacional quanto dos hiperparâmetros que parametrizam a distribuição a priori de $\boldsymbol{\theta}$.

De acordo com Blei, Kucukelbir e McAuliffe (2017), a escolha da família de densidades variacionais é um fator que pode afetar o processo de otimização do método. Isto é, a escolha

de famílias variacionais mais simples implicará em um processo de otimização mais rápido e simples também. Nesse sentido, uma abordagem bastante utilizada na literatura é considerar a chamada *família variacional do campo médio*, que parte da ideia de que o vetor de parâmetros é particionado em k parâmetros mutuamente independentes, tal que

$$q(\boldsymbol{\theta}) = \prod_{i=1}^k q_i(\theta_i). \quad (2.44)$$

Vale ressaltar que a utilização dessa estratégia para se obter $q(\boldsymbol{\theta})$ apresenta vantagens e desvantagens. Apesar do uso desta família variacional conseguir capturar o efeito das densidades marginais de cada θ_i , para $i = 1, \dots, k$, e simplificar a otimização, ela não consegue capturar possíveis efeitos de correlação que possam existir entre os parâmetros. Neste caso, a utilização do campo médio se torna inviável para modelos mais complexos com estruturas de dependência. Além disso, estudos mostram que supor independência entre os parâmetros pode levar a uma sub-estimação da variância marginal variacional quando comparado com as posterioris reais (BLEI; KUCUKELBIR; MCAULIFFE, 2017).

Em relação a forma da distribuição variacional de cada parâmetro sob a ótica da família do campo médio, cada parâmetro pode assumir qualquer forma paramétrica dependendo do seu comportamento. No entanto, Blei, Kucukelbir e McAuliffe (2017) descrevem formas mais simples de se obter essas distribuições considerando densidades pertencentes a família exponencial. Trabalhar com distribuições que estão nesta família podem simplificar ainda mais o processo de otimização, principalmente com o uso do algoritmo chamado *inferência variacional de ascensão por coordenadas* (IVAC), pois a densidade variacional para cada parâmetro de interesse pode ser otimizada como

$$q_i^*(\theta_i) \propto \exp \left\{ \mathbb{E}_{q_{-i}} [\log p(\theta_i | \boldsymbol{\theta}_{-i}, \mathbf{x})] \right\}, \quad (2.45)$$

em que $p(\theta_i | \boldsymbol{\theta}_{-i}, \mathbf{x})$ é a condicional completa a posteriori de θ_i e $\mathbb{E}_{q_{-i}}(\cdot)$ é a esperança variacional em relação a todos os parâmetros, exceto θ_i .

Assim como é usado no algoritmo do amostrador de Gibbs, as distribuições condicionais completas a posteriori também desempenham um papel importante na utilização do algoritmo IVAC, demonstrando a forte relação existente entre essas duas abordagens. No entanto, ao invés de se gerar amostras MCMC dessa condicional completa, o IVAC toma o valor esperado do logaritmo desta função (BLEI; KUCUKELBIR; MCAULIFFE, 2017).

Além do IVAC, existem outros algoritmos de otimização mais gerais que podem ser utilizados na inferência variacional, principalmente quando não se usa a família do campo médio para representar as densidades variacionais. São eles: o algoritmo do gradiente descendente e o gradiente descendente estocástico (HOFFMAN et al., 2013) (BLEI; KUCUKELBIR; MCAULIFFE, 2017), bem como os algoritmos de Quase-Newton, como o BFGS (Broyden-Fletcher-Goldfarb-Shanno) (GERSHMAN; HOFFMAN; BLEI, 2012).

A inferência variacional é um método de estimação que tem ganhado espaço nos últimos anos em análises Bayesianas nos últimos anos e que será utilizada nesta tese. Na próxima seção, será apresentada a estrutura inferencial desenvolvida no modelo proposto por Lloyd et al. (2015), que utiliza as técnicas variacionais para resolver seus problemas de estimação, cuja teoria serve como base para a modelagem proposta nesta tese.

2.4 Modelo proposto por Lloyd et al. (2015)

Lloyd et al. (2015) propuseram uma nova estrutura de modelagem para processos de Cox Gaussiano com a introdução da inferência Bayesiana variacional, o que permitiu um desenvolvimento de estimação com maior rapidez do ponto de vista computacional. Além disso, o método variacional estabelece uma forma fechada para a distribuição a posteriori do processo estocástico, permitindo assim que suas respectivas estimativas possam ser obtidas para qualquer ponto pertencente ao conjunto de indexação, não necessitando de uma discretização sob esse conjunto como é usual nos métodos MCMC. Além disso, o modelo buscou contornar um problema da chamada alta dimensionalidade dos dados, levando ao uso de um *processo Gaussiano esparso*, que será explicado um pouco mais a frente.

Apesar do seu modelo ser flexível para configurações pontuais tanto unidimensionais quanto espaciais (no espaço d -dimensional, para qualquer $d \geq 2$), esta seção tratará o modelo apenas para o caso unidimensional, uma vez que o objetivo de pesquisa é lidar com processos gerados na reta real.

Para a modelagem da função de intensidade, Lloyd et al. (2015) utilizaram uma função de ligação quadrática para conectar $\lambda(s)$ ao processo Gaussiano $f(s)$, levando a expressão

$$\lambda(s) = f^2(s). \quad (2.46)$$

Considerando o teorema de Bayes, a distribuição a posteriori para o modelo é dada por

$$p(f | \mathcal{S}, \Theta) = \frac{p(\mathcal{S} | f, \Theta) p(f | \Theta)}{p(\mathcal{S})}, \quad (2.47)$$

sendo $p(\mathcal{S} | f, \Theta)$ a verossimilhança da configuração pontual, condicionado ao processo Gaussiano f , expressa na forma de um processo de Poisson não homogêneo com função de intensidade, definida como (2.46),

$$p(\mathcal{S} | f, \Theta) = \exp \left\{ - \int_{\mathcal{T}} f^2(s) ds \right\} \times \prod_{i=1}^N f^2(s_i). \quad (2.48)$$

Já $p(f | \Theta)$ descreve a densidade a priori de f , que segue um processo Gaussiano (\mathcal{PG}) com função de média nula e função de covariância $K(s, s')$, definida por

$$\begin{aligned} K(s, s') &= \gamma \times c(s, s'; \alpha) \\ &= \gamma \times \exp \left\{ - \frac{(s - s')^2}{2\alpha} \right\}, \end{aligned} \quad (2.49)$$

em que γ é a variância do processo Gaussiano a priori comum para qualquer ponto $s \in \mathcal{T}$ ($K(s, s') = \gamma$, para $s = s'$) e $c(s, s'; \alpha)$ uma função de correlação exponencial quadrática, apresentada anteriormente na Tabela 2.1, sendo α um parâmetro de escala que controla a dependência do processo em dois pontos distintos, s e s' . Por fim, $p(\mathcal{S})$ é função de evidência (ou verossimilhança marginal) do modelo, difícil de ser obtida devido à uma dupla integração estocástica presente na expressão:

$$\begin{aligned} p(\mathcal{S}) &= \int p(\mathcal{S}, f | \Theta) df \\ &= \int p(f | \Theta) p(\mathcal{S} | f, \Theta) df \\ &= \int p(f | \Theta) \exp \left\{ - \int_{\mathcal{T}} f^2(s) ds \right\} \times \prod_{i=1}^N f^2(s_i) df, \end{aligned} \quad (2.50)$$

sendo $\Theta = \{\gamma, \alpha\}$ o conjunto de hiperparâmetros do modelo.

Para contornar o problema da obtenção de $p(\mathcal{S})$, Lloyd et al. (2015) propuseram o esquema inferencial variacional para o modelo em questão. Considerando uma densidade variacional para f , tem-se que $q(f) = \mathcal{PG}(\tilde{\mu}(s), \tilde{\Sigma}(s, s'))$, sendo $\tilde{\mu}(s)$ e $\tilde{\Sigma}(s, s')$ funções variacionais de média e covariância respectivamente. Por questão de simplificação na notação, será usado $\tilde{\mu}_s$ para representar $\tilde{\mu}(s)$ e, de modo análogo, $\tilde{\Sigma}_{ss'}$ para representar $\tilde{\Sigma}(s, s')$. Desse modo, tomando a divergência KL, definida em (2.42), para mensurar a similaridade entre $p(f | \mathcal{S}, \Theta)$ e $q(f)$, tem-se que o limite inferior da log evidência é dado por

$$\mathcal{L} = \mathbb{E}_q[\log p(\mathcal{S} | f, \Theta)] - KL(q(f) || p(f | \Theta)), \quad (2.51)$$

sendo $\mathbb{E}_q[\log p(\mathcal{S} | f, \Theta)]$ o valor esperado da função de log-verossimilhança em relação a $q(f)$ e $KL(q(f) || p(f | \Theta))$ a divergência de Kullback-Leibler entre a densidade variacional de f e a sua respectiva priori.

Aplicando os processos Gaussianos a priori e variacionais no conjunto de localizações dos N eventos da configuração pontual, a expressão para $KL(q(f) || p(f | \Theta))$ em (2.51) pode ser desenvolvida como

$$KL(q(f) || p(f | \Theta)) = \frac{1}{2} \left\{ \log \frac{\det \mathbf{K}_{ss'}}{\det \tilde{\Sigma}_{ss'}} - N + \text{tr} \{ \mathbf{K}_{ss'}^{-1} \tilde{\Sigma}_{ss'} \} + \tilde{\boldsymbol{\mu}}_s^\top \mathbf{K}_{ss'}^{-1} \tilde{\boldsymbol{\mu}}_s \right\}. \quad (2.52)$$

Uma vez que $KL(q(f) || p(f | \Theta))$ é um dos termos presente na função \mathcal{L} e sabendo que esta mesma função \mathcal{L} é usada como critério alternativo de otimização para aproximar $q(f)$ de $p(f | \mathcal{S}, \Theta)$, utilizar a abordagem variacional da forma como foi explicitada pode apresentar um custo computacional elevado. Isso se deve ao fato de que o termo $KL(q(f) || p(f | \Theta))$ em (2.52) inclui a inversão da matriz $\mathbf{K}_{ss'}$ de dimensão $N \times N$, sendo N o número de eventos observados na configuração pontual e que, em geral, é um número relativamente grande.

A fim de eliminar esse problema da inversão de matrizes de altas dimensões, os autores introduziram ao modelo a abordagem dos chamados processos Gaussianos esparsos. A ideia por traz do método esparsos em processos Gaussianos é introduzir um conjunto de M pontos de indução (também chamado de pseudo-pontos), denotados por $\mathcal{Z} = \{z_1, \dots, z_M\}$ para $M \ll N$, na qual esses pontos tem por objetivo refletir o comportamento dos dados originais, \mathcal{S} . Aplicando o processo Gaussiano nesses pontos de indução, um vetor de variáveis aleatórias *de indução* é obtido, denotado por $f(\mathcal{Z}) = \{f(z_1), \dots, f(z_M)\} = \{u_1, \dots, u_M\} = \mathbf{u}$, seguindo uma distribuição Gaussiana multivariada a priori: $\mathbf{u} \sim \mathcal{N}(\vec{\mathbf{1}}\bar{u}, \mathbf{K}_{zz})$, sendo $\vec{\mathbf{1}}$ um vetor coluna unitário de ordem M , \bar{u} o valor médio de \mathbf{u} comum para todos os pontos e \mathbf{K}_{zz} uma matriz de covariâncias de dimensão $M \times M$ cujas entradas são calculadas pela expressão do kernel em (2.49), considerando os pontos de indução \mathcal{Z} . Além disso, o modelo supõe que os parâmetros do kernel para calcular $\mathbf{K}_{ss'}$ e \mathbf{K}_{zz} são os mesmos, ressaltando assim que f e \mathbf{u} compartilham as mesmas características.

Deste modo, uma nova priori conjunta é determinada para o modelo, considerando tanto o processo estocástico Gaussiano f quando o vetor de indução Gaussiano multivariado u , que pode ser fatorada na forma

$$p(f, \mathbf{u} | \Theta) = p(f | \mathbf{u}, \Theta) p(\mathbf{u} | \Theta), \quad (2.53)$$

em que $f | \mathbf{u} \sim \mathcal{PG}(\mu(s), \Sigma(s, s'))$, sendo

$$\mu(s) = \mathbf{K}_{sz} \mathbf{K}_{zz}^{-1} \mathbf{u}, \quad (2.54)$$

$$\Sigma(s, s') = \mathbf{K}_{ss'} - \mathbf{K}_{sz} \mathbf{K}_{zz}^{-1} \mathbf{K}_{zs'}, \quad (2.55)$$

na qual \mathbf{K}_{sz} e $\mathbf{K}_{zs'}$ descrevem uma função matricial de covariâncias cruzadas entre f e \mathbf{u} , considerando os pontos em \mathcal{S} e \mathcal{Z} respectivamente, obtidas através da expressão (2.49). Além disso, Θ agora incorpora um novo hiperparâmetro, resultando em $\Theta = \{\gamma, \alpha, \bar{u}\}$. As expressões em (2.54) e (2.55) são obtidas considerando uma combinação linear de distribuições Gaussianas.

Sendo assim, a distribuição a posteriori do modelo em (2.47) pode ser reescrita como

$$p(f, \mathbf{u} | \mathcal{S}, \Theta) = \frac{p(\mathcal{S} | f, \Theta) p(f | \mathbf{u}, \Theta) p(\mathbf{u} | \Theta)}{p(\mathcal{S})}. \quad (2.56)$$

Para a formulação da inferência variacional, uma nova densidade variacional conjunta é proposta, na qual $q(f, \mathbf{u}) = q(f | \mathbf{u})q(\mathbf{u})$, em que $q(\mathbf{u})$ é uma função densidade de probabilidade Gaussiana multivariada, $\mathcal{N}(\mathbf{m}, \mathbf{S})$, sendo \mathbf{m} e \mathbf{S} um vetor de médias e uma matriz de covariâncias variacionais, respectivamente. Para simplificação dos cálculos a fim de garantir que alguns termos na divergência KL sejam cancelados, o modelo supõe que $q(f | \mathbf{u}) = p(f | \mathbf{u}, \Theta)$. Além disso, uma vez que $p(f | \mathbf{u}, \Theta)$ é conjugada com $q(\mathbf{u})$, é possível obter $q(f)$ na forma fechada a partir da integração da conjunta

$$\begin{aligned} q(f) &= \int q(f, \mathbf{u}) d\mathbf{u} \\ &= \int p(f | \mathbf{u}, \Theta) q(\mathbf{u}) d\mathbf{u} \\ &= \mathcal{PG}(\tilde{\mu}(s), \tilde{\Sigma}(s, s')), \end{aligned} \quad (2.57)$$

em que

$$\tilde{\mu}(s) = \mathbf{K}_{sz} \mathbf{K}_{zz}^{-1} \mathbf{m} \quad (2.58)$$

$$\tilde{\Sigma}(s, s') = \mathbf{K}_{ss'} - \mathbf{K}_{sz} \mathbf{K}_{zz}^{-1} \mathbf{K}_{zs'} + \mathbf{K}_{sz} \mathbf{K}_{zz}^{-1} \mathbf{S} \mathbf{K}_{zz}^{-1} \mathbf{K}_{zs'}. \quad (2.59)$$

Desse modo, a divergência KL pode ser obtida como

$$\begin{aligned}
KL(q(f, \mathbf{u}) \parallel p(f, \mathbf{u} \mid \mathcal{S}, \Theta)) &= \mathbb{E}_q \left[\log \frac{q(f, \mathbf{u})}{p(f, \mathbf{u} \mid \mathcal{S}, \Theta)} \right] \\
&= \mathbb{E}_q[\log q(f, \mathbf{u})] - \mathbb{E}_q[\log p(f, \mathbf{u} \mid \mathcal{S}, \Theta)] \\
&= \left\{ \mathbb{E}_q[\log p(f \mid \mathbf{u}, \Theta)] + \mathbb{E}_q[\log q(\mathbf{u})] \right\} - \left\{ \mathbb{E}_q[\log p(\mathcal{S} \mid f, \Theta)] \right. \\
&\quad \left. + \mathbb{E}_q[\log p(f \mid \mathbf{u}, \Theta)] + \mathbb{E}_q[\log p(\mathbf{u} \mid \Theta)] - \mathbb{E}_q[\log p(\mathcal{S})] \right\} \\
&= - \left\{ \mathbb{E}_q[\log p(\mathcal{S} \mid f, \Theta)] - \mathbb{E}_q \left[\log \frac{q(\mathbf{u})}{p(\mathbf{u} \mid \Theta)} \right] \right\} + \log p(\mathcal{S}) \\
&= - \left\{ \mathbb{E}_q[\log p(\mathcal{S} \mid f, \Theta)] - KL(q(\mathbf{u}) \parallel p(\mathbf{u} \mid \Theta)) \right\} + \log p(\mathcal{S}) \\
&= -\mathcal{L} + \log p(\mathcal{S}).
\end{aligned} \tag{2.60}$$

Assim, como exposto anteriormente, a divergência KL não é uma boa medida para ser usada como critério de otimização, uma vez que ela envolve o termo $\log p(\mathcal{S})$. Sendo assim, ao invés de minimizar a divergência KL, será maximizado o limite inferior da log evidência, dado por

$$\begin{aligned}
\mathcal{L} &= \mathbb{E}_q[\log p(\mathcal{S} \mid f)] - KL(q(\mathbf{u}) \parallel p(\mathbf{u}, \Theta)) \\
&= \mathbb{E}_q \left[- \int_{\mathcal{T}} f^2(s) ds + \sum_{i=1}^N \log f^2(s_i) \right] - KL(q(\mathbf{u}) \parallel p(\mathbf{u}, \Theta)) \\
&= - \int_{\mathcal{T}} \mathbb{E}_q[f^2(s)] ds + \sum_{i=1}^N \mathbb{E}_q[\log f^2(s_i)] - KL(q(\mathbf{u}) \parallel p(\mathbf{u}, \Theta)) \\
&= - \int_{\mathcal{T}} \mathbb{E}_q[f(s)]^2 ds - \int_{\mathcal{T}} Var_q[f(s)] ds + \sum_{i=1}^N \mathbb{E}_q[\log f^2(s_i)] - KL(q(\mathbf{u}) \parallel p(\mathbf{u}, \Theta)),
\end{aligned} \tag{2.61}$$

em que $KL(q(\mathbf{u}) \parallel p(\mathbf{u}, \Theta))$ é a divergência KL entre a densidade variacional de \mathbf{u} e a sua respectiva densidade a priori. Desenvolvendo sua expressão, tem-se que

$$KL(q(\mathbf{u}) \parallel p(\mathbf{u}, \Theta)) = \frac{1}{2} \left\{ \log \frac{\det \mathbf{K}_{zz}}{\det \mathbf{S}} - M + tr\{\mathbf{K}_{zz}^{-1} \mathbf{S}\} + (\vec{\mathbf{1}}\bar{u} - \mathbf{m})^\top \mathbf{K}_{zz}^{-1} (\vec{\mathbf{1}}\bar{u} - \mathbf{m}) \right\}. \tag{2.62}$$

Comparando as expressões (2.52) e (2.62), pode-se verificar que o problema da alta dimensionalidade da matriz $\mathbf{K}_{ss'}$, de dimensão $N \times N$, foi contornado com a inclusão das variáveis de indução ao modelo, uma vez que a inversão de $\mathbf{K}_{ss'}$ foi reduzida à uma inversão da matriz \mathbf{K}_{zz} com dimensão muito inferior, $M \times M$. Isso reduz a complexidade de tempo do algoritmo de otimização, levando a uma eficiência computacional maior.

De acordo com Lloyd et al. (2015), os termos de \mathcal{L} apresentam algumas vantagens que permitem que sejam calculadas esperanças em relação a $q(f)$ em qualquer ponto s , uma vez

que $q(f(s))$ também é um processo Gaussiano. Essas vantagens se devem a três motivos destacados pelos autores: (i) a formulação de um processo Gaussiano condicional, permitindo que esperanças tratáveis sejam tomadas em relação a $q(f)$; (ii) as variáveis de indução já foram integradas; (iii) o uso de uma função de ligação quadrática entre a intensidade estocástica e o processo Gaussiano.

Desenvolvendo o termo $\int_{\mathcal{T}} \mathbb{E}_q[f(s)]^2 ds$ em \mathcal{L} , a partir da expressão (2.58), tem-se que

$$\mathbb{E}_q[f(s)]^2 = \tilde{\mu}^2(s) = \mathbf{m}^\top \mathbf{K}_{zz}^{-1} \mathbf{K}_{zs} \mathbf{K}_{sz} \mathbf{K}_{zz}^{-1} \mathbf{m} \quad (2.63)$$

e, portanto,

$$\begin{aligned} \int_{\mathcal{T}} \mathbb{E}_q[f(s)]^2 ds &= \mathbf{m}^\top \mathbf{K}_{zz}^{-1} \int_{\mathcal{T}} \{\mathbf{K}_{zs} \mathbf{K}_{sz}\} ds \mathbf{K}_{zz}^{-1} \mathbf{m} \\ &= \mathbf{m}^\top \mathbf{K}_{zz}^{-1} \Psi \mathbf{K}_{zz}^{-1} \mathbf{m}, \end{aligned} \quad (2.64)$$

sendo Ψ uma matriz de dimensão $M \times M$, em que seus elementos são obtidos a partir da integração do produto entre $K(z, s)$ e $K(s, z')$, dados por

$$\begin{aligned} \Psi(z, z') &= \int_{\mathcal{T}} K(z, s) K(s, z') ds \\ &= -\gamma^2 \frac{\sqrt{\pi\alpha}}{2} \exp\left\{-\frac{(z - z')^2}{4\alpha}\right\} \times \left[\operatorname{erf}\left(\frac{(\bar{z} - \mathcal{T}_{max})}{\sqrt{\alpha}}\right) - \operatorname{erf}\left(\frac{(\bar{z} - \mathcal{T}_{min})}{\sqrt{\alpha}}\right) \right], \end{aligned} \quad (2.65)$$

cuja demonstração pode ser vista no Apêndice B.1, sendo $\bar{z} = \frac{z+z'}{2}$ e $\operatorname{erf}(\cdot)$ a função erro de Gauss.

Já o termo $\int_{\mathcal{T}} \operatorname{Var}_q[f(s)] ds$ em \mathcal{L} é obtido a partir da integração da expressão (2.59), considerando $s = s'$, de tal modo que

$$\operatorname{Var}_q[f(s)] = \tilde{\sigma}^2(s) = \gamma - \operatorname{Tr}(\mathbf{K}_{zz}^{-1} \mathbf{K}_{zs} \mathbf{K}_{sz}) + \operatorname{Tr}(\mathbf{K}_{zz}^{-1} \mathbf{S} \mathbf{K}_{zz}^{-1} \mathbf{K}_{zs} \mathbf{K}_{sz}) \quad (2.66)$$

e, portanto,

$$\begin{aligned} \int_{\mathcal{T}} \operatorname{Var}_q[f(s)] ds &= \int_{\mathcal{T}} \gamma ds - \operatorname{Tr}\left(\mathbf{K}_{zz}^{-1} \int_{\mathcal{T}} \{\mathbf{K}_{zs} \mathbf{K}_{sz}\} ds\right) + \operatorname{Tr}\left(\mathbf{K}_{zz}^{-1} \mathbf{S} \mathbf{K}_{zz}^{-1} \int_{\mathcal{T}} \{\mathbf{K}_{zs} \mathbf{K}_{sz}\} ds\right) \\ &= \gamma |\mathcal{T}| - \operatorname{Tr}(\mathbf{K}_{zz}^{-1} \Psi) + \operatorname{Tr}(\mathbf{K}_{zz}^{-1} \mathbf{S} \mathbf{K}_{zz}^{-1} \Psi). \end{aligned} \quad (2.67)$$

Já para $\sum_{i=1}^N \mathbb{E}_q[\log f^2(s_i)]$, Lloyd et al. (2015) apresentam uma solução analítica para a expressão envolvendo uma função hipergeométrica confluyente. A expressão é dada por

$$\begin{aligned}
\sum_{i=1}^N \mathbb{E}_q[\log f^2(s_i)] &= \sum_{i=1}^N \int_{-\infty}^{+\infty} \log(f^2(s_i)) \mathcal{N}(f(s_i); \tilde{\mu}(s_i), \tilde{\sigma}^2(s_i)) df(s_i) \\
&= \sum_{i=1}^N \left\{ -\tilde{G}\left(-\frac{\tilde{\mu}^2(s_i)}{2\tilde{\sigma}^2(s_i)}\right) + \log\left(\frac{\tilde{\sigma}^2(s_i)}{2}\right) - C \right\},
\end{aligned} \tag{2.68}$$

em que $C \approx 0,57721566$ é a constante de Euler-Mascheroni e $\tilde{G}(\cdot)$ é uma função definida a partir da derivada parcial de

$${}_1F_1(a, b, z) = \sum_{k=0}^{\infty} \frac{(a)_k z^k}{(b)_k k!},$$

em relação ao argumento “ a ”. A função ${}_1F_1(a, b, z)$ é chamada de *função hipergeométrica confluyente* e $(\cdot)_k$ descreve a crescente série de Pochhammer, tal que $(a)_0 = 1$ e $(a)_k = a(a+1)(a+2)\dots(a+k+1)$. Neste caso,

$$\tilde{G}(z) = \frac{\partial {}_1F_1}{\partial a}(a, b, z) = {}_1F_1^{(1,0,0)}\left(0, \frac{1}{2}, z\right) = 2z \sum_{j=0}^{\infty} \frac{j! z^j}{(2)_j (1\frac{1}{2})_j}$$

Uma vez que os termos que compõem \mathcal{L} foram especificados em (2.62), (2.64), (2.67) e em (2.68), o processo de inferência do modelo ocorre por meio da maximização de \mathcal{L} em relação aos hiperparâmetros, tanto a priori quanto variacional. Isto é, os hiperparâmetros do modelo vão sendo atualizados por meio de um processo iterativo de otimização até que se obtenha um ponto de máximo de \mathcal{L} . O processo de otimização é feito por meio de algoritmos baseados em gradientes, cujo os cálculos são apresentados no Apêndice B.2 e B.3. Dentre os vários algoritmos existentes na literatura destaca-se o algoritmo BFGS que, dentre os algoritmos quase-Newton, tem como vantagem a utilização apenas dos gradientes (derivadas parciais de primeira ordem) do modelo. O Apêndice A apresenta de forma mais detalhada a teoria do algoritmo de otimização BFGS.

Uma vez otimizados os hiperparâmetros do modelo, γ^* , α^* , \bar{u}^* , \mathbf{m}^* e \mathbf{S}^* , a distribuição a posteriori aproximada é obtida considerando $q^*(f) \approx p(f | \mathcal{S})$, sendo $q^*(f)$ um processo estocástico Gaussiano com funções de média e de covariância obtidas pelas expressões (2.58) e (2.59), respectivamente, considerando os hiperparâmetros otimizados. Com isso, é possível inferir sobre o processo Gaussiano em qualquer ponto $s \in \mathcal{T}$.

Considerando o processo de inferência sob f realizado, as estimativas da função de intensidade do processo pontual de Cox pode ser obtida por meio da relação em (2.46). Para isso, é necessário encontrar a distribuição de probabilidade de $\lambda(s)$. Sabendo de sua relação com

o processo Gaussiano e que, para cada ponto de indexação s específico, $f(s)$ segue uma distribuição Gaussiana univariada a posteriori, com média e variância denotadas respectivamente por $\tilde{\mu}^*(s)$ e $\tilde{\sigma}^{2*}(s)$, $\lambda(s)$ seguirá uma distribuição variacional Qui-quadrada não centralizada escalonada, com $\nu = 1$ grau de liberdade e um parâmetro δ de não centralidade dado pela razão entre $\tilde{\mu}^{2*}(s)$ e $\tilde{\sigma}^{2*}(s)$. Isto é,

$$\lambda(s) \sim \tilde{\sigma}^{2*}(s) \chi^2 \left(\nu = 1, \delta = \frac{\tilde{\mu}^{2*}(s)}{\tilde{\sigma}^{2*}(s)} \right), \quad (2.69)$$

com média variacional dada por

$$\begin{aligned} \mathbb{E}[\lambda(s)] &= \tilde{\sigma}^{2*}(s) \times (\nu + \delta) \\ &= \tilde{\sigma}^{2*}(s) + \tilde{\mu}^{2*}(s) \end{aligned} \quad (2.70)$$

e variância variacional

$$\begin{aligned} \text{Var}[\lambda(s)] &= \tilde{\sigma}^{2*}(s) \times (2\nu + 4\delta) \\ &= 2\tilde{\sigma}^{2*}(s) + 4\tilde{\mu}^{2*}(s). \end{aligned} \quad (2.71)$$

Uma vez que a distribuição a posteriori de $\lambda(s)$ foi obtida de maneira aproximada pela distribuição variacional, estimativas pontuais de $\lambda(s)$ podem ser estabelecidas por meio de sua média variacional, em (2.70), ou pela mediana variacional, definida por

$$\text{Med}[\lambda(s)] = \arg_{\lambda} \int_0^{\lambda} q^*(\lambda') d\lambda' = 0,5, \quad (2.72)$$

ou, ainda, por seu respectivo máximo variacional (ou moda variacional), definido por

$$\text{Moda}[\lambda(s)] = \arg \max_{\lambda} q^*(\lambda). \quad (2.73)$$

Já para a obtenção dos intervalos de credibilidade à um nível de $(1 - \alpha)100\%$ de probabilidade, é utilizado o método do HPD (sigla em inglês para *highest posterior density*). Tal método tem por finalidade obter um intervalo de valores a posteriori para λ no ponto s com $(1 - \alpha)100\%$ de probabilidade e que sejam os mais prováveis, isto é, o método visa obter aqueles intervalos de maior densidade a posteriori.

2.5 Modelo proposto por Liang, Carlin e Gelfand (2009)

Liang, Carlin e Gelfand (2009) propuseram uma extensão do modelo de Cox log Gaussiano para configurações pontuais espaciais indexadas no conjunto dos reais bidimensionais, \mathbb{R}^2 ,

o qual consiste em descrever o processo de intensidade em função de covariáveis tanto espaciais quanto não espaciais.

A motivação dos autores surge a partir do interesse em se investigar a distribuição espacial das ocorrências de casos de câncer de colon e reto no estado de Minnesota, no período de 1998 a 2002. Além do estudo das ocorrências no espaço, os autores estavam interessados em verificar a existência de possíveis associações entre os dois tipos de câncer, no qual foi incorporado ao modelo sob a estrutura de marcas categóricas. Isto é, eles estavam investigando se pessoas com câncer de reto estariam expostas ao risco de se ter câncer de colon devido à uma possível metástase. Além disso, outras informações externas foram introduzidas nas análises, como: taxa de pobreza local, sob a forma de covariável espacial; e fatores de risco do indivíduo, como idade e estágio do câncer, expresso na forma de covariável não espacial.

As covariáveis não espaciais foram classificadas, de acordo com os autores, em dois tipos. O primeiro tipo de covariável apresenta informações as quais são vistas como marcas (tipos de câncer) dos eventos na configuração pontual (modelagem de processos pontuais marcados). Já o segundo tipo é visto como variável auxiliar, no qual consiste em fornecer alguma informação adicional ao padrão de intensidade do processo (fatores de risco).

O processo de intensidade marcado é, portanto, descrito em função das posições espaciais, \mathbf{s} , e das covariáveis auxiliares (ou não espaciais), \mathbf{v} , gerando assim uma superfície no produto dos domínios do espaço geográfico, pelo espaço das marcas e, também, pelo espaço das covariáveis, $\mathcal{T} \times \mathcal{M} \times \mathcal{V}$. Desse modo, a intensidade estocástica para a k -ésima marca, para $k = 1, \dots, \mathcal{K}$, é definida como

$$\lambda_k(\mathbf{s}, \mathbf{v}) = r(\mathbf{s}) \exp\{\beta_{0k} + \mathbf{y}(\mathbf{s})'\boldsymbol{\beta}_k + \mathbf{v}'\boldsymbol{\delta}_k + (\mathbf{v} \otimes \mathbf{y}(\mathbf{s}))'\boldsymbol{\zeta}_k + f_k(\mathbf{s})\}, \quad (2.74)$$

em que $r(\mathbf{s})$ é um *offset* que descreve a densidade populacional de Minnesota, β_{0k} é o efeito global da k -ésima marca, $\boldsymbol{\beta}_k$ é um vetor de coeficientes de regressão associado às covariáveis espaciais $\mathbf{y}(\mathbf{s})$, $\boldsymbol{\delta}_k$ é um vetor de coeficientes associado ao conjunto de variáveis auxiliares \mathbf{v} , $\boldsymbol{\zeta}_k$ é um vetor de coeficientes associados à interação entre as covariáveis espaciais e não espaciais e, por fim, $f_k(\mathbf{s})$ é um processo estocástico Gaussiano para a k -ésima marca com média zero.

Além disso, o modelo permite que o processo Gaussiano varie de acordo com as marcas e que haja uma estrutura de dependência entre si. Essa dependência aborda a relação existente entre as marcas por meio da intensidade. Dessa forma, essa dependência estabelece se o aumento na intensidade na localização \mathbf{s} para uma determinada marca aumenta ou diminui a in-

tensidade para outra(s) marca(s) na mesma localização (LIANG; CARLIN; GELFAND, 2009). Assim, para dar suporte a essa teoria, define-se um processo Gaussiano multivariado, $f(\mathbf{s})$, com média zero e função de covariâncias $\Sigma(\mathbf{s}, \mathbf{s}'; \alpha, \boldsymbol{\gamma}, \boldsymbol{\rho})$ descrito como

$$\Sigma(\mathbf{s}, \mathbf{s}'; \alpha, \boldsymbol{\gamma}, \boldsymbol{\rho}) = c(\mathbf{s}, \mathbf{s}', \alpha) \begin{bmatrix} \gamma_1^2 & \rho_{12}\gamma_1\gamma_2 & \cdots & \rho_{1\mathcal{K}}\gamma_1\gamma_{\mathcal{K}} \\ \rho_{21}\gamma_2\gamma_1 & \gamma_2^2 & \cdots & \rho_{2\mathcal{K}}\gamma_2\gamma_{\mathcal{K}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{\mathcal{K}1}\gamma_{\mathcal{K}}\gamma_1 & \rho_{\mathcal{K}2}\gamma_{\mathcal{K}}\gamma_2 & \cdots & \gamma_{\mathcal{K}}^2 \end{bmatrix}, \quad (2.75)$$

sendo \mathbf{s} e \mathbf{s}' duas localizações espaciais, ρ_{ij} a correlação entre o i -ésimo e j -ésimo processo Gaussiano, para $i \neq j$, $i, j = 1, \dots, \mathcal{K}$ e $c(\cdot, \cdot; \alpha)$ uma função de correlação univariada, com um parâmetro de escala, α , comum para todas as marcas.

Casos particulares do modelo de intensidade, descrito em (2.74), podem ser obtidos a partir de determinadas simplificações. Por exemplo, formas separáveis podem ser estabelecidas em (2.74) quando $\boldsymbol{\zeta}_k = \mathbf{0}$, possibilitando, assim, um estudo do efeito das marcas na intensidade espacial. Outra simplificação seria considerar $\mathcal{K} = 1$, reduzindo o modelo para um processo pontual não marcado (LIANG; CARLIN; GELFAND, 2009).

A função de verossimilhança associada ao processo de intensidade em (2.74), dado um processo pontual marcado \mathcal{S}_m , é

$$\prod_k \exp \left\{ - \int_{\mathcal{T}} \int_{\mathcal{V}} \lambda_k(\mathbf{s}, \mathbf{v}) \, dsd\mathbf{v} \right\} \times \prod_k \prod_{\mathbf{s}_{ki}, \mathbf{v}_{ki}} \lambda_k(\mathbf{s}_{ki}, \mathbf{v}_{ki}). \quad (2.76)$$

No entanto, a função de verossimilhança, dada em (2.76), envolve uma integração estocástica da intensidade, sob o domínio \mathcal{T} , devido ao processo Gaussiano $f_k(\mathbf{s})$, no qual não possui forma fechada. Para contornar este problema, os autores utilizaram o método do processo preditivo, apresentado em Banerjee et al. (2008), para obter uma aproximação dessa integral.

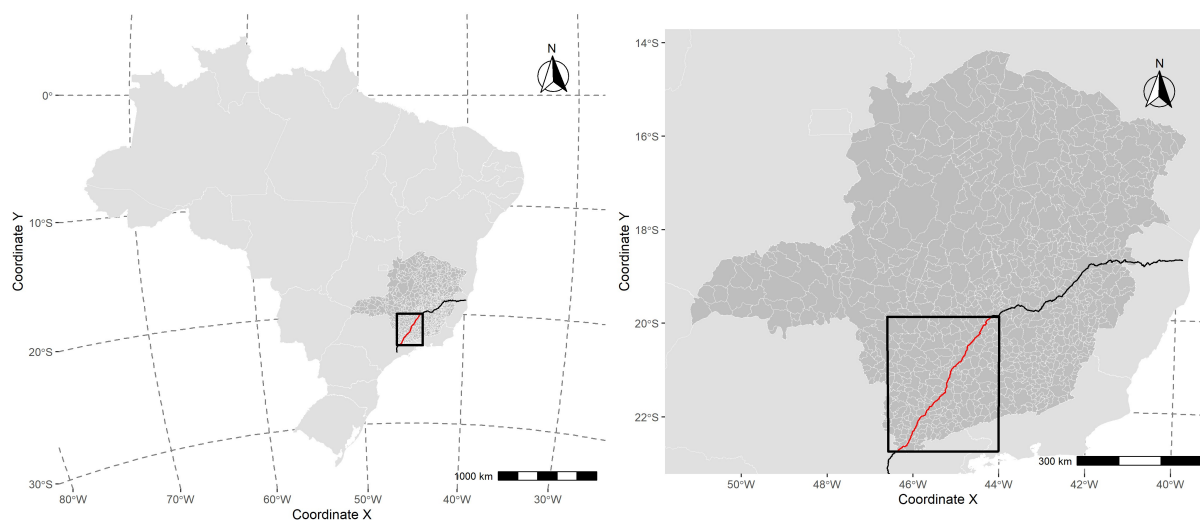
Uma vez obtida essa aproximação para a integral e, conseqüentemente, para a função de verossimilhança em (2.76), um modelo bayesiano pode ser especificado quando combinado com prioris para os parâmetros do modelo em (2.74). Além disso, amostras da posteriori podem ser obtidas usando o método MCMC.

3 MATERIAIS E MÉTODOS

3.1 Conjunto de dados reais

A base de dados que foi utilizada neste estudo se refere a ocorrências de acidentes registradas na rodovia Fernão Dias (BR-381) durante o período de 01 de janeiro de 2017 a 30 de setembro de 2021. Foi considerado apenas um trecho dessa rodovia, começando no município de Extrema e terminando em Betim. Este trecho se encontra, em grande parte, na região sul do estado de Minas Gerais, como mostra a Figura 3.1.

Figura 3.1 – Geolocalização de trecho da rodovia Fernão Dias (em vermelho), localizado na região sul do estado de Minas Gerais, no qual será estimada a intensidade das ocorrências de acidentes.



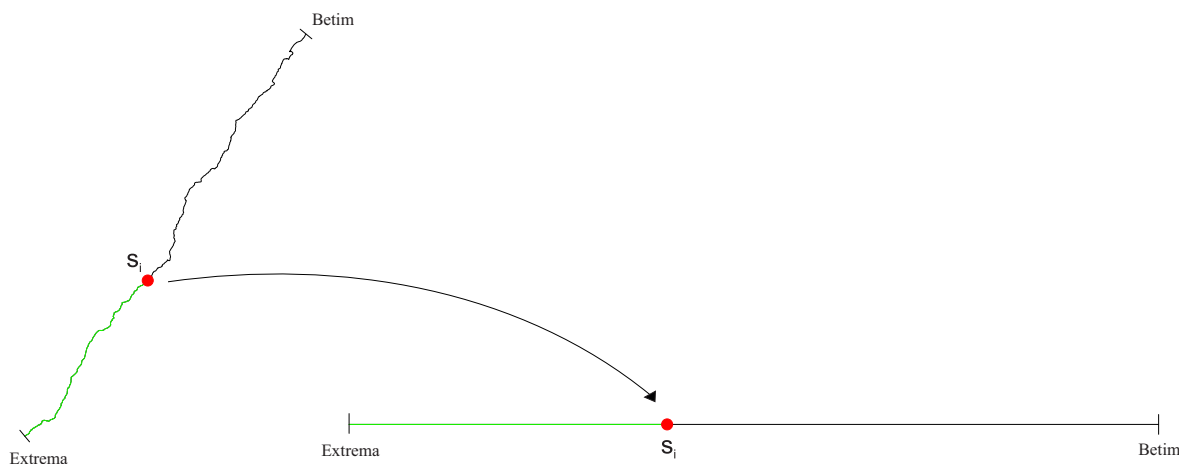
Fonte: Do autor (2021).

Os dados dessas ocorrências foram coletados no portal de Dados Abertos no site da Polícia Rodoviária Federal (POLÍCIA RODOVIÁRIA FEDERAL, 2021). Esse portal disponibiliza duas bases de dados associados aos acidentes. Para a primeira base, o registro é feito por ocorrência, associando a ocorrência à um ou mais veículos envolvidos no acidente. Já a segunda base disponibiliza o registro dos acidentes por pessoa, onde são especificadas informações relacionadas a todos os indivíduos envolvidos em uma determinada ocorrência de acidente na rodovia.

Os conjuntos de dados em questão consistem em duas informações importantes que podem ser usadas para a caracterizar a ocorrência como um evento de uma configuração pontual: o momento em que o acidente aconteceu e sua respectiva localização georreferenciada, no sistema de latitude e longitude. Apesar do modelo de intensidade em questão ter sido desen-

volvendo considerando configurações pontuais indexadas no tempo, esse estudo considerou uma indexação que levou em consideração apenas a informação espacial. Nesse caso, foi aplicada uma função sob o conjunto de coordenadas espaciais que calculasse a distância percorrida na rodovia entre o evento e um ponto inicial pré-estabelecido. Esse ponto inicial se refere ao início do trecho da rodovia Fernão Dias considerado aqui, que se inicia na divisa entre os estados de São Paulo e Minas Gerais, localizado no município de Extrema. A Figura 3.2 apresenta um esquema exemplificando a transformação feita para a construção do conjunto de indexação do processo pontual. Com isso, a configuração pontual em estudo pôde ser analisada considerando uma informação “espacial”, ou georreferenciada, mas usando modelos unidimensionais, que são comuns para processos indexados no tempo.

Figura 3.2 – Exemplificação da criação do conjunto de indexação. Distância percorrida (em verde) a partir do ponto inicial até a localização do evento.



Fonte: Do autor (2021).

Para uma primeira análise, foram utilizadas as informações presentes na primeira base de dados, acidentes por ocorrência. Nessa análise, o objetivo foi ajustar uma função de intensidade para os dados considerando apenas a informação dos pontos, desconsiderando qualquer informação adicional. Ou seja, o primeiro ajuste foi feito considerando um processo pontual não marcado, seguindo a proposta de Lloyd et al. (2015). Primeiramente, foi realizada uma filtragem no conjunto de dados a fim de se obter apenas ocorrências com uma causa específica de acidente. Considerou-se apenas os acidentes cuja a causa tenha sido devido à incompatibilidade da velocidade permitida na rodovia, em condições meteorológicas de: “chuva”, “garoa / chuveiro” e “nevoeiro / neblina”. A Figura 3.3 mostra parte das informações contidas na base de dados dos registros de acidentes por ocorrência que foram consideradas na primeira análise.

lise. A coluna “*DistKM*” mostra o valor das distâncias percorridas na rodovia até a ocorrência do acidente, calculadas por meio das coordenadas georreferenciadas nas colunas “*latitude*” e “*longitude*”.

Figura 3.3 – Base de dados de acidentes por ocorrência registrados no período de janeiro de 2017 a setembro de 2021 em um trecho da rodovia federal BR-381, considerando apenas os registros cuja a causa tenha sido “Velocidade Incompatível” e em condições meteorológicas de “chuva”, “garoa / chuvisco” e “nevoeiro / neblina”.

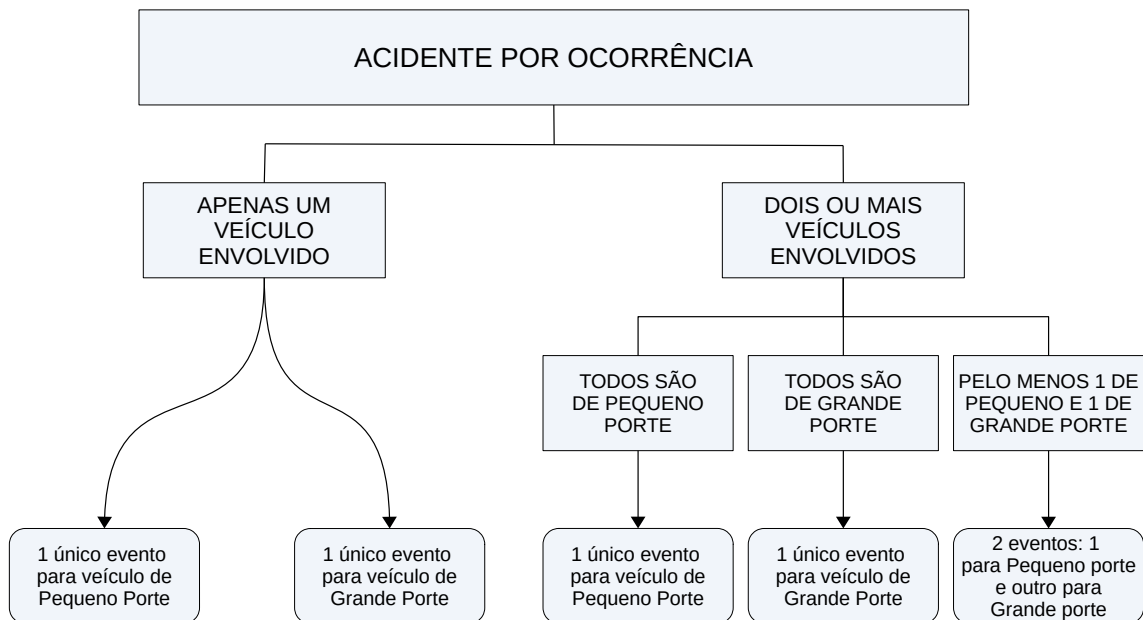
1	data_inversa	dia_semana	horario	uf	municipio	causa_acidente	condicao_meteorologica	pessoas	veiculos	latitude	longitude	DistKM
2	2017-01-03	terca-feira	15:20:00	MG	SAO GONCALO DO SAPUCAI	Velocidade Incompatível	Chuva	1		-1.21.90870839	-45.56908965	153.006161760012958
3	2017-01-03	terca-feira	16:35:00	MG	SANTO ANTONIO DO AMPARO	Velocidade Incompatível	Chuva	3		-1.20.94468944	-44.88101686	301.455706740238952
4	2017-01-04	quarta-feira	16:05:00	MG	PERDOES	Velocidade Incompatível	Chuva	1		-1.21.06056432	-45.04971909	276.756603217635700
5	2017-01-04	quarta-feira	16:50:00	MG	SANTO ANTONIO DO AMPARO	Velocidade Incompatível	Chuva	4		-1.21.01743832	-44.98967483	286.301740791597240
6	2017-01-04	quarta-feira	18:20:00	MG	CARMO DA CACHOEIRA	Velocidade Incompatível	Chuva	5		-1.21.37553671	-45.19232758	234.920519419340394
7	2017-01-04	quarta-feira	18:40:00	MG	LAVRAS	Velocidade Incompatível	Chuva	1		-1.21.31574237	-45.16718242	243.006689195761226
8	2017-01-04	quarta-feira	19:30:00	MG	LAVRAS	Velocidade Incompatível	Chuva	4		-1.21.31972088	-45.16669125	242.536457032579079
9	2017-01-04	quarta-feira	20:15:00	MG	OLIVEIRA	Velocidade Incompatível	Garoa/Chuvisco	4		-1.20.73730000	-44.77180000	329.381427549731370
10	2017-01-05	quinta-feira	15:45:00	MG	CAMBUI	Velocidade Incompatível	Chuva	4		-1.22.67449592	-46.09789697	41.868241326564799
11	2017-01-05	quinta-feira	16:40:00	MG	EXTREMA	Velocidade Incompatível	Chuva	1		-1.22.79291816	-46.25174311	16.998781032053859
12	2017-01-05	quinta-feira	18:00:00	MG	IGARAPE	Velocidade Incompatível	Garoa/Chuvisco	2		-1.20.10990465	-44.31611366	427.517201937424147
13	2017-01-06	sexta-feira	15:00:00	MG	IGARAPE	Velocidade Incompatível	Chuva	1		-1.20.10446047	-44.31901932	429.819196698141070
14	2017-01-06	sexta-feira	15:30:00	MG	CAMBUI	Velocidade Incompatível	Chuva	3		-2.22.55176526	-46.03949991	57.721587401207060
15	2017-01-07	sabado	17:10:00	MG	PERDOES	Velocidade Incompatível	Chuva	1		-1.21.06475285	-45.05118266	276.273432783345754
16	2017-01-07	sabado	17:30:00	MG	SANTO ANTONIO DO AMPARO	Velocidade Incompatível	Chuva	5		-2.20.86108760	-44.83981848	312.222692368744845
17	2017-01-07	sabado	18:00:00	MG	CAMBUI	Velocidade Incompatível	Chuva	2		-1.22.66877020	-46.09960555	42.860629153247785
18	2017-01-07	sabado	18:25:00	MG	ESTIVA	Velocidade Incompatível	Chuva	3		-1.22.42063165	-45.96580349	75.525511938274008
19	2017-01-07	sabado	23:50:00	MG	PERDOES	Velocidade Incompatível	Chuva	1		-1.21.06388938	-45.05106196	276.366562341229042
20	2017-01-09	segunda-feira	03:45:00	MG	CARMO DA CACHOEIRA	Velocidade Incompatível	Garoa/Chuvisco	3		-1.21.47868925	-45.20114422	221.43646022779492
21	2017-01-09	segunda-feira	04:15:00	MG	BRUMADINHO	Velocidade Incompatível	Chuva	1		-1.20.15080000	-44.32630000	420.640363054823240
22	2017-01-09	segunda-feira	07:10:00	MG	BRUMADINHO	Velocidade Incompatível	Chuva	1		-1.20.15079403	-44.32623982	420.641735402353049
23	2017-01-09	segunda-feira	07:40:00	MG	POUSO ALEGRE	Velocidade Incompatível	Chuva	2		-1.22.22707780	-45.93937160	96.910054263777496
24	2017-01-09	segunda-feira	08:35:00	MG	CARMO DA CACHOEIRA	Velocidade Incompatível	Chuva	1		-1.21.34566334	-45.17675693	239.141314588485557
25	2017-01-09	segunda-feira	08:40:00	MG	ITAGUARA	Velocidade Incompatível	Garoa/Chuvisco	1		-1.22.22707780	-45.93937160	96.910054263777496
26	2017-01-09	segunda-feira	13:15:00	MG	TRES CORACOES	Velocidade Incompatível	Chuva	1		-1.21.61838871	-45.25020226	203.485693383878868
27	2017-01-09	segunda-feira	13:50:00	MG	EXTREMA	Velocidade Incompatível	Chuva	1		-1.22.81224881	-46.28438115	12.725687481413642
28	2017-01-09	segunda-feira	16:05:00	MG	ESTIVA	Velocidade Incompatível	Chuva	4		-1.22.42063165	-45.96580349	75.525511938274008
29	2017-01-09	segunda-feira	17:08:00	MG	CAMBUI	Velocidade Incompatível	Chuva	2		-1.22.54315951	-46.03835637	58.716176039071527
30	2017-01-09	segunda-feira	17:50:00	MG	CAREACU	Velocidade Incompatível	Chuva	2		-1.21.99491704	-45.67410931	136.62797896265033
31	2017-01-09	segunda-feira	18:05:00	MG	CAMBUI	Velocidade Incompatível	Chuva	1		-1.22.64736832	-46.08905640	45.504204009131591
32	2017-01-09	segunda-feira	19:20:00	MG	LAVRAS	Velocidade Incompatível	Chuva	1		-1.21.17002091	-45.12613832	261.479064220869247
33	2017-01-10	terca-feira	16:30:00	MG	SANTO ANTONIO DO AMPARO	Velocidade Incompatível	Chuva	1		-1.20.86978269	-44.84342294	311.132056308915708
34	2017-01-10	terca-feira	17:00:00	MG	BETIM	Velocidade Incompatível	Chuva	2		-1.20.02563825	-44.23117161	443.188566891151538
35	2017-01-10	terca-feira	17:50:00	MG	LAVRAS	Velocidade Incompatível	Chuva	2		-1.21.28210000	-45.14020000	247.961946425301932
36	2017-01-12	quinta-feira	02:35:00	MG	LAVRAS	Velocidade Incompatível	Garoa/Chuvisco	1		-1.21.17852541	-45.12259331	260.353117643924179
37	2017-01-12	quinta-feira	17:35:00	MG	CAREACU	Velocidade Incompatível	Chuva	3		-1.21.98729247	-45.66110820	138.247318567479368
38	2017-01-12	quinta-feira	18:25:00	MG	RIO MANSO	Velocidade Incompatível	Chuva	2		-1.20.16461236	-44.36721861	414.68021836190971
39	2017-01-12	quinta-feira	20:40:00	MG	CARMO DA CACHOEIRA	Velocidade Incompatível	Chuva	1		-1.21.36678827	-45.18975019	236.04358550927997
40	2017-01-13	sexta-feira	04:35:00	MG	IGARAPE	Velocidade Incompatível	Chuva	1		-1.20.10907480	-44.32696402	428.824173102358145
41	2017-01-14	sabado	01:20:00	MG	SANTO ANTONIO DO AMPARO	Velocidade Incompatível	Chuva	5		-1.20.94912005	-44.88678306	300.611497750291903
42	2017-01-14	sabado	13:50:00	MG	OLIVEIRA	Velocidade Incompatível	Chuva	1		-1.20.81787139	-44.79791164	319.647887978583753
43	2017-01-14	sabado	21:30:00	MG	EXTREMA	Velocidade Incompatível	Chuva	3		-1.22.87500876	-46.36713266	0.643128312867279
44	2017-01-15	domingo	01:40:00	MG	SAO GONCALO DO SAPUCAI	Velocidade Incompatível	Chuva	1		-1.21.83846102	-45.51502705	163.219043695498925
45	2017-01-15	domingo	06:20:00	MG	ITAJAICU	Velocidade Incompatível	Chuva	1		-1.20.16420000	-44.36660000	414.688080046806306
46	2017-01-15	domingo	12:08:00	MG	CARMO DA CACHOEIRA	Velocidade Incompatível	Garoa/Chuvisco	1		-1.21.49886166	-45.21139115	218.862534652365355

Fonte: Do autor (2021).

Já para uma segunda análise, uma nova base de dados foi criada incorporando as informações presentes nas duas bases de dados disponibilizadas pela Polícia Rodoviária Federal, acidentes por ocorrência e por pessoa. Para essa análise, o objetivo foi obter a estimativa de intensidade para as ocorrências de acidentes, mas agora considerando uma informação adicional inerente ao evento, categorizando o acidente pelo tipo de veículo envolvido. A análise foi baseada na teoria de marcas de um processo pontual, levando à um ajuste de intensidade para acidentes envolvendo cada tipo de veículo. O modelo utilizado nessa análise se baseia na teoria desenvolvida nessa tese, apresentada no capítulo de Resultados Teóricos. Basicamente, foram consideradas duas categorias de veículos: grande porte e pequeno porte. De acordo com o Anexo 1 do Código de Trânsito Brasileiro, um veículo de grande porte é definido como um veículo automotor que se destina ao transporte de carga (com peso bruto total máximo superior à 10000 Kg) e de passageiros (superior a 20 passageiros). Desse modo, considerou-se como

grande porte apenas veículos do tipo “Caminhão”, “Caminhão-trator” e “Ônibus”. Os demais veículos foram classificados como pequeno porte. Considerando que um acidente pode envolver um ou mais veículos, a atribuição da marca foi realizada da seguinte maneira. Se o acidente envolveu apenas um veículo, então esse acidente foi considerado um evento da configuração pontual e rotulado de acordo com uma das categorias da marca (pequeno ou grande porte). Por outro lado, se o acidente envolveu dois ou mais veículos, três possíveis cenários foram considerados: (1) todos os veículos envolvidos foram de pequeno porte e, desse modo, o acidente passou a ser rotulado como um evento da configuração pontual para veículos de pequeno porte; (2) de maneira análoga, se todos os veículos envolvidos no acidente foram classificados como de grande porte, atribuiu-se um único evento para a configuração pontual com veículos dessa categoria; (3) por fim, se o acidente envolveu veículos de pequeno e grande porte simultaneamente, então foram criados dois eventos, sendo um para cada categoria da marca. A Figura 3.4 apresenta um esquema com a classificação das categorias da marca.

Figura 3.4 – Exemplificação da atribuição do tipo de marca (veículo de pequeno/ grande porte) associado ao evento (veículo(s) associado(s) à ocorrência do acidente).



Fonte: Do autor (2021).

A Figura 3.5 mostra parte das informações contidas na base de dados criada a partir das duas bases que são disponibilizadas pela Polícia Rodoviária Federal: acidentes por ocorrência e acidentes por pessoa. Da mesma forma como foi realizado na primeira análise, foram considerados apenas os registros feitos no período de janeiro de 2017 a setembro de 2021 em um

trecho da rodovia federal BR-381, cuja a causa do acidente tenha sido por “Velocidade Incompatível” e as condições meteorológicas tenham sido de “chuva”, “garoa / chuvisco” e “nevoeiro / neblina”

Figura 3.5 – Base de dados de acidentes utilizada na segunda análise, criada a partir da junção das duas bases disponibilizadas pela Polícia Rodoviária Federal, na qual o evento é rotulado de acordo com o tipo de veículo envolvido: *Pequeno porte* ou *Grande porte*.

id	data inversa	dia semana	horario	uf	km	município	causa acidente	condicao meteorologica	pessoas	veiculos	latitude	longitude	DistKM	tipo veiculo	
1	773	2017-01-03	terça-feira	16:35:00	MG	646.1	SANTO ANTONIO DO AMPARO	Velocidade Incompativel	Chuva	3	1	-20.94468944	-44.88101686	301.455706740238952	Pequeno Porte
3	842	2017-01-03	terça-feira	15:20:00	MG	794.6	SAO GONCALO DO SAPUCAI	Velocidade Incompativel	Chuva	1	1	-21.90870839	-45.56908965	153.006161760012958	Pequeno Porte
4	1019	2017-01-04	quarta-feira	16:05:00	MG	670.7	PERDOES	Velocidade Incompativel	Chuva	1	1	-21.06056432	-45.04971909	276.756603217635700	Pequeno Porte
5	1043	2017-01-04	quarta-feira	16:50:00	MG	661.4	SANTO ANTONIO DO AMPARO	Velocidade Incompativel	Chuva	4	1	-21.01743832	-44.98967483	286.301740791597240	Pequeno Porte
6	1082	2017-01-04	quarta-feira	18:40:00	MG	705.2	LAVRAS	Velocidade Incompativel	Chuva	1	1	-21.31574237	-45.16718242	243.006689195761226	Pequeno Porte
7	1092	2017-01-04	quarta-feira	19:30:00	MG	705.6	LAVRAS	Velocidade Incompativel	Chuva	4	1	-21.31972088	-45.16669125	242.536457032579079	Pequeno Porte
8	1099	2017-01-04	quarta-feira	20:15:00	MG	618.0	OLIVEIRA	Velocidade Incompativel	Garoa/Chuvisco	4	1	-20.73730000	-44.77180000	329.381427549731370	Pequeno Porte
9	1103	2017-01-04	quarta-feira	18:20:00	MG	713.2	CARMO DA CACHOEIRA	Velocidade Incompativel	Chuva	5	1	-21.37553671	-45.19232758	234.920519419340394	Pequeno Porte
10	1312	2017-01-05	quinta-feira	18:00:00	MG	519.1	IGARAPE	Velocidade Incompativel	Garoa/Chuvisco	2	1	-20.10990465	-44.31611366	427.517201937424147	Pequeno Porte
11	1315	2017-01-05	quinta-feira	15:45:00	MG	907.5	CAMBUI	Velocidade Incompativel	Chuva	4	1	-22.67449592	-46.0978697	41.864241326564799	Pequeno Porte
12	1318	2017-01-05	quinta-feira	16:40:00	MG	932.9	EXTREMA	Velocidade Incompativel	Chuva	1	1	-22.79291816	-46.25174311	16.998781032053859	Pequeno Porte
13	1509	2017-01-06	sexta-feira	15:30:00	MG	891.5	CAMBUI	Velocidade Incompativel	Chuva	3	2	-22.55176526	-46.03949991	57.721587401207060	Grande Porte
14	1509	2017-01-06	sexta-feira	15:30:00	MG	891.5	CAMBUI	Velocidade Incompativel	Chuva	3	2	-22.55176526	-46.03949991	57.721587401207060	Pequeno Porte
15	1520	2017-01-06	sexta-feira	15:10:00	MG	516.7	IGARAPE	Velocidade Incompativel	Chuva	1	1	-20.10446047	-44.31901932	429.819196698141070	Pequeno Porte
16	1832	2017-01-07	sábado	17:10:00	MG	671.4	PERDOES	Velocidade Incompativel	Chuva	1	1	-21.06475285	-45.05118266	276.273432783345754	Pequeno Porte
17	1846	2017-01-07	sábado	18:00:00	MG	907.4	CAMBUI	Velocidade Incompativel	Chuva	2	1	-22.66877020	-46.09960555	42.866291532477785	Pequeno Porte
18	1848	2017-01-07	sábado	18:25:00	MG	873.8	ESTIVA	Velocidade Incompativel	Chuva	3	1	-22.42063165	-45.96580349	75.525511938274008	Pequeno Porte
19	1925	2017-01-07	sábado	17:30:00	MG	635.7	SANTO ANTONIO DO AMPARO	Velocidade Incompativel	Chuva	5	2	-20.86108760	-44.83981848	312.222692368744845	Pequeno Porte
20	2072	2017-01-07	sábado	23:50:00	MG	671.3	PERDOES	Velocidade Incompativel	Chuva	1	1	-21.06388938	-45.05106196	276.365662341229042	Pequeno Porte
21	2280	2017-01-09	segunda-feira	03:45:00	MG	726.0	CARMO DA CACHOEIRA	Velocidade Incompativel	Garoa/Chuvisco	3	1	-21.47868925	-45.20114422	221.63460622779492	Pequeno Porte
22	2284	2017-01-09	segunda-feira	04:15:00	MG	526.0	BRUMADINHO	Velocidade Incompativel	Chuva	1	1	-20.15080000	-44.32630000	420.640363054823240	Pequeno Porte
23	2299	2017-01-09	segunda-feira	07:10:00	MG	525.3	BRUMADINHO	Velocidade Incompativel	Chuva	1	1	-20.15079403	-44.32623982	420.641735402353049	Pequeno Porte
24	2305	2017-01-09	segunda-feira	07:40:00	MG	707.0	POUSO ALEGRE	Velocidade Incompativel	Chuva	2	1	-22.22707780	-45.93937160	96.910054263777496	Pequeno Porte
25	2326	2017-01-09	segunda-feira	08:35:00	MG	708.9	CARMO DA CACHOEIRA	Velocidade Incompativel	Chuva	1	1	-21.34566334	-45.17675693	239.141314588485557	Pequeno Porte
26	2345	2017-01-09	segunda-feira	08:40:00	MG	564.0	ITAGUARA	Velocidade Incompativel	Garoa/Chuvisco	1	1	-22.22707780	-45.93937160	96.910054263777496	Pequeno Porte
27	2436	2017-01-09	segunda-feira	13:15:00	MG	744.7	TRES CORACOES	Velocidade Incompativel	Chuva	1	1	-21.61838871	-45.25020226	203.485693383878868	Grande Porte
28	2442	2017-01-09	segunda-feira	16:05:00	MG	873.8	ESTIVA	Velocidade Incompativel	Chuva	4	1	-22.42063165	-45.96580349	75.525511938274008	Pequeno Porte
29	2443	2017-01-09	segunda-feira	13:50:00	MG	937.0	EXTREMA	Velocidade Incompativel	Chuva	1	1	-22.81224881	-46.28438115	12.725687481413642	Pequeno Porte
30	2455	2017-01-09	segunda-feira	17:08:00	MG	892.0	CAMBUI	Velocidade Incompativel	Chuva	2	1	-22.54315951	-46.03835637	58.716176039071527	Pequeno Porte
31	2469	2017-01-09	segunda-feira	17:50:00	MG	808.0	CAREACU	Velocidade Incompativel	Chuva	2	1	-21.99491704	-45.67410931	136.627978966265033	Pequeno Porte
32	2478	2017-01-09	segunda-feira	18:05:00	MG	903.4	CAMBUI	Velocidade Incompativel	Chuva	1	1	-22.64736832	-46.08905640	45.504204009131591	Pequeno Porte
33	2493	2017-01-09	segunda-feira	19:20:00	MG	685.7	LAVRAS	Velocidade Incompativel	Chuva	1	1	-21.17002091	-45.12613832	261.479064220869247	Grande Porte
34	2656	2017-01-10	terça-feira	17:00:00	MG	501.8	BETIM	Velocidade Incompativel	Chuva	2	1	-20.02563825	-44.23117161	443.188566891151538	Grande Porte
35	2678	2017-01-10	terça-feira	17:50:00	MG	700.9	LAVRAS	Velocidade Incompativel	Chuva	2	1	-21.28210000	-45.14020000	247.961946425301932	Pequeno Porte
36	2682	2017-01-10	terça-feira	16:30:00	MG	636.8	SANTO ANTONIO DO AMPARO	Velocidade Incompativel	Chuva	1	1	-20.86978269	-44.84342294	311.1320666308915708	Pequeno Porte
37	3017	2017-01-12	quinta-feira	02:35:00	MG	686.9	LAVRAS	Velocidade Incompativel	Garoa/Chuvisco	1	1	-21.17852541	-45.12259331	260.353117643294179	Pequeno Porte
38	3182	2017-01-12	quinta-feira	17:35:00	MG	809.0	CAREACU	Velocidade Incompativel	Chuva	3	1	-21.98729247	-45.66110820	138.247318567479368	Pequeno Porte
39	3188	2017-01-12	quinta-feira	18:25:00	MG	532.1	RIO MANSO	Velocidade Incompativel	Chuva	2	1	-20.16461236	-44.36721861	414.680821836190971	Pequeno Porte
40	3220	2017-01-12	quinta-feira	20:40:00	MG	712.0	CARMO DA CACHOEIRA	Velocidade Incompativel	Chuva	1	1	-21.36678827	-45.18975019	236.043585850927997	Pequeno Porte
41	3280	2017-01-13	sexta-feira	04:35:00	MG	518.0	IGARAPE	Velocidade Incompativel	Chuva	1	1	-20.10907480	-44.32686402	428.824173102358145	Pequeno Porte
42	3575	2017-01-14	sábado	01:20:00	MG	646.2	SANTO ANTONIO DO AMPARO	Velocidade Incompativel	Chuva	5	1	-20.94912005	-44.88678306	300.611497750291903	Pequeno Porte
43	3713	2017-01-14	sábado	13:50:00	MG	627.9	OLIVEIRA	Velocidade Incompativel	Chuva	1	1	-20.81787139	-44.79791164	319.647287978833753	Pequeno Porte
44	3822	2017-01-14	sábado	21:30:00	MG	949.0	EXTREMA	Velocidade Incompativel	Chuva	3	1	-22.87500876	-46.36713266	0.643122831266729	Pequeno Porte
45	3877	2017-01-15	domingo	01:40:00	MG	784.3	SAO GONCALO DO SAPUCAI	Velocidade Incompativel	Chuva	1	1	-21.83846102	-45.51502705	163.219043695498925	Pequeno Porte
46	3928	2017-01-15	domingo	06:20:00	MG	532.1	ITAJAÍ	Velocidade Incompativel	Chuva	1	1	-20.16420000	-44.36660000	414.68808046806306	Pequeno Porte

Fonte: Do autor (2021).

3.2 Recursos computacionais

Todo o processo de preparação dos dados, bem como a programação dos modelos e sua respectiva estimação, foram feitos utilizando funções desenvolvidas pelo autor (vide Apêndice C) na linguagem de programação R (R CORE TEAM, 2021).

4 RESULTADOS TEÓRICOS

4.1 Modelo

Seja um processo estocástico pontual marcado $\{(s_1, A(s_1)), \dots : s_i \in \mathcal{T}, A(s_i) \in \mathcal{M}\}$ definido no produto cartesiano do conjunto indexador dos eventos com o espaço das marcas $\mathcal{T} \times \mathcal{M}$. Assume-se que processo pontual está definido em um conjunto de indexação pertencente aos reais unidimensionais ao passo que, o espaço das marcas é um conjunto finito composto por \mathcal{K} categorias. Logo, $\mathcal{T} \times \mathcal{M} = \mathbb{R}^1 \times \{1, \dots, \mathcal{K}\}$. Uma configuração pontual é definida como uma realização desse processo, que resulta em um conjunto finito de N eventos que se distribui sob o domínio dos reais unidimensionais, sendo cada evento associado à uma das \mathcal{K} categorias de uma marca. Desse modo, denota-se

$$\mathcal{S}_m = \{(s_1, A(s_1)), \dots, (s_N, A(s_N))\}$$

como sendo a realização de um processo pontual marcado.

De maneira análoga, pode-se denotar a configuração pontual marcada como um conjunto composto por \mathcal{K} padrões de pontos, $\mathcal{S}_m = \{\mathcal{S}^{(1)}, \mathcal{S}^{(2)}, \dots, \mathcal{S}^{(\mathcal{K})}\}$, cada um associado à uma categoria da marca, sendo

$$\mathcal{S}^{(k)} = \{(s_{1_k}, A(s_{1_k}) = k), \dots, (s_{N_k}, A(s_{N_k}) = k)\}$$

a configuração pontual associada à k -ésima categoria da marca, composta por N_k eventos, para todo $k = 1, \dots, \mathcal{K}$. Considerando todas as \mathcal{K} categorias da marca, o número total de eventos observados é dado por $N = \sum_{k=1}^{\mathcal{K}} N_k$.

Baseado na estrutura de um processo de Cox, assume-se que a medida de intensidade do processo pontual marcado é, também, uma componente estocástica, que pode ser representada em função de outra componente estocástica Gaussiana latente, expressa por

$$\begin{bmatrix} \lambda_1(s) \\ \lambda_2(s) \\ \vdots \\ \lambda_{\mathcal{K}}(s) \end{bmatrix} = \begin{bmatrix} f_1^2(s) \\ f_2^2(s) \\ \vdots \\ f_{\mathcal{K}}^2(s) \end{bmatrix},$$

e que pode ser resumida como

$$\boldsymbol{\lambda}(s) = \mathbf{f}^2(s), \quad (4.1)$$

sendo $\boldsymbol{\lambda}$ um vetor de \mathcal{K} intensidades estocásticas, descrito em função de um processo Gaussiano multivariado, $\mathbf{f}(\cdot)$.

Para o modelo proposto, será considerado uma configuração pontual marcada, restrita à uma marca composta por $\mathcal{K} = 2$ categorias, resultando em processo pontual marcado bivariado.

Seguindo a proposta de Lloyd et al. (2015), o modelo assume que o processo Gaussiano bivariado, $\mathbf{f}(\cdot)$, depende de um conjunto de M pontos de indução, denotados por $\mathcal{Z} = \{z_1, \dots, z_M\}$ e definidos no mesmo conjunto de indexação do processo pontual, $z_i \in \mathcal{T} \subset \mathbb{R}^1$ para $i = 1, \dots, M$. A avaliação do processo nestes pontos resulta em uma distribuição Normal multivariada, denotada por $\mathbf{u} = [\mathbf{u}_1^\top, \mathbf{u}_2^\top]^\top$. Deste modo,

$$\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \sim \mathcal{N} \left(\bar{\mathbf{u}} = \begin{bmatrix} \vec{\mathbf{I}} \bar{u}_1 \\ \vec{\mathbf{I}} \bar{u}_2 \end{bmatrix}, \mathbf{K}^{(z,z)} = \begin{bmatrix} \mathbf{K}_{11}^{(z,z)} & \mathbf{K}_{12}^{(z,z)} \\ \mathbf{K}_{21}^{(z,z)} & \mathbf{K}_{22}^{(z,z)} \end{bmatrix} \right), \quad (4.2)$$

em que $\bar{\mathbf{u}}$ é um vetor particionado de ordem $2M$, sendo \bar{u}_i a média associada ao i -ésimo vetor de variáveis de indução e $\vec{\mathbf{I}}$ um vetor unitário de ordem M . $\mathbf{K}^{(z,z)}$ é uma matriz particionada de dimensão $2M \times 2M$, sendo $\mathbf{K}_{ij}^{(z,z')}$ a matriz de covariâncias cruzada entre o i -ésimo e o j -ésimo vetor de variáveis de indução, \mathbf{u}_i e \mathbf{u}_j , de dimensão $M \times M$ e avaliado nos pontos $(z, z') \in \mathcal{Z}$.

Seguindo a regra da distribuição condicional entre subvetores Gaussianos multivariados, $\mathbf{f} \mid \mathbf{u}$ seguirá um processo Gaussiano bivariado a priori, com um vetor particionado de funções de médias

$$\boldsymbol{\mu} = \mathbf{K}^{(s,z)} \mathbf{K}^{(z,z)^{-1}} \mathbf{u} \quad (4.3)$$

e matriz particionada contendo funções matrizes de covariâncias

$$\boldsymbol{\Sigma} = \mathbf{K}^{(s,s')} - \mathbf{K}^{(s,z)} \mathbf{K}^{(z,z)^{-1}} \mathbf{K}^{(z,s')}, \quad (4.4)$$

sendo

$$\mathbf{K}^{(s,s')} = \begin{bmatrix} \mathbf{K}_{11}^{(s,s')} & \mathbf{K}_{12}^{(s,s')} \\ \mathbf{K}_{21}^{(s,s')} & \mathbf{K}_{22}^{(s,s')} \end{bmatrix} \quad \text{e} \quad \mathbf{K}^{(s,z)} = \begin{bmatrix} \mathbf{K}_{11}^{(s,z)} & \mathbf{K}_{12}^{(s,z)} \\ \mathbf{K}_{21}^{(s,z)} & \mathbf{K}_{22}^{(s,z)} \end{bmatrix}.$$

Os elementos de cada uma das submatrizes $\mathbf{K}_{ij}^{(s,s')}$, $\mathbf{K}_{ij}^{(z,z)}$ e $\mathbf{K}_{ij}^{(s,z)}$ são obtidos, respectivamente, a partir das seguintes funções, dados por

$$K_{ij}(s, s') = \rho_{ij} \times \gamma_i \gamma_j \times c(s, s'; \alpha), \quad (4.5)$$

$$K_{ij}(z, z') = \rho_{ij} \times \gamma_i \gamma_j \times c(z, z'; \alpha), \quad (4.6)$$

e

$$K_{ij}(s, z) = \rho_{ij} \times \gamma_i \gamma_j \times c(s, z; \alpha), \quad (4.7)$$

em que $-1 \leq \rho_{ij} \leq 1$ quantifica a correlação cruzada entre o i -ésimo e o j -ésimo processo Gaussiano, para $i, j = 1, 2$, sendo $\gamma_i > 0$ e $\gamma_j > 0$ seus respectivos desvios padrão; se $i = j$, então $\rho_{ii} = 1$ e $\gamma_i \gamma_i = \gamma_i^2$, levando à expressão da função de covariância para o i -ésimo processo Gaussiano. No entanto, se $i \neq j$, então $\rho_{ii} \neq 1$ levando à expressão de uma função de covariância cruzada entre o i -ésimo e o j -ésimo processo Gaussiano.

A função $c(\cdot, \cdot; \alpha)$, presente nas expressões (4.5), (4.6) e (4.7), quantifica a correlação existente entre dois pontos que pertencem ao conjunto de indexação do processo Gaussiano e é obtida através da função de correlação (ou kernel) exponencial quadrático, apresentado na Tabela 2.1 e que é dado por

$$c(s, s'; \alpha) = \exp \left\{ \frac{-(s - s')^2}{2\alpha} \right\}, \quad (4.8)$$

sendo $\alpha > 0$ um parâmetro de escala que caracteriza o alcance da dependência entre os pontos do conjunto de indexação. Assume-se que, tanto o parâmetro α , na equação (4.8), quanto o parâmetro de correlação cruzado, ρ_{ij} , bem como os desvios padrão, γ_i e γ_j , são os mesmos para calcular os elementos de $\mathbf{K}_{ij}^{(z,z)}$, $\mathbf{K}_{ij}^{(s,z)}$ e de $\mathbf{K}_{ij}^{(s,z)}$.

Na expressão (4.1), cada λ_i , para $i = 1, \dots, \mathcal{K}$, caracteriza a intensidade do processo pontual associada à i -ésima categoria da marca. Em geral, espera-se que as marcas tenham uma estrutura de dependência entre si e, conseqüentemente, entre suas respectivas intensidades. Deste modo, o processo Gaussiano multivariado se torna uma componente importante, pois ele consegue mensurar a estrutura de correlação cruzada entre os diferentes processos Gaussianos, descrita por ρ_{ij} , além da correlação presente no conjunto de indexação em cada um deles, descrita por $c(\cdot, \cdot; \alpha)$. Esta correlação cruzada é que irá determinar a relação existente entre as categorias da marca de um processo pontual, por meio de sua intensidade.

4.2 Estimação

Considerando uma marca composta de duas categorias, a função de verossimilhança para o modelo apresentado anteriormente é dada por

$$\begin{aligned}
 p(\mathcal{S}_m \mid \boldsymbol{\lambda} = \mathbf{f}^2, \Theta) &= \prod_{k=1}^2 p(\mathcal{S}^{(k)} \mid \lambda_k = f_k^2, \Theta) \\
 &= \prod_{k=1}^2 \left[\exp \left\{ - \int_{\mathcal{T}} \lambda_k(s) ds \right\} \times \prod_{i=1}^{N_k} \lambda_k(s_{ki}) \right] \\
 &= \prod_{k=1}^2 \left[\exp \left\{ - \int_{\mathcal{T}} f_k^2(s) ds \right\} \times \prod_{i=1}^{N_k} f_k^2(s_{ki}) \right],
 \end{aligned} \tag{4.9}$$

em que N_k corresponde ao número total de eventos observados para a k -ésima categoria da marca.

Com o modelo formulado, a distribuição a posteriori conjunta de \mathbf{f} e \mathbf{u} pode ser obtida através do teorema de Bayes, dado por

$$p(\mathbf{f}, \mathbf{u} \mid \mathcal{S}_m, \Theta) = \frac{p(\mathcal{S}_m \mid \boldsymbol{\lambda} = \mathbf{f}^2, \Theta) \times p(\mathbf{f} \mid \mathbf{u}, \Theta) \times p(\mathbf{u} \mid \Theta)}{p(\mathcal{S}_m)}, \tag{4.10}$$

em que $p(\mathbf{f} \mid \mathbf{u}, \Theta)$ é a priori condicional de $\mathbf{f} \mid \mathbf{u}$ que segue um processo Gaussiano com média e função de covariância dado pelas expressões em (4.3) e (4.4), respectivamente; $p(\mathbf{u} \mid \Theta)$ é a distribuição a priori do conjunto de variáveis indutoras, definida em (4.2); já $p(\mathcal{S}_m)$ é a função de verossimilhança marginal, também chamada de função de evidência, que pode ser obtida integrando o numerador da expressão em (4.10) em relação a \mathbf{f} e \mathbf{u} ; por fim, $\Theta = \{\rho_{12}, \alpha, \gamma_1, \gamma_2, \bar{u}_1, \bar{u}_2\}$ é o conjunto de todos os parâmetros do modelo.

Uma vez definida a estrutura do modelo e sua respectiva função de verossimilhança, um método inferencial se faz necessário a fim de obter estimativas dos parâmetros, bem como a realização de predições na função de intensidade. Para isso, será utilizado método Bayesiano variacional, seguindo o mesmo caminho do que foi proposto por Lloyd et al. (2015), uma vez que este método apresentou bom desempenho no processo inferencial.

Partindo da premissa de que a densidade a posteriori do modelo é desconhecida e não pode ser obtida de forma fechada, uma distribuição variacional, denotada por $q(\mathbf{f}, \mathbf{u})$ é considerada como forma de representar, de maneira aproximada, a verdadeira posteriori. Para isto, é necessário que haja um certo grau de similaridade entre $p(\mathbf{f}, \mathbf{u} \mid \mathcal{S}_m, \Theta)$ e $q(\mathbf{f}, \mathbf{u})$ para que, de fato, a distribuição variacional $q(\cdot)$ possa retratar adequadamente o padrão de aleatoriedade

presente na distribuição a posteriori. Para isso, utiliza-se a medida de divergência de Kullback-Leibler, ou simplesmente divergência KL, para avaliar o grau de similaridade entre as duas distribuições em questão: a posteriori e a variacional. Seguindo a proposta de Lloyd et al. (2015), assume-se que a distribuição variacional conjunta de \mathbf{f} e de \mathbf{u} pode ser fatorada como $q(\mathbf{f}, \mathbf{u}) = q(\mathbf{f} | \mathbf{u})q(\mathbf{u})$, sendo

$$q(\mathbf{u}) = \mathcal{N} \left(\mathbf{m} = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix}, \mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix} \right), \quad (4.11)$$

em que $\mathbf{S}_{21} = \mathbf{S}_{12}^\top$. Além disso, impõe-se que $q(\mathbf{f} | \mathbf{u}) = p(\mathbf{f} | \mathbf{u}, \Theta)$ para simplificar o processo inferencial.

Por meio dessas simplificações, a distribuição variacional marginal de \mathbf{f} pode ser obtida por meio da integração da densidade variacional conjunta, na qual

$$\begin{aligned} q(\mathbf{f}) &= \int q(\mathbf{f}, \mathbf{u}) \partial \mathbf{u} \\ &= \int p(\mathbf{f} | \mathbf{u}, \Theta) q(\mathbf{u}) \partial \mathbf{u}, \end{aligned} \quad (4.12)$$

sendo $\mathbf{u} = [\mathbf{u}_1^\top, \mathbf{u}_2^\top]^\top$. Nesse caso, $\mathbf{f}(s)$ seguirá um processo estocástico Gaussiano bivariado, com um vetor de funções de médias, $\tilde{\boldsymbol{\mu}}$, e uma matriz particionada, $\tilde{\boldsymbol{\Sigma}}$, contendo as funções de covariância para cada processo Gaussiano e a função de covariância cruzada entre eles, por meio da seguinte relação:

$$\begin{aligned} \tilde{\boldsymbol{\mu}} &= \begin{bmatrix} \tilde{\mu}_1(s) \\ \tilde{\mu}_2(s) \end{bmatrix} = \mathbf{K}^{(s,z)} (\mathbf{K}^{(z,z)})^{-1} \mathbf{m} \\ &= \begin{bmatrix} \mathbf{K}_{11}^{(s,z)} \mathbf{A} \mathbf{m}_1 + \mathbf{K}_{12}^{(s,z)} \mathbf{C} \mathbf{m}_1 + \mathbf{K}_{11}^{(s,z)} \mathbf{B} \mathbf{m}_2 + \mathbf{K}_{12}^{(s,z)} \mathbf{D} \mathbf{m}_2 \\ \mathbf{K}_{21}^{(s,z)} \mathbf{A} \mathbf{m}_1 + \mathbf{K}_{22}^{(s,z)} \mathbf{C} \mathbf{m}_1 + \mathbf{K}_{21}^{(s,z)} \mathbf{B} \mathbf{m}_2 + \mathbf{K}_{22}^{(s,z)} \mathbf{D} \mathbf{m}_2 \end{bmatrix}, \end{aligned} \quad (4.13)$$

em que \mathbf{A} , \mathbf{B} , \mathbf{C} e \mathbf{D} são submatrizes obtidas através da inversão de $\mathbf{K}^{(z,z)}$, seguindo a regra de inversas de matrizes particionadas apresentada em Searle (1982). Isto é,

$$(\mathbf{K}^{(z,z)})^{-1} = \left(\begin{bmatrix} \mathbf{K}_{11}^{(z,z)} & \mathbf{K}_{12}^{(z,z)} \\ \mathbf{K}_{21}^{(z,z)} & \mathbf{K}_{22}^{(z,z)} \end{bmatrix} \right)^{-1} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

sendo $\mathbf{A} = (\mathbf{K}_{11}^{(z,z)})^{-1} + (\mathbf{K}_{11}^{(z,z)})^{-1} \mathbf{K}_{12}^{(z,z)} \left(\mathbf{K}_{22}^{(z,z)} - \mathbf{K}_{21}^{(z,z)} (\mathbf{K}_{11}^{(z,z)})^{-1} \mathbf{K}_{12}^{(z,z)} \right)^{-1} \mathbf{K}_{21}^{(z,z)} (\mathbf{K}_{11}^{(z,z)})^{-1}$,
 $\mathbf{B} = -(\mathbf{K}_{11}^{(z,z)})^{-1} \mathbf{K}_{12}^{(z,z)} \left(\mathbf{K}_{22}^{(z,z)} - \mathbf{K}_{21}^{(z,z)} (\mathbf{K}_{11}^{(z,z)})^{-1} \mathbf{K}_{12}^{(z,z)} \right)^{-1}$, $\mathbf{C} = -\left(\mathbf{K}_{22}^{(z,z)} - \mathbf{K}_{21}^{(z,z)} (\mathbf{K}_{11}^{(z,z)})^{-1} \mathbf{K}_{12}^{(z,z)} \right)^{-1} \mathbf{K}_{21}^{(z,z)} (\mathbf{K}_{11}^{(z,z)})^{-1}$ e $\mathbf{D} = \left(\mathbf{K}_{22}^{(z,z)} - \mathbf{K}_{21}^{(z,z)} (\mathbf{K}_{11}^{(z,z)})^{-1} \mathbf{K}_{12}^{(z,z)} \right)^{-1}$. E $\tilde{\Sigma}$ é dado por

$$\begin{aligned} \tilde{\Sigma} &= \begin{bmatrix} \tilde{\Sigma}_{11}(s, s') & \tilde{\Sigma}_{12}(s, s') \\ \tilde{\Sigma}_{21}(s, s') & \tilde{\Sigma}_{22}(s, s') \end{bmatrix} \\ &= \mathbf{K}^{(s,s')} - \mathbf{K}^{(s,z)} (\mathbf{K}^{(z,z)})^{-1} \mathbf{K}^{(z,s')} + \mathbf{K}^{(s,z)} (\mathbf{K}^{(z,z)})^{-1} \mathbf{S} (\mathbf{K}^{(z,z)})^{-1} \mathbf{K}^{(z,s')}, \end{aligned} \quad (4.14)$$

sendo

- $\tilde{\Sigma}_{11}(s, s') = \mathbf{K}_{11}^{(s,s')} - (\mathbf{K}_{11}^{(s,z)} \mathbf{A} + \mathbf{K}_{12}^{(s,z)} \mathbf{C}) \mathbf{K}_{11}^{(z,s')} - (\mathbf{K}_{11}^{(s,z)} \mathbf{B} + \mathbf{K}_{12}^{(s,z)} \mathbf{D}) \mathbf{K}_{21}^{(z,s')} + \mathbf{K}_{11}^{(s,z)} (\mathbf{A} \mathbf{S}_{11} + \mathbf{B} \mathbf{S}_{21}) \mathbf{A} \mathbf{K}_{11}^{(z,s')} + \mathbf{K}_{12}^{(s,z)} (\mathbf{C} \mathbf{S}_{11} + \mathbf{D} \mathbf{S}_{21}) \mathbf{A} \mathbf{K}_{11}^{(z,s')} + \mathbf{K}_{11}^{(s,z)} (\mathbf{A} \mathbf{S}_{12} + \mathbf{B} \mathbf{S}_{22}) \mathbf{C} \mathbf{K}_{11}^{(z,s')} + \mathbf{K}_{12}^{(s,z)} (\mathbf{C} \mathbf{S}_{12} + \mathbf{D} \mathbf{S}_{22}) \mathbf{C} \mathbf{K}_{11}^{(z,s')} + \mathbf{K}_{11}^{(s,z)} (\mathbf{A} \mathbf{S}_{11} + \mathbf{B} \mathbf{S}_{21}) \mathbf{B} \mathbf{K}_{21}^{(z,s')} + \mathbf{K}_{12}^{(s,z)} (\mathbf{C} \mathbf{S}_{11} + \mathbf{D} \mathbf{S}_{21}) \mathbf{B} \mathbf{K}_{21}^{(z,s')} + \mathbf{K}_{11}^{(s,z)} (\mathbf{A} \mathbf{S}_{12} + \mathbf{B} \mathbf{S}_{22}) \mathbf{D} \mathbf{K}_{21}^{(z,s')} + \mathbf{K}_{12}^{(s,z)} (\mathbf{C} \mathbf{S}_{12} + \mathbf{D} \mathbf{S}_{22}) \mathbf{D} \mathbf{K}_{21}^{(z,s')}$;
- $\tilde{\Sigma}_{12}(s, s') = \mathbf{K}_{12}^{(s,s')} - (\mathbf{K}_{11}^{(s,z)} \mathbf{A} + \mathbf{K}_{12}^{(s,z)} \mathbf{C}) \mathbf{K}_{12}^{(z,s')} - (\mathbf{K}_{11}^{(s,z)} \mathbf{B} + \mathbf{K}_{12}^{(s,z)} \mathbf{D}) \mathbf{K}_{22}^{(z,s')} + \mathbf{K}_{11}^{(s,z)} (\mathbf{A} \mathbf{S}_{11} + \mathbf{B} \mathbf{S}_{21}) \mathbf{A} \mathbf{K}_{12}^{(z,s')} + \mathbf{K}_{12}^{(s,z)} (\mathbf{C} \mathbf{S}_{11} + \mathbf{D} \mathbf{S}_{21}) \mathbf{A} \mathbf{K}_{12}^{(z,s')} + \mathbf{K}_{11}^{(s,z)} (\mathbf{A} \mathbf{S}_{12} + \mathbf{B} \mathbf{S}_{22}) \mathbf{C} \mathbf{K}_{12}^{(z,s')} + \mathbf{K}_{12}^{(s,z)} (\mathbf{C} \mathbf{S}_{12} + \mathbf{D} \mathbf{S}_{22}) \mathbf{C} \mathbf{K}_{12}^{(z,s')} + \mathbf{K}_{11}^{(s,z)} (\mathbf{A} \mathbf{S}_{11} + \mathbf{B} \mathbf{S}_{21}) \mathbf{B} \mathbf{K}_{22}^{(z,s')} + \mathbf{K}_{12}^{(s,z)} (\mathbf{C} \mathbf{S}_{11} + \mathbf{D} \mathbf{S}_{21}) \mathbf{B} \mathbf{K}_{22}^{(z,s')} + \mathbf{K}_{11}^{(s,z)} (\mathbf{A} \mathbf{S}_{12} + \mathbf{B} \mathbf{S}_{22}) \mathbf{D} \mathbf{K}_{22}^{(z,s')} + \mathbf{K}_{12}^{(s,z)} (\mathbf{C} \mathbf{S}_{12} + \mathbf{D} \mathbf{S}_{22}) \mathbf{D} \mathbf{K}_{22}^{(z,s')}$;
- $\tilde{\Sigma}_{21}(s, s') = \mathbf{K}_{21}^{(s,s')} - (\mathbf{K}_{21}^{(s,z)} \mathbf{A} + \mathbf{K}_{22}^{(s,z)} \mathbf{C}) \mathbf{K}_{11}^{(z,s')} - (\mathbf{K}_{21}^{(s,z)} \mathbf{B} + \mathbf{K}_{22}^{(s,z)} \mathbf{D}) \mathbf{K}_{21}^{(z,s')} + \mathbf{K}_{21}^{(s,z)} (\mathbf{A} \mathbf{S}_{11} + \mathbf{B} \mathbf{S}_{21}) \mathbf{A} \mathbf{K}_{11}^{(z,s')} + \mathbf{K}_{22}^{(s,z)} (\mathbf{C} \mathbf{S}_{11} + \mathbf{D} \mathbf{S}_{21}) \mathbf{A} \mathbf{K}_{11}^{(z,s')} + \mathbf{K}_{21}^{(s,z)} (\mathbf{A} \mathbf{S}_{12} + \mathbf{B} \mathbf{S}_{22}) \mathbf{C} \mathbf{K}_{11}^{(z,s')} + \mathbf{K}_{22}^{(s,z)} (\mathbf{C} \mathbf{S}_{12} + \mathbf{D} \mathbf{S}_{22}) \mathbf{C} \mathbf{K}_{11}^{(z,s')} + \mathbf{K}_{21}^{(s,z)} (\mathbf{A} \mathbf{S}_{11} + \mathbf{B} \mathbf{S}_{21}) \mathbf{B} \mathbf{K}_{21}^{(z,s')} + \mathbf{K}_{22}^{(s,z)} (\mathbf{C} \mathbf{S}_{11} + \mathbf{D} \mathbf{S}_{21}) \mathbf{B} \mathbf{K}_{21}^{(z,s')} + \mathbf{K}_{21}^{(s,z)} (\mathbf{A} \mathbf{S}_{12} + \mathbf{B} \mathbf{S}_{22}) \mathbf{D} \mathbf{K}_{21}^{(z,s')} + \mathbf{K}_{22}^{(s,z)} (\mathbf{C} \mathbf{S}_{12} + \mathbf{D} \mathbf{S}_{22}) \mathbf{D} \mathbf{K}_{21}^{(z,s')}$;
- $\tilde{\Sigma}_{22}(s, s') = \mathbf{K}_{22}^{(s,s')} - (\mathbf{K}_{21}^{(s,z)} \mathbf{A} + \mathbf{K}_{22}^{(s,z)} \mathbf{C}) \mathbf{K}_{12}^{(z,s')} - (\mathbf{K}_{21}^{(s,z)} \mathbf{B} + \mathbf{K}_{22}^{(s,z)} \mathbf{D}) \mathbf{K}_{22}^{(z,s')} + \mathbf{K}_{21}^{(s,z)} (\mathbf{A} \mathbf{S}_{11} + \mathbf{B} \mathbf{S}_{21}) \mathbf{A} \mathbf{K}_{12}^{(z,s')} + \mathbf{K}_{22}^{(s,z)} (\mathbf{C} \mathbf{S}_{11} + \mathbf{D} \mathbf{S}_{21}) \mathbf{A} \mathbf{K}_{12}^{(z,s')} + \mathbf{K}_{21}^{(s,z)} (\mathbf{A} \mathbf{S}_{12} + \mathbf{B} \mathbf{S}_{22}) \mathbf{C} \mathbf{K}_{12}^{(z,s')} + \mathbf{K}_{22}^{(s,z)} (\mathbf{C} \mathbf{S}_{12} + \mathbf{D} \mathbf{S}_{22}) \mathbf{C} \mathbf{K}_{12}^{(z,s')} + \mathbf{K}_{21}^{(s,z)} (\mathbf{A} \mathbf{S}_{11} + \mathbf{B} \mathbf{S}_{21}) \mathbf{B} \mathbf{K}_{22}^{(z,s')} + \mathbf{K}_{22}^{(s,z)} (\mathbf{C} \mathbf{S}_{11} + \mathbf{D} \mathbf{S}_{21}) \mathbf{B} \mathbf{K}_{22}^{(z,s')} + \mathbf{K}_{21}^{(s,z)} (\mathbf{A} \mathbf{S}_{12} + \mathbf{B} \mathbf{S}_{22}) \mathbf{D} \mathbf{K}_{22}^{(z,s')} + \mathbf{K}_{22}^{(s,z)} (\mathbf{C} \mathbf{S}_{12} + \mathbf{D} \mathbf{S}_{22}) \mathbf{D} \mathbf{K}_{22}^{(z,s')}$.

Com base na distribuição a posteriori, explicitada na equação (4.10), e seguindo as definições de inferência variacional apresentadas na subseção 2.3.2, a divergência KL entre $q(\mathbf{f}, \mathbf{u})$ e $p(\mathbf{f}, \mathbf{u} \mid \mathcal{S}_m, \Theta)$ é dada por

$$\begin{aligned}
KL(q(\mathbf{f}, \mathbf{u}) \parallel p(\mathbf{f}, \mathbf{u} \mid \mathcal{S}_m, \Theta)) &= \mathbb{E}_q \left[\log \frac{q(\mathbf{f}, \mathbf{u})}{p(\mathbf{f}, \mathbf{u} \mid \mathcal{S}_m, \Theta)} \right] \\
&= \mathbb{E}_q \left[\log \frac{q(\mathbf{f}, \mathbf{u}) p(\mathcal{S}_m \mid \Theta)}{p(\mathcal{S}_m \mid \mathbf{f}, \Theta) p(\mathbf{f} \mid \mathbf{u}, \Theta) p(\mathbf{u} \mid \Theta)} \right] \\
&= \mathbb{E}_q[\log q(\mathbf{f}, \mathbf{u})] + \mathbb{E}_q[\log p(\mathcal{S}_m)] - \mathbb{E}_q[\log p(\mathcal{S}_m \mid \mathbf{f}, \Theta)] \\
&\quad - \mathbb{E}_q[\log p(\mathbf{f} \mid \mathbf{u}, \Theta)] - \mathbb{E}_q[\log p(\mathbf{u}, \Theta)] \\
&= \mathbb{E}_q[\log p(\mathbf{f} \mid \mathbf{u}, \Theta)] - \mathbb{E}_q[\log p(\mathbf{f} \mid \mathbf{u}, \Theta)] + \mathbb{E}_q[\log q(\mathbf{u})] \\
&\quad + \mathbb{E}_q[\log p(\mathcal{S}_m)] - \mathbb{E}_q[\log p(\mathcal{S}_m \mid \mathbf{f}, \Theta)] - \mathbb{E}_q[\log p(\mathbf{u}, \Theta)] \\
&= \mathbb{E}_q[\log q(\mathbf{u})] - \mathbb{E}_q[\log p(\mathbf{u} \mid \Theta)] - \mathbb{E}_q[\log p(\mathcal{S}_m \mid \mathbf{f}, \Theta)] \\
&\quad + \mathbb{E}_q[\log p(\mathcal{S}_m)] \\
&= \mathbb{E}_q \left[\log \frac{q(\mathbf{u})}{p(\mathbf{u} \mid \Theta)} \right] - \mathbb{E}_q[\log p(\mathcal{S}_m \mid \mathbf{f}, \Theta)] + \log p(\mathcal{S}_m) \\
&= KL(q(\mathbf{u}) \parallel p(\mathbf{u} \mid \Theta)) - \mathbb{E}_q[\log p(\mathcal{S}_m \mid \mathbf{f}, \Theta)] + \log p(\mathcal{S}_m) \\
&= -\mathcal{L} + \log p(\mathcal{S}_m),
\end{aligned} \tag{4.15}$$

em que $\mathcal{L} = \mathbb{E}_q[\log p(\mathcal{S}_m \mid \mathbf{f}, \Theta)] - KL(q(\mathbf{u}) \parallel p(\mathbf{u} \mid \Theta))$ é denominado de *limite inferior da log evidência*.

Como descrito anteriormente, a divergência KL é uma medida maior ou igual a zero. Deste modo, quanto mais parecido for o comportamento de $q(\mathbf{f}, \mathbf{u})$ em relação a $p(\mathbf{f}, \mathbf{u} \mid \mathcal{S}_m, \Theta)$, mais próximo a divergência KL estará de zero. Sendo assim, o método variacional tem por objetivo reduzir a discrepância existente entre as duas distribuições, isto é, minimizar a divergência KL. No entanto, pode-se observar na equação (4.15) que o logaritmo da função de evidência, $\log p(\mathcal{S}_m)$, é uma quantidade constante e que está somente em função da configuração pontual marcada observada e apenas \mathcal{L} depende do conjunto de parâmetros Θ . Uma vez que \mathcal{L} é inversamente proporcional à medida KL, minimizar a divergência de Kullback-Leibler, $KL(q \parallel p)$, é análogo a maximizar o limite inferior da log evidência, \mathcal{L} , o que torna o processo de inferência mais vantajoso, uma vez que o logaritmo da log evidência, difícil de ser obtido, não é levado em consideração.

O limite inferior da log evidência é composto por dois termos. O primeiro termo descreve o logaritmo da função de verossimilhança do processo pontual, mostrado na equação (4.9), ao passo que, o segundo termo descreve a divergência KL da distribuição variacional dos pontos de indução, $q(\mathbf{u})$, em relação a sua respectiva distribuição a priori, $p(\mathbf{u} \mid \Theta)$.

Aplicando o logaritmo na função de verossimilhança dos dados em (4.9), tem-se que

$$\mathbb{E}_q[\log p(\mathcal{S}_m | \mathbf{f}, \Theta)] = - \sum_{k=1}^2 \underbrace{\left\{ \int_{\mathcal{T}} \mathbb{E}_q[f_k^2(s)] ds \right\}}_{\text{Termo (*)}} + \sum_{k=1}^2 \sum_{i=1}^{N_k} \underbrace{\mathbb{E}_q[\log f_k^2(s_{ki})]}_{\text{Termo (**)}}, \quad (4.16)$$

sendo N_k o número total de eventos para a k -ésima categoria da marca. Desenvolvendo o Termo (*), tem-se que:

$$\begin{aligned} \int_{\mathcal{T}} \mathbb{E}_q[f_k^2(s)] ds &= \int_{\mathcal{T}} \left\{ \text{Var}_q[f_k(s)] + \mathbb{E}_q^2[f_k(s)] \right\} ds \\ &= \int_{\mathcal{T}} \text{Var}_q[f_k(s)] ds + \int_{\mathcal{T}} \mathbb{E}_q^2[f_k(s)] ds, \end{aligned} \quad (4.17)$$

na qual a esperança ao quadrado e a variância de $f_k(s)$ em relação a $q(\mathbf{f}, \mathbf{u})$ respectivamente, para $k = 1$, é dado por

$$\begin{aligned} \mathbb{E}_q^2[f_1(s)] &= \tilde{\mu}_1^2(s) = \tilde{\mu}_1^\top \tilde{\mu}_1 = \\ &= \left[\mathbf{K}_{11}^{(s,z)} \mathbf{A} \mathbf{m}_1 + \mathbf{K}_{12}^{(s,z)} \mathbf{C} \mathbf{m}_1 + \mathbf{K}_{11}^{(s,z)} \mathbf{B} \mathbf{m}_2 + \mathbf{K}_{12}^{(s,z)} \mathbf{D} \mathbf{m}_2 \right]^\top \times \\ &\quad \times \left[\mathbf{K}_{11}^{(s,z)} \mathbf{A} \mathbf{m}_1 + \mathbf{K}_{12}^{(s,z)} \mathbf{C} \mathbf{m}_1 + \mathbf{K}_{11}^{(s,z)} \mathbf{B} \mathbf{m}_2 + \mathbf{K}_{12}^{(s,z)} \mathbf{D} \mathbf{m}_2 \right] = \\ &= (\mathbf{A} \mathbf{m}_1 + \mathbf{B} \mathbf{m}_2)^\top \left[(\mathbf{K}_{11}^{(s,z)})^\top \mathbf{K}_{11}^{(s,z)} \mathbf{A} \mathbf{m}_1 + (\mathbf{K}_{11}^{(s,z)})^\top \mathbf{K}_{12}^{(s,z)} \mathbf{C} \mathbf{m}_1 + \right. \\ &\quad \left. + (\mathbf{K}_{11}^{(s,z)})^\top \mathbf{K}_{11}^{(s,z)} \mathbf{B} \mathbf{m}_2 + (\mathbf{K}_{11}^{(s,z)})^\top \mathbf{K}_{12}^{(s,z)} \mathbf{D} \mathbf{m}_2 \right] + \\ &\quad + (\mathbf{C} \mathbf{m}_1 + \mathbf{D} \mathbf{m}_2)^\top \left[(\mathbf{K}_{12}^{(s,z)})^\top \mathbf{K}_{11}^{(s,z)} \mathbf{A} \mathbf{m}_1 + (\mathbf{K}_{12}^{(s,z)})^\top \mathbf{K}_{12}^{(s,z)} \mathbf{C} \mathbf{m}_1 + \right. \\ &\quad \left. + (\mathbf{K}_{12}^{(s,z)})^\top \mathbf{K}_{11}^{(s,z)} \mathbf{B} \mathbf{m}_2 + (\mathbf{K}_{12}^{(s,z)})^\top \mathbf{K}_{12}^{(s,z)} \mathbf{D} \mathbf{m}_2 \right], \end{aligned}$$

e

$$\begin{aligned}
Var_q[f_1(s)] = \tilde{\sigma}_1^2 = & \gamma_1^2 - Tr\left(\mathbf{AK}_{11}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\right) - Tr\left(\mathbf{CK}_{11}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\right) - \\
& - Tr\left(\mathbf{BK}_{21}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\right) - Tr\left(\mathbf{DK}_{21}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\right) + \\
& + Tr\left(\mathbf{AS}_{11}\mathbf{AK}_{11}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\right) + Tr\left(\mathbf{BS}_{21}\mathbf{AK}_{11}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\right) + \\
& + Tr\left(\mathbf{CS}_{11}\mathbf{AK}_{11}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\right) + Tr\left(\mathbf{DS}_{21}\mathbf{AK}_{11}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\right) + \\
& + Tr\left(\mathbf{AS}_{12}\mathbf{CK}_{11}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\right) + Tr\left(\mathbf{BS}_{22}\mathbf{CK}_{11}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\right) + \\
& + Tr\left(\mathbf{CS}_{12}\mathbf{CK}_{11}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\right) + Tr\left(\mathbf{DS}_{22}\mathbf{CK}_{11}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\right) + \\
& + Tr\left(\mathbf{AS}_{11}\mathbf{BK}_{21}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\right) + Tr\left(\mathbf{BS}_{21}\mathbf{BK}_{21}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\right) + \\
& + Tr\left(\mathbf{CS}_{11}\mathbf{BK}_{21}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\right) + Tr\left(\mathbf{DS}_{21}\mathbf{BK}_{21}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\right) + \\
& + Tr\left(\mathbf{AS}_{12}\mathbf{DK}_{21}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\right) + Tr\left(\mathbf{BS}_{22}\mathbf{DK}_{21}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\right) + \\
& + Tr\left(\mathbf{CS}_{12}\mathbf{DK}_{21}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\right) + Tr\left(\mathbf{DS}_{22}\mathbf{DK}_{21}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\right),
\end{aligned}$$

Portanto, integrando estas medidas de momentos em relação a s , obtém-se

$$\begin{aligned}
\int_{\mathcal{T}} \mathbb{E}_q[f_1^2(s)] ds = & (\mathbf{Am}_1 + \mathbf{Bm}_2)^\top \left[\Psi_{11}^{11}\mathbf{Am}_1 + \Psi_{12}^{11}\mathbf{Cm}_1 + \Psi_{11}^{11}\mathbf{Bm}_2 + \Psi_{12}^{11}\mathbf{Dm}_2 \right] + \\
& + (\mathbf{Cm}_1 + \mathbf{Dm}_2)^\top \left[\Psi_{11}^{21}\mathbf{Am}_1 + \Psi_{12}^{21}\mathbf{Cm}_1 + \Psi_{11}^{21}\mathbf{Bm}_2 + \Psi_{12}^{21}\mathbf{Dm}_2 \right]
\end{aligned}$$

e

$$\begin{aligned}
\int_{\mathcal{T}} Var_q[f_1(s)] ds = & \gamma_1^2 |\mathcal{T}| - Tr\left(\mathbf{A}\Psi_{11}^{11}\right) - Tr\left(\mathbf{C}\Psi_{12}^{11}\right) - Tr\left(\mathbf{B}\Psi_{11}^{21}\right) - Tr\left(\mathbf{D}\Psi_{12}^{21}\right) + \\
& + Tr\left(\mathbf{AS}_{11}\mathbf{A}\Psi_{11}^{11}\right) + Tr\left(\mathbf{BS}_{21}\mathbf{A}\Psi_{11}^{11}\right) + Tr\left(\mathbf{CS}_{11}\mathbf{A}\Psi_{12}^{11}\right) + \\
& + Tr\left(\mathbf{DS}_{21}\mathbf{A}\Psi_{12}^{11}\right) + Tr\left(\mathbf{AS}_{12}\mathbf{C}\Psi_{11}^{11}\right) + Tr\left(\mathbf{BS}_{22}\mathbf{C}\Psi_{11}^{11}\right) + \\
& + Tr\left(\mathbf{CS}_{12}\mathbf{C}\Psi_{12}^{11}\right) + Tr\left(\mathbf{DS}_{22}\mathbf{C}\Psi_{12}^{11}\right) + Tr\left(\mathbf{AS}_{11}\mathbf{B}\Psi_{11}^{21}\right) + \\
& + Tr\left(\mathbf{BS}_{21}\mathbf{B}\Psi_{11}^{21}\right) + Tr\left(\mathbf{CS}_{11}\mathbf{B}\Psi_{12}^{21}\right) + Tr\left(\mathbf{DS}_{21}\mathbf{B}\Psi_{12}^{21}\right) + \\
& + Tr\left(\mathbf{AS}_{12}\mathbf{D}\Psi_{11}^{21}\right) + Tr\left(\mathbf{BS}_{22}\mathbf{D}\Psi_{11}^{21}\right) + Tr\left(\mathbf{CS}_{12}\mathbf{D}\Psi_{12}^{21}\right) + \\
& + Tr\left(\mathbf{DS}_{22}\mathbf{D}\Psi_{12}^{21}\right),
\end{aligned}$$

sendo o cálculo dos elementos da matriz Ψ_{ji}^{ij} , para todo $i, j = 1, 2$, apresentados no Apêndice C.1.

Já, para $k = 2$,

$$\begin{aligned}
\mathbb{E}_q^2[f_2(s)] &= \tilde{\mu}_2^2(s) = \tilde{\mu}_2^\top \tilde{\mu}_2 = \\
&= \left[\mathbf{K}_{21}^{(s,z)} \mathbf{A} \mathbf{m}_1 + \mathbf{K}_{22}^{(s,z)} \mathbf{C} \mathbf{m}_1 + \mathbf{K}_{21}^{(s,z)} \mathbf{B} \mathbf{m}_2 + \mathbf{K}_{22}^{(s,z)} \mathbf{D} \mathbf{m}_2 \right]^\top \times \\
&\quad \times \left[\mathbf{K}_{21}^{(s,z)} \mathbf{A} \mathbf{m}_1 + \mathbf{K}_{22}^{(s,z)} \mathbf{C} \mathbf{m}_1 + \mathbf{K}_{21}^{(s,z)} \mathbf{B} \mathbf{m}_2 + \mathbf{K}_{22}^{(s,z)} \mathbf{D} \mathbf{m}_2 \right] = \\
&= (\mathbf{A} \mathbf{m}_1 + \mathbf{B} \mathbf{m}_2)^\top \left[(\mathbf{K}_{21}^{(s,z)})^\top \mathbf{K}_{21}^{(s,z)} \mathbf{A} \mathbf{m}_1 + (\mathbf{K}_{21}^{(s,z)})^\top \mathbf{K}_{22}^{(s,z)} \mathbf{C} \mathbf{m}_1 + \right. \\
&\quad \left. + (\mathbf{K}_{21}^{(s,z)})^\top \mathbf{K}_{21}^{(s,z)} \mathbf{B} \mathbf{m}_2 + (\mathbf{K}_{21}^{(s,z)})^\top \mathbf{K}_{22}^{(s,z)} \mathbf{D} \mathbf{m}_2 \right] + \\
&\quad + (\mathbf{C} \mathbf{m}_1 + \mathbf{D} \mathbf{m}_2)^\top \left[(\mathbf{K}_{22}^{(s,z)})^\top \mathbf{K}_{21}^{(s,z)} \mathbf{A} \mathbf{m}_1 + (\mathbf{K}_{22}^{(s,z)})^\top \mathbf{K}_{22}^{(s,z)} \mathbf{C} \mathbf{m}_1 + \right. \\
&\quad \left. + (\mathbf{K}_{22}^{(s,z)})^\top \mathbf{K}_{21}^{(s,z)} \mathbf{B} \mathbf{m}_2 + (\mathbf{K}_{22}^{(s,z)})^\top \mathbf{K}_{22}^{(s,z)} \mathbf{D} \mathbf{m}_2 \right]
\end{aligned}$$

e

$$\begin{aligned}
\text{Var}_q[f_2(s)] &= \tilde{\sigma}_2^2 = \gamma_2^2 - \text{Tr} \left(\mathbf{A} \mathbf{K}_{12}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \right) - \text{Tr} \left(\mathbf{C} \mathbf{K}_{12}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \right) - \\
&\quad - \text{Tr} \left(\mathbf{B} \mathbf{K}_{22}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \right) - \text{Tr} \left(\mathbf{D} \mathbf{K}_{22}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \right) + \\
&\quad + \text{Tr} \left(\mathbf{A} \mathbf{S}_{11} \mathbf{A} \mathbf{K}_{12}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \right) + \text{Tr} \left(\mathbf{B} \mathbf{S}_{21} \mathbf{A} \mathbf{K}_{12}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \right) + \\
&\quad + \text{Tr} \left(\mathbf{C} \mathbf{S}_{11} \mathbf{A} \mathbf{K}_{12}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \right) + \text{Tr} \left(\mathbf{D} \mathbf{S}_{21} \mathbf{A} \mathbf{K}_{12}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \right) + \\
&\quad + \text{Tr} \left(\mathbf{A} \mathbf{S}_{12} \mathbf{C} \mathbf{K}_{12}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \right) + \text{Tr} \left(\mathbf{B} \mathbf{S}_{22} \mathbf{C} \mathbf{K}_{12}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \right) + \\
&\quad + \text{Tr} \left(\mathbf{C} \mathbf{S}_{12} \mathbf{C} \mathbf{K}_{12}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \right) + \text{Tr} \left(\mathbf{D} \mathbf{S}_{22} \mathbf{C} \mathbf{K}_{12}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \right) + \\
&\quad + \text{Tr} \left(\mathbf{A} \mathbf{S}_{11} \mathbf{B} \mathbf{K}_{22}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \right) + \text{Tr} \left(\mathbf{B} \mathbf{S}_{21} \mathbf{B} \mathbf{K}_{22}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \right) + \\
&\quad + \text{Tr} \left(\mathbf{C} \mathbf{S}_{11} \mathbf{B} \mathbf{K}_{22}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \right) + \text{Tr} \left(\mathbf{D} \mathbf{S}_{21} \mathbf{B} \mathbf{K}_{22}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \right) + \\
&\quad + \text{Tr} \left(\mathbf{A} \mathbf{S}_{12} \mathbf{D} \mathbf{K}_{22}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \right) + \text{Tr} \left(\mathbf{B} \mathbf{S}_{22} \mathbf{D} \mathbf{K}_{22}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \right) + \\
&\quad + \text{Tr} \left(\mathbf{C} \mathbf{S}_{12} \mathbf{D} \mathbf{K}_{22}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \right) + \text{Tr} \left(\mathbf{D} \mathbf{S}_{22} \mathbf{D} \mathbf{K}_{22}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \right),
\end{aligned}$$

e, portanto,

$$\int_{\mathcal{T}} \mathbb{E}_q[f_2^2(s)] ds = (\mathbf{A}\mathbf{m}_1 + \mathbf{B}\mathbf{m}_2)^\top \left[\Psi_{21}^{12}\mathbf{A}\mathbf{m}_1 + \Psi_{22}^{12}\mathbf{C}\mathbf{m}_1 + \Psi_{21}^{12}\mathbf{B}\mathbf{m}_2 + \Psi_{22}^{12}\mathbf{D}\mathbf{m}_2 \right] +$$

$$+ (\mathbf{C}\mathbf{m}_1 + \mathbf{D}\mathbf{m}_2)^\top \left[\Psi_{21}^{22}\mathbf{A}\mathbf{m}_1 + \Psi_{22}^{22}\mathbf{C}\mathbf{m}_1 + \Psi_{21}^{22}\mathbf{B}\mathbf{m}_2 + \Psi_{22}^{22}\mathbf{D}\mathbf{m}_2 \right]$$

e

$$\int_{\mathcal{T}} \text{Var}_q[f_2(s)] ds = \gamma_2^2 |\mathcal{T}| - \text{Tr}(\mathbf{A}\Psi_{21}^{12}) - \text{Tr}(\mathbf{C}\Psi_{22}^{12}) - \text{Tr}(\mathbf{B}\Psi_{21}^{22}) - \text{Tr}(\mathbf{D}\Psi_{22}^{22}) +$$

$$+ \text{Tr}(\mathbf{A}\mathbf{S}_{11}\mathbf{A}\Psi_{21}^{12}) + \text{Tr}(\mathbf{B}\mathbf{S}_{21}\mathbf{A}\Psi_{21}^{12}) + \text{Tr}(\mathbf{C}\mathbf{S}_{11}\mathbf{A}\Psi_{22}^{12}) +$$

$$+ \text{Tr}(\mathbf{D}\mathbf{S}_{21}\mathbf{A}\Psi_{22}^{12}) + \text{Tr}(\mathbf{A}\mathbf{S}_{12}\mathbf{C}\Psi_{21}^{12}) + \text{Tr}(\mathbf{B}\mathbf{S}_{22}\mathbf{C}\Psi_{21}^{12}) +$$

$$+ \text{Tr}(\mathbf{C}\mathbf{S}_{12}\mathbf{C}\Psi_{22}^{12}) + \text{Tr}(\mathbf{D}\mathbf{S}_{22}\mathbf{C}\Psi_{22}^{12}) + \text{Tr}(\mathbf{A}\mathbf{S}_{11}\mathbf{B}\Psi_{21}^{22}) +$$

$$+ \text{Tr}(\mathbf{B}\mathbf{S}_{21}\mathbf{B}\Psi_{21}^{22}) + \text{Tr}(\mathbf{C}\mathbf{S}_{11}\mathbf{B}\Psi_{22}^{22}) + \text{Tr}(\mathbf{D}\mathbf{S}_{21}\mathbf{B}\Psi_{22}^{22}) +$$

$$+ \text{Tr}(\mathbf{A}\mathbf{S}_{12}\mathbf{D}\Psi_{21}^{22}) + \text{Tr}(\mathbf{B}\mathbf{S}_{22}\mathbf{D}\Psi_{21}^{22}) + \text{Tr}(\mathbf{C}\mathbf{S}_{12}\mathbf{D}\Psi_{22}^{22}) +$$

$$+ \text{Tr}(\mathbf{D}\mathbf{S}_{22}\mathbf{D}\Psi_{22}^{22}).$$

Desse modo, tem-se as expressões para o Termo (*) considerando ambos os processos Gaussianos, $f_1(s)$ e $f_2(s)$.

Já o Termo (**) da equação (4.16) é desenvolvido de forma análoga à expressão (2.68). Nesse caso, para $k = 1$, tem-se que

$$\mathbb{E}_q[\log f_1^2(s_{1i})] = \int_{-\infty}^{+\infty} \log f_1^2(s_{1i}) \mathcal{N}\left(f_1(s_{1i}); \tilde{\mu}_1(s_{1i}), \tilde{\sigma}_1^2(s_{1i})\right) df_1(s_{1i})$$

$$= -\tilde{G}\left(\frac{-\tilde{\mu}_1^2(s_{1i})}{2\tilde{\sigma}_1^2(s_{1i})}\right) + \log\left(\frac{\tilde{\sigma}_1^2(s_{1i})}{2}\right) - C,$$

ao passo que, para $k = 2$,

$$\mathbb{E}_q[\log f_2^2(s_{2i})] = -\tilde{G}\left(\frac{-\tilde{\mu}_2^2(s_{2i})}{2\tilde{\sigma}_2^2(s_{2i})}\right) + \log\left(\frac{\tilde{\sigma}_2^2(s_{2i})}{2}\right) - C.$$

em que $C \approx 0,57721566$ é a constante de Euler-Mascheroni e $\tilde{G}(\cdot)$ é uma função definida a partir da derivada parcial de ${}_1F_1(a, b, z)$ em relação ao argumento “a”, como definido na subseção 2.4.

Outro termo que compõe a expressão \mathcal{L} é a divergência KL entre a densidade variacional de $\mathbf{u} = [\mathbf{u}_1^\top, \mathbf{u}_2^\top]^\top$, $q(\mathbf{u})$, e sua respectiva priori, $p(\mathbf{u} | \Theta)$. Sabendo que $\log q(\mathbf{u}) =$

$-\frac{1}{2}[2M \log(2\pi) + \log \det(\mathbf{S}) + (\mathbf{u} - \mathbf{m})^\top \mathbf{S}^{-1}(\mathbf{u} - \mathbf{m})]$ e que $\log p(\mathbf{u} | \Theta) = -\frac{1}{2}[2M \log(2\pi) + \log \det(\mathbf{K}^{(z,z')}) + (\mathbf{u} - \bar{\mathbf{u}})^\top \mathbf{K}^{(z,z')^{-1}}(\mathbf{u} - \bar{\mathbf{u}})]$ tem-se que

$$\begin{aligned}
KL(q(\mathbf{u}) \parallel p(\mathbf{u} | \Theta)) &= \mathbb{E}_q \left[\log \frac{q(\mathbf{u})}{p(\mathbf{u} | \Theta)} \right] \\
&= \mathbb{E}_q[\log q(\mathbf{u})] - \mathbb{E}_q[\log p(\mathbf{u} | \Theta)] \\
&= \mathbb{E}_q \left\{ \frac{1}{2} \left[-2M \log(2\pi) + 2M \log(2\pi) - \log \det(\mathbf{S}) + \log \det(\mathbf{K}^{(z,z')}) \right. \right. \\
&\quad \left. \left. - (\mathbf{u} - \mathbf{m})^\top \mathbf{S}^{-1}(\mathbf{u} - \mathbf{m}) + (\mathbf{u} - \bar{\mathbf{u}})^\top \mathbf{K}^{(z,z')^{-1}}(\mathbf{u} - \bar{\mathbf{u}}) \right] \right\} \\
&= \frac{1}{2} \left\{ \log \frac{\det(\mathbf{K}^{(z,z')})}{\det(\mathbf{S})} - \mathbb{E}_q[(\mathbf{u} - \mathbf{m})^\top \mathbf{S}^{-1}(\mathbf{u} - \mathbf{m}) \right. \\
&\quad \left. + (\mathbf{u} - \bar{\mathbf{u}})^\top \mathbf{K}^{(z,z')^{-1}}(\mathbf{u} - \bar{\mathbf{u}})] \right\} \\
&= \frac{1}{2} \left\{ \underbrace{\log \frac{\det(\mathbf{K}^{(z,z')})}{\det(\mathbf{S})}}_{\text{Expressão 1}} - \underbrace{\mathbb{E}_q[(\mathbf{u} - \mathbf{m})^\top \mathbf{S}^{-1}(\mathbf{u} - \mathbf{m})]}_{\text{Expressão 2}} \right. \\
&\quad \left. + \underbrace{\mathbb{E}_q[(\mathbf{u} - \bar{\mathbf{u}})^\top \mathbf{K}^{(z,z')^{-1}}(\mathbf{u} - \bar{\mathbf{u}})]}_{\text{Expressão 3}} \right\}
\end{aligned}$$

Para desenvolver a Expressão 1, foi utilizada a regra do determinante de matriz particionada. Desse modo, tem-se que para a matriz $\mathbf{K}^{(z,z')}$ seu determinante é dado por

$$\det(\mathbf{K}^{(z,z')}) = \det(\mathbf{K}_{11}^{(z,z')}) \times \det(\mathbf{K}_{22}^{(z,z')} - \mathbf{K}_{21}^{(z,z')} \mathbf{K}_{11}^{(z,z')^{-1}} \mathbf{K}_{12}^{(z,z')})$$

e, de forma análoga,

$$\det(\mathbf{S}) = \det(\mathbf{S}_{11}) \times \det(\mathbf{S}_{22} - \mathbf{S}_{21} \mathbf{S}_{11}^{-1} \mathbf{S}_{12}).$$

Logo, a Expressão 1 pode ser escrita por

$$\begin{aligned}
\log \frac{\det(\mathbf{K}^{(z,z')})}{\det(\mathbf{S})} &= \log \det(\mathbf{K}_{11}^{(z,z')}) + \log \det(\mathbf{K}_{22}^{(z,z')} - \mathbf{K}_{21}^{(z,z')} \mathbf{K}_{11}^{(z,z')^{-1}} \mathbf{K}_{12}^{(z,z')}) - \\
&\quad - \log \det(\mathbf{S}_{11}) - \log \det(\mathbf{S}_{22} - \mathbf{S}_{21} \mathbf{S}_{11}^{-1} \mathbf{S}_{12})
\end{aligned}$$

Já para a Expressão 2, usando a propriedade $\mathbb{E}(\mathbf{x}^\top \mathbf{A} \mathbf{x}) = \mathbb{E}[tr(\mathbf{x}^\top \mathbf{A} \mathbf{x})]$ e assumindo que a inversa da matriz particionada de covariâncias variacionais de \mathbf{u} é $\mathbf{S}^{-1} = \left(\begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix} \right)^{-1} =$

$\begin{bmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{bmatrix}$, em que $\mathbf{E} = \mathbf{S}_{11}^{-1} + \mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{H}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}$, $\mathbf{F} = -\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{H}$, $\mathbf{G} = -\mathbf{H}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}$ e $\mathbf{H} = (\mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12})^{-1}$, tem-se que

$$\begin{aligned}
\mathbb{E}_q[(\mathbf{u} - \mathbf{m})^\top \mathbf{S}^{-1}(\mathbf{u} - \mathbf{m})] &= \mathbb{E}_q[(\mathbf{u}_1 - \mathbf{m}_1)^\top \mathbf{E}(\mathbf{u}_1 - \mathbf{m}_1) + (\mathbf{u}_2 - \mathbf{m}_2)^\top \mathbf{G}(\mathbf{u}_1 - \mathbf{m}_1) + \\
&\quad (\mathbf{u}_1 - \mathbf{m}_1)^\top \mathbf{F}(\mathbf{u}_2 - \mathbf{m}_2) + (\mathbf{u}_2 - \mathbf{m}_2)^\top \mathbf{H}(\mathbf{u}_2 - \mathbf{m}_2)] \\
&= \text{tr}(\mathbf{E}\mathbf{S}_{11} + \mathbf{G}\mathbf{S}_{12} + \mathbf{F}\mathbf{S}_{21} + \mathbf{H}\mathbf{S}_{22}) \\
&= M + \text{tr}(-\mathbf{F}\mathbf{S}_{21} + \mathbf{G}\mathbf{S}_{12} + \mathbf{F}\mathbf{S}_{21} + \mathbf{H}\mathbf{S}_{22}) \\
&= M + \text{tr}(\mathbf{G}\mathbf{S}_{12} + \mathbf{H}\mathbf{S}_{22}) \\
&= M + \text{tr}(-\mathbf{H}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12} + \mathbf{H}\mathbf{S}_{22}) \\
&= M + \text{tr}[\mathbf{H}(\mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12})] = M + \text{tr}\{\mathbf{H}\mathbf{H}^{-1}\} \\
&= M + \text{tr}\{\mathbf{I}\} = 2M
\end{aligned}$$

De maneira análoga ao desenvolvimento da Expressão 2, assumindo que a inversa da matriz particionada de covariâncias a priori de \mathbf{u} é $\mathbf{K}^{(z,z')^{-1}} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$, tem-se que a Expressão 3 é dada por

$$\begin{aligned}
\mathbb{E}_q[(\mathbf{u} - \bar{\mathbf{u}})^\top \mathbf{K}^{(z,z')^{-1}}(\mathbf{u} - \bar{\mathbf{u}})] &= \mathbb{E}_q[(\mathbf{u}_1 - \vec{\mathbf{1}}\bar{u}_1)^\top \mathbf{A}(\mathbf{u}_1 - \vec{\mathbf{1}}\bar{u}_1) + (\mathbf{u}_2 - \vec{\mathbf{1}}\bar{u}_2)^\top \mathbf{C}(\mathbf{u}_1 - \vec{\mathbf{1}}\bar{u}_1) + \\
&\quad (\mathbf{u}_1 - \vec{\mathbf{1}}\bar{u}_1)^\top \mathbf{B}(\mathbf{u}_2 - \vec{\mathbf{1}}\bar{u}_2) + (\mathbf{u}_2 - \vec{\mathbf{1}}\bar{u}_2)^\top \mathbf{D}(\mathbf{u}_2 - \vec{\mathbf{1}}\bar{u}_2)] \\
&= \mathbb{E}_q\{\text{tr}[(\mathbf{u}_1 - \vec{\mathbf{1}}\bar{u}_1)^\top \mathbf{A}(\mathbf{u}_1 - \vec{\mathbf{1}}\bar{u}_1) + (\mathbf{u}_2 - \vec{\mathbf{1}}\bar{u}_2)^\top \mathbf{C}(\mathbf{u}_1 - \vec{\mathbf{1}}\bar{u}_1) \\
&\quad + (\mathbf{u}_1 - \vec{\mathbf{1}}\bar{u}_1)^\top \mathbf{B}(\mathbf{u}_2 - \vec{\mathbf{1}}\bar{u}_2) + (\mathbf{u}_2 - \vec{\mathbf{1}}\bar{u}_2)^\top \mathbf{D}(\mathbf{u}_2 - \vec{\mathbf{1}}\bar{u}_2)]\} \\
&= \text{tr}(\mathbf{A}\mathbf{S}_{11}) + \text{tr}(\mathbf{C}\mathbf{S}_{12}) + \text{tr}(\mathbf{B}\mathbf{S}_{21}) + \text{tr}(\mathbf{D}\mathbf{S}_{22}) + \\
&\quad + (\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1)^\top \mathbf{A}(\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1) + (\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2)^\top \mathbf{C}(\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1) + \\
&\quad + (\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1)^\top \mathbf{B}(\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2) + (\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2)^\top \mathbf{D}(\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2) \\
&= \text{tr}(\mathbf{A}\mathbf{S}_{11} + \mathbf{C}\mathbf{S}_{12} + \mathbf{B}\mathbf{S}_{21} + \mathbf{D}\mathbf{S}_{22}) + \\
&\quad + (\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1)^\top \mathbf{A}(\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1) + (\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2)^\top \mathbf{C}(\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1) + \\
&\quad + (\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1)^\top \mathbf{B}(\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2) + (\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2)^\top \mathbf{D}(\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2)
\end{aligned}$$

Logo, a divergência KL entre $q(\mathbf{u})$ e $p(\mathbf{u} | \Theta)$ é dada por

$$\begin{aligned}
KL\left(q(\mathbf{u}) \parallel p(\mathbf{u} \mid \Theta)\right) &= \frac{1}{2} \left[\log \det (\mathbf{K}_{11}^{(z,z')}) + \log \det (\mathbf{K}_{22}^{(z,z')} - \mathbf{K}_{21}^{(z,z')} \mathbf{K}_{11}^{(z,z')^{-1}} \mathbf{K}_{12}^{(z,z')}) - \right. \\
&\quad - \log \det (\mathbf{S}_{11}) - \log \det (\mathbf{S}_{22} - \mathbf{S}_{21} \mathbf{S}_{11}^{-1} \mathbf{S}_{12}) - 2M + \\
&\quad + \text{tr}(\mathbf{A} \mathbf{S}_{11} + \mathbf{C} \mathbf{S}_{12} + \mathbf{B} \mathbf{S}_{21} + \mathbf{D} \mathbf{S}_{22}) + \\
&\quad + (\mathbf{m}_1 - \vec{\mathbf{1}} \bar{u}_1)^\top \mathbf{A} (\mathbf{m}_1 - \vec{\mathbf{1}} \bar{u}_1) + (\mathbf{m}_2 - \vec{\mathbf{1}} \bar{u}_2)^\top \mathbf{C} (\mathbf{m}_1 - \vec{\mathbf{1}} \bar{u}_1) + \\
&\quad \left. + (\mathbf{m}_1 - \vec{\mathbf{1}} \bar{u}_1)^\top \mathbf{B} (\mathbf{m}_2 - \vec{\mathbf{1}} \bar{u}_2) + (\mathbf{m}_2 - \vec{\mathbf{1}} \bar{u}_2)^\top \mathbf{D} (\mathbf{m}_2 - \vec{\mathbf{1}} \bar{u}_2) \right] \quad (4.18)
\end{aligned}$$

Uma vez definidas as expressões dos termos que compõem \mathcal{L} , o processo de inferência pode ser realizado com base em um método de otimização. Como foi descrito anteriormente, minimizar a divergência KL entre $q(\mathbf{f}, \mathbf{u})$ e $p(\mathbf{f}, \mathbf{u} \mid \Theta)$ é análogo a maximizar o limite inferior da log evidência, \mathcal{L} . Desse modo, a inferência variacional para o presente modelo tem por finalidade obter os valores de seus hiperparâmetros que maximizam \mathcal{L} . Na literatura, existem vários métodos de otimização que podem ser utilizados no processo de estimação, entre eles o algoritmo BFGS (Apêndice A), que necessita do cálculo do gradiente de \mathcal{L} . Para fazer uso desse método, os gradientes de \mathcal{L} foram calculados e estão apresentados no Apêndice C, especificamente nas partes C.2, C.3, C.4, C.5 e C.6.

Dado que os hiperparâmetros do modelo foram otimizados, denotados por α^* , ρ_{12}^* , γ_1^* , γ_2^* , \bar{u}_1^* , \bar{u}_2^* , \mathbf{m}_1^* , \mathbf{m}_2^* , \mathbf{S}_{11}^* , \mathbf{S}_{12}^* , \mathbf{S}_{21}^* e \mathbf{S}_{22}^* , a distribuição a posteriori para o processo Gaussiano bivariado $\mathbf{f}(s)$ pode ser aproximada por meio de sua distribuição marginal variacional otimizada, isto é, $q^*(\mathbf{f}) \approx p(\mathbf{f} \mid \mathcal{S}_m, \Theta)$. Seguindo a expressão em (4.12), tem-se que

$$q^*(\mathbf{f}) = q^*(f_1, f_2) = \mathcal{P}\mathcal{G} \left(\tilde{\boldsymbol{\mu}}^* = \begin{bmatrix} \tilde{\mu}_1^*(s) \\ \tilde{\mu}_2^*(s) \end{bmatrix}, \tilde{\boldsymbol{\Sigma}}^* = \begin{bmatrix} \tilde{\Sigma}_{11}^*(s, s') & \tilde{\Sigma}_{12}^*(s, s') \\ \tilde{\Sigma}_{21}^*(s, s') & \tilde{\Sigma}_{22}^*(s, s') \end{bmatrix} \right), \quad (4.19)$$

sendo $\tilde{\boldsymbol{\mu}}^*$ o vetor de funções de médias e $\tilde{\boldsymbol{\Sigma}}^*$ uma matriz de funções de covariâncias cruzadas definidas, respectivamente, pelas expressões (4.13) e (4.14), considerando os hiperparâmetros otimizados.

Por meio da expressão (4.19), é possível obter a distribuição marginal variacional para cada processo Gaussiano através da integração. Desse modo, tem-se que

$$q^*(f_1) = \int q^*(f_1, f_2) df_2 = \mathcal{P}\mathcal{G} \left(\tilde{\mu}_1^*(s), \tilde{\Sigma}_{11}^*(s, s') \right) \quad (4.20)$$

sendo $\tilde{\mu}_1^*(s)$ e $\tilde{\Sigma}_{11}^*(s, s')$ as funções de média e de covariância para o processo Gaussiano $f_1(s)$, para qualquer $s, s' \in \mathcal{T} \subset \mathbb{R}^1$. De maneira análoga, a distribuição marginal variacional de $f_2(s)$ será

$$q^*(f_2) = \int q^*(f_1, f_2) df_1 = \mathcal{P}\mathcal{G}\left(\tilde{\mu}_2^*(s), \tilde{\Sigma}_{22}^*(s, s')\right) \quad (4.21)$$

em que $\tilde{\mu}_2^*(s)$ e $\tilde{\Sigma}_{22}^*(s, s')$ suas respectivas funções de média e de covariância, para qualquer $s, s' \in \mathcal{T} \subset \mathbb{R}^1$.

Uma vez obtidas as distribuições marginais variacionais de $f_1(s)$ e $f_2(s)$, é possível inferir sob as intensidades de cada categoria da marca do processo de Cox Gaussiano, definido em (4.1). Sabendo que, para cada ponto de indexação $s \in \mathcal{T}$, $f_i(s)$ pode ser vista como uma variável aleatória univariada, seguindo uma distribuição variacional Normal com média $\tilde{\mu}_i^*(s)$ e variância $\tilde{\Sigma}_{ii}^*(s, s) = \tilde{\sigma}_i^{2*}(s)$, para $i = 1, 2$, então $\lambda_i(s) = [f_i^*(s)]^2$ pode ser inferida por meio de uma distribuição Qui-quadrada não centralizada escalonada, com $\nu_i = 1$ grau de liberdade e com um parâmetro δ_i de não centralidade, que é obtido pela razão entre $\tilde{\mu}_i^*(s)$ e $\tilde{\sigma}_i^{2*}(s)$. Ou seja,

$$\lambda_i(s) \sim \tilde{\sigma}_i^{2*}(s) \chi^2\left(\nu_i = 1, \delta_i = \frac{\tilde{\mu}_i^{*2}(s)}{\tilde{\sigma}_i^{2*}(s)}\right), \quad \text{para } s \in \mathcal{T} \text{ e } i = 1, 2, \quad (4.22)$$

na qual a média variacional de $\lambda_i(s)$ é dada por

$$\mathbb{E}[\lambda_i(s)] = \tilde{\sigma}_i^{2*}(s) + \tilde{\mu}_i^{*2}(s) \quad (4.23)$$

e sua variância variacional é dada por

$$Var[\lambda_i(s)] = 2\tilde{\sigma}_i^{2*}(s) + 4\tilde{\mu}_i^{*2}(s). \quad (4.24)$$

Para a obtenção de uma estimativa pontual de $\lambda_i(s)$, para $i = 1, 2$, pode-se utilizar algumas de suas medidas de tendência central, como a média, mediana ou a moda variacional, seguindo o mesmo raciocínio de Lloyd et al. (2015). Já para a obtenção dos intervalos de credibilidade de $\lambda_i(s)$ à um nível de $(1 - \alpha)100\%$ de probabilidade, a ideia também é similar. Utiliza-se o método do HPD (sigla em inglês para *highest posterior density*) para obter aquele intervalo de valores de $\lambda_i(s)$ com maior densidade variacional.

5 APLICAÇÃO EM DADOS DE ACIDENTES RODOVIÁRIOS

Os acidentes de trânsito figuram como uma das principais causas de morte no mundo, estando na nona posição no ranking mundial entre as causas de óbito. De acordo com a Organização Mundial de Saúde, cerca de 1,2 milhões de pessoas morrem anualmente em decorrência de tal fatalidade e cerca de 3% do PIB mundial é destinado para esta finalidade. Além disso, os acidentes de trânsito são considerados a principal causa de morte mundial entre os jovens de 15 à 29 anos de idade (WORLD HEALTH ORGANIZATION, 2015).

Diversos fatores podem agravar o risco de acidentes. A World Health Organization (2015) destaca cinco fatores chave que aumentam tal risco, como a velocidade inadequada, uso de bebida alcoólica, falha no uso de capacetes para motociclistas, falta do uso de cinto de segurança e de restrições para crianças.

Segundo Mesquita Filho, Carvalho e Garcia (2017), o Brasil é considerado o país que apresenta a maior taxa de mortalidade decorrente de acidentes com veículos automotores no continente americano. De acordo com dados divulgados pelo Ministério da Saúde, a região Sudeste se encontra na primeira posição dentre as regiões brasileiras que mais se destaca em relação à frequência de mortes associadas a acidentes de trânsito. Para o ano de 2015, cerca de 34% dos óbitos por acidente de trânsito no Brasil ocorreram no Sudeste, sendo fortemente contribuído pelos estados de São Paulo (47%) e Minas Gerais(30%) (MINISTÉRIO DA SAÚDE, 2018).

Uma das principais causas de acidentes está associada a imprudências cometidas por condutores desses veículos. Este fato já foi discutido por Freitas (2013), que destaca a influência do ser humano como um dos principais causadores para que um acidente, de fato, aconteça. Dentre as várias causas que podem ser classificadas como imprudência, o excesso de velocidade tem sido a causa de maior destaque, uma vez que o uso inadequado da velocidade nestas vias aumentam diretamente o risco de acidente e, conseqüentemente, a chance de uma pessoa vir a óbito (WORLD HEALTH ORGANIZATION, 2018).

De acordo com a base de dados apresentada pela Polícia Rodoviária Federal, durante o período de janeiro de 2017 à setembro de 2021, foram registrados 337584 acidentes em todas as rodovias federais em todo o país. Extratificando essas ocorrências por cada estado brasileiro, verificou-se que Minas Gerais, estado com a maior malha rodoviária do Brasil, foi o estado que apresentou o maior número de acidentes registrados no período, com 13,3% das ocorrências

totais ($N = 44983$). Dentre todas as rodovias que passam por Minas Gerais, a rodovia BR-381 foi a rodovia que apresentou o maior número absoluto de acidentes, com 31,1% das notificações registradas no estado ($N = 14008$).

A rodovia Fernão Dias, parte da BR-381, é considerada uma das principais rodovias federais do Brasil, uma vez que ela está situada na região de maior desenvolvimento industrial e econômico do país, responsável por interligar dois dos mais importantes estados do país: São Paulo e Minas Gerais. Por conter uma grande circulação de veículos, essa rodovia se torna uma das mais expressivas em termos de ocorrências de acidentes. Entre os anos de 2015 e 2019, a BR-381 foi a terceira rodovia com o maior registro de acidentes em relação as demais (POLÍCIA RODOVIÁRIA FEDERAL, 2021).

Diversos fatores podem potencializar a exposição ao risco em um acidente de trânsito. Alguns estudos indicam que condições climáticas ruins (CALEFFI et al., 2016), geometria da rodovia (ANDRIOLA; TORRES; GARCÍA, 2019), entre outros, exercem grande influência para o aumento do risco de tais acontecimentos, sendo, a imprudência por parte do condutor do veículo a mais expressiva (TEIXEIRA; CASTANHO; NETO, 2017).

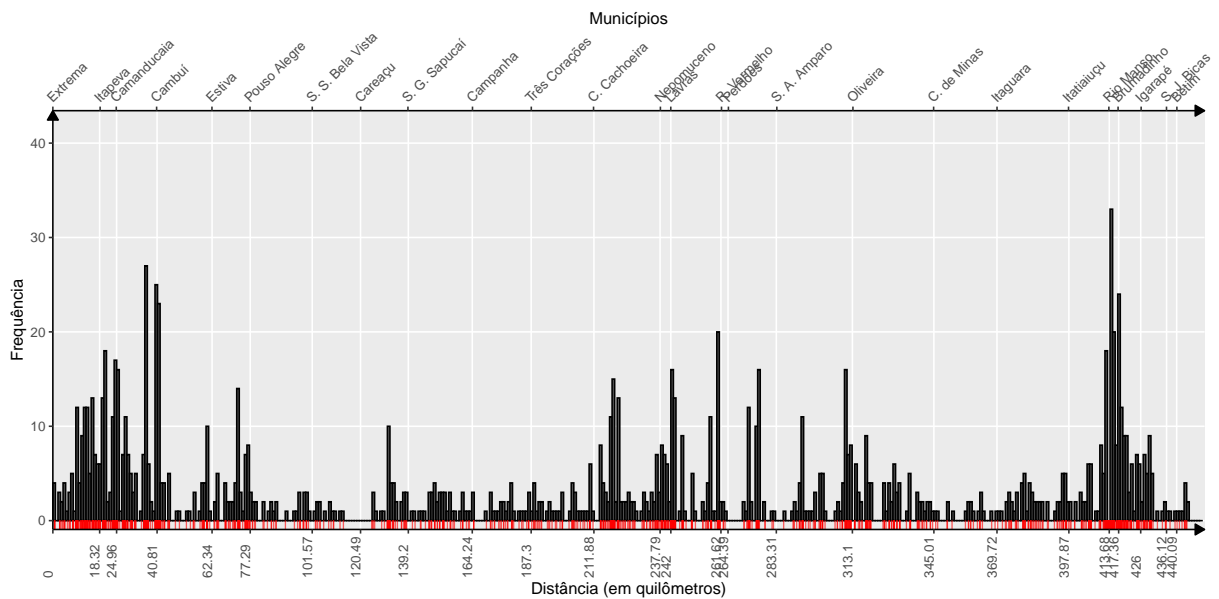
Em concordância com Teixeira, Castanho e Neto (2017), observou-se que as principais causas de acidentes na BR-381 dentro do estado de Minas Gerais decorre da imprudência do condutor, sendo a incompatibilidade da velocidade permitida na rodovia a causa de acidente mais frequente registrada, correspondendo à 28,1% ($N = 3933$) de todos os acidentes registrados nesse estado e rodovia. Considerando um cenário meteorológico adverso, observou-se que cerca de 44,4% dos acidentes em toda a BR-381 no estado de Minas Gerais ($N = 1746$) ocorreram em situações de chuva, garoa, chuvisto ou granizo.

Do ponto de vista estatístico, os acidentes podem ser vistos como fenômenos aleatórios de natureza pontual, ou seja, as ocorrências estão associadas, especificamente, a uma localização no espaço e em um instante de tempo, representando um “ponto” indexado no domínio espaço-temporal. Desse modo, um estudo sob o comportamento dos acidentes ao longo de um determinado trecho de uma rodovia pode ser realizado a partir de uma metodologia envolvendo processos pontuais, inclusive com base na teoria apresentada nessa tese.

Considerando exclusivamente a informação referenciada da distância percorrida na rodovia BR-381 a partir do ponto inicial (0 Km no município de Extrema) até a localização exata de ocorrência do acidente, um conjunto de $N = 1357$ “pontos indexadores” unidimensionais foi obtido, formando assim uma configuração pontual. A Figura 5.1 mostra o histograma com a

frequência absoluta dos acidentes durante o período de janeiro de 2017 à setembro de 2021, ao longo do trecho da rodovia em estudo (Extrema à Betim), totalizando um percurso de 448,6025 quilômetros percorridos. O histograma foi obtido considerando a frequência absoluta dos acidentes a cada 1 Km de distância percorrida.

Figura 5.1 – Histograma com a frequência absoluta do número de ocorrências de acidentes por quilômetro em um trecho da rodovia Fernão Dias (BR-381) entre os municípios de Extrema e Betim. Os traços em vermelho, **I**, representam as posições exatas (distância em quilômetros) de cada ocorrência.



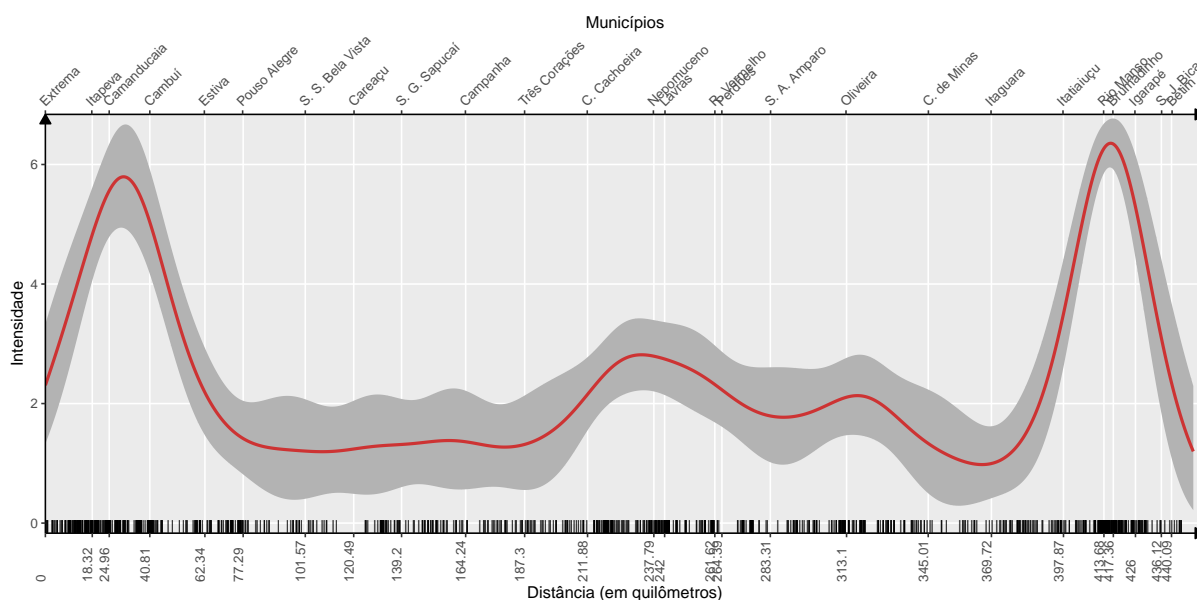
Fonte: Do autor (2021).

Com base no histograma, observou-se basicamente dois trechos da rodovia que apresentaram uma frequência alta de acidentes. São eles: entre os municípios de Extrema e Estiva e entre os municípios de Itatiaiuçu e Igarapé. Outros trechos, como entre Carmo da Cachoeira e Oliveira, apresentaram um leve aumento das ocorrências, mas não tão elevado quanto os dois trechos citados anteriormente.

Uma vez que o conjunto de dados de ocorrências de acidentes pode ser estudado a partir da teoria de processos pontuais, foi aplicado o modelo proposto por Lloyd et al. (2015), descrito anteriormente na subseção 2.4. Utilizando o algoritmo BFGS (Broyden-Fletcher-Goldfarb-Shanno) para a otimização no processo inferencial do modelo, utilizou-se $\alpha = 200$, $\gamma = 0,01$ e $\bar{u} = 1$ como valores iniciais para os hiperparâmetros e $M = 15$ pontos de indução. Com relação ao vetor de médias variacionais associado à distribuição Gaussiana nos pontos de indução, \mathbf{m} , foi utilizado um vetor unitário de ordem M . Já os valores iniciais da matriz de variâncias de covariâncias variacional de dimensão $M \times M$, \mathbf{S} , foram obtidos utilizando o modelo de co-

variância exponencial quadrático, expressão (2.49), considerando os valores iniciais de α e γ descritos anteriormente. A função de intensidade estimada para a configuração pontual de acidentes, bem como seu respectivo intervalo com 95% de credibilidade são mostrados na Figura 5.2.

Figura 5.2 – Gráfico da estimativa da função de intensidade dos registros de acidentes ao longo do trecho da BR-381 (entre Extrema e Betim), durante janeiro de 2017 à setembro de 2021, considerando apenas os acidentes cuja a causa principal tenha sido por “velocidade incompatível” e tenham ocorrido em situações de chuva, chuvisco, garoa ou granizo. A faixa na cor cinza em torno da estimativa de intensidade mostra o intervalo de credibilidade com 95% de probabilidade.



Fonte: Do autor (2021).

Com base no gráfico na Figura 5.2, observou-se que a função de intensidade estimada apresentou uma característica não homogênea em seu comportamento, na qual em alguns trechos a taxa de ocorrência se mostrou mais elevada do que em outros. Verificou-se que nos primeiros 33 quilômetros de distância aproximadamente, entre os municípios de Extrema e de Camanducaia, houve um aumento substancial na intensidade de acidentes na rodovia. Após atingir seu nível máximo no município de Camanducaia, a função de intensidade apresentou um comportamento de queda. A partir do município de Pouso Alegre, a função de intensidade apresentou um comportamento de estabilidade, com taxas de acidente relativamente baixas, que perpetuou-se até o município de Campanha. Mais adiante, a intensidade de acidentes na rodovia apresentou um novo aumento, porém menos expressivo em relação ao trecho inicial, a partir do município de Três Corações e que perdurou até, aproximadamente, o município de Oliveira. Nesse segundo trecho em destaque, é possível verificar que a função de intensidade

das ocorrências atinge seu máximo dentro dos limites do município de Carmo da Cachoeira, trecho bastante conhecido por conter muitas curvas sinuosas. Novamente, as taxas voltam a crescer bruscamente a partir do município de Itaguara, atingindo seu ápice entre Rio Manso e Brumadinho. De acordo com o modelo estimado, esse trecho foi considerado o de maior risco em relação a todo o percurso da rodovia.

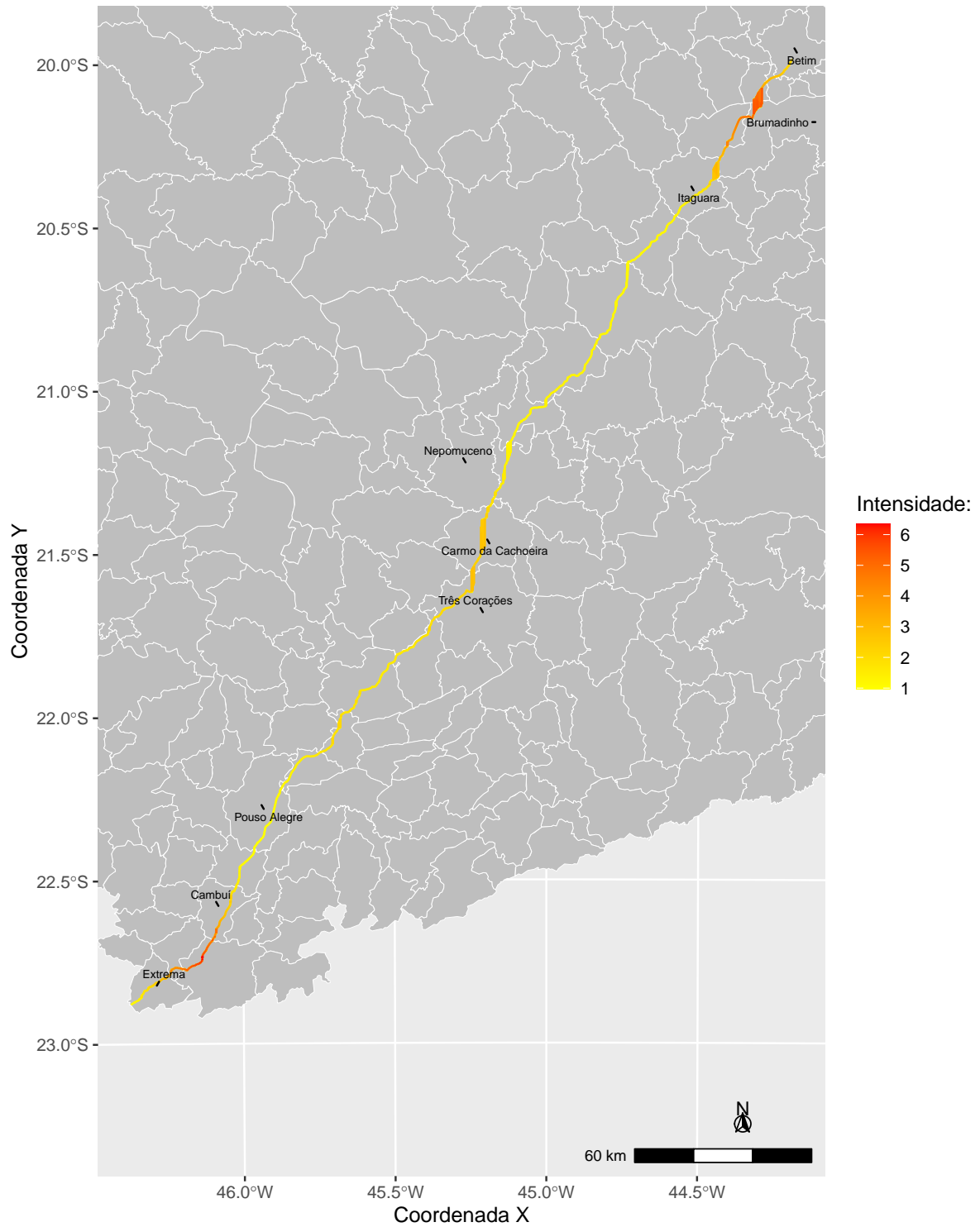
A Figura 5.3 mostra a projeção da mesma intensidade estimada, discutida anteriormente e apresentada na Figura 5.2, no trecho da rodovia BR-381, entre os municípios de Extrema e de Betim. Para isso foi criado um mapa de calor ao longo do trajeto em estudo, de tal modo que áreas mais “quentes”, com cores mais avermelhadas, refletem intensidades mais altas.

Uma das principais causas de acidentes está associada a imprudências cometidas por condutores desses veículos. Este fato já foi discutido por Freitas (2013), que destaca a influência do ser humano como um dos principais causadores para que um acidente, de fato, aconteça. Dentre as várias causas que podem ser classificadas como imprudência, o excesso de velocidade tem sido a causa de maior destaque, uma vez que o uso inadequado da velocidade nestas vias aumentam diretamente o risco de acidente e, conseqüentemente, a chance de uma pessoa vir a óbito (WORLD HEALTH ORGANIZATION, 2018).

Outro fator importante que pôde ser notado neste estudo foi a influência das condições meteorológicas nas ocorrências de acidentes. Para dias de chuva, o risco de acidentes tende a se agravar, principalmente quando a causa se deve ao excesso de velocidade permitido na rodovia. Especificamente para o conjunto de dados analisados neste estudo, percebeu-se o impacto que as condições meteorológicas desfavoráveis tem nos acidentes por incompatibilidade de velocidade, uma vez que 45% do total dos acidentes registrados nesse trecho ocorreram em situações adversas (chuva, chuvisco, granizo ou garoa). Em geral, veículos de grande porte passam longas horas nessas estradas, o que aumenta a probabilidade de apresentarem problemas mecânicos, como vazamentos de combustível nas pistas. Associando esse problema, a presença de chuvas, o cansaço dos motoristas e o excesso de velocidade nessas rodovias, também aumentam a probabilidade de potencializar o risco dos acidentes. Outros fatores também podem ser cruciais para o aumento da gravidade desses acidentes, como o consumo de bebidas alcoólicas e drogas, como destaca Bringmann et al. (2014).

De acordo com os resultados obtidos, basicamente três trechos apresentaram intensidades elevadas em diferentes situações meteorológicas e fases do dia, sendo considerados potenciais trechos de risco. São eles os percursos entre Extrema e Estiva, Três Corações e Oliveira e

Figura 5.3 – Mapa de calor da projeção da função de intensidade estimada no trecho da rodovia federal BR-381 (entre Extrema e Betim). Áreas mais quentes (mais avermelhadas) representam áreas de maior intensidade.



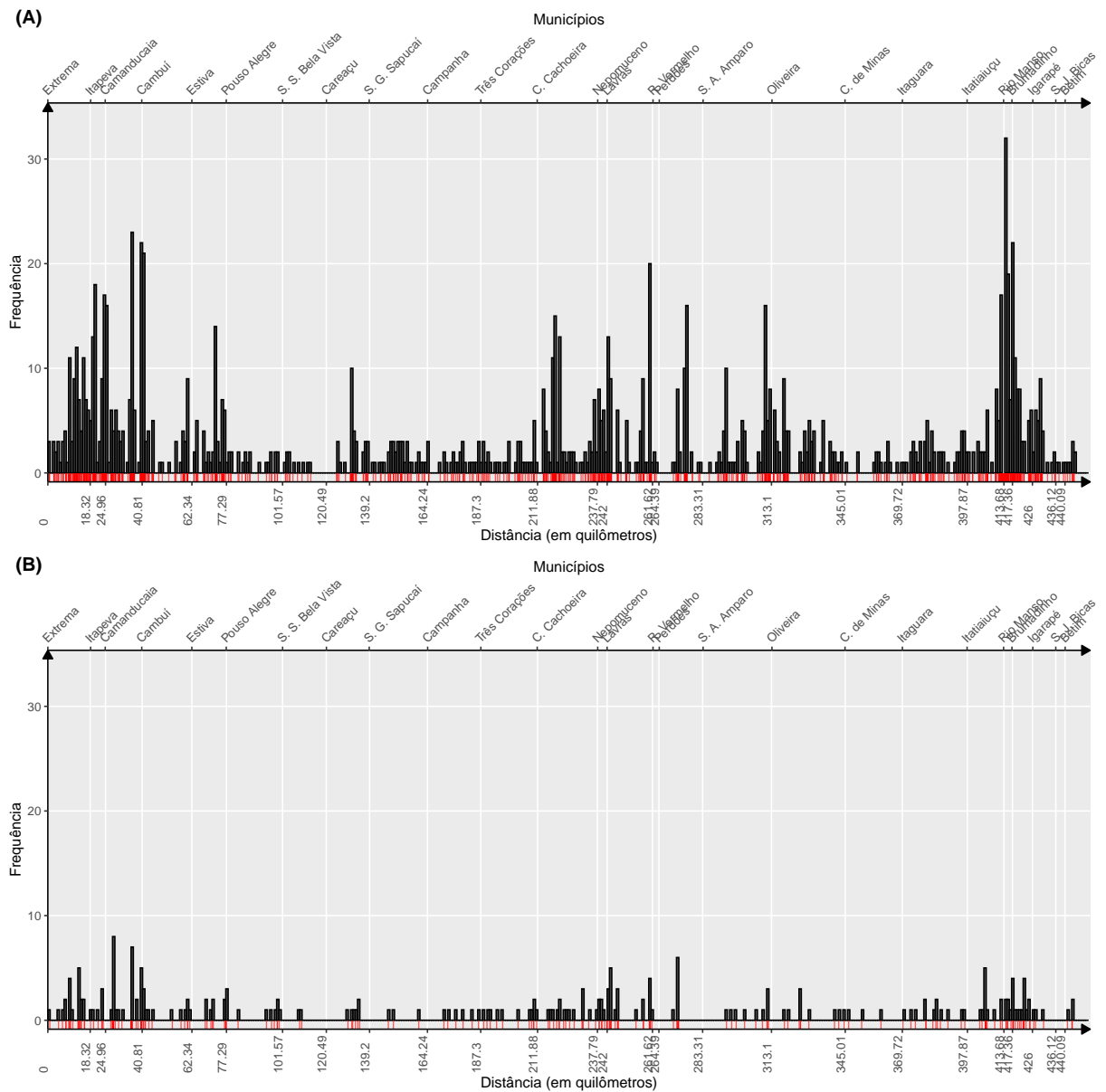
Fonte: Do autor (2021).

Itaguara e Betim. Estes trechos são considerados bastante perigosos, uma vez que elas apresentam curvas bastante sinuosas, o que aliado ao excesso de velocidade, aumentam as chances de que um acidente ocorra. Com base nas informações associadas aos dados em análise, observou-se que grande parte dos registros de acidentes nestes trechos em decorrência da velocidade excedida ocorreram em trajetos da via compostos por curvas. Para o primeiro trecho, ocorreram 302 acidentes em curvas, o que corresponde a 72,4% das ocorrências totais registradas. Para o segundo, observou-se 299 acidentes (67,5%), ao passo que no terceiro trecho, foram notificadas 264 ocorrências (77,6%). Diversos estudos relatam que a condição geométrica de uma rodovia pode estar associada aos acidentes que ocorrem nela (ANDRIOLA; TORRES; GARCÍA, 2019; CARMO; JUNIOR, 2019; OLIVEIRA; BORGES; GODINHO, 2018). Isto é, trechos mais sinuosos tendem a aumentar a probabilidade de acidentes nas rodovias. Desse modo, os resultados obtidos no presente estudo se mostra coerente com o que vem sendo discutido na literatura.

Considera-se agora uma configuração pontual marcada, sendo essas marcas de natureza qualitativa e composta por $\mathcal{K} = 2$ categorias. Os eventos representam agora ocorrências de acidentes por incompatibilidade de velocidade em um trecho da rodovia BR-381 associados à um determinado tipo de veículo, em que o veículo classificado é como sendo de grande ou de pequeno porte. O número de observações de ocorrências de acidentes envolvendo veículos de pequeno porte foi de $N_1 = 1210$ ao passo que, para veículos de grande porte, o total de acidentes foi de $N_2 = 225$. Vale ressaltar que a base de dados é a mesma utilizada no estudo feito anteriormente, na qual desconsiderou o tipo de veículo envolvido no acidente. À princípio, observa-se que o número de eventos considerando a marca “tipo de veículo” é maior do que o número de registros de acidentes usado no estudo anterior. Isso se deve ao fato de que uma determinada ocorrência pode envolver veículos de pequeno e de grande porte simultaneamente. Nesse caso, a ocorrência é contabilizada duas vezes, sendo uma para cada tipo de categoria da marca. A Figura 5.4 mostra o histograma da frequência de acidentes por quilômetro ao longo do trecho da rodovia em estudo, considerando o tipo de veículo envolvido no acidente, sendo a Figura 5.4 (A) o histograma para acidentes com veículos de pequeno porte e a Figura 5.4 (B) o histograma considerando os veículos de grande porte.

Com base nos histogramas da Figura 5.4, observou-se que os acidentes envolvendo veículos de pequeno porte é muito mais frequente do que os acidentes de veículos de grande porte. Ao comparar os dois histogramas, percebeu-se que ambos apresentaram o mesmo padrão de

Figura 5.4 – Histograma com a frequência absoluta do número de ocorrências de acidentes envolvendo veículos de pequeno porte (A) e grande porte (B) por quilômetro em um trecho da rodovia Fernão Dias (BR-381) entre os municípios de Extrema e Betim. Os traços em vermelho, |, representam as posições exatas (distância em quilômetros) de cada ocorrência.



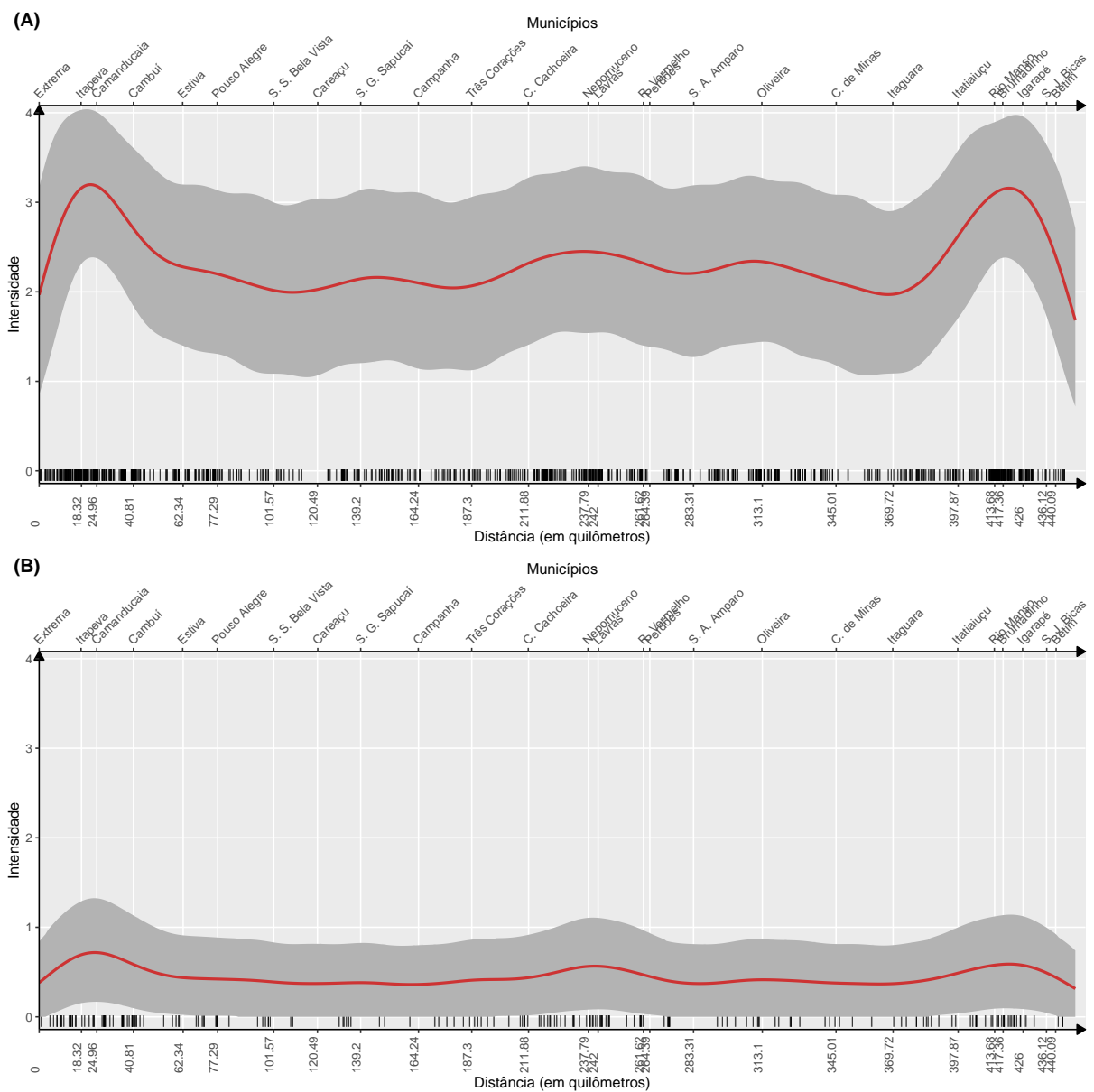
Fonte: Do autor (2021).

ocorrência, isto é, ambos tiveram mais ocorrências de acidentes praticamente nos mesmos trechos da rodovia. Esse comportamento indica uma forte evidência de que ambas as configurações pontuais podem estar correlacionadas positivamente. Ou seja, o aumento no número de acidentes envolvendo um tipo de veículo para um determinado trecho da rodovia pode estar associado ao aumento de ocorrências envolvendo o outro tipo de veículo.

Com base nisso, foi ajustado à esses dados de acidentes o modelo desenvolvido anteriormente considerando o porte do veículo envolvido no acidente como sendo a marca que rotula o evento da configuração pontual. Para isso, considerou-se novamente o algoritmo BFGS como otimizador necessário para realizar a inferência variacional do modelo. Além disso, considerou-se $M = 20$ pontos de indução e, para os valores iniciais dos hiperparâmetros, utilizou-se $\alpha = 350$, $\rho_{12} = 0,75$, $\gamma_1 = 0,1$, $\gamma_2 = 0,015$, $\bar{u}_1 = 1,5$ e $\bar{u}_2 = 0,5$. Para os hiperparâmetros variacionais associados aos pontos de indução, considerou-se um vetor de médias constante com os valores de 1,5 e 0,5 para \mathbf{m}_1 e \mathbf{m}_2 , respectivamente. Já os valores iniciais das matrizes de covariâncias variacionais cruzadas, \mathbf{S}_{ij} para $i, j = 1, 2$, foram obtidos utilizando a função de covariância exponencial quadrática com os mesmos valores iniciais dos hiperparâmetros do modelo, α , ρ_{12} , γ_1 e γ_2 .

Baseado nas Figuras 5.5(A) e 5.5(B), observou-se que, de fato, a função de intensidade para os acidentes envolvendo veículos de pequeno porte no trecho em estudo foi superior à intensidade de acidentes envolvendo veículos de grande porte. Foi possível verificar também que o modelo conseguiu detectar os trechos em que houve uma maior taxa de acidentes e, além disso, corroborou com os resultados obtidos por meio do ajuste do modelo de Lloyd et al. (2015), detectando basicamente os mesmos trechos: entre os municípios de Extrema e Camanducaia; Três Corações e Oliveira; e no trecho final do percurso da rodovia, entre Itaguara e Betim. No entanto, verificou-se que o modelo apresentou pouca flexibilidade em modelar possíveis oscilações no comportamento de queda das funções de intensidade, quando comparado com o resultado usando modelo de Lloyd et al. (2015), na Figura 5.2. Uma possível causa que poderia justificar tal comportamento se deve ao fato de que o modelo de covariância cruzado, expressões (4.5), (4.6) e (4.7), compartilham de um mesmo parâmetro de escala, α , para descrever a correlação de ambos os processos Gaussianos. Desse modo, limitar o modelo à ter um mesmo parâmetro α para ambos os processos poderia resultar em intensidades menos flexíveis. No entanto, um estudo mais aprofundado precisaria ser realizado para analisar a influência dessa suposição feita ao modelo.

Figura 5.5 – Estimativas das funções de intensidade dos registros de acidentes em trecho da BR-381, entre os municípios de Extrema e Betim, durante janeiro de 2017 à setembro de 2021, considerando os acidentes envolvendo veículos de pequeno porte (A) e grande porte (B), cuja a causa principal tenha sido por “velocidade incompatível” ocorridos em situações de chuva, chuvisco, garoa ou granizo. A faixa na cor cinza em torno das estimativas descrevem os intervalos de credibilidade com 95% de probabilidade.

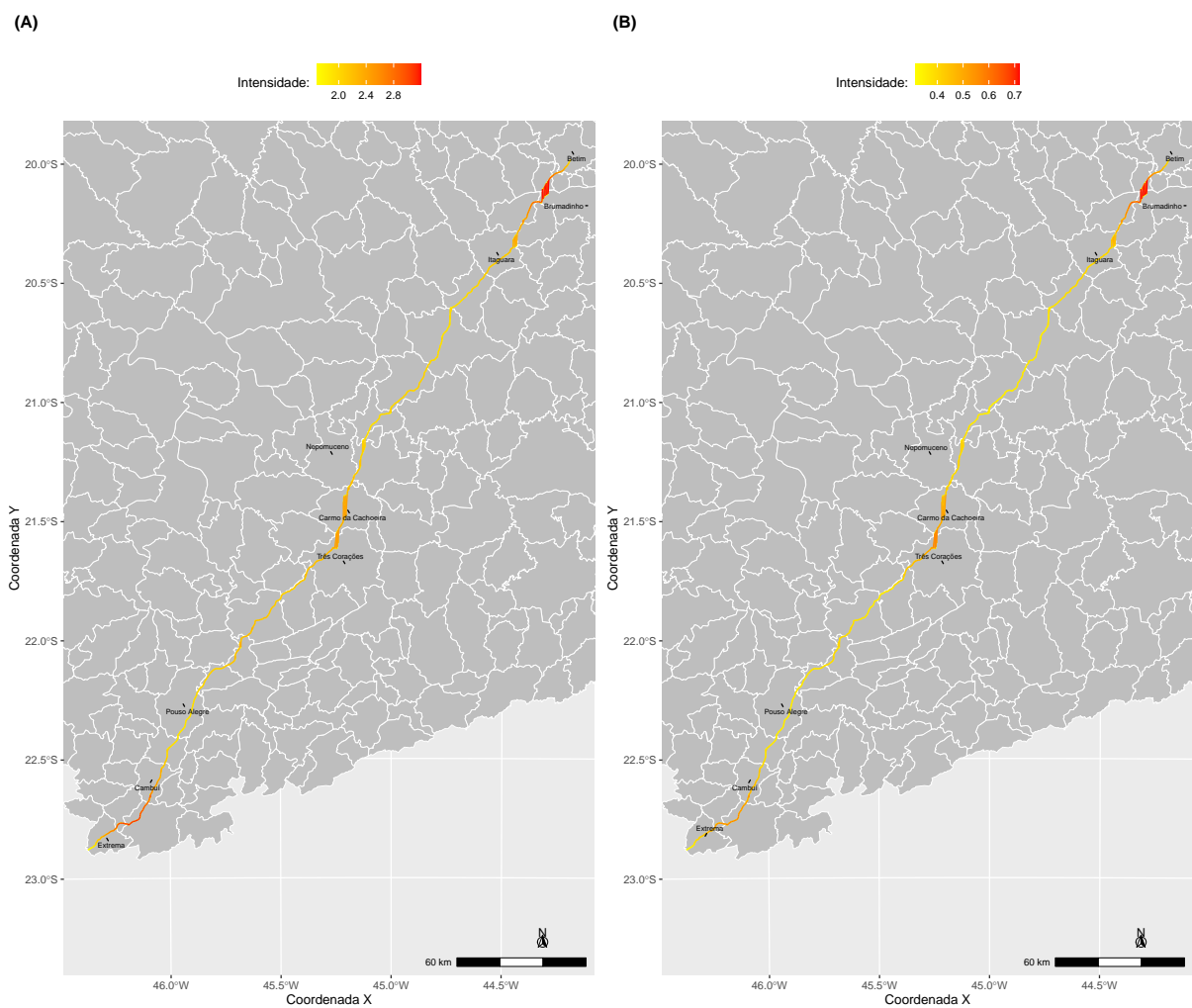


Fonte: Do autor (2021).

Outra característica que pôde ser verificada foi que ambas as funções de intensidade apresentaram um padrão similar em detectar trechos da rodovia com altas intensidades. Isso se deve ao fato dos processos Gaussianos terem apresentado uma estimativa para ρ_{12} de 0,67, aproximadamente. Esse resultado revelou uma correlação positiva entre os processos Gaussianos e, conseqüentemente, entre as funções de intensidade para cada porte de veículo. No entanto, não se pode afirmar se o aumento na intensidade de acidentes envolvendo um determinado tipo de veículo influenciou no aumento da intensidade de acidentes para o outro tipo de veículo, pois a correlação existente não reflete a relação de causa e efeito entre essas quantidades aleatórias. Além disso, outros fatores externos podem também estar influenciando nas intensidades de acidentes das categorias das marcas, como por exemplo as características que são inerentes à rodovia.

A Figura 5.6 apresenta a projeção das intensidades estimadas no trecho da rodovia BR-381, entre os municípios de Extrema e de Betim, para cada porte de veículos envolvidos nos acidentes. Da mesma forma como apresentado nos resultados usando o modelo de Lloyd et al. (2015), um mapa de calor ao longo do trajeto em estudo foi novamente gerado, a fim de evidenciar os trechos da rodovia com maior risco de acidentes.

Figura 5.6 – Projeção das funções de intensidades estimadas para as ocorrências de acidentes no trecho da rodovia federal BR-381, entre Extrema e Betim, considerando o porte do(s) veículo(s) envolvido(s), sendo (A) a intensidade de acidentes para veículos de pequeno porte e (B) para veículos de grande porte.



Fonte: Do autor (2021).

6 LIMITAÇÕES DO MODELO E FUTUROS TRABALHOS

Com base na proposta de modelagem desenvolvida nessa tese e na aplicação feita aos dados de acidentes na rodovia federal brasileira BR-381, é possível discutir alguns pontos em relação às possíveis limitações que o modelo apresenta.

O primeiro ponto a ser discutido é a informação da indexação do processo estocástico. Muitos fenômenos aleatórios que são tratados pela teoria de processos estocásticos, em geral, são observados com base em algum momento no tempo e em alguma posição espacial. Em alguns casos, estudos de tais fenômenos podem ser feitos separadamente de acordo com a indexação, isto é, uma análise puramente espacial e outra puramente temporal. No entanto, em algumas situações, ambas as informações de indexação do processo estocástico são importantes e precisam ser levadas em consideração no estudo para que o comportamento aleatório presente no fenômeno possa ser compreendido de maneira clara e que sua estimativa expresse seu real comportamento. O modelo de Lloyd et al. (2015), bem como sua extensão para marcas categóricas desenvolvida nessa tese, levam em consideração apenas uma única informação de indexação. Considerando a aplicação do modelo proposto nessa tese em dados de acidentes, é possível verificar que esse estudo de fato necessita não apenas de uma informação de indexação georeferenciada (ou, de maneira análoga, a transformação feita por meio de distâncias), mas também da informação do momento exato em que o acidente ocorreu. Ambas as informações se fazem necessárias no estudo de modelagem em acidentes uma vez que, a intensidade de acidentes envolvendo veículos de uma determinada categoria da marca (pequeno ou grande porte) tende a influenciar na intensidade de novos acidentes com veículos de outra categoria da marca, para curtas distâncias de tempo e de espaço. Desse modo, para que o modelo consiga estimar de forma adequada essa influência, a função correlação cruzada entre os processos Gaussianos nas expressões (4.5), (4.6) e (4.7) precisam levar em consideração tanto as defasagens temporais quanto as defasagens referenciadas baseadas na distância. A falta da informação de tempo pode levar a um viés nas estimativas dos hiperparâmetros do modelo, principalmente na estimativa de correlação cruzada entre os processos Gaussianos, ρ_{ij} , pois o modelo pode estar correlacionando fortemente dois eventos que estejam próximos espacialmente, mas com uma defasagem tempo muito grande.

Replicações de um processo estocástico, apesar de não serem frequentes, também podem ocorrer. Assim como diversas amostras de uma variável aleatória podem ser obtidas, di-

versas realizações de um processo aleatório também podem ser passíveis de serem observadas. No entanto, isso raramente ocorre para alguns fenômenos aleatórios, uma vez que é esperado que as diferentes amostras sejam independentes umas das outras. A configuração pontual que descreve os acidentes em rodovias apresenta uma característica interessante que atende a esse enfoque. As ocorrências de acidentes ao longo da rodovia em um determinado dia, à princípio, não irá influenciar em novas ocorrências para o dia seguinte. A dependência estocástica temporal presente entre os eventos de acidente existe para curtos períodos de tempo. Ou seja, a ocorrência de um acidente na rodovia pode aumentar a probabilidade de novos acidentes nos próximos segundos ou minutos, mas não terá impacto em possíveis acidentes que ocorrerão nos dias subsequentes. Dessa forma, pode-se pensar que o conjunto de acidentes ocorridos para cada dia representa uma replicação, ou seja, uma amostra do processo estocástico pontual que gera esses eventos. No entanto, o modelo desenvolvido na presente tese não abrange esse tipo de informação.

Deve-se ressaltar que, para a presente tese, o modelo foi desenvolvido unicamente para o ajuste de intensidades envolvendo configurações pontuais indexadas na reta real, \mathbb{R}^1 . Porém, é possível estender o modelo para processos pontuais com indexação em dimensões mais elevadas, assim como foi apresentado por Lloyd et al. (2015).

No que se refere ao número de categorias de uma marca, o modelo desenvolvido se restringe à modelagem de intensidade para configurações pontuais marcadas envolvendo apenas $\mathcal{K} = 2$ categorias, dado a complexidade dos cálculos envolvidos na sua inferência. No entanto, muitos fenômenos aleatórios de natureza pontual podem não satisfazer essa condição, tornando o modelo bastante restrito para a representação de fenômenos na prática.

Com base no que foi exposto anteriormente, estrutura e limitações que o modelo apresenta, é possível propor diversas extensões na estrutura de modelagem para trabalhos futuros. Dentre essas possíveis propostas de extensão, destacam-se: a inclusão da informação “tempo” como indexador adicional do processo, levando à um processo pontual com dupla indexação; atribuição de uma estrutura de modelagem para processos com múltiplas observações, isto é, um processo pontual replicado com marcas; uma estrutura de modelagem para processos pontuais com marcas, indexados em qualquer dimensão \mathbb{R}^d , para $d \geq 1$; um estudo envolvendo a modelagem da intensidade para outras funções de ligação diferentes da quadrática entre o processo de intensidade e o processo Gaussiano, bem como um estudo com outras funções de correlação para o processo Gaussiano, além da que foi utilizada nessa tese.

7 CONCLUSÃO

O objetivo dessa tese foi propor um novo esquema de modelagem para configurações pontuais marcadas na qual a indexação pudesse ser qualquer conjunto pertencente aos reais unidimensionais, seja ele o tempo ou não. A proposta surgiu como uma extensão do modelo *Variational Bayes for point process*, proposto por Lloyd et al. (2015), considerando uma marca de natureza qualitativa nominal, composta por duas categorias.

Como contribuição, o modelo proposto considerou uma estrutura de processos Gaussianos multivariados associados ao vetor de intensidades do processo de Cox, possibilitando assim que as intensidades associadas a cada atributo da marca pudessem estar correlacionadas entre si. Com base nos resultados obtidos, foi possível verificar que o modelo proposto conseguiu estimar a intensidade de configurações pontuais marcadas, a partir de um estudo de aplicação em um conjunto de dados reais sobre ocorrências de acidentes em rodovias federais brasileiras. Além disso, a proposta dessa tese sugere que novos avanços metodológicos possam ser feitos, a fim de que o modelo consiga descrever um número muito maior de fenômenos aleatórios de natureza pontual.

REFERÊNCIAS

- ADAMS, R. P.; MURRAY, I.; MACKAY, D. J. C. Tractable nonparametric bayesian inference in poisson processes with gaussian process intensities. In: **Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09**. New York: ACM Press, 2009. p. 9–16.
- ALBUQUERQUE, J. P. d. A. e.; FORTES, J. M. P.; FINAMORE, W. A. **Probabilidade, Variáveis Aleatórias e Processos Estocásticos**. Rio de Janeiro: Interciência, 2008. ISBN 9788571931909.
- ANDRIOLA, C. L.; TORRES, T. B.; GARCÍA, D. S. P. Influência dos parâmetros geométricos de rodovias na frequência e severidade de acidentes viários em curvas: o caso da BR 116. **TRANSPORTES**, Lepidus Tecnologia, v. 27, n. 3, p. 1–14, nov 2019.
- BADDELEY, A.; BÁRÁNY, I.; SCHNEIDER, R. Spatial point processes and their applications. In: WEIL, W. (Ed.). **Stochastic Geometry**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. cap. 1, p. 1–75. ISBN 978-3-540-38175-4.
- BANERJEE, S. et al. Gaussian predictive process models for large spatial data sets. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley, v. 70, n. 4, p. 825–848, set. 2008.
- BARTLETT, M. S. The spectral analysis of two-dimensional point processes. **Biometrika**, [Oxford University Press, Biometrika Trust], v. 51, n. 3/4, p. 299–311, 1964. ISSN 00063444. Disponível em: <<http://www.jstor.org/stable/2334136>>.
- BASU, A. K. **Introduction to stochastic processes**. Pangbourne, England: Alpha Science International, 2003. ISBN 9781842651056.
- BERNARDO, J. M.; SMITH, A. F. M. **Bayesian Theory**. New York: John Wiley & Sons, 1994. ISBN 0-471-92416-4.
- BISHOP, C. M. **Pattern Recognition and Machine Learning**. New York: Springer-Verlag New York Inc., 2011. ISBN 978-0-387-31073-2.
- BLEI, D. M.; KUCUKELBIR, A.; MCAULIFFE, J. D. Variational inference: A review for statisticians. **Journal of the American Statistical Association**, Informa UK Limited, v. 112, n. 518, p. 859–877, apr 2017.
- BOLSTAD, W. M. **Introduction to Bayesian Statistics**. New Jersey: John Wiley & Sons, Inc., 2004. ISBN 9780471722120.
- BRINGMANN, P. B. et al. A pattern of adult involvement in highway acidentes. **Ciência & Saúde Coletiva**, v. 19, n. 12, p. 4861–4868, dec 2014.
- BRIX, A.; DIGGLE, P. J. Spatiotemporal prediction for log-Gaussian Cox processes. **Journal of the Royal Statistical Society. Series B**, v. 63, n. 4, p. 823–841, 2001.
- CALEFFI, F. et al. Influência das condições climáticas e de acidentes na caracterização do comportamento do tráfego em rodovias. **TRANSPORTES**, Lepidus Tecnologia, v. 24, n. 4, p. 57, dec 2016.

- CARMO, C. L. do; JUNIOR, A. A. R. Segurança em rodovias inseridas em áreas urbanas na região sul do Brasil. *FapUNIFESP (SciELO)*, v. 11, 2019.
- CHIU, S. N. et al. **Stochastic Geometry and its Applications**. 3. ed. Chichester: John Wiley & Sons, Inc., 2013.
- CHONG, E. K. P.; ZAK, S. H. **An introduction to optimization**. 2. ed. New York: Wiley, 2001. ISBN 978-0-471-65400-1.
- CONGDON, P. **Applied Bayesian modelling**. Chichester, West Sussex, England Hoboken, NJ: Wiley, 2003. ISBN 0-471-48695-7.
- COX, D. R. Some statistical methods connected with series of events. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, v. 17, n. 2, p. 129–164, 1955. ISSN 00359246.
- COX, D. R.; ISHAM, V. **Point Processes**. New York: Chapman and Hall/CRC, 1980. ISBN 0412219107.
- CRESSIE, N. A. **Statistics for spatial data**. Hoboken: John Wiley & Sons, 1993.
- DALEY, D. J.; VERE-JONES, D. **An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods**. 2. ed. New York: Springer Science & Business Media, 2003.
- DIGGLE, P. J. et al. Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm. **Statistical Science**, v. 28, n. 4, p. 542–563, 2013.
- DIGGLE, P. J.; ROWLINGSON, B.; SU, T. li. Point process methodology for on-line spatio-temporal disease surveillance. **Environmetrics**, v. 16, n. 5, p. 423–434, 2005.
- DOBROW, R. P. **Introduction to Stochastic Processes with R**. New Jersey: John Wiley & Sons, 2016. ISBN 978-1-118-74065-1.
- DU, N. et al. Recurrent marked temporal point processes: Embedding event history to vector. In: **Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: ACM, 2016. (KDD '16), p. 1555–1564. ISBN 978-1-4503-4232-2. Disponível em: <<http://doi.acm.org/10.1145/2939672.2939875>>.
- FREITAS, J. P. P. **Análise dos acidentes de trânsito envolvendo crianças na cidade de Uberlândia: subsídios para as políticas públicas**. Tese (Doutorado em Ciências Humanas) — Universidade Federal de Uberlândia, Uberlândia, dez. 2013. Disponível em: <<https://repositorio.ufu.br/handle/123456789/15972>>.
- GELFAND, A. E.; SMITH, A. F. M. Sampling-based approaches to calculating marginal densities. **Journal of the American Statistical Association**, Informa UK Limited, v. 85, n. 410, p. 398–409, jun 1990.
- GELMAN, A. et al. **Bayesian Data Analysis**. 3. ed. Boca Raton: Chapman and Hall/CRC, 2014. ISBN 9781439840955.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Institute of Electrical and Electronics Engineers (IEEE), PAMI-6, n. 6, p. 721–741, nov 1984.

GERSHMAN, S.; HOFFMAN, M.; BLEI, D. **Nonparametric variational inference**. 2012.

GHOSH, J. K.; DELAMPADY, M.; SAMANTA, T. **An introduction to Bayesian Analysis: Theory and Methods**. New York: Springer Science & Business Media, 2007. ISBN 9780387354330.

HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, Biometrika Trust, v. 57, n. 1, p. 97–109, 1970.

HOFFMAN, M. D. et al. Stochastic variational inference. **Journal of Machine Learning Research**, v. 14, n. 4, p. 1303–1347, 2013. Disponível em: <<http://jmlr.org/papers/v14/hoffman13a.html>>.

JOHNSON, D. H. Point process models of single-neuron discharges. **Journal of Computational Neuroscience**, Springer Nature, v. 3, n. 4, p. 275–299, dec 1996. ISSN 1573-6873.

KELSALL, J.; WAKEFIELD, J. Modeling spatial variation in disease risk: A geostatistical approach. **Journal of the American Statistical Association**, [American Statistical Association, Taylor & Francis, Ltd.], v. 97, n. 459, p. 692–701, 2002. ISSN 01621459. Disponível em: <<http://www.jstor.org/stable/3085705>>.

KINGMAN, J. F. C. **Poisson Processes**. New York: Oxford University Press, 1996. ISBN 0198536933.

LIANG, S.; CARLIN, B. P.; GELFAND, A. E. Analysis of minnesota colon and rectum cancer point patterns with spatial and nonspatial covariate information. **The Annals of Applied Statistics**, NIH Public Access, v. 3, n. 3, p. 943 – 962, out. 2009.

LLOYD, C. et al. Variational inference for gaussian process modulated poisson processes. In: BACH, F.; BLEI, D. (Ed.). **Proceedings of the 32nd International Conference on Machine Learning**. Lille, France: PMLR, 2015. (Proceedings of Machine Learning Research, v. 37), p. 1814–1822. Disponível em: <<http://proceedings.mlr.press/v37/lloyd15.html>>.

LUNDBERG, O. **On random processes and their application to sickness and accident statistics**. Tese (Doutorado) — Almqvist & Wiksell, 1940.

Mesquita Filho, M.; CARVALHO, C. R. D.; GARCIA, E. D. P. Fatores associados à ocorrência de acidentes de trânsito entre universitários. **Ciência & Saúde**, EDIPUCRS, v. 10, n. 2, p. 62, may 2017. Acesso em: 04 abr. 2018.

METROPOLIS, N. et al. Simulated annealing. **Journal of Chemical Physics**, v. 21, p. 1087–1092, 1953.

MINISTÉRIO DA SAÚDE. **Departamento de Informática do SUS (DATASUS)**. 2018. Acesso em: 07 abr. 2018. Disponível em: <<http://datasus.saude.gov.br/>>.

MØLLER, J. Shot noise cox processes. **Advances in Applied Probability**, Applied Probability Trust, v. 35, n. 3, p. 614–640, 2003. ISSN 00018678. Disponível em: <<http://www.jstor.org/stable/1428211>>.

MØLLER, J.; SYVERSVEEN, A. R.; WAAGEPETERSEN, R. P. Log-Gaussian Cox processes. **Scandinavian journal of statistics**, Wiley Online Library, v. 25, n. 3, p. 451–482, 1998.

MØLLER, J.; TORRISI, G. L. Generalised shot noise cox processes. **Advances in Applied Probability**, Applied Probability Trust, v. 37, n. 1, p. 48–74, 2005. ISSN 00018678. Disponível em: <<http://www.jstor.org/stable/30037315>>.

MORETTIN, P.; TOLOI, C. M. C. **Análise de séries temporais**. 2. ed. São Paulo: Edgard Blucher, 2006. ISBN 9788521203896.

NOCEDAL, J.; WRIGHT, S. **Numerical Optimization**. 2. ed. New York: Springer, 2006. ISBN 1431-8598.

OLIVEIRA, M. da P. B.; BORGES, P. R. H.; GODINHO, D. P. Uma visão crítica sobre a relação:acidentes e projeto geométrico rodoviário, estudo na BR-381. **Revista Petra**, v. 4, n. 1, p. 114–132, 2018. ISSN 2359-5302.

PAULINO, C. D. M.; TURKMAN, M. A. A.; MURTEIRA, B. **Estatística Bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2003. ISBN 972-31-1043-1.

POLÍCIA RODOVIÁRIA FEDERAL. **Acidentes - Portal de Dados Abertos**. 2021. Disponível em: <<https://www.gov.br/prf/pt-br/aceso-a-informacao/dados-abertos/dados-abertos-acidentes>>.

PRESS, S. J. **Subjective and Objective Bayesian Statistics**. 2. ed. Hoboken: John Wiley & Sons, Inc., 2003. ISBN 0-471-34843-0.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2021. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>.

RASMUSSEN, C. K. I. W. C. E. **Gaussian Processes for Machine Learning**. Massachusetts: MIT Press Ltd, 2006. ISBN 026218253X. Disponível em: <https://www.ebook.de/de/product/5192092/carl_edward_university_of_cambridge_rasmussen_christopher_k_i_university_of_edinburgh_williams_gaussian_processes_for_machine_learning.html>.

RASMUSSEN, J. G. **Aspects of temporal and spatio-temporal processes**. Tese (Doutorado) — Department of Mathematical Science, Aalborg University, Denmark, ago. 2006.

RASMUSSEN, J. G. Lecture notes: Temporal point processes and the conditional intensity function. **arXiv preprint arXiv:1806.00221**, 2018.

RIPLEY, B. D. **Spatial Statistics**. Hoboken, New Jersey: John Wiley & Sons, Inc., 1981.

ROBERT, C.; CASELLA, G. **Introducing Monte Carlo Methods with R**. New York: Springer, 2009.

RUSTAGI, J. S. **Variational Methods in Statistics**. New York: Elsevier Science & Techn., 1976. ISBN 0-12-604560-7.

SCHABENBERGER, O.; GOTWAY, C. A. **Statistical methods for spatial data analysis**. Boca Raton: Chapman and Hall/CRC, 2004.

SEARLE, S. R. **Matrix algebra useful for statistics**. New York: Wiley, 1982. ISBN 0-471-86681-4.

SILVEIRA, D. D. B. **Modelos de Tópicos baseados em Autocodificadores Variacionais utilizando as distribuições Gumbel-Softmax e mistura de Normais-Logísticas**. Tese (Mestrado em Informática) — Universidade Federal do Amazonas, Manaus, dez. 2018. Disponível em: <<https://tede.ufam.edu.br/handle/tede/7439>>.

TEIXEIRA, M. E. S.; CASTANHO, R. B.; NETO, F. A. A. As geotecnologias aplicadas ao mapeamento de acidentes de trânsito urbano. **Espaço em Revista**, v. 19, n. 1, p. 38–55, 2017. ISSN 1519-7816.

TERRIEN, C. **Probability and random processes for electrical and computer engineers**. 2. ed. Boca Raton, FL: CRC Press, 2012. ISBN 978-1439826980.

WORLD HEALTH ORGANIZATION. **Global Status Report on Road Safety 2015**. Geneva, 2015. Acesso em: 04 abr. 2018. Disponível em: <http://www.who.int/violence_injury_prevention/road_safety_status/2015/en/>.

WORLD HEALTH ORGANIZATION. **Global Status Report on Road Safety 2018**. Geneva, 2018. Acesso em: 04 dez. 2020. Disponível em: <http://www.who.int/violence_injury_prevention/road_safety_status/2015/en/>.

ZHANG, C. et al. Advances in variational inference. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Institute of Electrical and Electronics Engineers (IEEE), v. 41, n. 8, p. 2008–2026, aug 2019.

APÊNDICE A – Algoritmo Broyden–Fletcher–Goldfarb–Shanno (BFGS)

A otimização é um problema recorrente em estudos matemáticos e estatísticos, na qual se tem por objetivo conhecer o valor das variáveis onde uma determinada função atinge seu valor extremo, seja ele o máximo ou o mínimo. Isto é, a otimização se preocupa em minimizar ou maximizar uma função com base nos valores de suas variáveis. Matematicamente, o problema da otimização é dado por

$$\min_{\Theta} \mathcal{L}(\Theta),$$

ou, de maneira análoga,

$$\max_{\Theta} \mathcal{L}(\Theta),$$

sendo $\Theta = [\theta_1, \dots, \theta_P]^\top \in \mathbb{R}^P$ um vetor composto por $P \geq 1$ variáveis reais e $\mathcal{L} : \mathbb{R}^P \rightarrow \mathbb{R}$ uma função que associa um valor do vetor Θ a um valor real. Para o modelo de Lloyd et al. (2015), bem como o modelo proposto nessa tese, a função \mathcal{L} descreve o limite inferior da log-evidência, presente na divergência de KL entre $q(\mathbf{f}, \mathbf{u})$ e $p(\mathbf{f}, \mathbf{u} \mid \mathcal{S}, \Theta)$, ao passo que Θ é o conjunto de hiperparâmetros do modelo.

De acordo com Chong e Zak (2001), a maximização e a minimização de uma função podem ser vistas de maneira análoga, uma vez que maximizar a função $\mathcal{L}(\Theta)$ é a mesma coisa que minimizar a função $-\mathcal{L}(\Theta)$. Uma vez que o problema de otimização tratado nessa tese se resume em obter o valor máximo do limite inferior da log-evidência, os algoritmos aqui apresentados terão por finalidade maximizar $\mathcal{L}(\Theta)$.

Algoritmos de otimização se fazem necessários quando a função na qual se deseja maximizar for muito complexa ou o número de variáveis em Θ é muito grande, dificultando a obtenção de uma solução analítica. De acordo com Nocedal e Wright (2006), existem várias propostas de algoritmos de otimização na literatura, cada um com suas características e particularidades. A escolha de um algoritmo adequado para a solução do problema é muito importante, pois ele determinará a rapidez com que a otimização será alcançada.

Dentre os diversos algoritmos existentes, o método de Newton, também conhecido como Newton-Raphson, é uma das abordagens mais conhecidas. Esse algoritmo tem por objetivo obter uma aproximação quadrática para a função $\mathcal{L}(\Theta)$ usando expansão em série de Taylor de

ordem 2. Além disso, assume-se que essa função de aproximação, que será denotada por $\mathcal{J}(\Theta)$, é contínua e duas vezes diferenciável. Desse modo, define-se

$$\mathcal{J}(\Theta) = \mathcal{L}(\Theta^{(k)}) + (\Theta - \Theta^{(k)})^\top \nabla \mathcal{L}(\Theta^{(k)}) + \frac{1}{2}(\Theta - \Theta^{(k)})^\top \nabla^2 \mathcal{L}(\Theta^{(k)}) (\Theta - \Theta^{(k)}), \quad (1)$$

em que $\nabla \mathcal{L}(\Theta^{(k)}) = \left[\frac{\partial \mathcal{L}(\Theta^{(k)})}{\partial \theta_1}, \dots, \frac{\partial \mathcal{L}(\Theta^{(k)})}{\partial \theta_P} \right]^\top$ é o vetor gradiente da função $\mathcal{L}(\Theta)$ aplicado no ponto $\Theta^{(k)}$, composto pelas derivadas parciais de primeira ordem em relação ao conjunto de variáveis Θ da função. Já $\nabla^2 \mathcal{L}(\Theta^{(k)})$ é definida como a matriz Hessiana composta pelas derivadas parciais de segunda ordem da função \mathcal{L} relação às variáveis, aplicado em $\Theta^{(k)}$, sendo

$$\nabla^2 \mathcal{L}(\Theta^{(k)}) = \begin{bmatrix} \frac{\partial^2 \mathcal{L}(\Theta^{(k)})}{\partial \theta_1^2} & \frac{\partial^2 \mathcal{L}(\Theta^{(k)})}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \mathcal{L}(\Theta^{(k)})}{\partial \theta_1 \partial \theta_P} \\ \frac{\partial^2 \mathcal{L}(\Theta^{(k)})}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \mathcal{L}(\Theta^{(k)})}{\partial \theta_2^2} & \cdots & \frac{\partial^2 \mathcal{L}(\Theta^{(k)})}{\partial \theta_2 \partial \theta_P} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \mathcal{L}(\Theta^{(k)})}{\partial \theta_P \partial \theta_1} & \frac{\partial^2 \mathcal{L}(\Theta^{(k)})}{\partial \theta_P \partial \theta_2} & \cdots & \frac{\partial^2 \mathcal{L}(\Theta^{(k)})}{\partial \theta_P^2} \end{bmatrix}.$$

É possível verificar que a função de aproximação em (1) coincide exatamente com a função de interesse \mathcal{L} considerando $\Theta = \Theta^{(k)}$. Da mesma forma, $\nabla \mathcal{J}(\Theta^{(k)}) = \nabla \mathcal{L}(\Theta^{(k)})$ e $\nabla^2 \mathcal{J}(\Theta^{(k)}) = \nabla^2 \mathcal{L}(\Theta^{(k)})$. Sendo assim, otimizar a função $\mathcal{J}(\Theta)$ é muito mais fácil do que otimizar a função de interesse, $\mathcal{L}(\Theta)$.

Uma forma de garantir que a função atinja seu ponto de máximo, é garantindo que seu vetor gradiente seja um vetor nulo, isto é, $\nabla \mathcal{J}(\Theta) = \mathbf{0}$. Desse modo, derivando a função $\mathcal{J}(\Theta)$ em relação a Θ e igualando-o a $\mathbf{0}$, tem-se que

$$\begin{aligned} \mathbf{0} &= \nabla \mathcal{J}(\Theta) \\ &= \nabla \mathcal{L}(\Theta^{(k)}) + \nabla^2 \mathcal{L}(\Theta^{(k)}) (\Theta - \Theta^{(k)}). \end{aligned} \quad (2)$$

Tomando $\Theta = \Theta^{(k+1)}$ e garantindo que $\nabla^2 \mathcal{L}(\Theta^{(k)}) > 0$, a função \mathcal{J} atinge seu ponto ótimo em

$$\Theta^{(k+1)} = \Theta^{(k)} - \{\nabla^2 \mathcal{L}(\Theta^{(k)})\}^{-1} \nabla \mathcal{L}(\Theta^{(k)}), \quad (3)$$

em que $\{\nabla^2 \mathcal{L}(\Theta^{(k)})\}^{-1}$ é a inversa da matriz Hessiana.

De acordo com Chong e Zak (2001), apesar do algoritmo de Newton apresentar boas propriedades de convergência, o método pode apresentar falhas se a escolha do ponto inicial,

Θ^0 , estiver longe do ponto de solução. Além disso, o método pode não maximizar a função, fazendo com que $\mathcal{L}(\Theta^{(k+1)}) \leq \mathcal{L}(\Theta^{(k)})$. Para contornar esse problema, uma modificação do algoritmo é imposta a fim de garantir que a direção do gradiente de \mathcal{L} seja mantida na direção e subida, fazendo assim com que a função seja maximizada. Desse modo, a equação (3) é reescrita como

$$\Theta^{(k+1)} = \Theta^{(k)} - \vartheta_k \{\nabla^2 \mathcal{L}(\Theta^{(k)})\}^{-1} \nabla \mathcal{L}(\Theta^{(k)}), \quad (4)$$

em que

$$\vartheta_k = \arg \max_{\vartheta > 0} \mathcal{L}(\Theta^{(k)} - \vartheta \{\nabla^2 \mathcal{L}(\Theta^{(k)})\}^{-1} \nabla \mathcal{L}(\Theta^{(k)})) \quad (5)$$

sendo ϑ_k o tamanho do passo de subida em direção ao ponto de máximo que, a cada iteração do algoritmo, seu valor é atualizado a fim de buscar um tamanho ótimo de passo para que o método de Newton garanta que o ponto de máximo da função seja obtido.

No entanto, o método de Newton, apesar de apresentar rápida taxa de convergência, apresenta algumas desvantagens, sendo a principal delas a necessidade de se obter a matriz Hessiana, $\nabla^2 \mathcal{L}(\Theta^{(k)})$, e, conseqüentemente, a solução da equação (4), que pode ter um elevado custo computacional para ser calculada (NOCEDAL; WRIGHT, 2006).

Como forma de contornar esse problema, uma nova classe de métodos de otimização surgiu, denominada como *métodos quasi-Newton*, com a finalidade de obter a solução em (4) sem a necessidade de calcular a inversa da matriz Hessiana. Como alternativa, os métodos quasi-Newton se utilizam de uma aproximação de $\{\nabla^2 \mathcal{L}(\Theta^{(k)})\}^{-1}$ ao invés de utilizar sua verdadeira inversa (CHONG; ZAK, 2001). O cálculo da aproximação dessa matriz requer apenas o uso da avaliação dos gradientes e da própria função em cada iteração. E justamente por não haver mais a necessidade de se obter derivadas de segunda ordem da função \mathcal{L} é que os métodos quasi-Newton se mostram mais eficientes do que os métodos de Newton (NOCEDAL; WRIGHT, 2006).

Considerando a abordagem dos métodos quasi-Newton, a expressão (4) pode ser reescrita como

$$\Theta^{(k+1)} = \Theta^{(k)} - \vartheta_k H_k \nabla \mathcal{L}(\Theta^{(k)}), \quad (6)$$

sendo H_k uma matriz de dimensão $P \times P$ definida como uma aproximação da inversa da verdadeira matriz Hessiana.

Na literatura, muitas técnicas para aproximar a inversa da matriz Hessiana tem sido propostas, gerando diferentes novos algoritmos de otimização. Segundo Nocedal e Wright (2006), um dos mais populares algoritmos quasi-Newton foi proposto por Boyden, Fletcher, Goldfarb e Shanno em 1970, o qual recebeu o nome de fórmula BFGS, cuja aproximação da inversa da Hessiana é dada por:

$$H_{k+1} = H_k + \left(1 + \frac{\Delta\mathcal{L}^{(k)\top} H_k \Delta\mathcal{L}^{(k)}}{\Delta\mathcal{L}^{(k)\top} \Delta\Theta^{(k)}} \right) \frac{\Delta\Theta^{(k)} \Delta\Theta^{(k)\top}}{\Delta\Theta^{(k)\top} \Delta\mathcal{L}^{(k)}} - \frac{H_k \Delta\mathcal{L}^{(k)} \Delta\Theta^{(k)\top} + (H_k \Delta\mathcal{L}^{(k)} \Delta\Theta^{(k)\top})^\top}{\Delta\mathcal{L}^{(k)\top} \Delta\Theta^{(k)}}, \quad (7)$$

em que $\Delta\mathcal{L}^{(k)} = \nabla\mathcal{L}(\Theta^{(k+1)}) - \nabla\mathcal{L}(\Theta^{(k)})$ e $\Delta\Theta^{(k)} = \Theta^{(k+1)} - \Theta^{(k)}$. É possível verificar que a expressão (7) depende apenas da informação dos gradientes, isto é, das derivadas de primeira ordem da função $\mathcal{L}(\Theta)$ em relação a Θ . Desse modo, o método BFGS não necessita da utilização da matriz Hessiana.

Portanto, o algoritmo BFGS é dado da seguinte forma:

Algoritmo 1: Algoritmo BFGS

Entrada: Valor inicial para $\Theta^{(0)}$ e para H_0 ; Tolerância de convergência $\epsilon > 0$;

$k \leftarrow 0$;

while $\|\nabla\mathcal{L}(\Theta)\| > \epsilon$ **do**

1. Calcular:

$$d_k = -H_k \nabla\mathcal{L}(\Theta^{(k)})$$

2. Atualizar o valor de Θ :

$$\Theta^{(k+1)} = \Theta^{(k)} + \vartheta_k d_k,$$

sendo ϑ_k calculado pela expressão (5).

3. Atualizar a matriz de aproximação H :

$$H_{k+1} = H_k + \left(1 + \frac{\Delta\mathcal{L}^{(k)\top} H_k \Delta\mathcal{L}^{(k)}}{\Delta\mathcal{L}^{(k)\top} \Delta\Theta^{(k)}} \right) \frac{\Delta\Theta^{(k)} \Delta\Theta^{(k)\top}}{\Delta\Theta^{(k)\top} \Delta\mathcal{L}^{(k)}} - \frac{H_k \Delta\mathcal{L}^{(k)} \Delta\Theta^{(k)\top} + (H_k \Delta\mathcal{L}^{(k)} \Delta\Theta^{(k)\top})^\top}{\Delta\mathcal{L}^{(k)\top} \Delta\Theta^{(k)}}$$

4. $k \leftarrow k + 1$

end

APÊNDICE B – Alguns resultados importantes para o modelo de Lloyd et al. (2015)

Neste Apêndice são apresentados o desenvolvimento de alguns resultados de cálculos importantes e necessários para a compreensão da estrutura de modelagem e de inferência proposto por Lloyd et al. (2015).

B.1 Obtenção da expressão das entradas da matriz Ψ

Os elementos da matriz Ψ podem ser obtidos a partir da integração do produto de $K(z, s)$ e $K(s, z')$ em relação a s . Nesse caso, tem-se que

$$\begin{aligned}
 \Psi(z, z') &= \int_{\mathcal{T}} K(z, s) \times K(s, z') ds \\
 &= \int_{\mathcal{T}} \gamma^2 \times \exp \left\{ \frac{-(z-s)^2}{2\alpha} - \frac{(s-z')^2}{2\alpha} \right\} ds \\
 &= \gamma^2 \int_{\mathcal{T}} \exp \left\{ \frac{-z^2 + 2sz - s^2 - s^2 + 2sz' - z'^2}{2\alpha} \right\} ds \\
 &= \gamma^2 \int_{\mathcal{T}} \exp \left\{ \frac{-z^2 - z'^2 + 2s(z+z') - 2s^2}{2\alpha} \right\} ds \\
 &= \gamma^2 \int_{\mathcal{T}} \exp \left\{ \frac{-z^2 - z'^2 + 2zz' - 2zz' + 4s\bar{z} - 2s^2 + 2\bar{z}^2 - 2\bar{z}^2}{2\alpha} \right\} ds \\
 &= \gamma^2 \int_{\mathcal{T}} \exp \left\{ \frac{-(z-z')^2 - 2zz' + 2\bar{z}^2}{2\alpha} - \frac{2(s-\bar{z})^2}{2\alpha} \right\} ds \\
 &= \gamma^2 \int_{\mathcal{T}} \exp \left\{ \frac{-(z-z')^2 - 2zz' + \frac{2}{4}(z+z')^2}{2\alpha} - \frac{(s-\bar{z})^2}{\alpha} \right\} ds \\
 &= \gamma^2 \int_{\mathcal{T}} \exp \left\{ \frac{-(z-z')^2}{4\alpha} - \frac{(s-\bar{z})^2}{\alpha} \right\} ds \\
 &= \gamma^2 \exp \left\{ \frac{-(z-z')^2}{4\alpha} \right\} \int_{\mathcal{T}} \exp \left\{ -\frac{(s-\bar{z})^2}{\alpha} \right\} ds \\
 &= \gamma^2 \sqrt{\pi\alpha} \exp \left\{ \frac{-(z-z')^2}{4\alpha} \right\} \int_{\mathcal{T}_{min}}^{\mathcal{T}_{max}} \frac{1}{\sqrt{\pi\alpha}} \exp \left\{ -\frac{(s-\bar{z})^2}{\alpha} \right\} ds \\
 &= \gamma^2 \sqrt{\pi\alpha} \exp \left\{ \frac{-(z-z')^2}{4\alpha} \right\} \left[\Phi \left(\frac{\mathcal{T}_{max} - \bar{z}}{\sqrt{\frac{\alpha}{2}}} \right) - \Phi \left(\frac{\mathcal{T}_{min} - \bar{z}}{\sqrt{\frac{\alpha}{2}}} \right) \right] \\
 &= \gamma^2 \sqrt{\pi\alpha} \exp \left\{ \frac{-(z-z')^2}{4\alpha} \right\} \left[\frac{1}{2} + \frac{1}{2} \operatorname{erf} \left(\frac{\mathcal{T}_{max} - \bar{z}}{\sqrt{2}\sqrt{\frac{\alpha}{2}}} \right) - \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left(\frac{\mathcal{T}_{min} - \bar{z}}{\sqrt{2}\sqrt{\frac{\alpha}{2}}} \right) \right] \\
 &= -\gamma^2 \frac{\sqrt{\pi\alpha}}{2} \exp \left\{ \frac{-(z-z')^2}{4\alpha} \right\} \left[\operatorname{erf} \left(\frac{\bar{z} - \mathcal{T}_{max}}{\sqrt{\alpha}} \right) - \operatorname{erf} \left(\frac{\bar{z} - \mathcal{T}_{min}}{\sqrt{\alpha}} \right) \right],
 \end{aligned}$$

em que $\bar{z} = \frac{z+z'}{2}$ e $\Phi(x) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right)$, sendo $\operatorname{erf}(\cdot)$ denominado como a função erro de Gauss.

B.2 Derivadas parciais de \mathcal{L} em relação aos parâmetros

O Limite inferior do logaritmo da evidência, \mathcal{L} , é dado pela expressão abaixo.

$$\begin{aligned} \mathcal{L} = & \underbrace{- \int_{\mathcal{T}} \mathbb{E}_{q(f)}[f_s]^2 ds}_{\text{TERMO 1}} - \underbrace{\int_{\mathcal{T}} \text{Var}_{q(f)}[f_s] ds}_{\text{TERMO 2}} - \underbrace{\sum_{i=1}^N \tilde{G} \left(- \frac{\tilde{\mu}^2(s_i)}{2\tilde{\sigma}^2(s_i)} \right)}_{\text{TERMO 3}} \\ & + \underbrace{\sum_{i=1}^N \log(\tilde{\sigma}^2(s_i)) - \log(2) - C}_{\text{TERMO 4}} - \underbrace{KL(q(u) \parallel p(u \mid \Theta))}_{\text{TERMO 5}} \end{aligned}$$

Para fins didáticos, a expressão foi dividida em cinco termos. Para derivar o Termo 1, em relação a cada um dos parâmetros do modelo, seja ele a priori ou variacional, foi aplicada a regra do produto, resultando nas seguintes expressões:

$$\begin{aligned} \frac{\partial \text{Termo 1}}{\partial \alpha} &= \frac{-\partial [\mathbf{m}^\top \mathbf{K}_{zz}^{-1} \Psi \mathbf{K}_{zz}^{-1} \mathbf{m}]}{\partial \alpha} \\ &= -2 \left[\frac{\partial \mathbf{K}_{zz}^{-1}}{\partial \alpha} \mathbf{m} \right]^\top \Psi [\mathbf{K}_{zz}^{-1} \mathbf{m}] - [\mathbf{K}_{zz}^{-1} \mathbf{m}]^\top \frac{\partial \Psi}{\partial \alpha} [\mathbf{K}_{zz}^{-1} \mathbf{m}] \end{aligned}$$

$$\frac{\partial \text{Termo 1}}{\partial \gamma} = -2 \left[\frac{\partial \mathbf{K}_{zz}^{-1}}{\partial \gamma} \mathbf{m} \right]^\top \Psi [\mathbf{K}_{zz}^{-1} \mathbf{m}] - [\mathbf{K}_{zz}^{-1} \mathbf{m}]^\top \frac{\partial \Psi}{\partial \gamma} [\mathbf{K}_{zz}^{-1} \mathbf{m}]$$

$$\frac{\partial \text{Termo 1}}{\partial \bar{u}} = 0$$

$$\frac{\partial \text{Termo 1}}{\partial \mathbf{m}} = -2 \mathbf{K}_{zz}^{-1} \Psi \mathbf{K}_{zz}^{-1} \mathbf{m}$$

$$\frac{\partial \text{Termo 1}}{\partial \mathbf{S}} = 0$$

De forma análoga, para derivar o Termo 2, aplicou-se a regra de derivada para traços de matriz. Como o traço envolve produtos de matrizes, foi utilizada a regra do produto para derivar os termos em relação aos parâmetros, levando as seguintes expressões:

$$\begin{aligned} \frac{\partial \text{Termo 2}}{\partial \alpha} &= \frac{-\partial [\gamma \mid \mathcal{T} \mid -\text{Tr}(\mathbf{K}_{zz}^{-1} \Psi) + \text{Tr}(\mathbf{K}_{zz}^{-1} \mathbf{S} \mathbf{K}_{zz}^{-1} \Psi)]}{\partial \alpha} \\ &= \text{Tr} \left(\mathbf{K}_{zz}^{-1} \frac{\partial \Psi}{\partial \alpha} + \Psi \frac{\partial \mathbf{K}_{zz}^{-1}}{\partial \alpha} \right) - \text{Tr} \left(\mathbf{S} \left[2 \mathbf{K}_{zz}^{-1} \Psi \frac{\partial \mathbf{K}_{zz}^{-1}}{\partial \alpha} + \mathbf{K}_{zz}^{-1} \frac{\partial \Psi}{\partial \alpha} \mathbf{K}_{zz}^{-1} \right] \right) \end{aligned}$$

$$\frac{\partial \text{Termo 2}}{\partial \gamma} = \text{Tr} \left(\mathbf{K}_{zz}^{-1} \frac{\partial \Psi}{\partial \gamma} + \Psi \frac{\partial \mathbf{K}_{zz}^{-1}}{\partial \gamma} \right) - \text{Tr} \left(\mathbf{S} \left[2\mathbf{K}_{zz}^{-1} \Psi \frac{\partial \mathbf{K}_{zz}^{-1}}{\partial \gamma} + \mathbf{K}_{zz}^{-1} \frac{\partial \Psi}{\partial \gamma} \mathbf{K}_{zz}^{-1} \right] \right)$$

$$\frac{\partial \text{Termo 2}}{\partial \bar{u}} = 0$$

$$\frac{\partial \text{Termo 2}}{\partial \mathbf{m}} = 0$$

$$\frac{\partial \text{Termo 2}}{\partial \mathbf{S}} = -\mathbf{K}_{zz}^{-1} \Psi \mathbf{K}_{zz}^{-1}$$

Já para derivar o Termo 3, foi aplicada a regra da cadeia e, posteriormente, a regra do quociente, o que leva as seguintes expressões:

$$\frac{\partial \text{Termo 3}}{\partial \alpha} = \sum_{i=1}^N \tilde{G}' \left(-\frac{\tilde{\mu}^2(s_i)}{2\tilde{\sigma}^2(s_i)} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}(s_i) \frac{\partial \tilde{\mu}(s_i)}{\partial \alpha} \tilde{\sigma}^2(s_i)) + (\tilde{\mu}^2(s_i) \frac{\partial \tilde{\sigma}^2(s_i)}{\partial \alpha})}{(\tilde{\sigma}^2(s_i))^2} \right]$$

$$\frac{\partial \text{Termo 3}}{\partial \gamma} = \sum_{i=1}^N \tilde{G}' \left(-\frac{\tilde{\mu}^2(s_i)}{2\tilde{\sigma}^2(s_i)} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}(s_i) \frac{\partial \tilde{\mu}(s_i)}{\partial \gamma} \tilde{\sigma}^2(s_i)) + (\tilde{\mu}^2(s_i) \frac{\partial \tilde{\sigma}^2(s_i)}{\partial \gamma})}{(\tilde{\sigma}^2(s_i))^2} \right]$$

$$\frac{\partial \text{Termo 3}}{\partial \bar{u}} = 0$$

$$\frac{\partial \text{Termo 3}}{\partial \mathbf{m}} = \sum_{i=1}^N \tilde{G}' \left(-\frac{\tilde{\mu}^2(s_i)}{2\tilde{\sigma}^2(s_i)} \right) \frac{\tilde{\mu}(s_i)}{\tilde{\sigma}^2(s_i)} \frac{\partial \tilde{\mu}(s_i)}{\partial \mathbf{m}}$$

$$\frac{\partial \text{Termo 3}}{\partial \mathbf{S}} = -\sum_{i=1}^N \tilde{G}' \left(-\frac{\tilde{\mu}^2(s_i)}{2\tilde{\sigma}^2(s_i)} \right) \frac{\tilde{\mu}^2(s_i)}{2(\tilde{\sigma}^2(s_i))^2} \frac{\partial \tilde{\sigma}^2(s_i)}{\partial \mathbf{S}},$$

sendo as derivadas parciais de $\tilde{\mu}(s_i)$ e $\tilde{\sigma}^2(s_i)$ em relação aos parâmetros são apresentadas no Apêndice B.3. Já para o Termo 4, as derivadas parciais são dadas por

$$\frac{\partial \text{Termo 4}}{\partial \alpha} = \sum_{i=1}^N \frac{1}{\tilde{\sigma}^2(s_i)} \frac{\partial \tilde{\sigma}^2(s_i)}{\partial \alpha}$$

$$\frac{\partial \text{Termo 4}}{\partial \gamma} = \sum_{i=1}^N \frac{1}{\tilde{\sigma}^2(s_i)} \frac{\partial \tilde{\sigma}^2(s_i)}{\partial \gamma}$$

$$\frac{\partial \text{Termo 4}}{\partial \bar{u}} = 0$$

$$\frac{\partial \text{Termo 4}}{\partial \mathbf{m}} = \mathbf{0}$$

$$\frac{\partial \text{Termo 4}}{\partial \mathbf{S}} = \sum_{i=1}^N \frac{1}{\tilde{\sigma}^2(s_i)} \frac{\partial \tilde{\sigma}^2(s_i)}{\partial \mathbf{S}}$$

Por fim, o Termo 5 é derivado como

$$\begin{aligned} \frac{\partial \text{Termo 5}}{\partial \alpha} &= \frac{\partial \left\{ -\frac{1}{2} \left[\text{tr}(\mathbf{K}_{zz}^{-1} \mathbf{S}) - \log \frac{\det(\mathbf{K}_{zz})}{\det(\mathbf{S})} - M + (\vec{\mathbf{1}}\bar{u} - \mathbf{m})^\top \mathbf{K}_{zz}^{-1} (\vec{\mathbf{1}}\bar{u} - \mathbf{m}) \right] \right\}}{\partial \alpha} \\ &= -\frac{1}{2} \left[\text{tr} \left(\frac{\partial \mathbf{K}_{zz}^{-1}}{\partial \alpha} \mathbf{S} \right) - \text{tr} \left(\mathbf{K}_{zz}^{-1} \frac{\partial \mathbf{K}_{zz}}{\partial \alpha} \right) + (\vec{\mathbf{1}}\bar{u} - \mathbf{m})^\top \frac{\partial \mathbf{K}_{zz}^{-1}}{\partial \alpha} (\vec{\mathbf{1}}\bar{u} - \mathbf{m}) \right] \end{aligned}$$

$$\frac{\partial \text{Termo 5}}{\partial \gamma} = -\frac{1}{2} \left[\text{tr} \left(\frac{\partial \mathbf{K}_{zz}^{-1}}{\partial \gamma} \mathbf{S} \right) - \text{tr} \left(\mathbf{K}_{zz}^{-1} \frac{\partial \mathbf{K}_{zz}}{\partial \gamma} \right) + (\vec{\mathbf{1}}\bar{u} - \mathbf{m})^\top \frac{\partial \mathbf{K}_{zz}^{-1}}{\partial \gamma} (\vec{\mathbf{1}}\bar{u} - \mathbf{m}) \right]$$

$$\frac{\partial \text{Termo 5}}{\partial \bar{u}} = -\vec{\mathbf{1}}^\top (\mathbf{K}_{zz}^{-1} \vec{\mathbf{1}}\bar{u} - \mathbf{K}_{zz}^{-1} \mathbf{m})$$

$$\frac{\partial \text{Termo 5}}{\partial \mathbf{m}} = -\mathbf{K}_{zz}^{-1} \mathbf{m} + \mathbf{K}_{zz}^{-1} \vec{\mathbf{1}}\bar{u}$$

$$\frac{\partial \text{Termo 5}}{\partial \mathbf{S}} = -\frac{1}{2} (\mathbf{K}_{zz}^{-1} - \mathbf{S}^{-1})$$

B.3 Derivadas parciais de $\tilde{\mu}(s_i)$ e $\tilde{\sigma}^2(s_i)$ em relação aos parâmetros

A seguir, são apresentadas as derivadas em relação aos parâmetros da média variacional, $\tilde{\mu}(s_i)$, avaliado i -ésimo evento observado da configuração pontual, para $i = 1, \dots, N$. Deste modo,

$$\begin{aligned} \frac{\partial \tilde{\mu}(s_i)}{\partial \alpha} &= \frac{\partial [\mathbf{K}_{sz} \mathbf{K}_{zz}^{-1} \mathbf{m}]}{\partial \alpha} \\ &= \frac{\partial \mathbf{K}_{sz}}{\partial \alpha} \mathbf{K}_{zz}^{-1} \mathbf{m} + \mathbf{K}_{sz} \frac{\partial \mathbf{K}_{zz}^{-1}}{\partial \alpha} \mathbf{m} \end{aligned}$$

$$\frac{\partial \tilde{\mu}(s_i)}{\partial \gamma} = \frac{\partial \mathbf{K}_{sz}}{\partial \gamma} \mathbf{K}_{zz}^{-1} \mathbf{m} + \mathbf{K}_{sz} \frac{\partial \mathbf{K}_{zz}^{-1}}{\partial \gamma} \mathbf{m}$$

$$\frac{\partial \tilde{\mu}(s_i)}{\partial \bar{u}} = 0$$

$$\frac{\partial \tilde{\mu}(s_i)}{\partial \mathbf{m}} = \mathbf{K}_{sz} \mathbf{K}_{zz}^{-1}$$

$$\frac{\partial \tilde{\mu}(s_i)}{\partial \mathbf{S}} = \mathbf{0}$$

Do mesmo modo, as derivadas parciais da variância variacional do processo Gaussiano, $\tilde{\sigma}^2(s_i)$, em cada evento observado s_i foram obtidas, de tal modo que

$$\begin{aligned} \frac{\partial \tilde{\sigma}^2(s_i)}{\partial \alpha} &= \frac{\partial [\gamma - Tr(\mathbf{K}_{zz}^{-1} \mathbf{K}_{zs} \mathbf{K}_{sz}) + Tr(\mathbf{K}_{zz}^{-1} \mathbf{S} \mathbf{K}_{zz}^{-1} \mathbf{K}_{zs} \mathbf{K}_{sz})]}{\partial \alpha} \\ &= -2 Tr\left(\frac{\partial \mathbf{K}_{sz}}{\partial \alpha} \mathbf{K}_{zz}^{-1} \mathbf{K}_{zs}\right) - Tr\left(\mathbf{K}_{sz} \frac{\partial \mathbf{K}_{zz}^{-1}}{\partial \alpha} \mathbf{K}_{zs}\right) + 2 Tr\left(\frac{\partial \mathbf{K}_{sz}}{\partial \alpha} \mathbf{K}_{zz}^{-1} \mathbf{S} \mathbf{K}_{zz}^{-1} \mathbf{K}_{zs}\right) + \\ &\quad + 2 Tr\left(\mathbf{K}_{sz} \frac{\partial \mathbf{K}_{zz}^{-1}}{\partial \alpha} \mathbf{S} \mathbf{K}_{zz}^{-1} \mathbf{K}_{zs}\right) \end{aligned}$$

$$\begin{aligned} \frac{\partial \tilde{\sigma}^2(s_i)}{\partial \gamma} &= -2 Tr\left(\frac{\partial \mathbf{K}_{sz}}{\partial \gamma} \mathbf{K}_{zz}^{-1} \mathbf{K}_{zs}\right) - Tr\left(\mathbf{K}_{sz} \frac{\partial \mathbf{K}_{zz}^{-1}}{\partial \gamma} \mathbf{K}_{zs}\right) + 2 Tr\left(\frac{\partial \mathbf{K}_{sz}}{\partial \gamma} \mathbf{K}_{zz}^{-1} \mathbf{S} \mathbf{K}_{zz}^{-1} \mathbf{K}_{zs}\right) + \\ &\quad + 2 Tr\left(\mathbf{K}_{sz} \frac{\partial \mathbf{K}_{zz}^{-1}}{\partial \gamma} \mathbf{S} \mathbf{K}_{zz}^{-1} \mathbf{K}_{zs}\right) \end{aligned}$$

$$\frac{\partial \tilde{\sigma}^2(s_i)}{\partial \bar{u}} = 0$$

$$\frac{\partial \tilde{\sigma}^2(s_i)}{\partial \mathbf{m}} = \mathbf{0}$$

$$\frac{\partial \tilde{\sigma}^2(s_i)}{\partial \mathbf{S}} = \mathbf{K}_{zz}^{-1} \mathbf{K}_{sz} \mathbf{K}_{zs} \mathbf{K}_{zz}^{-1}$$

APÊNDICE C – Alguns resultados importantes para o modelo proposto

Neste segundo Apêndice serão apresentados alguns resultados relacionados a estrutura do modelo proposto, bem como alguns cálculos que são usados no processo inferencial, apresentado no capítulo de “Resultados”, que são importantes e necessários para a compreensão do texto principal.

C.1 Cálculo dos elementos das matrizes Ψ

Nessa subseção, são apresentados os cálculos necessários para a obtenção dos elementos de todas as matrizes Ψ_{ij}^{ij} , para $i, j = 1, 2$.

$$\begin{aligned}
\Psi_{11}^{11}(z, z') &= \int_{\mathcal{T}} K_{11}(z, s) \times K_{11}(s, z') ds \\
&= (\gamma_1^2)^2 \times \int_{\mathcal{T}} \exp \left\{ \frac{-(z-s)^2}{2\alpha} - \frac{(s-z')^2}{2\alpha} \right\} ds \\
&= \gamma_1^4 \int_{\mathcal{T}} \exp \left\{ \frac{-z^2 + 2sz - s^2 - s^2 + 2sz' - z'^2}{2\alpha} \right\} ds \\
&= \gamma_1^4 \int_{\mathcal{T}} \exp \left\{ \frac{-z^2 - z'^2 + 2s(z+z') - 2s^2}{2\alpha} \right\} ds \\
&= \gamma_1^4 \int_{\mathcal{T}} \exp \left\{ \frac{-z^2 - z'^2 + 2zz' - 2zz' + 4s\bar{z} - 2s^2 + 2\bar{z}^2 - 2\bar{z}^2}{2\alpha} \right\} ds \\
&= \gamma_1^4 \int_{\mathcal{T}} \exp \left\{ \frac{-(z-z')^2 - 2zz' + 2\bar{z}^2}{2\alpha} - \frac{2(s-\bar{z})^2}{2\alpha} \right\} ds \\
&= \gamma_1^4 \int_{\mathcal{T}} \exp \left\{ \frac{-(z-z')^2 - 2zz' + \frac{2}{4}(z+z')^2}{2\alpha} - \frac{(s-\bar{z})^2}{\alpha} \right\} ds \\
&= \gamma_1^4 \int_{\mathcal{T}} \exp \left\{ \frac{-(z-z')^2}{4\alpha} - \frac{(s-\bar{z})^2}{\alpha} \right\} ds \\
&= \gamma_1^4 \exp \left\{ \frac{-(z-z')^2}{4\alpha} \right\} \int_{\mathcal{T}} \exp \left\{ -\frac{(s-\bar{z})^2}{\alpha} \right\} ds \\
&= \gamma_1^4 \sqrt{\pi\alpha} \exp \left\{ \frac{-(z-z')^2}{4\alpha} \right\} \int_{\mathcal{T}_{min}}^{\mathcal{T}_{max}} \frac{1}{\sqrt{\pi\alpha}} \exp \left\{ -\frac{(s-\bar{z})^2}{\alpha} \right\} ds \\
&= \gamma_1^4 \sqrt{\pi\alpha} \exp \left\{ \frac{-(z-z')^2}{4\alpha} \right\} \left[\Phi \left(\frac{\mathcal{T}_{max} - \bar{z}}{\sqrt{\frac{\alpha}{2}}} \right) - \Phi \left(\frac{\mathcal{T}_{min} - \bar{z}}{\sqrt{\frac{\alpha}{2}}} \right) \right] \\
&= \gamma_1^4 \sqrt{\pi\alpha} \exp \left\{ \frac{-(z-z')^2}{4\alpha} \right\} \left[\frac{1}{2} + \frac{1}{2} \operatorname{erf} \left(\frac{\mathcal{T}_{max} - \bar{z}}{\sqrt{2}\sqrt{\frac{\alpha}{2}}} \right) - \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left(\frac{\mathcal{T}_{min} - \bar{z}}{\sqrt{2}\sqrt{\frac{\alpha}{2}}} \right) \right] \\
&= -\gamma_1^4 \frac{\sqrt{\pi\alpha}}{2} \exp \left\{ \frac{-(z-z')^2}{4\alpha} \right\} \left[\operatorname{erf} \left(\frac{\bar{z} - \mathcal{T}_{max}}{\sqrt{\alpha}} \right) - \operatorname{erf} \left(\frac{\bar{z} - \mathcal{T}_{min}}{\sqrt{\alpha}} \right) \right],
\end{aligned}$$

em que $\bar{z} = \frac{z+z'}{2}$ e $\Phi(x) = \frac{1}{2} + \frac{1}{2}\text{erf}\left(\frac{x}{\sqrt{2}}\right)$, sendo $\text{erf}(\cdot)$ denominado como a função erro de Gauss.

$$\begin{aligned}\Psi_{11}^{21}(z, z') &= \Psi_{11}^{12}(z, z') = \Psi_{21}^{11}(z, z') = \Psi_{12}^{11}(z, z') = \\ &= \int_{\mathcal{T}} K_{21}(z, s) \times K_{11}(s, z') ds \\ &= \rho_{12}\gamma_1^3\gamma_2 \times \int_{\mathcal{T}} \exp\left\{\frac{-(z-s)^2}{2\alpha} - \frac{(s-z')^2}{2\alpha}\right\} ds \\ &= -\rho_{12}\gamma_1^3\gamma_2 \frac{\sqrt{\pi\alpha}}{2} \exp\left\{\frac{-(z-z')^2}{4\alpha}\right\} \left[\text{erf}\left(\frac{\bar{z}-\mathcal{T}_{max}}{\sqrt{\alpha}}\right) - \text{erf}\left(\frac{\bar{z}-\mathcal{T}_{min}}{\sqrt{\alpha}}\right)\right],\end{aligned}$$

$$\begin{aligned}\Psi_{12}^{12}(z, z') &= \Psi_{21}^{21}(z, z') = \Psi_{21}^{12}(z, z') = \Psi_{12}^{21}(z, z') = \\ &= \int_{\mathcal{T}} K_{12}(z, s) \times K_{12}(s, z') ds \\ &= \rho_{12}^2\gamma_1^2\gamma_2^2 \times \int_{\mathcal{T}} \exp\left\{\frac{-(z-s)^2}{2\alpha} - \frac{(s-z')^2}{2\alpha}\right\} ds \\ &= -\rho_{12}^2\gamma_1^2\gamma_2^2 \frac{\sqrt{\pi\alpha}}{2} \exp\left\{\frac{-(z-z')^2}{4\alpha}\right\} \left[\text{erf}\left(\frac{\bar{z}-\mathcal{T}_{max}}{\sqrt{\alpha}}\right) - \text{erf}\left(\frac{\bar{z}-\mathcal{T}_{min}}{\sqrt{\alpha}}\right)\right],\end{aligned}$$

$$\begin{aligned}\Psi_{22}^{11}(z, z') &= \Psi_{11}^{22}(z, z') = \int_{\mathcal{T}} K_{11}(z, s) \times K_{22}(s, z') ds \\ &= \gamma_1^2\gamma_2^2 \times \int_{\mathcal{T}} \exp\left\{\frac{-(z-s)^2}{2\alpha} - \frac{(s-z')^2}{2\alpha}\right\} ds \\ &= -\gamma_1^2\gamma_2^2 \frac{\sqrt{\pi\alpha}}{2} \exp\left\{\frac{-(z-z')^2}{4\alpha}\right\} \left[\text{erf}\left(\frac{\bar{z}-\mathcal{T}_{max}}{\sqrt{\alpha}}\right) - \text{erf}\left(\frac{\bar{z}-\mathcal{T}_{min}}{\sqrt{\alpha}}\right)\right],\end{aligned}$$

$$\begin{aligned}\Psi_{22}^{12}(z, z') &= \Psi_{22}^{21}(z, z') = \Psi_{12}^{22}(z, z') = \Psi_{21}^{22}(z, z') = \\ &= \int_{\mathcal{T}} K_{12}(z, s) \times K_{22}(s, z') ds \\ &= \rho_{12}\gamma_1\gamma_2^3 \times \int_{\mathcal{T}} \exp\left\{\frac{-(z-s)^2}{2\alpha} - \frac{(s-z')^2}{2\alpha}\right\} ds \\ &= -\rho_{12}\gamma_1\gamma_2^3 \frac{\sqrt{\pi\alpha}}{2} \exp\left\{\frac{-(z-z')^2}{4\alpha}\right\} \left[\text{erf}\left(\frac{\bar{z}-\mathcal{T}_{max}}{\sqrt{\alpha}}\right) - \text{erf}\left(\frac{\bar{z}-\mathcal{T}_{min}}{\sqrt{\alpha}}\right)\right],\end{aligned}$$

$$\begin{aligned}\Psi_{22}^{22}(z, z') &= \int_{\mathcal{T}} K_{22}(z, s) \times K_{22}(s, z') ds \\ &= \gamma_2^4 \times \int_{\mathcal{T}} \exp\left\{\frac{-(z-s)^2}{2\alpha} - \frac{(s-z')^2}{2\alpha}\right\} ds \\ &= -\gamma_2^4 \frac{\sqrt{\pi\alpha}}{2} \exp\left\{\frac{-(z-z')^2}{4\alpha}\right\} \left[\text{erf}\left(\frac{\bar{z}-\mathcal{T}_{max}}{\sqrt{\alpha}}\right) - \text{erf}\left(\frac{\bar{z}-\mathcal{T}_{min}}{\sqrt{\alpha}}\right)\right],\end{aligned}$$

C.2 Derivadas parciais de \mathcal{L} em relação aos parâmetros

O Limite inferior do logaritmo da evidência, \mathcal{L} , obtido pelo modelo proposto nessa tese, é dado pela expressão abaixo. Assim como foi apresentado no Apêndice B, esse Apêndice tem por objetivo expor os cálculos dos gradientes de \mathcal{L} em relação aos hiperparâmetros do modelo. Esses cálculos serão importantes para o processo de otimização, a fim de minimizar a divergência KL entre a distribuição variacional e a posteriori. A expressão \mathcal{L} foi dividida em 9 termos para que os gradientes pudessem ser melhor apresentados.

$$\begin{aligned}
\mathcal{L} = & \underbrace{- \int_{\mathcal{T}} \mathbb{E}_{q(f)} [f_1(s)]^2 ds}_{\text{TERMO 1}} - \underbrace{\int_{\mathcal{T}} \mathbb{E}_{q(f)} [f_2(s)]^2 ds}_{\text{TERMO 2}} - \underbrace{\int_{\mathcal{T}} \text{Var}_{q(f)} [f_1(s)] ds}_{\text{TERMO 3}} \\
& - \underbrace{\int_{\mathcal{T}} \text{Var}_{q(f)} [f_2(s)] ds}_{\text{TERMO 4}} - \underbrace{\sum_{i=1}^{N_1} \tilde{G} \left(-\frac{\tilde{\mu}_1^2(s_{1i})}{2\tilde{\sigma}_1^2(s_{1i})} \right)}_{\text{TERMO 5}} - \underbrace{\sum_{i=1}^{N_2} \tilde{G} \left(-\frac{\tilde{\mu}_2^2(s_{2i})}{2\tilde{\sigma}_2^2(s_{2i})} \right)}_{\text{TERMO 6}} \\
& + \underbrace{\sum_{i=1}^{N_1} \log(\tilde{\sigma}_1^2(s_{1i})) - \log(2) - C}_{\text{TERMO 7}} + \underbrace{\sum_{i=1}^{N_2} \log(\tilde{\sigma}_2^2(s_{2i})) - \log(2) - C}_{\text{TERMO 8}} \\
& - \underbrace{KL(q(\mathbf{u}) \parallel p(\mathbf{u} \mid \Theta))}_{\text{TERMO 9}}
\end{aligned}$$

Aplicando a regra do produto para as derivadas parciais do Termo 1 em relação aos hiperparâmetros do modelo, tem-se que

$$\begin{aligned}
\frac{\partial \text{Termo 1}}{\partial \alpha} = & - \left\{ \left[\frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{m}_1 + \frac{\partial \mathbf{B}}{\partial \alpha} \mathbf{m}_2 \right]^\top [\Psi_{11}^{11} \mathbf{A} \mathbf{m}_1 + \Psi_{12}^{11} \mathbf{C} \mathbf{m}_1 + \Psi_{11}^{11} \mathbf{B} \mathbf{m}_2 + \Psi_{12}^{11} \mathbf{D} \mathbf{m}_2] + \right. \\
& + [\mathbf{A} \mathbf{m}_1 + \mathbf{B} \mathbf{m}_2]^\top \left[\left(\frac{\partial \Psi_{11}^{11}}{\partial \alpha} \mathbf{A} + \Psi_{11}^{11} \frac{\partial \mathbf{A}}{\partial \alpha} \right) \mathbf{m}_1 + \left(\frac{\partial \Psi_{12}^{11}}{\partial \alpha} \mathbf{C} + \Psi_{12}^{11} \frac{\partial \mathbf{C}}{\partial \alpha} \right) \mathbf{m}_1 + \right. \\
& \quad \left. + \left(\frac{\partial \Psi_{11}^{11}}{\partial \alpha} \mathbf{B} + \Psi_{11}^{11} \frac{\partial \mathbf{B}}{\partial \alpha} \right) \mathbf{m}_2 + \left(\frac{\partial \Psi_{12}^{11}}{\partial \alpha} \mathbf{D} + \Psi_{12}^{11} \frac{\partial \mathbf{D}}{\partial \alpha} \right) \mathbf{m}_2 \right] + \\
& + \left[\frac{\partial \mathbf{C}}{\partial \alpha} \mathbf{m}_1 + \frac{\partial \mathbf{D}}{\partial \alpha} \mathbf{m}_2 \right]^\top [\Psi_{11}^{21} \mathbf{A} \mathbf{m}_1 + \Psi_{12}^{21} \mathbf{C} \mathbf{m}_1 + \Psi_{11}^{21} \mathbf{B} \mathbf{m}_2 + \Psi_{12}^{21} \mathbf{D} \mathbf{m}_2] + \\
& + [\mathbf{C} \mathbf{m}_1 + \mathbf{D} \mathbf{m}_2]^\top \left[\left(\frac{\partial \Psi_{11}^{21}}{\partial \alpha} \mathbf{A} + \Psi_{11}^{21} \frac{\partial \mathbf{A}}{\partial \alpha} \right) \mathbf{m}_1 + \left(\frac{\partial \Psi_{12}^{21}}{\partial \alpha} \mathbf{C} + \Psi_{12}^{21} \frac{\partial \mathbf{C}}{\partial \alpha} \right) \mathbf{m}_1 + \right. \\
& \quad \left. + \left(\frac{\partial \Psi_{11}^{21}}{\partial \alpha} \mathbf{B} + \Psi_{11}^{21} \frac{\partial \mathbf{B}}{\partial \alpha} \right) \mathbf{m}_2 + \left(\frac{\partial \Psi_{12}^{21}}{\partial \alpha} \mathbf{D} + \Psi_{12}^{21} \frac{\partial \mathbf{D}}{\partial \alpha} \right) \mathbf{m}_2 \right] \left. \right\}
\end{aligned}$$

$$\begin{aligned} \frac{\partial \text{Termo 1}}{\partial \mathbf{m}_1} = & - \left\{ \mathbf{A}^\top [\Psi_{11}^{11} \mathbf{A} \mathbf{m}_1 + \Psi_{12}^{11} \mathbf{C} \mathbf{m}_1 + \Psi_{11}^{11} \mathbf{B} \mathbf{m}_2 + \Psi_{12}^{11} \mathbf{D} \mathbf{m}_2] + \right. \\ & + \mathbf{A}^\top \Psi_{11}^{11} (\mathbf{A} \mathbf{m}_1 + \mathbf{B} \mathbf{m}_2) + \mathbf{C}^\top \Psi_{12}^{11} (\mathbf{A} \mathbf{m}_1 + \mathbf{B} \mathbf{m}_2) + \\ & + \mathbf{C}^\top [\Psi_{11}^{21} \mathbf{A} \mathbf{m}_1 + \Psi_{12}^{21} \mathbf{C} \mathbf{m}_1 + \Psi_{11}^{21} \mathbf{B} \mathbf{m}_2 + \Psi_{12}^{21} \mathbf{D} \mathbf{m}_2] + \\ & \left. + \mathbf{A}^\top \Psi_{11}^{21} (\mathbf{C} \mathbf{m}_1 + \mathbf{D} \mathbf{m}_2) + \mathbf{C}^\top \Psi_{12}^{21} (\mathbf{C} \mathbf{m}_1 + \mathbf{D} \mathbf{m}_2) \right\} \end{aligned}$$

$$\begin{aligned} \frac{\partial \text{Termo 1}}{\partial \mathbf{m}_2} = & - \left\{ \mathbf{B}^\top [\Psi_{11}^{11} \mathbf{A} \mathbf{m}_1 + \Psi_{12}^{11} \mathbf{C} \mathbf{m}_1 + \Psi_{11}^{11} \mathbf{B} \mathbf{m}_2 + \Psi_{12}^{11} \mathbf{D} \mathbf{m}_2] + \right. \\ & + \mathbf{B}^\top \Psi_{11}^{11} (\mathbf{A} \mathbf{m}_1 + \mathbf{B} \mathbf{m}_2) + \mathbf{D}^\top \Psi_{12}^{11} (\mathbf{A} \mathbf{m}_1 + \mathbf{B} \mathbf{m}_2) + \\ & + \mathbf{D}^\top [\Psi_{11}^{21} \mathbf{A} \mathbf{m}_1 + \Psi_{12}^{21} \mathbf{C} \mathbf{m}_1 + \Psi_{11}^{21} \mathbf{B} \mathbf{m}_2 + \Psi_{12}^{21} \mathbf{D} \mathbf{m}_2] + \\ & \left. + \mathbf{B}^\top \Psi_{11}^{21} (\mathbf{C} \mathbf{m}_1 + \mathbf{D} \mathbf{m}_2) + \mathbf{D}^\top \Psi_{12}^{21} (\mathbf{C} \mathbf{m}_1 + \mathbf{D} \mathbf{m}_2) \right\} \end{aligned}$$

As derivadas do Termo 1 em relação à \bar{u}_1 , \bar{u}_2 , \mathbf{S}_{11} , \mathbf{S}_{12} e \mathbf{S}_{22} foram nulas, uma vez que a expressão não depende desses parâmetros. Já as derivadas parciais das matrizes \mathbf{A} , \mathbf{B} , \mathbf{C} e \mathbf{D} estão apresentadas na subseção C.3 desse Apêndice. Já as derivadas parciais dos Ψ 's estão descritas na subseção C.6. Seguindo o mesmo raciocínio, as expressões das derivadas parciais do Termo 2 são dadas abaixo.

$$\begin{aligned} \frac{\partial \text{Termo 2}}{\partial \alpha} = & - \left\{ \left[\frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{m}_1 + \frac{\partial \mathbf{B}}{\partial \alpha} \mathbf{m}_2 \right]^\top [\Psi_{21}^{12} \mathbf{A} \mathbf{m}_1 + \Psi_{22}^{12} \mathbf{C} \mathbf{m}_1 + \Psi_{21}^{12} \mathbf{B} \mathbf{m}_2 + \Psi_{22}^{12} \mathbf{D} \mathbf{m}_2] + \right. \\ & + [\mathbf{A} \mathbf{m}_1 + \mathbf{B} \mathbf{m}_2]^\top \left[\left(\frac{\partial \Psi_{21}^{12}}{\partial \alpha} \mathbf{A} + \Psi_{21}^{12} \frac{\partial \mathbf{A}}{\partial \alpha} \right) \mathbf{m}_1 + \left(\frac{\partial \Psi_{22}^{12}}{\partial \alpha} \mathbf{C} + \Psi_{22}^{12} \frac{\partial \mathbf{C}}{\partial \alpha} \right) \mathbf{m}_1 + \right. \\ & \quad \left. + \left(\frac{\partial \Psi_{21}^{12}}{\partial \alpha} \mathbf{B} + \Psi_{21}^{12} \frac{\partial \mathbf{B}}{\partial \alpha} \right) \mathbf{m}_2 + \left(\frac{\partial \Psi_{22}^{12}}{\partial \alpha} \mathbf{D} + \Psi_{22}^{12} \frac{\partial \mathbf{D}}{\partial \alpha} \right) \mathbf{m}_2 \right] + \\ & + \left[\frac{\partial \mathbf{C}}{\partial \alpha} \mathbf{m}_1 + \frac{\partial \mathbf{D}}{\partial \alpha} \mathbf{m}_2 \right]^\top [\Psi_{21}^{22} \mathbf{A} \mathbf{m}_1 + \Psi_{22}^{22} \mathbf{C} \mathbf{m}_1 + \Psi_{21}^{22} \mathbf{B} \mathbf{m}_2 + \Psi_{22}^{22} \mathbf{D} \mathbf{m}_2] + \\ & + [\mathbf{C} \mathbf{m}_1 + \mathbf{D} \mathbf{m}_2]^\top \left[\left(\frac{\partial \Psi_{21}^{22}}{\partial \alpha} \mathbf{A} + \Psi_{21}^{22} \frac{\partial \mathbf{A}}{\partial \alpha} \right) \mathbf{m}_1 + \left(\frac{\partial \Psi_{22}^{22}}{\partial \alpha} \mathbf{C} + \Psi_{22}^{22} \frac{\partial \mathbf{C}}{\partial \alpha} \right) \mathbf{m}_1 + \right. \\ & \quad \left. + \left(\frac{\partial \Psi_{21}^{22}}{\partial \alpha} \mathbf{B} + \Psi_{21}^{22} \frac{\partial \mathbf{B}}{\partial \alpha} \right) \mathbf{m}_2 + \left(\frac{\partial \Psi_{22}^{22}}{\partial \alpha} \mathbf{D} + \Psi_{22}^{22} \frac{\partial \mathbf{D}}{\partial \alpha} \right) \mathbf{m}_2 \right] \left. \right\} \end{aligned}$$

$$\begin{aligned} \frac{\partial \text{Termo 2}}{\partial \mathbf{m}_1} = & - \left\{ \mathbf{A}^\top [\Psi_{21}^{12} \mathbf{A} \mathbf{m}_1 + \Psi_{22}^{12} \mathbf{C} \mathbf{m}_1 + \Psi_{21}^{12} \mathbf{B} \mathbf{m}_2 + \Psi_{22}^{12} \mathbf{D} \mathbf{m}_2] + \right. \\ & + \mathbf{A}^\top \Psi_{21}^{12} (\mathbf{A} \mathbf{m}_1 + \mathbf{B} \mathbf{m}_2) + \mathbf{C}^\top \Psi_{21}^{22} (\mathbf{A} \mathbf{m}_1 + \mathbf{B} \mathbf{m}_2) + \\ & + \mathbf{C}^\top [\Psi_{21}^{22} \mathbf{A} \mathbf{m}_1 + \Psi_{22}^{22} \mathbf{C} \mathbf{m}_1 + \Psi_{21}^{22} \mathbf{B} \mathbf{m}_2 + \Psi_{22}^{22} \mathbf{D} \mathbf{m}_2] + \\ & \left. + \mathbf{A}^\top \Psi_{22}^{12} (\mathbf{C} \mathbf{m}_1 + \mathbf{D} \mathbf{m}_2) + \mathbf{C}^\top \Psi_{22}^{22} (\mathbf{C} \mathbf{m}_1 + \mathbf{D} \mathbf{m}_2) \right\} \end{aligned}$$

$$\begin{aligned} \frac{\partial \text{Termo 2}}{\partial \mathbf{m}_2} = & - \left\{ \mathbf{B}^\top [\Psi_{21}^{12} \mathbf{A} \mathbf{m}_1 + \Psi_{22}^{12} \mathbf{C} \mathbf{m}_1 + \Psi_{21}^{12} \mathbf{B} \mathbf{m}_2 + \Psi_{22}^{12} \mathbf{D} \mathbf{m}_2] + \right. \\ & + \mathbf{B}^\top \Psi_{21}^{12} (\mathbf{A} \mathbf{m}_1 + \mathbf{B} \mathbf{m}_2) + \mathbf{D}^\top \Psi_{21}^{22} (\mathbf{A} \mathbf{m}_1 + \mathbf{B} \mathbf{m}_2) + \\ & + \mathbf{D}^\top [\Psi_{21}^{22} \mathbf{A} \mathbf{m}_1 + \Psi_{22}^{22} \mathbf{C} \mathbf{m}_1 + \Psi_{21}^{22} \mathbf{B} \mathbf{m}_2 + \Psi_{22}^{22} \mathbf{D} \mathbf{m}_2] + \\ & \left. + \mathbf{B}^\top \Psi_{22}^{12} (\mathbf{C} \mathbf{m}_1 + \mathbf{D} \mathbf{m}_2) + \mathbf{D}^\top \Psi_{22}^{22} (\mathbf{C} \mathbf{m}_1 + \mathbf{D} \mathbf{m}_2) \right\} \end{aligned}$$

Da mesma forma como no Termo 1, o Termo 2 de \mathcal{L} não depende de \bar{u}_1 , \bar{u}_2 , \mathbf{S}_{11} , \mathbf{S}_{12} e \mathbf{S}_{22} . Portanto, sua derivada em relação à esses parâmetros também serão nulas. Derivando o Termo 3, tem-se

$$\begin{aligned}
\frac{\partial \text{Termo 3}}{\partial \gamma_1} = & - \left[2\gamma_1 | \mathcal{T} | - \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \gamma_1} \Psi_{11}^{11} + \mathbf{A} \frac{\partial \Psi_{11}^{11}}{\partial \gamma_1} \right) - \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \gamma_1} \Psi_{12}^{11} + \mathbf{C} \frac{\partial \Psi_{12}^{11}}{\partial \gamma_1} \right) - \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \gamma_1} \Psi_{11}^{21} + \right. \\
& + \mathbf{B} \frac{\partial \Psi_{11}^{21}}{\partial \gamma_1} \left. \right) - \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \gamma_1} \Psi_{12}^{21} + \mathbf{D} \frac{\partial \Psi_{12}^{21}}{\partial \gamma_1} \right) + \text{Tr} \left(2 \frac{\partial \mathbf{A}}{\partial \gamma_1} \mathbf{S}_{11} \mathbf{A} \Psi_{11}^{11} + \mathbf{A} \mathbf{S}_{11} \mathbf{A} \frac{\partial \Psi_{11}^{11}}{\partial \gamma_1} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \gamma_1} \mathbf{S}_{12}^\top \mathbf{A} \Psi_{11}^{11} + \mathbf{B} \mathbf{S}_{12}^\top \frac{\partial \mathbf{A}}{\partial \gamma_1} \Psi_{11}^{11} + \mathbf{B} \mathbf{S}_{12}^\top \mathbf{A} \frac{\partial \Psi_{11}^{11}}{\partial \gamma_1} \right) + \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \gamma_1} \mathbf{S}_{11} \mathbf{A} \Psi_{12}^{11} + \right. \\
& + \mathbf{C} \mathbf{S}_{11} \frac{\partial \mathbf{A}}{\partial \gamma_1} \Psi_{12}^{11} + \mathbf{C} \mathbf{S}_{11} \mathbf{A} \frac{\partial \Psi_{12}^{11}}{\partial \gamma_1} \left. \right) \left. \right] + \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \gamma_1} \mathbf{S}_{12}^\top \mathbf{A} \Psi_{12}^{11} + \mathbf{D} \mathbf{S}_{12}^\top \frac{\partial \mathbf{A}}{\partial \gamma_1} \Psi_{12}^{11} + \right. \\
& + \mathbf{D} \mathbf{S}_{12}^\top \mathbf{A} \frac{\partial \Psi_{12}^{11}}{\partial \gamma_1} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \gamma_1} \mathbf{S}_{12} \mathbf{C} \Psi_{11}^{11} + \mathbf{A} \mathbf{S}_{12} \frac{\partial \mathbf{C}}{\partial \gamma_1} \Psi_{11}^{11} + \mathbf{A} \mathbf{S}_{12} \mathbf{C} \frac{\partial \Psi_{11}^{11}}{\partial \gamma_1} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \gamma_1} \mathbf{S}_{22} \mathbf{C} \Psi_{11}^{11} + \mathbf{B} \mathbf{S}_{22} \frac{\partial \mathbf{C}}{\partial \gamma_1} \Psi_{11}^{11} + \mathbf{B} \mathbf{S}_{22} \mathbf{C} \frac{\partial \Psi_{11}^{11}}{\partial \gamma_1} \right) + \text{Tr} \left(2 \frac{\partial \mathbf{C}}{\partial \gamma_1} \mathbf{S}_{12} \mathbf{C} \Psi_{12}^{11} + \right. \\
& + \mathbf{C} \mathbf{S}_{12} \mathbf{C} \frac{\partial \Psi_{12}^{11}}{\partial \gamma_1} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \gamma_1} \mathbf{S}_{22} \mathbf{C} \Psi_{12}^{11} + \mathbf{D} \mathbf{S}_{22} \frac{\partial \mathbf{C}}{\partial \gamma_1} \Psi_{12}^{11} + \mathbf{D} \mathbf{S}_{22} \mathbf{C} \frac{\partial \Psi_{12}^{11}}{\partial \gamma_1} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \gamma_1} \mathbf{S}_{11} \mathbf{B} \Psi_{11}^{21} + \mathbf{A} \mathbf{S}_{11} \frac{\partial \mathbf{B}}{\partial \gamma_1} \Psi_{11}^{21} + \mathbf{A} \mathbf{S}_{11} \mathbf{B} \frac{\partial \Psi_{11}^{21}}{\partial \gamma_1} \right) + \text{Tr} \left(2 \frac{\partial \mathbf{B}}{\partial \gamma_1} \mathbf{S}_{12}^\top \mathbf{B} \Psi_{11}^{21} + \right. \\
& + \mathbf{B} \mathbf{S}_{12}^\top \mathbf{B} \frac{\partial \Psi_{11}^{21}}{\partial \gamma_1} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \gamma_1} \mathbf{S}_{11} \mathbf{B} \Psi_{12}^{21} + \mathbf{C} \mathbf{S}_{11} \frac{\partial \mathbf{B}}{\partial \gamma_1} \Psi_{12}^{21} + \mathbf{C} \mathbf{S}_{11} \mathbf{B} \frac{\partial \Psi_{12}^{21}}{\partial \gamma_1} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \gamma_1} \mathbf{S}_{12}^\top \mathbf{B} \Psi_{12}^{21} + \mathbf{D} \mathbf{S}_{12}^\top \frac{\partial \mathbf{B}}{\partial \gamma_1} \Psi_{12}^{21} + \mathbf{D} \mathbf{S}_{12}^\top \mathbf{B} \frac{\partial \Psi_{12}^{21}}{\partial \gamma_1} \right) + \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \gamma_1} \mathbf{S}_{12} \mathbf{D} \Psi_{11}^{21} + \right. \\
& + \mathbf{A} \mathbf{S}_{12} \frac{\partial \mathbf{D}}{\partial \gamma_1} \Psi_{11}^{21} + \mathbf{A} \mathbf{S}_{12} \mathbf{D} \frac{\partial \Psi_{11}^{21}}{\partial \gamma_1} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \gamma_1} \mathbf{S}_{22} \mathbf{D} \Psi_{11}^{21} + \mathbf{B} \mathbf{S}_{22} \frac{\partial \mathbf{D}}{\partial \gamma_1} \Psi_{11}^{21} + \right. \\
& + \mathbf{B} \mathbf{S}_{22} \mathbf{D} \frac{\partial \Psi_{11}^{21}}{\partial \gamma_1} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \gamma_1} \mathbf{S}_{12} \mathbf{D} \Psi_{12}^{21} + \mathbf{C} \mathbf{S}_{12} \frac{\partial \mathbf{D}}{\partial \gamma_1} \Psi_{12}^{21} + \mathbf{C} \mathbf{S}_{12} \mathbf{D} \frac{\partial \Psi_{12}^{21}}{\partial \gamma_1} \right) + \\
& + \text{Tr} \left(2 \frac{\partial \mathbf{D}}{\partial \gamma_1} \mathbf{S}_{22} \mathbf{D} \Psi_{12}^{21} + \mathbf{D} \mathbf{S}_{22} \mathbf{D} \frac{\partial \Psi_{12}^{21}}{\partial \gamma_1} \right) \left. \right]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \text{Termo 3}}{\partial \gamma_2} = & - \left[- \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \gamma_2} \Psi_{11}^{11} + \mathbf{A} \frac{\partial \Psi_{11}^{11}}{\partial \gamma_2} \right) - \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \gamma_2} \Psi_{12}^{11} + \mathbf{C} \frac{\partial \Psi_{12}^{11}}{\partial \gamma_2} \right) - \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \gamma_2} \Psi_{11}^{21} + \right. \\
& + \mathbf{B} \frac{\partial \Psi_{11}^{21}}{\partial \gamma_2} \left. \right) - \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \gamma_2} \Psi_{12}^{21} + \mathbf{D} \frac{\partial \Psi_{12}^{21}}{\partial \gamma_2} \right) + \text{Tr} \left(2 \frac{\partial \mathbf{A}}{\partial \gamma_2} \mathbf{S}_{11} \mathbf{A} \Psi_{11}^{11} + \mathbf{A} \mathbf{S}_{11} \mathbf{A} \frac{\partial \Psi_{11}^{11}}{\partial \gamma_2} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \gamma_2} \mathbf{S}_{12}^\top \mathbf{A} \Psi_{11}^{11} + \mathbf{B} \mathbf{S}_{12}^\top \frac{\partial \mathbf{A}}{\partial \gamma_2} \Psi_{11}^{11} + \mathbf{B} \mathbf{S}_{12}^\top \mathbf{A} \frac{\partial \Psi_{11}^{11}}{\partial \gamma_2} \right) + \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \gamma_2} \mathbf{S}_{11} \mathbf{A} \Psi_{12}^{11} + \right. \\
& + \mathbf{C} \mathbf{S}_{11} \frac{\partial \mathbf{A}}{\partial \gamma_2} \Psi_{12}^{11} + \mathbf{C} \mathbf{S}_{11} \mathbf{A} \frac{\partial \Psi_{12}^{11}}{\partial \gamma_2} \left. \right) \left. \right] + \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \gamma_2} \mathbf{S}_{12}^\top \mathbf{A} \Psi_{12}^{11} + \mathbf{D} \mathbf{S}_{12}^\top \frac{\partial \mathbf{A}}{\partial \gamma_2} \Psi_{12}^{11} + \right. \\
& + \mathbf{D} \mathbf{S}_{12}^\top \mathbf{A} \frac{\partial \Psi_{12}^{11}}{\partial \gamma_2} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \gamma_2} \mathbf{S}_{12} \mathbf{C} \Psi_{11}^{11} + \mathbf{A} \mathbf{S}_{12} \frac{\partial \mathbf{C}}{\partial \gamma_2} \Psi_{11}^{11} + \mathbf{A} \mathbf{S}_{12} \mathbf{C} \frac{\partial \Psi_{11}^{11}}{\partial \gamma_2} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \gamma_2} \mathbf{S}_{22} \mathbf{C} \Psi_{11}^{11} + \mathbf{B} \mathbf{S}_{22} \frac{\partial \mathbf{C}}{\partial \gamma_2} \Psi_{11}^{11} + \mathbf{B} \mathbf{S}_{22} \mathbf{C} \frac{\partial \Psi_{11}^{11}}{\partial \gamma_2} \right) + \text{Tr} \left(2 \frac{\partial \mathbf{C}}{\partial \gamma_2} \mathbf{S}_{12} \mathbf{C} \Psi_{12}^{11} + \right. \\
& + \mathbf{C} \mathbf{S}_{12} \mathbf{C} \frac{\partial \Psi_{12}^{11}}{\partial \gamma_2} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \gamma_2} \mathbf{S}_{22} \mathbf{C} \Psi_{12}^{11} + \mathbf{D} \mathbf{S}_{22} \frac{\partial \mathbf{C}}{\partial \gamma_2} \Psi_{12}^{11} + \mathbf{D} \mathbf{S}_{22} \mathbf{C} \frac{\partial \Psi_{12}^{11}}{\partial \gamma_2} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \gamma_2} \mathbf{S}_{11} \mathbf{B} \Psi_{11}^{21} + \mathbf{A} \mathbf{S}_{11} \frac{\partial \mathbf{B}}{\partial \gamma_2} \Psi_{11}^{21} + \mathbf{A} \mathbf{S}_{11} \mathbf{B} \frac{\partial \Psi_{11}^{21}}{\partial \gamma_2} \right) + \text{Tr} \left(2 \frac{\partial \mathbf{B}}{\partial \gamma_2} \mathbf{S}_{12}^\top \mathbf{B} \Psi_{11}^{21} + \right. \\
& + \mathbf{B} \mathbf{S}_{12}^\top \mathbf{B} \frac{\partial \Psi_{11}^{21}}{\partial \gamma_2} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \gamma_2} \mathbf{S}_{11} \mathbf{B} \Psi_{12}^{21} + \mathbf{C} \mathbf{S}_{11} \frac{\partial \mathbf{B}}{\partial \gamma_2} \Psi_{12}^{21} + \mathbf{C} \mathbf{S}_{11} \mathbf{B} \frac{\partial \Psi_{12}^{21}}{\partial \gamma_2} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \gamma_2} \mathbf{S}_{12}^\top \mathbf{B} \Psi_{12}^{21} + \mathbf{D} \mathbf{S}_{12}^\top \frac{\partial \mathbf{B}}{\partial \gamma_2} \Psi_{12}^{21} + \mathbf{D} \mathbf{S}_{12}^\top \mathbf{B} \frac{\partial \Psi_{12}^{21}}{\partial \gamma_2} \right) + \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \gamma_2} \mathbf{S}_{12} \mathbf{D} \Psi_{11}^{21} + \right. \\
& + \mathbf{A} \mathbf{S}_{12} \frac{\partial \mathbf{D}}{\partial \gamma_2} \Psi_{11}^{21} + \mathbf{A} \mathbf{S}_{12} \mathbf{D} \frac{\partial \Psi_{11}^{21}}{\partial \gamma_2} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \gamma_2} \mathbf{S}_{22} \mathbf{D} \Psi_{11}^{21} + \mathbf{B} \mathbf{S}_{22} \frac{\partial \mathbf{D}}{\partial \gamma_2} \Psi_{11}^{21} + \right. \\
& + \mathbf{B} \mathbf{S}_{22} \mathbf{D} \frac{\partial \Psi_{11}^{21}}{\partial \gamma_2} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \gamma_2} \mathbf{S}_{12} \mathbf{D} \Psi_{12}^{21} + \mathbf{C} \mathbf{S}_{12} \frac{\partial \mathbf{D}}{\partial \gamma_2} \Psi_{12}^{21} + \mathbf{C} \mathbf{S}_{12} \mathbf{D} \frac{\partial \Psi_{12}^{21}}{\partial \gamma_2} \right) + \\
& + \text{Tr} \left(2 \frac{\partial \mathbf{D}}{\partial \gamma_2} \mathbf{S}_{22} \mathbf{D} \Psi_{12}^{21} + \mathbf{D} \mathbf{S}_{22} \mathbf{D} \frac{\partial \Psi_{12}^{21}}{\partial \gamma_2} \right) \left. \right]
\end{aligned}$$

$$\frac{\partial \text{Termo 3}}{\partial \mathbf{S}_{11}} = - \left\{ (\mathbf{A} \Psi_{11}^{11} \mathbf{A})^\top + (\mathbf{A} \Psi_{12}^{11} \mathbf{C})^\top + (\mathbf{B} \Psi_{11}^{21} \mathbf{A})^\top + (\mathbf{B} \Psi_{12}^{21} \mathbf{C})^\top \right\}$$

$$\begin{aligned}
\frac{\partial \text{Termo 3}}{\partial \mathbf{S}_{12}} = & - \left\{ (\mathbf{C} \Psi_{11}^{11} \mathbf{A})^\top + (\mathbf{C} \Psi_{12}^{11} \mathbf{C})^\top + (\mathbf{D} \Psi_{11}^{21} \mathbf{A})^\top + (\mathbf{D} \Psi_{12}^{21} \mathbf{C})^\top + \right. \\
& \left. + (\mathbf{A} \Psi_{11}^{11} \mathbf{B}) + (\mathbf{A} \Psi_{12}^{11} \mathbf{D}) + (\mathbf{B} \Psi_{11}^{21} \mathbf{B}) + (\mathbf{B} \Psi_{12}^{21} \mathbf{D}) \right\}
\end{aligned}$$

$$\frac{\partial \text{Termo 3}}{\partial \mathbf{S}_{22}} = - \left\{ (\mathbf{C} \Psi_{11}^{11} \mathbf{B})^\top + (\mathbf{C} \Psi_{12}^{11} \mathbf{D})^\top + (\mathbf{D} \Psi_{11}^{21} \mathbf{B})^\top + (\mathbf{D} \Psi_{12}^{21} \mathbf{D})^\top \right\}$$

Derivando o Termo 4, tem-se as seguintes expressões:

$$\begin{aligned}
\frac{\partial \text{Termo 4}}{\partial \alpha} = & - \left[- \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \alpha} \Psi_{21}^{12} + \mathbf{A} \frac{\partial \Psi_{21}^{12}}{\partial \alpha} \right) - \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \alpha} \Psi_{22}^{12} + \mathbf{C} \frac{\partial \Psi_{22}^{12}}{\partial \alpha} \right) - \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \alpha} \Psi_{21}^{22} + \right. \\
& + \mathbf{B} \frac{\partial \Psi_{21}^{22}}{\partial \alpha} \left. \right) - \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \alpha} \Psi_{22}^{22} + \mathbf{D} \frac{\partial \Psi_{22}^{22}}{\partial \alpha} \right) + \text{Tr} \left(2 \frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{S}_{11} \mathbf{A} \Psi_{21}^{12} + \mathbf{A} \mathbf{S}_{11} \mathbf{A} \frac{\partial \Psi_{21}^{12}}{\partial \alpha} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \alpha} \mathbf{S}_{12}^{\top} \mathbf{A} \Psi_{21}^{12} + \mathbf{B} \mathbf{S}_{12}^{\top} \frac{\partial \mathbf{A}}{\partial \alpha} \Psi_{21}^{12} + \mathbf{B} \mathbf{S}_{12}^{\top} \mathbf{A} \frac{\partial \Psi_{21}^{12}}{\partial \alpha} \right) + \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \alpha} \mathbf{S}_{11} \mathbf{A} \Psi_{22}^{12} + \right. \\
& + \mathbf{C} \mathbf{S}_{11} \frac{\partial \mathbf{A}}{\partial \alpha} \Psi_{22}^{12} + \mathbf{C} \mathbf{S}_{11} \mathbf{A} \frac{\partial \Psi_{22}^{12}}{\partial \alpha} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \alpha} \mathbf{S}_{12}^{\top} \mathbf{A} \Psi_{22}^{12} + \mathbf{D} \mathbf{S}_{12}^{\top} \frac{\partial \mathbf{A}}{\partial \alpha} \Psi_{22}^{12} + \right. \\
& + \mathbf{D} \mathbf{S}_{12}^{\top} \mathbf{A} \frac{\partial \Psi_{22}^{12}}{\partial \alpha} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{S}_{12} \mathbf{C} \Psi_{21}^{12} + \mathbf{A} \mathbf{S}_{12} \frac{\partial \mathbf{C}}{\partial \alpha} \Psi_{21}^{12} + \mathbf{A} \mathbf{S}_{12} \mathbf{C} \frac{\partial \Psi_{21}^{12}}{\partial \alpha} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \alpha} \mathbf{S}_{22} \mathbf{C} \Psi_{21}^{12} + \mathbf{B} \mathbf{S}_{22} \frac{\partial \mathbf{C}}{\partial \alpha} \Psi_{21}^{12} + \mathbf{B} \mathbf{S}_{22} \mathbf{C} \frac{\partial \Psi_{21}^{12}}{\partial \alpha} \right) + \text{Tr} \left(2 \frac{\partial \mathbf{C}}{\partial \alpha} \mathbf{S}_{12} \mathbf{C} \Psi_{22}^{12} + \right. \\
& + \mathbf{C} \mathbf{S}_{12} \mathbf{C} \frac{\partial \Psi_{22}^{12}}{\partial \alpha} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \alpha} \mathbf{S}_{22} \mathbf{C} \Psi_{22}^{12} + \mathbf{D} \mathbf{S}_{22} \frac{\partial \mathbf{C}}{\partial \alpha} \Psi_{22}^{12} + \mathbf{D} \mathbf{S}_{22} \mathbf{C} \frac{\partial \Psi_{22}^{12}}{\partial \alpha} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{S}_{11} \mathbf{B} \Psi_{21}^{22} + \mathbf{A} \mathbf{S}_{11} \frac{\partial \mathbf{B}}{\partial \alpha} \Psi_{21}^{22} + \mathbf{A} \mathbf{S}_{11} \mathbf{B} \frac{\partial \Psi_{21}^{22}}{\partial \alpha} \right) + \text{Tr} \left(2 \frac{\partial \mathbf{B}}{\partial \alpha} \mathbf{S}_{12}^{\top} \mathbf{B} \Psi_{21}^{22} + \right. \\
& + \mathbf{B} \mathbf{S}_{12}^{\top} \mathbf{B} \frac{\partial \Psi_{21}^{22}}{\partial \alpha} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \alpha} \mathbf{S}_{11} \mathbf{B} \Psi_{22}^{22} + \mathbf{C} \mathbf{S}_{11} \frac{\partial \mathbf{B}}{\partial \alpha} \Psi_{22}^{22} + \mathbf{C} \mathbf{S}_{11} \mathbf{B} \frac{\partial \Psi_{22}^{22}}{\partial \alpha} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \alpha} \mathbf{S}_{12}^{\top} \mathbf{B} \Psi_{22}^{22} + \mathbf{D} \mathbf{S}_{12}^{\top} \frac{\partial \mathbf{B}}{\partial \alpha} \Psi_{22}^{22} + \mathbf{D} \mathbf{S}_{12}^{\top} \mathbf{B} \frac{\partial \Psi_{22}^{22}}{\partial \alpha} \right) + \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{S}_{12} \mathbf{D} \Psi_{21}^{22} + \right. \\
& + \mathbf{A} \mathbf{S}_{12} \frac{\partial \mathbf{D}}{\partial \alpha} \Psi_{21}^{22} + \mathbf{A} \mathbf{S}_{12} \mathbf{D} \frac{\partial \Psi_{21}^{22}}{\partial \alpha} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \alpha} \mathbf{S}_{22} \mathbf{D} \Psi_{21}^{22} + \mathbf{B} \mathbf{S}_{22} \frac{\partial \mathbf{D}}{\partial \alpha} \Psi_{21}^{22} + \right. \\
& + \mathbf{B} \mathbf{S}_{22} \mathbf{D} \frac{\partial \Psi_{21}^{22}}{\partial \alpha} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \alpha} \mathbf{S}_{12} \mathbf{D} \Psi_{22}^{22} + \mathbf{C} \mathbf{S}_{12} \frac{\partial \mathbf{D}}{\partial \alpha} \Psi_{22}^{22} + \mathbf{C} \mathbf{S}_{12} \mathbf{D} \frac{\partial \Psi_{22}^{22}}{\partial \alpha} \right) + \\
& + \text{Tr} \left(2 \frac{\partial \mathbf{D}}{\partial \alpha} \mathbf{S}_{22} \mathbf{D} \Psi_{22}^{22} + \mathbf{D} \mathbf{S}_{22} \mathbf{D} \frac{\partial \Psi_{22}^{22}}{\partial \alpha} \right) \left. \right]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \text{Termo 4}}{\partial \gamma_1} = & - \left[- \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \gamma_1} \Psi_{21}^{12} + \mathbf{A} \frac{\partial \Psi_{21}^{12}}{\partial \gamma_1} \right) - \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \gamma_1} \Psi_{22}^{12} + \mathbf{C} \frac{\partial \Psi_{22}^{12}}{\partial \gamma_1} \right) - \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \gamma_1} \Psi_{21}^{22} + \right. \\
& + \mathbf{B} \frac{\partial \Psi_{21}^{22}}{\partial \gamma_1} \left. \right) - \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \gamma_1} \Psi_{22}^{22} + \mathbf{D} \frac{\partial \Psi_{22}^{22}}{\partial \gamma_1} \right) + \text{Tr} \left(2 \frac{\partial \mathbf{A}}{\partial \gamma_1} \mathbf{S}_{11} \mathbf{A} \Psi_{21}^{12} + \mathbf{A} \mathbf{S}_{11} \mathbf{A} \frac{\partial \Psi_{21}^{12}}{\partial \gamma_1} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \gamma_1} \mathbf{S}_{12}^\top \mathbf{A} \Psi_{21}^{12} + \mathbf{B} \mathbf{S}_{12}^\top \frac{\partial \mathbf{A}}{\partial \gamma_1} \Psi_{21}^{12} + \mathbf{B} \mathbf{S}_{12}^\top \mathbf{A} \frac{\partial \Psi_{21}^{12}}{\partial \gamma_1} \right) + \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \gamma_1} \mathbf{S}_{11} \mathbf{A} \Psi_{22}^{12} + \right. \\
& + \mathbf{C} \mathbf{S}_{11} \frac{\partial \mathbf{A}}{\partial \gamma_1} \Psi_{22}^{12} + \mathbf{C} \mathbf{S}_{11} \mathbf{A} \frac{\partial \Psi_{22}^{12}}{\partial \gamma_1} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \gamma_1} \mathbf{S}_{12}^\top \mathbf{A} \Psi_{22}^{12} + \mathbf{D} \mathbf{S}_{12}^\top \frac{\partial \mathbf{A}}{\partial \gamma_1} \Psi_{22}^{12} + \right. \\
& + \mathbf{D} \mathbf{S}_{12}^\top \mathbf{A} \frac{\partial \Psi_{22}^{12}}{\partial \gamma_1} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \gamma_1} \mathbf{S}_{12} \mathbf{C} \Psi_{21}^{12} + \mathbf{A} \mathbf{S}_{12} \frac{\partial \mathbf{C}}{\partial \gamma_1} \Psi_{21}^{12} + \mathbf{A} \mathbf{S}_{12} \mathbf{C} \frac{\partial \Psi_{21}^{12}}{\partial \gamma_1} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \gamma_1} \mathbf{S}_{22} \mathbf{C} \Psi_{21}^{12} + \mathbf{B} \mathbf{S}_{22} \frac{\partial \mathbf{C}}{\partial \gamma_1} \Psi_{21}^{12} + \mathbf{B} \mathbf{S}_{22} \mathbf{C} \frac{\partial \Psi_{21}^{12}}{\partial \gamma_1} \right) + \text{Tr} \left(2 \frac{\partial \mathbf{C}}{\partial \gamma_1} \mathbf{S}_{12} \mathbf{C} \Psi_{22}^{12} + \right. \\
& + \mathbf{C} \mathbf{S}_{12} \mathbf{C} \frac{\partial \Psi_{22}^{12}}{\partial \gamma_1} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \gamma_1} \mathbf{S}_{22} \mathbf{C} \Psi_{22}^{12} + \mathbf{D} \mathbf{S}_{22} \frac{\partial \mathbf{C}}{\partial \gamma_1} \Psi_{22}^{12} + \mathbf{D} \mathbf{S}_{22} \mathbf{C} \frac{\partial \Psi_{22}^{12}}{\partial \gamma_1} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \gamma_1} \mathbf{S}_{11} \mathbf{B} \Psi_{21}^{22} + \mathbf{A} \mathbf{S}_{11} \frac{\partial \mathbf{B}}{\partial \gamma_1} \Psi_{21}^{22} + \mathbf{A} \mathbf{S}_{11} \mathbf{B} \frac{\partial \Psi_{21}^{22}}{\partial \gamma_1} \right) + \text{Tr} \left(2 \frac{\partial \mathbf{B}}{\partial \gamma_1} \mathbf{S}_{12}^\top \mathbf{B} \Psi_{21}^{22} + \right. \\
& + \mathbf{B} \mathbf{S}_{12}^\top \mathbf{B} \frac{\partial \Psi_{21}^{22}}{\partial \gamma_1} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \gamma_1} \mathbf{S}_{11} \mathbf{B} \Psi_{22}^{22} + \mathbf{C} \mathbf{S}_{11} \frac{\partial \mathbf{B}}{\partial \gamma_1} \Psi_{22}^{22} + \mathbf{C} \mathbf{S}_{11} \mathbf{B} \frac{\partial \Psi_{22}^{22}}{\partial \gamma_1} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \gamma_1} \mathbf{S}_{12}^\top \mathbf{B} \Psi_{22}^{22} + \mathbf{D} \mathbf{S}_{12}^\top \frac{\partial \mathbf{B}}{\partial \gamma_1} \Psi_{22}^{22} + \mathbf{D} \mathbf{S}_{12}^\top \mathbf{B} \frac{\partial \Psi_{22}^{22}}{\partial \gamma_1} \right) + \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \gamma_1} \mathbf{S}_{12} \mathbf{D} \Psi_{21}^{22} + \right. \\
& + \mathbf{A} \mathbf{S}_{12} \frac{\partial \mathbf{D}}{\partial \gamma_1} \Psi_{21}^{22} + \mathbf{A} \mathbf{S}_{12} \mathbf{D} \frac{\partial \Psi_{21}^{22}}{\partial \gamma_1} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \gamma_1} \mathbf{S}_{22} \mathbf{D} \Psi_{21}^{22} + \mathbf{B} \mathbf{S}_{22} \frac{\partial \mathbf{D}}{\partial \gamma_1} \Psi_{21}^{22} + \right. \\
& + \mathbf{B} \mathbf{S}_{22} \mathbf{D} \frac{\partial \Psi_{21}^{22}}{\partial \gamma_1} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \gamma_1} \mathbf{S}_{12} \mathbf{D} \Psi_{22}^{22} + \mathbf{C} \mathbf{S}_{12} \frac{\partial \mathbf{D}}{\partial \gamma_1} \Psi_{22}^{22} + \mathbf{C} \mathbf{S}_{12} \mathbf{D} \frac{\partial \Psi_{22}^{22}}{\partial \gamma_1} \right) + \\
& + \text{Tr} \left(2 \frac{\partial \mathbf{D}}{\partial \gamma_1} \mathbf{S}_{22} \mathbf{D} \Psi_{22}^{22} + \mathbf{D} \mathbf{S}_{22} \mathbf{D} \frac{\partial \Psi_{22}^{22}}{\partial \gamma_1} \right) \left. \right]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \text{Termo 4}}{\partial \gamma_2} = & - \left[2\gamma_2 |\mathcal{T}| - \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \gamma_2} \Psi_{21}^{12} + \mathbf{A} \frac{\partial \Psi_{21}^{12}}{\partial \gamma_2} \right) - \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \gamma_2} \Psi_{22}^{12} + \mathbf{C} \frac{\partial \Psi_{22}^{12}}{\partial \gamma_2} \right) - \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \gamma_2} \Psi_{21}^{22} + \right. \\
& + \mathbf{B} \frac{\partial \Psi_{21}^{22}}{\partial \gamma_2} \left. \right) - \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \gamma_2} \Psi_{22}^{22} + \mathbf{D} \frac{\partial \Psi_{22}^{22}}{\partial \gamma_2} \right) + \text{Tr} \left(2 \frac{\partial \mathbf{A}}{\partial \gamma_2} \mathbf{S}_{11} \mathbf{A} \Psi_{21}^{12} + \mathbf{A} \mathbf{S}_{11} \mathbf{A} \frac{\partial \Psi_{21}^{12}}{\partial \gamma_2} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \gamma_2} \mathbf{S}_{12}^\top \mathbf{A} \Psi_{21}^{12} + \mathbf{B} \mathbf{S}_{12}^\top \frac{\partial \mathbf{A}}{\partial \gamma_2} \Psi_{21}^{12} + \mathbf{B} \mathbf{S}_{12}^\top \mathbf{A} \frac{\partial \Psi_{21}^{12}}{\partial \gamma_2} \right) + \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \gamma_2} \mathbf{S}_{11} \mathbf{A} \Psi_{22}^{12} + \right. \\
& + \mathbf{C} \mathbf{S}_{11} \frac{\partial \mathbf{A}}{\partial \gamma_2} \Psi_{22}^{12} + \mathbf{C} \mathbf{S}_{11} \mathbf{A} \frac{\partial \Psi_{22}^{12}}{\partial \gamma_2} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \gamma_2} \mathbf{S}_{12}^\top \mathbf{A} \Psi_{22}^{12} + \mathbf{D} \mathbf{S}_{12}^\top \frac{\partial \mathbf{A}}{\partial \gamma_2} \Psi_{22}^{12} + \right. \\
& + \mathbf{D} \mathbf{S}_{12}^\top \mathbf{A} \frac{\partial \Psi_{22}^{12}}{\partial \gamma_2} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \gamma_2} \mathbf{S}_{12} \mathbf{C} \Psi_{21}^{12} + \mathbf{A} \mathbf{S}_{12} \frac{\partial \mathbf{C}}{\partial \gamma_2} \Psi_{21}^{12} + \mathbf{A} \mathbf{S}_{12} \mathbf{C} \frac{\partial \Psi_{21}^{12}}{\partial \gamma_2} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \gamma_2} \mathbf{S}_{22} \mathbf{C} \Psi_{21}^{12} + \mathbf{B} \mathbf{S}_{22} \frac{\partial \mathbf{C}}{\partial \gamma_2} \Psi_{21}^{12} + \mathbf{B} \mathbf{S}_{22} \mathbf{C} \frac{\partial \Psi_{21}^{12}}{\partial \gamma_2} \right) + \text{Tr} \left(2 \frac{\partial \mathbf{C}}{\partial \gamma_2} \mathbf{S}_{12} \mathbf{C} \Psi_{22}^{12} + \right. \\
& + \mathbf{C} \mathbf{S}_{12} \mathbf{C} \frac{\partial \Psi_{22}^{12}}{\partial \gamma_2} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \gamma_2} \mathbf{S}_{22} \mathbf{C} \Psi_{22}^{12} + \mathbf{D} \mathbf{S}_{22} \frac{\partial \mathbf{C}}{\partial \gamma_2} \Psi_{22}^{12} + \mathbf{D} \mathbf{S}_{22} \mathbf{C} \frac{\partial \Psi_{22}^{12}}{\partial \gamma_2} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \gamma_2} \mathbf{S}_{11} \mathbf{B} \Psi_{21}^{22} + \mathbf{A} \mathbf{S}_{11} \frac{\partial \mathbf{B}}{\partial \gamma_2} \Psi_{21}^{22} + \mathbf{A} \mathbf{S}_{11} \mathbf{B} \frac{\partial \Psi_{21}^{22}}{\partial \gamma_2} \right) + \text{Tr} \left(2 \frac{\partial \mathbf{B}}{\partial \gamma_2} \mathbf{S}_{12}^\top \mathbf{B} \Psi_{21}^{22} + \right. \\
& + \mathbf{B} \mathbf{S}_{12}^\top \mathbf{B} \frac{\partial \Psi_{21}^{22}}{\partial \gamma_2} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \gamma_2} \mathbf{S}_{11} \mathbf{B} \Psi_{22}^{22} + \mathbf{C} \mathbf{S}_{11} \frac{\partial \mathbf{B}}{\partial \gamma_2} \Psi_{22}^{22} + \mathbf{C} \mathbf{S}_{11} \mathbf{B} \frac{\partial \Psi_{22}^{22}}{\partial \gamma_2} \right) + \\
& + \text{Tr} \left(\frac{\partial \mathbf{D}}{\partial \gamma_2} \mathbf{S}_{12}^\top \mathbf{B} \Psi_{22}^{22} + \mathbf{D} \mathbf{S}_{12}^\top \frac{\partial \mathbf{B}}{\partial \gamma_2} \Psi_{22}^{22} + \mathbf{D} \mathbf{S}_{12}^\top \mathbf{B} \frac{\partial \Psi_{22}^{22}}{\partial \gamma_2} \right) + \text{Tr} \left(\frac{\partial \mathbf{A}}{\partial \gamma_2} \mathbf{S}_{12} \mathbf{D} \Psi_{21}^{22} + \right. \\
& + \mathbf{A} \mathbf{S}_{12} \frac{\partial \mathbf{D}}{\partial \gamma_2} \Psi_{21}^{22} + \mathbf{A} \mathbf{S}_{12} \mathbf{D} \frac{\partial \Psi_{21}^{22}}{\partial \gamma_2} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{B}}{\partial \gamma_2} \mathbf{S}_{22} \mathbf{D} \Psi_{21}^{22} + \mathbf{B} \mathbf{S}_{22} \frac{\partial \mathbf{D}}{\partial \gamma_2} \Psi_{21}^{22} + \right. \\
& + \mathbf{B} \mathbf{S}_{22} \mathbf{D} \frac{\partial \Psi_{21}^{22}}{\partial \gamma_2} \left. \right) + \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \gamma_2} \mathbf{S}_{12} \mathbf{D} \Psi_{22}^{22} + \mathbf{C} \mathbf{S}_{12} \frac{\partial \mathbf{D}}{\partial \gamma_2} \Psi_{22}^{22} + \mathbf{C} \mathbf{S}_{12} \mathbf{D} \frac{\partial \Psi_{22}^{22}}{\partial \gamma_2} \right) + \\
& + \text{Tr} \left(2 \frac{\partial \mathbf{D}}{\partial \gamma_2} \mathbf{S}_{22} \mathbf{D} \Psi_{22}^{22} + \mathbf{D} \mathbf{S}_{22} \mathbf{D} \frac{\partial \Psi_{22}^{22}}{\partial \gamma_2} \right) \left. \right]
\end{aligned}$$

$$\frac{\partial \text{Termo 4}}{\partial \mathbf{S}_{11}} = - \left\{ (\mathbf{A} \Psi_{21}^{12} \mathbf{A})^\top + (\mathbf{A} \Psi_{22}^{12} \mathbf{C})^\top + (\mathbf{B} \Psi_{21}^{22} \mathbf{A})^\top + (\mathbf{B} \Psi_{22}^{22} \mathbf{C})^\top \right\}$$

$$\begin{aligned}
\frac{\partial \text{Termo 4}}{\partial \mathbf{S}_{12}} = & - \left\{ (\mathbf{C} \Psi_{21}^{12} \mathbf{A})^\top + (\mathbf{C} \Psi_{22}^{12} \mathbf{C})^\top + (\mathbf{D} \Psi_{21}^{22} \mathbf{A})^\top + (\mathbf{D} \Psi_{22}^{22} \mathbf{C})^\top + \right. \\
& + (\mathbf{A} \Psi_{21}^{12} \mathbf{B}) + (\mathbf{A} \Psi_{22}^{12} \mathbf{D}) + (\mathbf{B} \Psi_{21}^{22} \mathbf{B}) + (\mathbf{B} \Psi_{22}^{22} \mathbf{D}) \left. \right\}
\end{aligned}$$

$$\frac{\partial \text{Termo 4}}{\partial \mathbf{S}_{22}} = - \left\{ (\mathbf{C} \Psi_{21}^{12} \mathbf{B})^\top + (\mathbf{C} \Psi_{22}^{12} \mathbf{D})^\top + (\mathbf{D} \Psi_{21}^{22} \mathbf{B})^\top + (\mathbf{D} \Psi_{22}^{22} \mathbf{D})^\top \right\}$$

Os Termos 3 e 4 não dependem de \bar{u}_1 , \bar{u}_2 , \mathbf{m}_1 e \mathbf{m}_2 e, portanto, suas derivadas também serão nulas. As derivadas parciais do o Termo 5, aplicando a regra da cadeia, são dadas por:

$$\frac{\partial \text{Termo 5}}{\partial \alpha} = \sum_{j=1}^R \sum_{i=1}^{N_1} \tilde{G}' \left(-\frac{\tilde{\mu}_1^2(s_{1i})}{2\tilde{\sigma}_1^2(s_{1i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_1(s_{1i}) \frac{\partial \tilde{\mu}_1(s_{1i})}{\partial \alpha} \tilde{\sigma}_1^2(s_{1i})) - (\tilde{\mu}_1^2(s_{1i}) \frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \alpha})}{(\tilde{\sigma}_1^2(s_{1i}))^2} \right]$$

$$\frac{\partial \text{Termo 5}}{\partial \rho_{12}} = \sum_{i=1}^{N_1} \tilde{G}' \left(-\frac{\tilde{\mu}_1^2(s_{1i})}{2\tilde{\sigma}_1^2(s_{1i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_1(s_{1i}) \frac{\partial \tilde{\mu}_1(s_{1i})}{\partial \rho_{12}} \tilde{\sigma}_1^2(s_{1i})) - (\tilde{\mu}_1^2(s_{1i}) \frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \rho_{12}})}{(\tilde{\sigma}_1^2(s_{1i}))^2} \right]$$

$$\frac{\partial \text{Termo 5}}{\partial \gamma_1} = \sum_{i=1}^{N_1} \tilde{G}' \left(-\frac{\tilde{\mu}_1^2(s_{1i})}{2\tilde{\sigma}_1^2(s_{1i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_1(s_{1i}) \frac{\partial \tilde{\mu}_1(s_{1i})}{\partial \gamma_1} \tilde{\sigma}_1^2(s_{1i})) - (\tilde{\mu}_1^2(s_{1i}) \frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \gamma_1})}{(\tilde{\sigma}_1^2(s_{1i}))^2} \right]$$

$$\frac{\partial \text{Termo 5}}{\partial \gamma_2} = \sum_{i=1}^{N_1} \tilde{G}' \left(-\frac{\tilde{\mu}_1^2(s_{1i})}{2\tilde{\sigma}_1^2(s_{1i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_1(s_{1i}) \frac{\partial \tilde{\mu}_1(s_{1i})}{\partial \gamma_2} \tilde{\sigma}_1^2(s_{1i})) - (\tilde{\mu}_1^2(s_{1i}) \frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \gamma_2})}{(\tilde{\sigma}_1^2(s_{1i}))^2} \right]$$

$$\frac{\partial \text{Termo 5}}{\partial \mathbf{m}_1} = \sum_{i=1}^{N_1} \tilde{G}' \left(-\frac{\tilde{\mu}_1^2(s_{1i})}{2\tilde{\sigma}_1^2(s_{1i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_1(s_{1i}) \frac{\partial \tilde{\mu}_1(s_{1i})}{\partial \mathbf{m}_1} \tilde{\sigma}_1^2(s_{1i})) - (\tilde{\mu}_1^2(s_{1i}) \frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \mathbf{m}_1})}{(\tilde{\sigma}_1^2(s_{1i}))^2} \right]$$

$$\frac{\partial \text{Termo 5}}{\partial \mathbf{m}_2} = \sum_{i=1}^{N_1} \tilde{G}' \left(-\frac{\tilde{\mu}_1^2(s_{1i})}{2\tilde{\sigma}_1^2(s_{1i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_1(s_{1i}) \frac{\partial \tilde{\mu}_1(s_{1i})}{\partial \mathbf{m}_2} \tilde{\sigma}_1^2(s_{1i})) - (\tilde{\mu}_1^2(s_{1i}) \frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \mathbf{m}_2})}{(\tilde{\sigma}_1^2(s_{1i}))^2} \right]$$

$$\frac{\partial \text{Termo 5}}{\partial \mathbf{S}_{11}} = \sum_{i=1}^{N_1} \tilde{G}' \left(-\frac{\tilde{\mu}_1^2(s_{1i})}{2\tilde{\sigma}_1^2(s_{1i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_1(s_{1i}) \frac{\partial \tilde{\mu}_1(s_{1i})}{\partial \mathbf{S}_{11}} \tilde{\sigma}_1^2(s_{1i})) - (\tilde{\mu}_1^2(s_{1i}) \frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \mathbf{S}_{11}})}{(\tilde{\sigma}_1^2(s_{1i}))^2} \right]$$

$$\frac{\partial \text{Termo 5}}{\partial \mathbf{S}_{12}} = \sum_{i=1}^{N_1} \tilde{G}' \left(-\frac{\tilde{\mu}_1^2(s_{1i})}{2\tilde{\sigma}_1^2(s_{1i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_1(s_{1i}) \frac{\partial \tilde{\mu}_1(s_{1i})}{\partial \mathbf{S}_{12}} \tilde{\sigma}_1^2(s_{1i})) - (\tilde{\mu}_1^2(s_{1i}) \frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \mathbf{S}_{12}})}{(\tilde{\sigma}_1^2(s_{1i}))^2} \right]$$

$$\frac{\partial \text{Termo 5}}{\partial \mathbf{S}_{22}} = \sum_{i=1}^{N_1} \tilde{G}' \left(-\frac{\tilde{\mu}_1^2(s_{1i})}{2\tilde{\sigma}_1^2(s_{1i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_1(s_{1i}) \frac{\partial \tilde{\mu}_1(s_{1i})}{\partial \mathbf{S}_{22}} \tilde{\sigma}_1^2(s_{1i})) - (\tilde{\mu}_1^2(s_{1i}) \frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \mathbf{S}_{22}})}{(\tilde{\sigma}_1^2(s_{1i}))^2} \right]$$

As derivadas parciais de $\tilde{\mu}_1(s_{1i})$, $\tilde{\mu}_2(s_{2i})$ e $\tilde{\sigma}_1^2(s_{1i})$, $\tilde{\sigma}_2^2(s_{2i})$ em relação aos hiperparâmetros do modelo estão apresentadas na subseção C.4 e C.5, respectivamente. Seguindo a mesma regra aplicada ao Termo 5, as derivadas parciais do Termo 6 são:

$$\frac{\partial \text{Termo 6}}{\partial \alpha} = \sum_{i=1}^{N_2} \tilde{G}' \left(-\frac{\tilde{\mu}_2^2(s_{2i})}{2\tilde{\sigma}_2^2(s_{2i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_2(s_{2i}) \frac{\partial \tilde{\mu}_2(s_{2i})}{\partial \alpha} \tilde{\sigma}_2^2(s_{2i})) - (\tilde{\mu}_2^2(s_{2i}) \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \alpha})}{(\tilde{\sigma}_2^2(s_{2i}))^2} \right]$$

$$\frac{\partial \text{Termo 6}}{\partial \rho_{12}} = \sum_{i=1}^{N_2} \tilde{G}' \left(-\frac{\tilde{\mu}_2^2(s_{2i})}{2\tilde{\sigma}_2^2(s_{2i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_2(s_{2i}) \frac{\partial \tilde{\mu}_2(s_{2i})}{\partial \rho_{12}} \tilde{\sigma}_2^2(s_{2i})) - (\tilde{\mu}_2^2(s_{2i}) \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \rho_{12}})}{(\tilde{\sigma}_2^2(s_{2i}))^2} \right]$$

$$\frac{\partial \text{Termo 6}}{\partial \gamma_1} = \sum_{i=1}^{N_2} \tilde{G}' \left(-\frac{\tilde{\mu}_2^2(s_{2i})}{2\tilde{\sigma}_2^2(s_{2i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_2(s_{2i}) \frac{\partial \tilde{\mu}_2(s_{2i})}{\partial \gamma_1} \tilde{\sigma}_2^2(s_{2i})) - (\tilde{\mu}_2^2(s_{2i}) \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \gamma_1})}{(\tilde{\sigma}_2^2(s_{2i}))^2} \right]$$

$$\frac{\partial \text{Termo 6}}{\partial \gamma_2} = \sum_{i=1}^{N_2} \tilde{G}' \left(-\frac{\tilde{\mu}_2^2(s_{2i})}{2\tilde{\sigma}_2^2(s_{2i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_2(s_{2i}) \frac{\partial \tilde{\mu}_2(s_{2i})}{\partial \gamma_2} \tilde{\sigma}_2^2(s_{2i})) - (\tilde{\mu}_2^2(s_{2i}) \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \gamma_2})}{(\tilde{\sigma}_2^2(s_{2i}))^2} \right]$$

$$\frac{\partial \text{Termo 6}}{\partial \mathbf{m}_1} = \sum_{i=1}^{N_2} \tilde{G}' \left(-\frac{\tilde{\mu}_2^2(s_{2i})}{2\tilde{\sigma}_2^2(s_{2i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_2(s_{2i}) \frac{\partial \tilde{\mu}_2(s_{2i})}{\partial \mathbf{m}_1} \tilde{\sigma}_2^2(s_{2i})) - (\tilde{\mu}_2^2(s_{2i}) \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \mathbf{m}_1})}{(\tilde{\sigma}_2^2(s_{2i}))^2} \right]$$

$$\frac{\partial \text{Termo 6}}{\partial \mathbf{m}_2} = \sum_{i=1}^{N_2} \tilde{G}' \left(-\frac{\tilde{\mu}_2^2(s_{2i})}{2\tilde{\sigma}_2^2(s_{2i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_2(s_{2i}) \frac{\partial \tilde{\mu}_2(s_{2i})}{\partial \mathbf{m}_2} \tilde{\sigma}_2^2(s_{2i})) - (\tilde{\mu}_2^2(s_{2i}) \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \mathbf{m}_2})}{(\tilde{\sigma}_2^2(s_{2i}))^2} \right]$$

$$\frac{\partial \text{Termo 6}}{\partial \mathbf{S}_{11}} = \sum_{i=1}^{N_2} \tilde{G}' \left(-\frac{\tilde{\mu}_2^2(s_{2i})}{2\tilde{\sigma}_2^2(s_{2i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_2(s_{2i}) \frac{\partial \tilde{\mu}_2(s_{2i})}{\partial \mathbf{S}_{11}} \tilde{\sigma}_2^2(s_{2i})) - (\tilde{\mu}_2^2(s_{2i}) \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \mathbf{S}_{11}})}{(\tilde{\sigma}_2^2(s_{2i}))^2} \right]$$

$$\frac{\partial \text{Termo 6}}{\partial \mathbf{S}_{12}} = \sum_{i=1}^{N_2} \tilde{G}' \left(-\frac{\tilde{\mu}_2^2(s_{2i})}{2\tilde{\sigma}_2^2(s_{2i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_2(s_{2i}) \frac{\partial \tilde{\mu}_2(s_{2i})}{\partial \mathbf{S}_{12}} \tilde{\sigma}_2^2(s_{2i})) - (\tilde{\mu}_2^2(s_{2i}) \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \mathbf{S}_{12}})}{(\tilde{\sigma}_2^2(s_{2i}))^2} \right]$$

$$\frac{\partial \text{Termo 6}}{\partial \mathbf{S}_{22}} = \sum_{i=1}^{N_2} \tilde{G}' \left(-\frac{\tilde{\mu}_2^2(s_{2i})}{2\tilde{\sigma}_2^2(s_{2i})} \right) \left[\frac{1}{2} \frac{(2\tilde{\mu}_2(s_{2i}) \frac{\partial \tilde{\mu}_2(s_{2i})}{\partial \mathbf{S}_{22}} \tilde{\sigma}_2^2(s_{2i})) - (\tilde{\mu}_2^2(s_{2i}) \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \mathbf{S}_{22}})}{(\tilde{\sigma}_2^2(s_{2i}))^2} \right]$$

Aplicando a regra da cadeia, as derivadas parciais do Termo 7 em relação aos hiperparâmetros do modelo são:

$$\frac{\partial \text{Termo 7}}{\partial \alpha} = \sum_{i=1}^{N_1} \frac{1}{\tilde{\sigma}_1^2(s_{1i})} \frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \alpha}$$

$$\frac{\partial \text{Termo 7}}{\partial \rho_{12}} = \sum_{i=1}^{N_1} \frac{1}{\tilde{\sigma}_1^2(s_{1i})} \frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \rho_{12}}$$

$$\frac{\partial \text{Termo 7}}{\partial \gamma_1} = \sum_{i=1}^{N_1} \frac{1}{\tilde{\sigma}_1^2(s_{1i})} \frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \gamma_1}$$

$$\frac{\partial \text{Termo 7}}{\partial \gamma_2} = \sum_{i=1}^{N_1} \frac{1}{\tilde{\sigma}_1^2(s_{1i})} \frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \gamma_2}$$

$$\frac{\partial \text{Termo 7}}{\partial \mathbf{S}_{11}} = \sum_{i=1}^{N_1} \frac{1}{\tilde{\sigma}_1^2(s_{1i})} \frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \mathbf{S}_{11}}$$

$$\frac{\partial \text{Termo 7}}{\partial \mathbf{S}_{12}} = \sum_{i=1}^{N_1} \frac{1}{\tilde{\sigma}_1^2(s_{1i})} \frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \mathbf{S}_{12}}$$

$$\frac{\partial \text{Termo 7}}{\partial \mathbf{S}_{22}} = \sum_{i=1}^{N_1} \frac{1}{\tilde{\sigma}_1^2(s_{1i})} \frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \mathbf{S}_{22}}$$

O Termo 7 não depende dos hiperparâmetros \bar{u}_1 , \bar{u}_2 , \mathbf{m}_1 e \mathbf{m}_2 e, desse modo, suas derivadas parciais serão 0. De maneira análoga, as derivadas do Termo 8 são dadas por

$$\frac{\partial \text{Termo 8}}{\partial \alpha} = \sum_{i=1}^{N_2} \frac{1}{\tilde{\sigma}_2^2(s_{2i})} \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \alpha}$$

$$\frac{\partial \text{Termo 8}}{\partial \rho_{12}} = \sum_{i=1}^{N_2} \frac{1}{\tilde{\sigma}_2^2(s_{2i})} \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \rho_{12}}$$

$$\frac{\partial \text{Termo 8}}{\partial \gamma_1} = \sum_{i=1}^{N_2} \frac{1}{\tilde{\sigma}_2^2(s_{2i})} \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \gamma_1}$$

$$\frac{\partial \text{Termo 8}}{\partial \gamma_2} = \sum_{i=1}^{N_2} \frac{1}{\tilde{\sigma}_2^2(s_{2i})} \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \gamma_2}$$

$$\frac{\partial \text{Termo 8}}{\partial \mathbf{S}_{11}} = \sum_{i=1}^{N_2} \frac{1}{\tilde{\sigma}_2^2(s_{2i})} \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \mathbf{S}_{11}}$$

$$\frac{\partial \text{Termo 8}}{\partial \mathbf{S}_{12}} = \sum_{i=1}^{N_2} \frac{1}{\tilde{\sigma}_2^2(s_{2i})} \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \mathbf{S}_{12}}$$

$$\frac{\partial \text{Termo 8}}{\partial \mathbf{S}_{22}} = \sum_{i=1}^{N_2} \frac{1}{\tilde{\sigma}_2^2(s_{2i})} \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \mathbf{S}_{22}}$$

Por fim, as derivadas do Termo 9 são dadas por

$$\begin{aligned}
\frac{\partial \text{Termo 9}}{\partial \alpha} = & -\frac{1}{2} \left\{ Tr \left[\mathbf{K}_{11}^{(z,z)^{-1}} \frac{\partial \mathbf{K}_{11}^{(z,z)}}{\partial \alpha} \right] + Tr \left[\mathbf{D} \left(\frac{\partial \mathbf{K}_{22}^{(z,z)}}{\partial \alpha} - \frac{\partial \mathbf{K}_{21}^{(z,z)}}{\partial \alpha} \mathbf{K}_{11}^{(z,z)^{-1}} \mathbf{K}_{12}^{(z,z)} - \right. \right. \\
& \left. \left. - \mathbf{K}_{21}^{(z,z)} \frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \alpha} \mathbf{K}_{12}^{(z,z)} - \mathbf{K}_{21}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} \frac{\partial \mathbf{K}_{12}^{(z,z)}}{\partial \alpha} \right) \right] + Tr \left[\frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{S}_{11} + \right. \\
& \left. + \frac{\partial \mathbf{C}}{\partial \alpha} \mathbf{S}_{12} + \frac{\partial \mathbf{B}}{\partial \alpha} \mathbf{S}_{12}^\top + \frac{\partial \mathbf{D}}{\partial \alpha} \mathbf{S}_{22} \right] + (\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1)^\top \frac{\partial \mathbf{A}}{\partial \alpha} (\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1) + \\
& + (\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2)^\top \frac{\partial \mathbf{C}}{\partial \alpha} (\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1) + (\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1)^\top \frac{\partial \mathbf{B}}{\partial \alpha} (\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2) + \\
& \left. (\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2)^\top \frac{\partial \mathbf{D}}{\partial \alpha} (\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2) \right\}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \text{Termo 9}}{\partial \rho_{12}} = & -\frac{1}{2} \left\{ Tr \left[\mathbf{K}_{11}^{(z,z)^{-1}} \frac{\partial \mathbf{K}_{11}^{(z,z)}}{\partial \rho_{12}} \right] + Tr \left[\mathbf{D} \left(\frac{\partial \mathbf{K}_{22}^{(z,z)}}{\partial \rho_{12}} - \frac{\partial \mathbf{K}_{21}^{(z,z)}}{\partial \rho_{12}} \mathbf{K}_{11}^{(z,z)^{-1}} \mathbf{K}_{12}^{(z,z)} - \right. \right. \\
& \left. \left. - \mathbf{K}_{21}^{(z,z)} \frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \rho_{12}} \mathbf{K}_{12}^{(z,z)} - \mathbf{K}_{21}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} \frac{\partial \mathbf{K}_{12}^{(z,z)}}{\partial \rho_{12}} \right) \right] + Tr \left[\frac{\partial \mathbf{A}}{\partial \rho_{12}} \mathbf{S}_{11} + \right. \\
& \left. + \frac{\partial \mathbf{C}}{\partial \rho_{12}} \mathbf{S}_{12} + \frac{\partial \mathbf{B}}{\partial \rho_{12}} \mathbf{S}_{12}^\top + \frac{\partial \mathbf{D}}{\partial \rho_{12}} \mathbf{S}_{22} \right] + (\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1)^\top \frac{\partial \mathbf{A}}{\partial \rho_{12}} (\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1) + \\
& + (\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2)^\top \frac{\partial \mathbf{C}}{\partial \rho_{12}} (\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1) + (\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1)^\top \frac{\partial \mathbf{B}}{\partial \rho_{12}} (\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2) + \\
& \left. (\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2)^\top \frac{\partial \mathbf{D}}{\partial \rho_{12}} (\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2) \right\}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \text{Termo 9}}{\partial \gamma_1} = & -\frac{1}{2} \left\{ Tr \left[\mathbf{K}_{11}^{(z,z)^{-1}} \frac{\partial \mathbf{K}_{11}^{(z,z)}}{\partial \gamma_1} \right] + Tr \left[\mathbf{D} \left(\frac{\partial \mathbf{K}_{22}^{(z,z)}}{\partial \gamma_1} - \frac{\partial \mathbf{K}_{21}^{(z,z)}}{\partial \gamma_1} \mathbf{K}_{11}^{(z,z)^{-1}} \mathbf{K}_{12}^{(z,z)} - \right. \right. \\
& \left. \left. - \mathbf{K}_{21}^{(z,z)} \frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \gamma_1} \mathbf{K}_{12}^{(z,z)} - \mathbf{K}_{21}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} \frac{\partial \mathbf{K}_{12}^{(z,z)}}{\partial \gamma_1} \right) \right] + Tr \left[\frac{\partial \mathbf{A}}{\partial \gamma_1} \mathbf{S}_{11} + \right. \\
& \left. + \frac{\partial \mathbf{C}}{\partial \gamma_1} \mathbf{S}_{12} + \frac{\partial \mathbf{B}}{\partial \gamma_1} \mathbf{S}_{12}^\top + \frac{\partial \mathbf{D}}{\partial \gamma_1} \mathbf{S}_{22} \right] + (\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1)^\top \frac{\partial \mathbf{A}}{\partial \gamma_1} (\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1) + \\
& + (\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2)^\top \frac{\partial \mathbf{C}}{\partial \gamma_1} (\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1) + (\mathbf{m}_1 - \vec{\mathbf{1}}\bar{u}_1)^\top \frac{\partial \mathbf{B}}{\partial \gamma_1} (\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2) + \\
& \left. (\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2)^\top \frac{\partial \mathbf{D}}{\partial \gamma_1} (\mathbf{m}_2 - \vec{\mathbf{1}}\bar{u}_2) \right\}
\end{aligned}$$

$$\begin{aligned} \frac{\partial \text{Termo 9}}{\partial \gamma_2} = & -\frac{1}{2} \left\{ \text{Tr} \left[\mathbf{K}_{11}^{(z,z)^{-1}} \frac{\partial \mathbf{K}_{11}^{(z,z)}}{\partial \gamma_2} \right] + \text{Tr} \left[\mathbf{D} \left(\frac{\partial \mathbf{K}_{22}^{(z,z)}}{\partial \gamma_2} - \frac{\partial \mathbf{K}_{21}^{(z,z)}}{\partial \gamma_2} \mathbf{K}_{11}^{(z,z)^{-1}} \mathbf{K}_{12}^{(z,z)} - \right. \right. \\ & \left. \left. - \mathbf{K}_{21}^{(z,z)} \frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \gamma_2} \mathbf{K}_{12}^{(z,z)} - \mathbf{K}_{21}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} \frac{\partial \mathbf{K}_{12}^{(z,z)}}{\partial \gamma_2} \right) \right] + \text{Tr} \left[\frac{\partial \mathbf{A}}{\partial \gamma_2} \mathbf{S}_{11} + \right. \\ & \left. + \frac{\partial \mathbf{C}}{\partial \gamma_2} \mathbf{S}_{12} + \frac{\partial \mathbf{B}}{\partial \gamma_2} \mathbf{S}_{12}^\top + \frac{\partial \mathbf{D}}{\partial \gamma_2} \mathbf{S}_{22} \right] + (\mathbf{m}_1 - \vec{\mathbf{1}} \bar{u}_1)^\top \frac{\partial \mathbf{A}}{\partial \gamma_2} (\mathbf{m}_1 - \vec{\mathbf{1}} \bar{u}_1) + \\ & + (\mathbf{m}_2 - \vec{\mathbf{1}} \bar{u}_2)^\top \frac{\partial \mathbf{C}}{\partial \gamma_2} (\mathbf{m}_1 - \vec{\mathbf{1}} \bar{u}_1) + (\mathbf{m}_1 - \vec{\mathbf{1}} \bar{u}_1)^\top \frac{\partial \mathbf{B}}{\partial \gamma_2} (\mathbf{m}_2 - \vec{\mathbf{1}} \bar{u}_2) + \\ & \left. (\mathbf{m}_2 - \vec{\mathbf{1}} \bar{u}_2)^\top \frac{\partial \mathbf{D}}{\partial \gamma_2} (\mathbf{m}_2 - \vec{\mathbf{1}} \bar{u}_2) \right\} \end{aligned}$$

$$\frac{\partial \text{Termo 9}}{\partial \bar{u}_1} = -\frac{1}{2} \left\{ \bar{u}_2 (\vec{\mathbf{1}}^\top \mathbf{C} \vec{\mathbf{1}} + \vec{\mathbf{1}}^\top \mathbf{B} \vec{\mathbf{1}}) - \mathbf{m}_2^\top \mathbf{C} \vec{\mathbf{1}} - \vec{\mathbf{1}}^\top \mathbf{B} \mathbf{m}_2 - 2 \mathbf{m}_1^\top \mathbf{A} \vec{\mathbf{1}} + 2 \bar{u}_1 \vec{\mathbf{1}}^\top \mathbf{A} \vec{\mathbf{1}} \right\}$$

$$\frac{\partial \text{Termo 9}}{\partial \bar{u}_2} = -\frac{1}{2} \left\{ \bar{u}_1 (\vec{\mathbf{1}}^\top \mathbf{C} \vec{\mathbf{1}} + \vec{\mathbf{1}}^\top \mathbf{B} \vec{\mathbf{1}}) - \mathbf{m}_1^\top \mathbf{B} \vec{\mathbf{1}} - \vec{\mathbf{1}}^\top \mathbf{C} \mathbf{m}_1 - 2 \mathbf{m}_2^\top \mathbf{D} \vec{\mathbf{1}} + 2 \bar{u}_2 \vec{\mathbf{1}}^\top \mathbf{D} \vec{\mathbf{1}} \right\}$$

$$\frac{\partial \text{Termo 9}}{\partial \mathbf{m}_1} = -\frac{1}{2} \left\{ [\mathbf{A} (\mathbf{m}_1 - \vec{\mathbf{1}} \bar{u}_1) + \mathbf{A}^\top (\mathbf{m}_1 - \vec{\mathbf{1}} \bar{u}_1)] + \mathbf{C}^\top (\mathbf{m}_2 - \vec{\mathbf{1}} \bar{u}_2) + \mathbf{B} (\mathbf{m}_2 - \vec{\mathbf{1}} \bar{u}_2) \right\}$$

$$\frac{\partial \text{Termo 9}}{\partial \mathbf{m}_2} = -\frac{1}{2} \left\{ \mathbf{C} (\mathbf{m}_1 - \vec{\mathbf{1}} \bar{u}_1) + \mathbf{B}^\top (\mathbf{m}_1 - \vec{\mathbf{1}} \bar{u}_1) + [\mathbf{D} (\mathbf{m}_2 - \vec{\mathbf{1}} \bar{u}_2) + \mathbf{D}^\top (\mathbf{m}_2 - \vec{\mathbf{1}} \bar{u}_2)] \right\}$$

$$\frac{\partial \text{Termo 9}}{\partial \mathbf{S}_{11}} = -\frac{1}{2} \left\{ \mathbf{A}^\top - \mathbf{S}_{11}^{-1} + \mathbf{S}_{11}^{-1} \mathbf{S}_{12} (\mathbf{S}_{22} - \mathbf{S}_{12}^\top \mathbf{S}_{11}^{-1} \mathbf{S}_{12})^{-1} \mathbf{S}_{12}^\top \mathbf{S}_{11}^{-1} \right\}$$

$$\frac{\partial \text{Termo 9}}{\partial \mathbf{S}_{12}} = -\frac{1}{2} \left\{ \mathbf{C}^\top + \mathbf{B} + 2 \times \left(\mathbf{S}_{11}^{-1} \mathbf{S}_{12} (\mathbf{S}_{22} - \mathbf{S}_{12}^\top \mathbf{S}_{11}^{-1} \mathbf{S}_{12})^{-1} \right) \right\}$$

$$\frac{\partial \text{Termo 9}}{\partial \mathbf{S}_{22}} = -\frac{1}{2} \left\{ \mathbf{D}^\top - (\mathbf{S}_{22} - \mathbf{S}_{12}^\top \mathbf{S}_{11}^{-1} \mathbf{S}_{12})^{-1} \right\}$$

C.3 Derivadas parciais da inversa da matriz particionada $\mathbf{K}^{(z,z)}$ em relação aos parâmetros

Nessa subseção, são apresentadas as derivadas das submatrizes que compõe a inversa da matriz particionada $\mathbf{K}^{(z,z)}$, sendo $\mathbf{K}^{(z,z)^{-1}} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$ e $\mathbf{A} = (\mathbf{K}_{11}^{(z,z)})^{-1} + (\mathbf{K}_{11}^{(z,z)})^{-1} \mathbf{K}_{12}^{(z,z)} (\mathbf{K}_{22}^{(z,z)} - \mathbf{K}_{21}^{(z,z)} (\mathbf{K}_{11}^{(z,z)})^{-1} \mathbf{K}_{12}^{(z,z)})^{-1} \mathbf{K}_{21}^{(z,z)} (\mathbf{K}_{11}^{(z,z)})^{-1}$, $\mathbf{B} = -(\mathbf{K}_{11}^{(z,z)})^{-1} \mathbf{K}_{12}^{(z,z)} (\mathbf{K}_{22}^{(z,z)} - \mathbf{K}_{21}^{(z,z)} (\mathbf{K}_{11}^{(z,z)})^{-1} \mathbf{K}_{12}^{(z,z)})^{-1}$,

$$\mathbf{C} = -\left(\mathbf{K}_{22}^{(z,z)} - \mathbf{K}_{21}^{(z,z)} (\mathbf{K}_{11}^{(z,z)})^{-1} \mathbf{K}_{12}^{(z,z)}\right)^{-1} \mathbf{K}_{21}^{(z,z)} (\mathbf{K}_{11}^{(z,z)})^{-1} \mathbf{e} \mathbf{D} = \left(\mathbf{K}_{22}^{(z,z)} - \mathbf{K}_{21}^{(z,z)} (\mathbf{K}_{11}^{(z,z)})^{-1} \mathbf{K}_{12}^{(z,z)}\right)^{-1}.$$

Essas matrizes não dependem de \bar{u}_1 , \bar{u}_2 , \mathbf{m}_1 , \mathbf{m}_2 , \mathbf{S}_{11} , \mathbf{S}_{12} , \mathbf{S}_{12}^\top e \mathbf{S}_{22} . Portanto, suas derivadas serão nulas.

$$\begin{aligned} \frac{\partial \mathbf{A}}{\partial \alpha} &= \frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \alpha} + 2 \frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \alpha} \mathbf{K}_{12}^{(z,z)} \mathbf{D} \mathbf{K}_{21}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} + 2 \mathbf{K}_{11}^{(z,z)^{-1}} \frac{\partial \mathbf{K}_{12}^{(z,z)}}{\partial \alpha} \mathbf{D} \mathbf{K}_{21}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} + \\ &+ 2 \mathbf{K}_{11}^{(z,z)^{-1}} \mathbf{K}_{12}^{(z,z)} \frac{\partial \mathbf{D}}{\partial \alpha} \mathbf{K}_{21}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} \end{aligned}$$

$$\begin{aligned} \frac{\partial \mathbf{A}}{\partial \rho_{12}} &= \frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \rho_{12}} + 2 \frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \rho_{12}} \mathbf{K}_{12}^{(z,z)} \mathbf{D} \mathbf{K}_{21}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} + 2 \mathbf{K}_{11}^{(z,z)^{-1}} \frac{\partial \mathbf{K}_{12}^{(z,z)}}{\partial \rho_{12}} \mathbf{D} \mathbf{K}_{21}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} + \\ &+ 2 \mathbf{K}_{11}^{(z,z)^{-1}} \mathbf{K}_{12}^{(z,z)} \frac{\partial \mathbf{D}}{\partial \rho_{12}} \mathbf{K}_{21}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} \end{aligned}$$

$$\begin{aligned} \frac{\partial \mathbf{A}}{\partial \gamma_1} &= \frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \gamma_1} + 2 \frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \gamma_1} \mathbf{K}_{12}^{(z,z)} \mathbf{D} \mathbf{K}_{21}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} + 2 \mathbf{K}_{11}^{(z,z)^{-1}} \frac{\partial \mathbf{K}_{12}^{(z,z)}}{\partial \gamma_1} \mathbf{D} \mathbf{K}_{21}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} + \\ &+ 2 \mathbf{K}_{11}^{(z,z)^{-1}} \mathbf{K}_{12}^{(z,z)} \frac{\partial \mathbf{D}}{\partial \gamma_1} \mathbf{K}_{21}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} \end{aligned}$$

$$\begin{aligned} \frac{\partial \mathbf{A}}{\partial \gamma_2} &= \frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \gamma_2} + 2 \frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \gamma_2} \mathbf{K}_{12}^{(z,z)} \mathbf{D} \mathbf{K}_{21}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} + 2 \mathbf{K}_{11}^{(z,z)^{-1}} \frac{\partial \mathbf{K}_{12}^{(z,z)}}{\partial \gamma_2} \mathbf{D} \mathbf{K}_{21}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} + \\ &+ 2 \mathbf{K}_{11}^{(z,z)^{-1}} \mathbf{K}_{12}^{(z,z)} \frac{\partial \mathbf{D}}{\partial \gamma_2} \mathbf{K}_{21}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} \end{aligned}$$

$$\frac{\partial \mathbf{B}}{\partial \alpha} = -\frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \alpha} \mathbf{K}_{12}^{(z,z)} \mathbf{D} - \mathbf{K}_{11}^{(z,z)^{-1}} \frac{\partial \mathbf{K}_{12}^{(z,z)}}{\partial \alpha} \mathbf{D} - \mathbf{K}_{11}^{(z,z)^{-1}} \mathbf{K}_{12}^{(z,z)} \frac{\partial \mathbf{D}}{\partial \alpha}$$

$$\frac{\partial \mathbf{B}}{\partial \rho_{12}} = -\frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \rho_{12}} \mathbf{K}_{12}^{(z,z)} \mathbf{D} - \mathbf{K}_{11}^{(z,z)^{-1}} \frac{\partial \mathbf{K}_{12}^{(z,z)}}{\partial \rho_{12}} \mathbf{D} - \mathbf{K}_{11}^{(z,z)^{-1}} \mathbf{K}_{12}^{(z,z)} \frac{\partial \mathbf{D}}{\partial \rho_{12}}$$

$$\frac{\partial \mathbf{B}}{\partial \gamma_1} = -\frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \gamma_1} \mathbf{K}_{12}^{(z,z)} \mathbf{D} - \mathbf{K}_{11}^{(z,z)^{-1}} \frac{\partial \mathbf{K}_{12}^{(z,z)}}{\partial \gamma_1} \mathbf{D} - \mathbf{K}_{11}^{(z,z)^{-1}} \mathbf{K}_{12}^{(z,z)} \frac{\partial \mathbf{D}}{\partial \gamma_1}$$

$$\frac{\partial \mathbf{B}}{\partial \gamma_2} = -\frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \gamma_2} \mathbf{K}_{12}^{(z,z)} \mathbf{D} - \mathbf{K}_{11}^{(z,z)^{-1}} \frac{\partial \mathbf{K}_{12}^{(z,z)}}{\partial \gamma_2} \mathbf{D} - \mathbf{K}_{11}^{(z,z)^{-1}} \mathbf{K}_{12}^{(z,z)} \frac{\partial \mathbf{D}}{\partial \gamma_2}$$

$$\frac{\partial \mathbf{C}}{\partial \alpha} = -\frac{\partial \mathbf{D}}{\partial \alpha} \mathbf{K}_{12}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} - \mathbf{D} \frac{\partial \mathbf{K}_{21}^{(z,z)}}{\partial \alpha} \mathbf{K}_{11}^{(z,z)^{-1}} - \mathbf{D} \mathbf{K}_{21}^{(z,z)} \frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \alpha}$$

$$\frac{\partial \mathbf{C}}{\partial \rho_{12}} = -\frac{\partial \mathbf{D}}{\partial \rho_{12}} \mathbf{K}_{12}^{(z,z)} \mathbf{K}_{11}^{(z,z)^{-1}} - \mathbf{D} \frac{\partial \mathbf{K}_{21}^{(z,z)}}{\partial \rho_{12}} \mathbf{K}_{11}^{(z,z)^{-1}} - \mathbf{D} \mathbf{K}_{21}^{(z,z)} \frac{\partial \mathbf{K}_{11}^{(z,z)^{-1}}}{\partial \rho_{12}}$$

$$\frac{\partial \mathbf{C}}{\partial \gamma_1} = -\frac{\partial \mathbf{D}}{\partial \gamma_1} \mathbf{K}_{12}^{(z,z)} \mathbf{K}_{11}^{(z,z)-1} - \mathbf{D} \frac{\partial \mathbf{K}_{21}^{(z,z)}}{\partial \gamma_1} \mathbf{K}_{11}^{(z,z)-1} - \mathbf{D} \mathbf{K}_{21}^{(z,z)} \frac{\partial \mathbf{K}_{11}^{(z,z)-1}}{\partial \gamma_1}$$

$$\frac{\partial \mathbf{C}}{\partial \gamma_2} = -\frac{\partial \mathbf{D}}{\partial \gamma_2} \mathbf{K}_{12}^{(z,z)} \mathbf{K}_{11}^{(z,z)-1} - \mathbf{D} \frac{\partial \mathbf{K}_{21}^{(z,z)}}{\partial \gamma_2} \mathbf{K}_{11}^{(z,z)-1} - \mathbf{D} \mathbf{K}_{21}^{(z,z)} \frac{\partial \mathbf{K}_{11}^{(z,z)-1}}{\partial \gamma_2}$$

$$\frac{\partial \mathbf{D}}{\partial \alpha} = \mathbf{D} \frac{\partial \mathbf{D}^*}{\partial \alpha} \mathbf{D}, \quad \text{em que} \quad \frac{\partial \mathbf{D}^*}{\partial \alpha} = 2 \frac{\partial \mathbf{K}_{21}^{(z,z)}}{\partial \alpha} \mathbf{K}_{11}^{(z,z)-1} \mathbf{K}_{12}^{(z,z)} + \mathbf{K}_{21}^{(z,z)} \frac{\partial \mathbf{K}_{11}^{(z,z)-1}}{\partial \alpha} \mathbf{K}_{12}^{(z,z)}$$

$$\frac{\partial \mathbf{D}}{\partial \rho_{12}} = \mathbf{D} \frac{\partial \mathbf{D}^*}{\partial \rho_{12}} \mathbf{D}, \quad \text{em que} \quad \frac{\partial \mathbf{D}^*}{\partial \rho_{12}} = 2 \frac{\partial \mathbf{K}_{21}^{(z,z)}}{\partial \rho_{12}} \mathbf{K}_{11}^{(z,z)-1} \mathbf{K}_{12}^{(z,z)} + \mathbf{K}_{21}^{(z,z)} \frac{\partial \mathbf{K}_{11}^{(z,z)-1}}{\partial \rho_{12}} \mathbf{K}_{12}^{(z,z)}$$

$$\frac{\partial \mathbf{D}}{\partial \gamma_1} = \mathbf{D} \frac{\partial \mathbf{D}^*}{\partial \gamma_1} \mathbf{D}, \quad \text{em que} \quad \frac{\partial \mathbf{D}^*}{\partial \gamma_1} = 2 \frac{\partial \mathbf{K}_{21}^{(z,z)}}{\partial \gamma_1} \mathbf{K}_{11}^{(z,z)-1} \mathbf{K}_{12}^{(z,z)} + \mathbf{K}_{21}^{(z,z)} \frac{\partial \mathbf{K}_{11}^{(z,z)-1}}{\partial \gamma_1} \mathbf{K}_{12}^{(z,z)}$$

$$\frac{\partial \mathbf{D}}{\partial \gamma_2} = \mathbf{D} \frac{\partial \mathbf{D}^*}{\partial \gamma_2} \mathbf{D}, \quad \text{em que} \quad \frac{\partial \mathbf{D}^*}{\partial \gamma_2} = 2 \frac{\partial \mathbf{K}_{21}^{(z,z)}}{\partial \gamma_2} \mathbf{K}_{11}^{(z,z)-1} \mathbf{K}_{12}^{(z,z)} + \mathbf{K}_{21}^{(z,z)} \frac{\partial \mathbf{K}_{11}^{(z,z)-1}}{\partial \gamma_2} \mathbf{K}_{12}^{(z,z)}$$

C.4 Cálculo das derivadas parciais de $\tilde{\mu}_1(s_{1i})$ e $\tilde{\mu}_2(s_{2i})$ em relação aos parâmetros do modelo

Nessa subsecção são apresentadas as derivadas parciais de $\tilde{\mu}_1^2(s_{1i})$ e $\tilde{\mu}_2^2(s_{2i})$ em relação aos hiperparâmetros a priori e variacional do modelo. As derivadas de $\tilde{\mu}_1^2(s_{1i})$, necessárias para o cálculo das derivadas parciais do Termo 5 de \mathcal{L} são dadas por

$$\begin{aligned} \frac{\partial \tilde{\mu}_1(s_{1i})}{\partial \alpha} &= \left(\frac{\partial \mathbf{K}_{11}^{(s,z)}}{\partial \alpha} \mathbf{A} + \mathbf{K}_{11}^{(s,z)} \frac{\partial \mathbf{A}}{\partial \alpha} \right) \mathbf{m}_1 + \left(\frac{\partial \mathbf{K}_{12}^{(s,z)}}{\partial \alpha} \mathbf{C} + \mathbf{K}_{12}^{(s,z)} \frac{\partial \mathbf{C}}{\partial \alpha} \right) \mathbf{m}_1 + \\ &+ \left(\frac{\partial \mathbf{K}_{11}^{(s,z)}}{\partial \alpha} \mathbf{B} + \mathbf{K}_{11}^{(s,z)} \frac{\partial \mathbf{B}}{\partial \alpha} \right) \mathbf{m}_2 + \left(\frac{\partial \mathbf{K}_{12}^{(s,z)}}{\partial \alpha} \mathbf{D} + \mathbf{K}_{12}^{(s,z)} \frac{\partial \mathbf{D}}{\partial \alpha} \right) \mathbf{m}_2 \end{aligned}$$

$$\begin{aligned} \frac{\partial \tilde{\mu}_1(s_{1i})}{\partial \rho_{12}} &= \left(\frac{\partial \mathbf{K}_{11}^{(s,z)}}{\partial \rho_{12}} \mathbf{A} + \mathbf{K}_{11}^{(s,z)} \frac{\partial \mathbf{A}}{\partial \rho_{12}} \right) \mathbf{m}_1 + \left(\frac{\partial \mathbf{K}_{12}^{(s,z)}}{\partial \rho_{12}} \mathbf{C} + \mathbf{K}_{12}^{(s,z)} \frac{\partial \mathbf{C}}{\partial \rho_{12}} \right) \mathbf{m}_1 + \\ &+ \left(\frac{\partial \mathbf{K}_{11}^{(s,z)}}{\partial \rho_{12}} \mathbf{B} + \mathbf{K}_{11}^{(s,z)} \frac{\partial \mathbf{B}}{\partial \rho_{12}} \right) \mathbf{m}_2 + \left(\frac{\partial \mathbf{K}_{12}^{(s,z)}}{\partial \rho_{12}} \mathbf{D} + \mathbf{K}_{12}^{(s,z)} \frac{\partial \mathbf{D}}{\partial \rho_{12}} \right) \mathbf{m}_2 \end{aligned}$$

$$\begin{aligned} \frac{\partial \tilde{\mu}_1(s_{1i})}{\partial \gamma_1} &= \left(\frac{\partial \mathbf{K}_{11}^{(s,z)}}{\partial \gamma_1} \mathbf{A} + \mathbf{K}_{11}^{(s,z)} \frac{\partial \mathbf{A}}{\partial \gamma_1} \right) \mathbf{m}_1 + \left(\frac{\partial \mathbf{K}_{12}^{(s,z)}}{\partial \gamma_1} \mathbf{C} + \mathbf{K}_{12}^{(s,z)} \frac{\partial \mathbf{C}}{\partial \gamma_1} \right) \mathbf{m}_1 + \\ &+ \left(\frac{\partial \mathbf{K}_{11}^{(s,z)}}{\partial \gamma_1} \mathbf{B} + \mathbf{K}_{11}^{(s,z)} \frac{\partial \mathbf{B}}{\partial \gamma_1} \right) \mathbf{m}_2 + \left(\frac{\partial \mathbf{K}_{12}^{(s,z)}}{\partial \gamma_1} \mathbf{D} + \mathbf{K}_{12}^{(s,z)} \frac{\partial \mathbf{D}}{\partial \gamma_1} \right) \mathbf{m}_2 \end{aligned}$$

$$\begin{aligned} \frac{\partial \tilde{\mu}_1(s_{1i})}{\partial \gamma_2} &= \left(\frac{\partial \mathbf{K}_{11}^{(s,z)}}{\partial \gamma_2} \mathbf{A} + \mathbf{K}_{11}^{(s,z)} \frac{\partial \mathbf{A}}{\partial \gamma_2} \right) \mathbf{m}_1 + \left(\frac{\partial \mathbf{K}_{12}^{(s,z)}}{\partial \gamma_2} \mathbf{C} + \mathbf{K}_{12}^{(s,z)} \frac{\partial \mathbf{C}}{\partial \gamma_2} \right) \mathbf{m}_1 + \\ &+ \left(\frac{\partial \mathbf{K}_{11}^{(s,z)}}{\partial \gamma_2} \mathbf{B} + \mathbf{K}_{11}^{(s,z)} \frac{\partial \mathbf{B}}{\partial \gamma_2} \right) \mathbf{m}_2 + \left(\frac{\partial \mathbf{K}_{12}^{(s,z)}}{\partial \gamma_2} \mathbf{D} + \mathbf{K}_{12}^{(s,z)} \frac{\partial \mathbf{D}}{\partial \gamma_2} \right) \mathbf{m}_2 \end{aligned}$$

$$\frac{\partial \tilde{\mu}_1(s_{1i})}{\partial \mathbf{m}_1} = \mathbf{K}_{11}^{(s,z)} \mathbf{A} + \mathbf{K}_{12}^{(s,z)} \mathbf{C}$$

$$\frac{\partial \tilde{\mu}_1(s_{1i})}{\partial \mathbf{m}_2} = \mathbf{K}_{11}^{(s,z)} \mathbf{B} + \mathbf{K}_{12}^{(s,z)} \mathbf{D}$$

As derivadas parciais de $\tilde{\mu}_1^2(s_{1i})$ em relação à \bar{u}_1 , \bar{u}_2 , \mathbf{S}_{11} , \mathbf{S}_{12} e \mathbf{S}_{22} são nulas e, portanto, não foram apresentadas. Já as derivadas de $\tilde{\mu}_2^2(s_{2i})$ são dadas por:

$$\begin{aligned} \frac{\partial \tilde{\mu}_2(s_{2i})}{\partial \alpha} &= \left(\frac{\partial \mathbf{K}_{21}^{(s,z)}}{\partial \alpha} \mathbf{A} + \mathbf{K}_{21}^{(s,z)} \frac{\partial \mathbf{A}}{\partial \alpha} \right) \mathbf{m}_1 + \left(\frac{\partial \mathbf{K}_{22}^{(s,z)}}{\partial \alpha} \mathbf{C} + \mathbf{K}_{22}^{(s,z)} \frac{\partial \mathbf{C}}{\partial \alpha} \right) \mathbf{m}_1 + \\ &+ \left(\frac{\partial \mathbf{K}_{21}^{(s,z)}}{\partial \alpha} \mathbf{B} + \mathbf{K}_{21}^{(s,z)} \frac{\partial \mathbf{B}}{\partial \alpha} \right) \mathbf{m}_2 + \left(\frac{\partial \mathbf{K}_{22}^{(s,z)}}{\partial \alpha} \mathbf{D} + \mathbf{K}_{22}^{(s,z)} \frac{\partial \mathbf{D}}{\partial \alpha} \right) \mathbf{m}_2 \end{aligned}$$

$$\begin{aligned} \frac{\partial \tilde{\mu}_2(s_{2i})}{\partial \rho_{12}} &= \left(\frac{\partial \mathbf{K}_{21}^{(s,z)}}{\partial \rho_{12}} \mathbf{A} + \mathbf{K}_{21}^{(s,z)} \frac{\partial \mathbf{A}}{\partial \rho_{12}} \right) \mathbf{m}_1 + \left(\frac{\partial \mathbf{K}_{22}^{(s,z)}}{\partial \rho_{12}} \mathbf{C} + \mathbf{K}_{22}^{(s,z)} \frac{\partial \mathbf{C}}{\partial \rho_{12}} \right) \mathbf{m}_1 + \\ &+ \left(\frac{\partial \mathbf{K}_{21}^{(s,z)}}{\partial \rho_{12}} \mathbf{B} + \mathbf{K}_{21}^{(s,z)} \frac{\partial \mathbf{B}}{\partial \rho_{12}} \right) \mathbf{m}_2 + \left(\frac{\partial \mathbf{K}_{22}^{(s,z)}}{\partial \rho_{12}} \mathbf{D} + \mathbf{K}_{22}^{(s,z)} \frac{\partial \mathbf{D}}{\partial \rho_{12}} \right) \mathbf{m}_2 \end{aligned}$$

$$\begin{aligned} \frac{\partial \tilde{\mu}_2(s_{2i})}{\partial \gamma_1} &= \left(\frac{\partial \mathbf{K}_{21}^{(s,z)}}{\partial \gamma_1} \mathbf{A} + \mathbf{K}_{21}^{(s,z)} \frac{\partial \mathbf{A}}{\partial \gamma_1} \right) \mathbf{m}_1 + \left(\frac{\partial \mathbf{K}_{22}^{(s,z)}}{\partial \gamma_1} \mathbf{C} + \mathbf{K}_{22}^{(s,z)} \frac{\partial \mathbf{C}}{\partial \gamma_1} \right) \mathbf{m}_1 + \\ &+ \left(\frac{\partial \mathbf{K}_{21}^{(s,z)}}{\partial \gamma_1} \mathbf{B} + \mathbf{K}_{21}^{(s,z)} \frac{\partial \mathbf{B}}{\partial \gamma_1} \right) \mathbf{m}_2 + \left(\frac{\partial \mathbf{K}_{22}^{(s,z)}}{\partial \gamma_1} \mathbf{D} + \mathbf{K}_{22}^{(s,z)} \frac{\partial \mathbf{D}}{\partial \gamma_1} \right) \mathbf{m}_2 \end{aligned}$$

$$\begin{aligned} \frac{\partial \tilde{\mu}_2(s_{2i})}{\partial \gamma_2} &= \left(\frac{\partial \mathbf{K}_{21}^{(s,z)}}{\partial \gamma_2} \mathbf{A} + \mathbf{K}_{21}^{(s,z)} \frac{\partial \mathbf{A}}{\partial \gamma_2} \right) \mathbf{m}_1 + \left(\frac{\partial \mathbf{K}_{22}^{(s,z)}}{\partial \gamma_2} \mathbf{C} + \mathbf{K}_{22}^{(s,z)} \frac{\partial \mathbf{C}}{\partial \gamma_2} \right) \mathbf{m}_1 + \\ &+ \left(\frac{\partial \mathbf{K}_{21}^{(s,z)}}{\partial \gamma_2} \mathbf{B} + \mathbf{K}_{21}^{(s,z)} \frac{\partial \mathbf{B}}{\partial \gamma_2} \right) \mathbf{m}_2 + \left(\frac{\partial \mathbf{K}_{22}^{(s,z)}}{\partial \gamma_2} \mathbf{D} + \mathbf{K}_{22}^{(s,z)} \frac{\partial \mathbf{D}}{\partial \gamma_2} \right) \mathbf{m}_2 \end{aligned}$$

$$\frac{\partial \tilde{\mu}_2(s_{2i})}{\partial \mathbf{m}_1} = \mathbf{K}_{21}^{(s,z)} \mathbf{A} + \mathbf{K}_{22}^{(s,z)} \mathbf{C}$$

$$\frac{\partial \tilde{\mu}_2(s_{2i})}{\partial \mathbf{m}_2} = \mathbf{K}_{21}^{(s,z)} \mathbf{B} + \mathbf{K}_{22}^{(s,z)} \mathbf{D}$$

C.5 Cálculo das derivadas parciais de $\tilde{\sigma}_1^2(s_{1i})$ e $\tilde{\sigma}_2^2(s_{2i})$ em relação aos parâmetros do modelo

Nessa subseção são mostrados os resultados das derivadas parciais de $\tilde{\sigma}_1^2(s_{1i})$ e $\tilde{\sigma}_2^2(s_{2i})$, respectivamente, em relação aos hiperparâmetros α , ρ_{12} , γ_1 , γ_2 , \mathbf{S}_{11} , \mathbf{S}_{12} e \mathbf{S}_{22} . Para os demais hiperparâmetros, \bar{u}_1 , \bar{u}_2 , \mathbf{m}_1 e \mathbf{m}_2 , as derivadas são nulas. Desse modo, calculando as derivadas parciais de $\tilde{\sigma}_1^2(s_{1i})$, necessárias para o cálculo das derivadas dos Termos 5 e 7 de \mathcal{L} , tem-se

$$\frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \mathbf{S}_{11}} = (\mathbf{A}\mathbf{K}_{11}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\mathbf{A})^\top + (\mathbf{A}\mathbf{K}_{11}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\mathbf{C})^\top + (\mathbf{B}\mathbf{K}_{21}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\mathbf{A})^\top + (\mathbf{B}\mathbf{K}_{21}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\mathbf{C})^\top$$

$$\begin{aligned} \frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \mathbf{S}_{12}} &= (\mathbf{C}\mathbf{K}_{11}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\mathbf{A})^\top + (\mathbf{C}\mathbf{K}_{11}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\mathbf{C})^\top + (\mathbf{D}\mathbf{K}_{21}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\mathbf{A})^\top + (\mathbf{D}\mathbf{K}_{21}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\mathbf{C})^\top + \\ &+ (\mathbf{A}\mathbf{K}_{11}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\mathbf{B}) + (\mathbf{A}\mathbf{K}_{11}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\mathbf{D}) + (\mathbf{B}\mathbf{K}_{21}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\mathbf{B}) + (\mathbf{B}\mathbf{K}_{21}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\mathbf{D}) \end{aligned}$$

$$\frac{\partial \tilde{\sigma}_1^2(s_{1i})}{\partial \mathbf{S}_{22}} = (\mathbf{C}\mathbf{K}_{11}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\mathbf{B})^\top + (\mathbf{C}\mathbf{K}_{11}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\mathbf{D})^\top + (\mathbf{D}\mathbf{K}_{21}^{(z,s)}\mathbf{K}_{11}^{(s,z)}\mathbf{B})^\top + (\mathbf{D}\mathbf{K}_{21}^{(z,s)}\mathbf{K}_{12}^{(s,z)}\mathbf{D})^\top$$

Da mesma maneira, o cálculo das derivadas parciais de $\tilde{\sigma}_2^2(s_{2i})$, necessárias para o cálculo das derivadas dos Termos 6 e 8 de \mathcal{L} , é dado por:

$$\begin{aligned} \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \mathbf{S}_{11}} &= (\mathbf{AK}_{12}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \mathbf{A})^\top + (\mathbf{AK}_{12}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \mathbf{C})^\top + (\mathbf{BK}_{22}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \mathbf{A})^\top + (\mathbf{BK}_{22}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \mathbf{C})^\top \\ \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \mathbf{S}_{12}} &= (\mathbf{CK}_{12}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \mathbf{A})^\top + (\mathbf{CK}_{12}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \mathbf{C})^\top + (\mathbf{DK}_{22}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \mathbf{A})^\top + (\mathbf{DK}_{22}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \mathbf{C})^\top + \\ &\quad + (\mathbf{AK}_{12}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \mathbf{B}) + (\mathbf{AK}_{12}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \mathbf{D}) + (\mathbf{BK}_{22}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \mathbf{B}) + (\mathbf{BK}_{22}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \mathbf{D}) \\ \frac{\partial \tilde{\sigma}_2^2(s_{2i})}{\partial \mathbf{S}_{22}} &= (\mathbf{CK}_{12}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \mathbf{B})^\top + (\mathbf{CK}_{12}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \mathbf{D})^\top + (\mathbf{DK}_{22}^{(z,s)} \mathbf{K}_{21}^{(s,z)} \mathbf{B})^\top + (\mathbf{DK}_{22}^{(z,s)} \mathbf{K}_{22}^{(s,z)} \mathbf{D})^\top \end{aligned}$$

C.6 Cálculo das derivadas parciais dos Ψ 's em relação aos parâmetros do modelo

Nessa seção serão apresentados os resultados referentes às derivadas parciais de todas as matrizes Ψ 's, apresentadas anteriormente na seção C.1, em relação aos parâmetros α , ρ_{12} , γ_1 e γ_2 . Os Ψ 's não dependem dos demais parâmetros e, conseqüentemente, suas derivadas parciais são nulas. Para a realizar os cálculos das derivadas parciais, será necessário destacar um termo comum para todos os Ψ 's, que é denotado por

$$\text{Núcleo} = \frac{\sqrt{\pi\alpha}}{2} \exp\left\{-\frac{(z-z')^2}{4\alpha}\right\} \left[\text{erf}\left(\frac{\bar{z} - \mathcal{T}_{max}}{\sqrt{\alpha}}\right) - \text{erf}\left(\frac{\bar{z} - \mathcal{T}_{min}}{\sqrt{\alpha}}\right) \right].$$

Uma vez que o termo “Núcleo” depende do parâmetro α , sua derivada é dada por

$$\begin{aligned} \frac{\partial \text{Núcleo}}{\partial \alpha} &= -\frac{\sqrt{\pi}(z-z')^2}{8\alpha^{3/2}} \exp\left\{-\frac{(z-z')^2}{4\alpha}\right\} \left[\text{erf}\left(\frac{\bar{z} - \mathcal{T}_{max}}{\sqrt{\alpha}}\right) - \text{erf}\left(\frac{\bar{z} - \mathcal{T}_{min}}{\sqrt{\alpha}}\right) \right] - \\ &\quad - \frac{\sqrt{\pi}}{4\sqrt{\alpha}} \exp\left\{-\frac{(z-z')^2}{4\alpha}\right\} \left[\text{erf}\left(\frac{\bar{z} - \mathcal{T}_{max}}{\sqrt{\alpha}}\right) - \text{erf}\left(\frac{\bar{z} - \mathcal{T}_{min}}{\sqrt{\alpha}}\right) \right] - \\ &\quad - \frac{\sqrt{\pi}\sqrt{\alpha}}{2} \exp\left\{-\frac{(z-z')^2}{4\alpha}\right\} \left[\frac{(\bar{z} - \mathcal{T}_{min})}{\sqrt{\pi}\alpha^{3/2}} \exp\left\{-\frac{(\bar{z} - \mathcal{T}_{min})^2}{\alpha}\right\} - \right. \\ &\quad \left. - \frac{(\bar{z} - \mathcal{T}_{max})}{\sqrt{\pi}\alpha^{3/2}} \exp\left\{-\frac{(\bar{z} - \mathcal{T}_{max})^2}{\alpha}\right\} \right]. \end{aligned}$$

Desse modo, as derivadas parciais de $\Psi_{11}^{11}(z, z')$ são dadas por

$$\frac{\partial \Psi_{11}^{11}(z, z')}{\partial \alpha} = \gamma_1^4 \times \frac{\partial \text{Núcleo}}{\partial \alpha}$$

$$\frac{\partial \Psi_{11}^{11}(z, z')}{\partial \rho_{12}} = 0$$

$$\frac{\partial \Psi_{11}^{11}(z, z')}{\partial \gamma_1} = 4\gamma_1^3 \times \text{Núcleo}$$

$$\frac{\partial \Psi_{11}^{11}(z, z')}{\partial \gamma_2} = 0$$

Como as expressões de $\Psi_{11}^{21}(z, z')$, $\Psi_{11}^{12}(z, z')$, $\Psi_{21}^{11}(z, z')$ e $\Psi_{12}^{11}(z, z')$ são as mesmas, suas respectivas derivadas parciais também serão. Desse modo,

$$\begin{aligned} \frac{\partial \Psi_{11}^{21}(z, z')}{\partial \alpha} &= \frac{\partial \Psi_{11}^{12}(z, z')}{\partial \alpha} = \frac{\partial \Psi_{21}^{11}(z, z')}{\partial \alpha} = \frac{\partial \Psi_{12}^{11}(z, z')}{\partial \alpha} = \\ &= \rho_{12}\gamma_1^3\gamma_2 \times \frac{\partial \text{Núcleo}}{\partial \alpha} \end{aligned}$$

$$\begin{aligned} \frac{\partial \Psi_{11}^{21}(z, z')}{\partial \rho_{12}} &= \frac{\partial \Psi_{11}^{12}(z, z')}{\partial \rho_{12}} = \frac{\partial \Psi_{21}^{11}(z, z')}{\partial \rho_{12}} = \frac{\partial \Psi_{12}^{11}(z, z')}{\partial \rho_{12}} = \\ &= \gamma_1^3\gamma_2 \times \text{Núcleo} \end{aligned}$$

$$\begin{aligned} \frac{\partial \Psi_{11}^{21}(z, z')}{\partial \gamma_1} &= \frac{\partial \Psi_{11}^{12}(z, z')}{\partial \gamma_1} = \frac{\partial \Psi_{21}^{11}(z, z')}{\partial \gamma_1} = \frac{\partial \Psi_{12}^{11}(z, z')}{\partial \gamma_1} = \\ &= 3\rho_{12}\gamma_1^2\gamma_2 \times \text{Núcleo} \end{aligned}$$

$$\begin{aligned} \frac{\partial \Psi_{11}^{21}(z, z')}{\partial \gamma_2} &= \frac{\partial \Psi_{11}^{12}(z, z')}{\partial \gamma_2} = \frac{\partial \Psi_{21}^{11}(z, z')}{\partial \gamma_2} = \frac{\partial \Psi_{12}^{11}(z, z')}{\partial \gamma_2} = \\ &= \rho_{12}\gamma_1^3 \times \text{Núcleo} \end{aligned}$$

Já os resultados das derivadas de $\Psi_{11}^{22}(z, z')$ e $\Psi_{22}^{11}(z, z')$ são dadas por

$$\frac{\partial \Psi_{11}^{22}(z, z')}{\partial \alpha} = \frac{\partial \Psi_{22}^{11}(z, z')}{\partial \alpha} = \gamma_1^2\gamma_2^2 \times \frac{\partial \text{Núcleo}}{\partial \alpha}$$

$$\frac{\partial \Psi_{11}^{22}(z, z')}{\partial \rho_{12}} = \frac{\partial \Psi_{22}^{11}(z, z')}{\partial \rho_{12}} = 0$$

$$\frac{\partial \Psi_{11}^{22}(z, z')}{\partial \gamma_1} = \frac{\partial \Psi_{22}^{11}(z, z')}{\partial \gamma_1} = 2\gamma_1\gamma_2^2 \times \text{Núcleo}$$

$$\frac{\partial \Psi_{11}^{22}(z, z')}{\partial \gamma_2} = \frac{\partial \Psi_{22}^{11}(z, z')}{\partial \gamma_2} = 2\gamma_1^2\gamma_2 \times \text{Núcleo}$$

De maneira análoga, as derivadas parciais de $\Psi_{21}^{12}(z, z')$, $\Psi_{12}^{21}(z, z')$, $\Psi_{21}^{21}(z, z')$ e $\Psi_{12}^{12}(z, z')$ serão

$$\begin{aligned} \frac{\partial \Psi_{21}^{12}(z, z')}{\partial \alpha} &= \frac{\partial \Psi_{12}^{21}(z, z')}{\partial \alpha} = \frac{\partial \Psi_{21}^{21}(z, z')}{\partial \alpha} = \frac{\partial \Psi_{12}^{12}(z, z')}{\partial \alpha} = \\ &= \rho_{12} \gamma_1^2 \gamma_2^2 \times \frac{\partial \text{Núcleo}}{\partial \alpha} \end{aligned}$$

$$\begin{aligned} \frac{\partial \Psi_{21}^{12}(z, z')}{\partial \rho_{12}} &= \frac{\partial \Psi_{12}^{21}(z, z')}{\partial \rho_{12}} = \frac{\partial \Psi_{21}^{21}(z, z')}{\partial \rho_{12}} = \frac{\partial \Psi_{12}^{12}(z, z')}{\partial \rho_{12}} = \\ &= \gamma_1^2 \gamma_2^2 \times \text{Núcleo} \end{aligned}$$

$$\begin{aligned} \frac{\partial \Psi_{21}^{12}(z, z')}{\partial \gamma_1} &= \frac{\partial \Psi_{12}^{21}(z, z')}{\partial \gamma_1} = \frac{\partial \Psi_{21}^{21}(z, z')}{\partial \gamma_1} = \frac{\partial \Psi_{12}^{12}(z, z')}{\partial \gamma_1} = \\ &= 2\rho_{12} \gamma_1 \gamma_2^2 \times \text{Núcleo} \end{aligned}$$

$$\begin{aligned} \frac{\partial \Psi_{21}^{12}(z, z')}{\partial \gamma_2} &= \frac{\partial \Psi_{12}^{21}(z, z')}{\partial \gamma_2} = \frac{\partial \Psi_{21}^{21}(z, z')}{\partial \gamma_2} = \frac{\partial \Psi_{12}^{12}(z, z')}{\partial \gamma_2} = \\ &= 2\rho_{12} \gamma_1^2 \gamma_2 \times \text{Núcleo} \end{aligned}$$

Considerando agora as funções $\Psi_{22}^{12}(z, z')$, $\Psi_{22}^{21}(z, z')$, $\Psi_{12}^{22}(z, z')$ e $\Psi_{21}^{22}(z, z')$, tem-se que

$$\begin{aligned} \frac{\partial \Psi_{22}^{12}(z, z')}{\partial \alpha} &= \frac{\partial \Psi_{22}^{21}(z, z')}{\partial \alpha} = \frac{\partial \Psi_{12}^{22}(z, z')}{\partial \alpha} = \frac{\partial \Psi_{21}^{22}(z, z')}{\partial \alpha} = \\ &= \rho_{12} \gamma_1 \gamma_2^3 \times \frac{\partial \text{Núcleo}}{\partial \alpha} \end{aligned}$$

$$\begin{aligned} \frac{\partial \Psi_{22}^{12}(z, z')}{\partial \rho_{12}} &= \frac{\partial \Psi_{22}^{21}(z, z')}{\partial \rho_{12}} = \frac{\partial \Psi_{12}^{22}(z, z')}{\partial \rho_{12}} = \frac{\partial \Psi_{21}^{22}(z, z')}{\partial \rho_{12}} = \\ &= \gamma_1 \gamma_2^3 \times \text{Núcleo} \end{aligned}$$

$$\begin{aligned} \frac{\partial \Psi_{22}^{12}(z, z')}{\partial \gamma_1} &= \frac{\partial \Psi_{22}^{21}(z, z')}{\partial \gamma_1} = \frac{\partial \Psi_{12}^{22}(z, z')}{\partial \gamma_1} = \frac{\partial \Psi_{21}^{22}(z, z')}{\partial \gamma_1} = \\ &= \rho_{12} \gamma_2^3 \times \text{Núcleo} \end{aligned}$$

$$\begin{aligned} \frac{\partial \Psi_{22}^{12}(z, z')}{\partial \gamma_2} &= \frac{\partial \Psi_{22}^{21}(z, z')}{\partial \gamma_2} = \frac{\partial \Psi_{12}^{22}(z, z')}{\partial \gamma_2} = \frac{\partial \Psi_{21}^{22}(z, z')}{\partial \gamma_2} = \\ &= 3\rho_{12} \gamma_1 \gamma_2^2 \times \text{Núcleo} \end{aligned}$$

Por fim, tem-se que as derivadas parciais de $\Psi_{22}^{22}(z, z')$ serão

$$\frac{\partial \Psi_{22}^{22}(z, z')}{\partial \alpha} = \gamma_2^4 \times \frac{\partial \text{Núcleo}}{\partial \alpha}$$

$$\frac{\partial \Psi_{22}^{22}(z, z')}{\partial \rho_{12}} = 0$$

$$\frac{\partial \Psi_{22}^{22}(z, z')}{\partial \gamma_1} = 0$$

$$\frac{\partial \Psi_{22}^{22}(z, z')}{\partial \gamma_2} = 4\gamma_2^3 \times \text{Núcleo}$$