

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA**

Douglas Mateus da Silva

**MODELOS APROXIMADOS E EXATOS PARA
GEOESTATÍSTICA COM AMOSTRAGEM PREFERENCIAL**

BELO HORIZONTE - MG

2021

Douglas Mateus da Silva

**MODELOS APROXIMADOS E EXATOS PARA
GEOESTATÍSTICA COM AMOSTRAGEM PREFERENCIAL**

Tese apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutor em Estatística

Orientador: Dani Gamerman
Coorientadora: Lourdes Coral
Contreras Montenegro

BELO HORIZONTE - MG
2021

	Silva ,Douglas Mateus da
S586m	Modelos aproximados e exatos para geoestatística com amostragem preferencial [manuscrito] / Douglas Mateus da Silva. – 2021. 216 f. il. Orientador: Dani Gamerman. Coorientadora: Lourdes Coral Contreras Montenegro Tese (doutorado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística. Referências: f. 201–203. 1. Estatística – Teses. 2. Geologia – Métodos estatísticos – Teses. 3. Amostragem (Estatística) – Teses. 4 Teoria bayesiana de decisão estatística– Teses. I. Gamerman, Dani. II. Montenegro, Lourdes Coral Contreras III. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística. IV. Título.
	CDU 519.2 (043)



ATA DA DEFESA DE TESE DE DOUTORADO DO ALUNO DOUGLAS MATEUS DA SILVA, MATRICULADO, SOB O N° 2017.668.782, NO PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA, DO INSTITUTO DE CIÊNCIAS EXATAS, DA UNIVERSIDADE FEDERAL DE MINAS GERAIS, REALIZADA NO DIA 03 DE SETEMBRO DE 2021.

Aos 03 dias do mês de Setembro de 2021, às 10h00, em reunião pública virtual 70 (conforme orientações para a atividade de defesa de tese durante a vigência da Portaria PRPG n° 1819) OU na sala <https://us06web.zoom.us/j/84650441978?pwd=Vmo2bkl2Tk1Tb1JvdXhtNVBtMG9lQT09> do Instituto de Ciências Exatas da UFMG, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pelo Colegiado do Programa de Pós-Graduação em Estatística, para julgar a defesa de tese do aluno Douglas Mateus da Silva, n° matrícula 2017668782, intitulada: "*Modelos Aproximados e Exatos para Geoestatística com Amostragem Preferencial*", requisito final para obtenção do Grau de doutor em Estatística. Abrindo a sessão, o Senhor Presidente da Comissão, Prof(a). Dani Gamerman - Orientador (DEST/UFMG), passou a palavra ao(à) aluno(a) para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do(a) aluno(a). Após a defesa, os membros da banca examinadora reuniram-se reservadamente sem a presença do(a) aluno(a) e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação:

Aprovada.

Reprovada com resubmissão do texto em ____ dias.

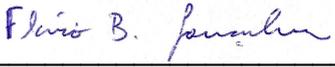
Reprovada com resubmissão do texto e nova defesa em ____ dias.

Reprovada.


Prof. Dani Gamerman - Orientador (DEST/UFMG)


Profa. Lourdes Coral Contreras Montenegro Coorientadora (DEST/UFMG)


Prof. Marcos Oliveira Prates (DEST/UFMG)


Prof. Flávio B. Bambirra Gonçalves (DEST/UFMG)


Prof. Hélio dos Santos Migon (DME/UFRJ)


Profa. Raquel Menezes (Universidade do Minho, Portugal)

O resultado final foi comunicado publicamente ao aluno pelo Senhor Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 03 de setembro de 2021.

Observações:

1. No caso de aprovação da tese, a banca pode solicitar modificações a serem feitas na versão final do texto. Neste caso, o texto final deve ser aprovado pelo orientador da tese. O pedido de expedição do diploma do candidato fica condicionado à submissão e aprovação, pelo orientador, da versão final do texto.
2. No caso de reprovação da tese com resubmissão do texto, o candidato deve submeter o novo texto dentro do prazo estipulado pela banca, que deve ser de no máximo 6 (seis) meses. O novo texto deve ser avaliado por todos os membros da banca que então decidirão pela aprovação ou reprovação da tese.
3. No caso de reprovação da tese com resubmissão do texto e nova defesa, o candidato deve submeter o novo texto com a antecedência à nova defesa que o orientador julgar adequada. A nova defesa, mediante todos os membros da banca, deve ser realizada dentro do prazo estipulado pela banca, que deve ser de no máximo 6 (seis) meses. O novo texto deve ser avaliado por todos os membros da banca. Baseada no novo texto e na nova defesa, a banca decidirá pela aprovação ou reprovação da tese.

*À minha mãe Elisabete, ao meu pai Geraldo,
à minha irmã Andressa e à minha avó Maria
(in memoriam).*

Agradecimentos

- Em primeiro lugar, agradeço à minha família pelo apoio e incentivo constantes.
- Ao meu orientador Professor Dani Gamerman, pela orientação e paciência nessa trajetória e por ter aceitado o convite de orientação em uma etapa final do meu doutorado, o que foi de suma importância para o desenvolvimento desse trabalho.
- À minha co-orientador, Professora Lourdes Coral Contreras Montenegro, pelo apoio e orientação em todo esse trabalho, bem como na minha jornada acadêmica desde a graduação.
- Aos professores do Departamento de Estatística pelos ensinamentos concedidos e disposição para ajudar.
- Ao meu namorado Welbert Alves, pelo apoio constante e incentivo durante todo meu trajeto no doutorado, principalmente nos momentos mais difíceis.
- Aos meus amigos que sempre me incentivaram a continuar e seguir em frente.
- À CAPES pelo apoio financeiro.
- Ao CNPq e à FAPEMIG pelo apoio financeiro ao departamento, auxiliando na infraestrutura dos laboratórios para o desenvolvimento dos trabalhos dos alunos e professores da pós-graduação.
- E agradeço a Deus pela oportunidade de realizar esse sonho.

Resumo

Modelos geoestatísticos são utilizados para analisar dados espacialmente correlacionados sob um domínio contínuo. A metodologia tradicional de geoestatística supõe que o desenho amostral é realizado de forma independente do fenômeno em estudo e denominamos esse processo de amostragem como não preferencial. Quando o desenho amostral é construído com base em informações da variável de interesse, certos locais tem maior chance de serem escolhidos em detrimento à outros. A esse tipo de amostragem denominamos preferencial. Sob essa amostragem, o processo latente que modela os dados e o processo pontual que modela a configuração espacial da amostra não são independentes. Não podemos ignorar a distribuição conjunta desses processos, pois obteremos resultados viesados na estimação e na predição. Sendo o algoritmo EM um método que trata o problema de dados faltantes, é natural pensar em utilizar essa ferramenta como um método de estimação para esse modelo. Dessa forma, o primeiro ponto abordado nesta tese é o desenvolvimento de um algoritmo SAEM para estimação dos parâmetros do modelo, sob abordagem clássica. Além disso, na presença de dados atípicos, o modelo Gaussiano não é adequado e temos que recorrer a distribuições de caudas mais pesadas que a normal. Desenvolvemos o modelo com amostragem preferencial com distribuição t-Student para lidar com esse problema, sendo o segundo ponto abordado nesta tese. Na literatura existente, é assumida como função de intensidade para o processo pontual a função exponencial, tendo como consequência a intratabilidade da função de verossimilhança. Nesse caso, a inferência é realizada através de aproximações por discretização da região de estudo, o que pode levar a erros difíceis de serem mensurados. Dessa forma, como terceiro ponto abordado nessa tese, desenvolvemos o modelo Bayesiano exato para os dados de geoestatística com amostragem preferencial. Estudos de simulações são realizados para avaliar a metodologia proposta pelos três pontos considerados, assim como a aplicação em conjunto de dados reais, sendo um deles os dados de poluição em Galícia, conjunto de dados bastante conhecido na literatura. Os modelos propostos apresentaram melhor estimação dos parâmetros do modelo e melhor predição da variável de interesse em locais não observados comparados aos métodos já existentes, principalmente em regiões distantes de pontos da amostra cuja quantidade de informação é baixa.

Palavras chaves: Geoestatística. Amostragem Preferencial. Inferência Clássica. Inferência Bayesiana. *Outlier*. t-Student.

Abstract

Geostatistical models are used to analyze spatially correlated data under a continuous domain. The traditional methodology of geostatistics supposes that the sampling design is constructed independently of the phenomena in the study and we denominate this sampling process as non preferential. When the sampling design is constructed with information about the variable of interest, some locals have more chance of being selected in comparison to others. This kind of sampling we denominate preferential. Under the preferential sampling, the latent process that models data and the pontual process that models the spacial configuration of the sample are not independent. We can not ignore the joint distribution of those processes, because we will obtain biased results under estimation and prediction. Since the EM algorithm is a method that deals with latent data, it is natural to consider it as a tool of estimation for this model. In this way, the first aim in this thesis is the development of an SAEM algorithm for parameter estimation of the model under preferential sampling, in the classic approach. Besides this, in presence of outliers, the Gaussian model is not adequate and we have to consider heavier tails distributions. We developed a geostatistical model under preferential sampling with t-Student distribution to deal with this problem, beign this the second aim of this thesis. In the existent literature, it assumes the exponential function as the intensity function of the pontual process, which has intractable likelihood function. In this case, the inference is realized considering discrete approximation of the region in study, which can lead to errors that are difficult to measure. In this way, as the third aim of this thesis, we developed an exact Bayesian model for geostatistical data under preferential sampling. Simulated examples are considered to evaluate the methodology proposed for the three aims of this thesis, as well application on real data, one of them being the moss data of Galicia, a dataset well known in the literature. The proposed models provided better results in parameter estimation and prediction of the variable of interest on locals that were not sampled compared to the existing methods, mainly in regions that are distant of locals sampled which the quantity of information is low.

Keywords: Geostatistics. Preferential Sampling. Classic Inference. Bayesian Inference. Outlier. t-Sudent.

Sumário

1	Introdução	11
1.1	Contexto	11
1.2	Objetivos da tese	16
1.3	Organização da tese	17
2	Preliminares	20
2.1	Simulação via cadeias de Markov	20
2.2	Algoritmo EM e variações	23
2.3	Aproximação de Laplace	28
2.4	Processos pontuais	29
2.4.1	Processo de Poisson	29
2.4.2	Simulação de processo de Poisson	33
2.4.3	Processo de Cox	35
2.4.4	Inferência em processos pontuais	36
3	Geoestatística com Amostragem Preferencial	38
3.1	Processo geoestatístico	38
3.2	Covariograma e variograma	40
3.3	Modelo geoestatístico	45
3.4	Estimação via método de máxima verossimilhança	46
3.5	Estimação via inferência Bayesiana	47
3.6	Predição	49
3.7	Modelo geoestatístico com amostragem preferencial	51
4	Inferência clássica - Modelagem e computação	57
4.1	Modelo Gaussiano	57
4.1.1	Estimação via MCLA	57
4.1.2	Estimação via aproximação de Laplace	60
4.1.3	Estimação via SAEM	62
4.1.4	Predição	67
4.2	Modelo t-Student	68
4.2.1	Estimação via aproximação de Laplace	70

4.2.2	Estimação via SAEM	70
4.2.3	Predição	73
5	Inferência clássica - Estudo de simulação	74
5.1	Distribuição preditiva de $S \mathbf{x}, \mathbf{y}$	74
5.2	Efeito da blocagem na estimação dos parâmetros do modelo	85
5.3	Efeito de fixar o parâmetro de correlação	86
5.4	Estimação dos parâmetros do modelo	91
5.5	Comparação dos métodos de predição	96
6	Inferência Bayesiana - Modelagem e computação	103
6.1	Modelo Bayesiano Exato	103
6.1.1	Computação	106
6.1.2	Predição	112
6.2	Modelo com intercepto na função de intensidade	114
6.3	Modelo t-Student	116
7	Inferência Bayesiana - Estudo de simulação	122
7.1	Amostragem preferencial e não preferencial	123
7.2	Efeito de fixar o parâmetro de correlação	140
7.3	Desempenho do modelo para poucos dados	145
7.4	Especificação incorreta da função de intensidade	149
7.5	Modelo com intercepto na função de intensidade	156
7.6	Modelo com discretização	162
7.7	Modelo t-Student	167
8	Aplicações a dados reais	171
8.1	Aplicação 1: Dados de poluição em Galícia	171
8.2	Aplicação 2: Dados de radiação na Alemanha	184
8.3	Aplicação 3: Dados de precipitação pluviométrica no Rio de Janeiro	189
9	Discussão e trabalhos futuros	195
	Referências Bibliográficas	202

APÊNDICE A — Especificação incorreta da função de intensidade: cenário	
2	205
APÊNDICE B — Trajetórias das iterações das cadeias de Markov por	
estimação Bayesiana	210

Capítulo 1

Introdução

Neste capítulo, apresentamos o tema da presente tese, introduzindo o problema de amostragem preferencial no contexto geoestatístico. Discutimos os trabalhos já realizados nessa área e alguns problemas que não foram abordados. A partir dessa discussão, apresentamos os objetivos desse trabalho, evidenciando nossa contribuição para a sociedade científica. Por fim, descrevemos a organização desta tese.

1.1 Contexto

Na área de geoestatística, estudamos fenômenos que estão espacialmente correlacionados em uma região contínua. A dependência espacial é modelada através de uma função de covariância que determina a lei comportamental da associação da variável em estudo. Podemos citar como exemplos: o teor de um elemento na terra, a temperatura de uma noite em uma determinada região, medidas de poluição monitoradas por estações em uma cidade, entre outros.

Em seu trabalho sobre crescimento populacional urbano (Tobler, 1970), Tobler invoca a primeira lei da geografia: “Todas as coisas estão relacionadas, mas coisas mais próximas são mais parecidas do que coisas mais distantes”. É nesse sentido que a associação espacial se caracteriza. Por exemplo, se estamos interessados em estudar o teor de um elemento presente no solo em uma determinada região, é provável que pontos próximos a um local com alto teor apresentarão valores elevados, ao passo que pontos próximos a um local com baixo teor também apresentarão valores reduzidos, e pontos distantes não fornecerão valores tão semelhantes.

Em seu trabalho “Princípios da geoestatística”, Matheron (1963) critica o uso de métodos que não consideram a dependência espacial em estudos na área da indústria de mineração (por exemplo, concentração de minério). Dessa forma, o autor desenvolve a teoria das variáveis regionalizadas com o objetivo de considerar em sua análise o aspecto aleatório e estruturado do fenômeno em estudo. Apesar do estudo ter sido desenvolvido sobre problemas de mineração, a geoestatística tem uma vasta área de aplicação, por exemplo, em estudo de fatores ambientais, como poluição, temperatura e precipitação.

No modelo tradicional de geoestatística, a variável de interesse é modelada pela soma de três componentes: uma determinística referente à media do processo, uma aleatória espacialmente correlacionada (que chamamos de processo latente) e uma representando um ruído aleatório. É comum considerar que a variável de interesse seja normalmente distribuída e Diggle and Ribeiro (2007) apresentam resultados de inferência e predição do modelo geoestatístico Gaussiano. Nesse modelo, assume-se que tanto o processo latente quanto o erro aleatório são normalmente distribuídos. Exemplos de aplicações podem ser consultados em López-Granados et al. (2005), que utilizam o modelo para mapear características de solo (como matéria orgânica e pH), e em Biondi et al. (1994), que abordam o problema de mapeamento de recursos florestais, modelando tamanho e crescimento de árvores.

Em estudos geoestatísticos, é comum considerar que o desenho amostral seja realizado de forma independente do fenômeno em estudo, o que significa que nenhuma informação sobre o fenômeno é utilizada na escolha dos locais amostrados. Esse tipo de amostragem é denominado não preferencial pois, a priori, nenhuma localização possui preferência em ser escolhida. Um exemplo seria escolher de forma totalmente aleatória os pontos na região de estudo que farão parte da amostra.

Há casos em que se opta por escolher locais na amostragem considerando o fenômeno em questão, por exemplo, escolher regiões em que se espera alto valor da variável de interesse. Como exemplo, podemos citar a alocação de estações de monitoramento de poluentes atmosféricos próximas a possíveis fontes de poluição ou alocar máquinas de extração de minério onde se espera maior concentração do minério de interesse. Nesse caso, o desenho amostral é construído com base em informações do problema e esse tipo de amostragem é denominado amostragem preferencial. Os métodos usuais de análise geoestatística para esse tipo de amostragem não são adequados, sendo necessária outra abordagem. Nesse contexto, Diggle et al. (2010) apresentam uma metodologia para tratar o modelo geoestatístico com amostragem preferencial, introduzindo um processo pontual (PP) para modelar o desenho amostral. Essa modelagem é realizada através de processos de Cox. No caso de amostragem não preferencial, o PP que gera os locais amostrados não depende do processo latente e a inferência pode ser realizada sem a informação do mesmo. Na presença de amostragem preferencial, o PP depende do processo latente e sua distribuição conjunta não pode ser ignorada. Diggle et al. (2010) mostram que, quando há

preferência de locais amostrados, as estimativas dos parâmetros do modelo são viesadas e, mais importante, a predição também se torna viesada. Visto que o principal objetivo em geoestatística é a predição em regiões não amostradas (normalmente para construção de mapas), o impacto da amostragem preferencial nos valores preditos da variável resposta se torna um problema grave. Dessa forma, a informação da configuração espacial da amostra não deve ser ignorada.

Um exemplo real de amostragem preferencial pode ser observado na Figura 1.1, cujos dados são referentes à amostras de musgo em Galícia para biomonitoramento de metais pesados. A pesquisa realizada no ano de 1997 teve como planejamento amostral escolher locais em que se esperava maior concentração de chumbo (Fernández et al., 2000), o que torna a amostragem potencialmente preferencial. Em relação ao ano 2000, os dados foram coletados em uma malha aproximadamente regular, o que torna a amostragem não preferencial e o uso das técnicas convencionais de geoestatística são adequadas. Os dados foram analisados por Diggle et al. (2010) e Dinsdale and Salibian-Barrera (2019) e também serão analisados neste trabalho sob a nossa proposta.

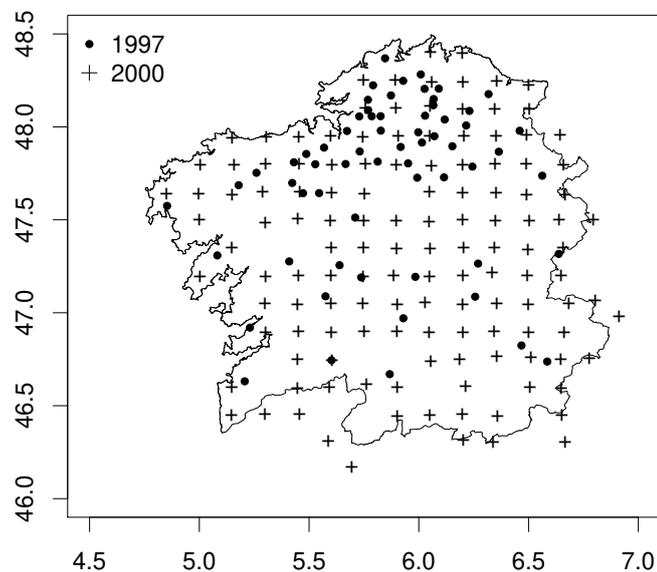


Figura 1.1: Aplicação 1: Mapa de Galícia e locais amostrados em 1997 e 2000 para os dados de biomonitoramento de metais pesados.

Diggle et al. (2010) propõe modelar a configuração espacial da amostra através

de um processo de Cox log-Gaussiano (Møller and Waagepetersen, 2003) cuja função de intensidade depende do processo latente do modelo geoestatístico. O grau de preferência na amostragem é controlado por um parâmetro na função de intensidade, sendo positivo se as regiões com maiores valores da variável de interesse tem maior chance de pertencer à amostra, negativo caso contrário e zero indica amostragem não preferencial. Sem obter forma fechada para os estimadores de máxima verossimilhança, os autores utilizam a abordagem de aproximação de verossimilhança por Monte Carlo. Através desse trabalho, outros tem sido desenvolvidos na área de amostragem preferencial. Pati et al. (2011) iniciam o estudo do modelo na área de inferência Bayesiana, apresentando propriedades teóricas do modelo. Em particular, os autores provam que uma posteriori própria para o parâmetro de preferência é obtida mesmo sobre uma priori imprópria não informativa, indicando que os dados possuem informação sobre esse parâmetro. Gelfand et al. (2012) avaliam o efeito da amostragem preferencial na predição espacial do modelo em dados de concentração de ozônio na Califórnia, também na abordagem Bayesiana, comparando os mapas de predição obtidos pelo modelo que considera a amostragem preferencial e o modelo usual de geoestatística. Shaddick and Zidek (2014) aplicam a metodologia em dados de poluição de ar no Reino Unido, discutindo o efeito da escolha dos locais para adição ou remoção de estações de monitoramento na estimação da exposição de poluição do ar ao longo do tempo. Ferreira and Gamerman (2015) analisam o efeito de amostragem preferencial no problema de planejamento amostral ótimo, isto é, quando a escolha de novos locais é o principal interesse do pesquisador.

Em relação à estatística clássica, há poucas metodologias para a estimação dos parâmetros do modelo geoestatístico com amostragem preferencial. Além do método de estimação de Diggle et al. (2010), Dinsdale and Salibian-Barrera (2019) propõe realizar a estimação dos parâmetros do modelo e predição da variável de interesse através da aproximação de Laplace. Visto que a componente responsável por modelar a estrutura espacial é um processo latente, o algoritmo EM (*Expectation–Maximization*) (Dempster et al., 1977) se torna uma ferramenta possível para estimação dos parâmetros. Como veremos posteriormente, a primeira etapa do algoritmo EM não poderá ser realizada de forma fechada e recorreremos aos algoritmos MCEM (*Monte Carlo EM*) (Wei and Tanner, 1990) e SAEM (*Stochastic Approximation EM*) (Delyon et al., 1999). Ambos algoritmos são iterativos e baseados em simulações da distribuição preditiva do processo latente

condicionado aos dados do modelo e possuem propriedades distintas. Os estimadores do MCEM convergem em distribuição para um máximo local ao passo que os estimadores do SAEM convergem quase certamente para os estimadores de máxima verossimilhança. Dessa forma, o primeiro objetivo da presente tese é desenvolver os algoritmos de estimação MCEM e SAEM para o modelo geoestatístico com amostragem preferencial, proporcionando mais uma ferramenta para solução desse problema. Mostramos que os resultados obtidos para ambos algoritmos são similares, sendo computacionalmente mais eficiente o MCEM. Além disso, aprimoramos o algoritmo de geração de amostras da distribuição preditiva do processo latente presente em Ferreira and Gamerman (2015), que utiliza um algoritmo MCMC (*Markov chain Monte Carlo*) para gerar de cada elemento do vetor do processo latente individualmente, para uma amostragem em blocos, reduzindo o tempo computacional sem perder qualidade preditiva.

Apesar de muito utilizado, o modelo Gaussiano não é adequado para todos os fenômenos. Isso ocorre pela própria característica da variável de interesse, podendo ser contínua mas não possuir distribuição normal ou ser discreta. Em relação ao modelo sem amostragem preferencial, há uma vasta literatura de alternativas para tratar dados não Gaussianos. Uma possibilidade é realizar uma transformação a partir da qual a distribuição dos dados transformados pode ser considerada normal e Diggle and Ribeiro (2007) apresentam uma análise sobre modelos Gaussianos transformados. Outra alternativa, presente no artigo de Diggle et al. (1998), é utilizar modelos lineares generalizados em geoestatística para acomodar distribuições não Gaussianas.

A presença de observações atípicas também podem comprometer a análise geoestatística quando utilizado o modelo Gaussiano. As estimativas podem se tornar viesadas e outra abordagem se faz necessária. Como alternativa, De Bastiani et al. (2015) propõe a classe de distribuição da família de contornos elípticos no modelo geoestatístico como uma extensão robusta do modelo normal para acomodar tais observações. Para o modelo com amostragem preferencial, não identificamos estudos com o objetivo de propor soluções para esse problema. Dada a importância de termos uma modelagem que forneça resultados confiáveis e não viesados e motivados pela falta de literatura para o problema de dados atípicos, o segundo objetivo dessa tese é desenvolver o modelo geoestatístico com amostragem preferencial utilizando a distribuição t-Student.

A função de densidade do processo de Cox log-Gaussiano possui uma integral in-

tratável com domínio em toda a região de estudo. Dessa forma, esse termo precisa ser aproximado e o usual é realizar uma discretização da região, o que acarreta em trabalharmos com uma aproximação do processo latente do modelo geoestatístico. Essa alternativa é utilizada em todos os trabalhos mencionados acima que envolvem amostragem preferencial. Apesar de ser uma solução para se realizar inferência sobre esse tipo de processo, a discretização acarreta uma fonte de erros difícil de mensurar, podendo produzir resultados viesados (Simpson et al., 2016). Outro problema está no custo computacional, pois a simulação de processos Gaussianos são da ordem de n^3 e rapidamente se tornam ineficientes. Dessa forma, é necessário ter um balanço entre custo computacional e a qualidade da aproximação do processo latente na região de estudo. Gonçalves and Gamerman (2018) apresentam uma metodologia para realizar inferência exata em processos de Cox sob o enfoque Bayesiano, sendo que o termo “exata” é utilizado no sentido de que nenhuma aproximação por discretização é realizada. Para isso, os autores consideram que a função de intensidade é limitada superiormente e que o PP que gera a configuração dos locais da amostra é parte de um PP maior com função de intensidade homogênea. Os autores também apresentam um algoritmo MCMC capaz de realizar a geração do processo latente nos locais exatos amostrados. Dessa forma, a inferência é livre de erros de discretização. Motivados pela ideia dos autores, o terceiro objetivo dessa tese é desenvolver uma metodologia para realizar inferência exata sobre o modelo geoestatístico com amostragem preferencial, sob o enfoque Bayesiano. A partir de simulações, mostramos o efeito de se ignorar a amostragem preferencial na estimação e predição do modelo, obtendo melhores resultados quando a amostragem preferencial é considerada.

1.2 Objetivos da tese

Motivados pela falta de estudos e trabalhos que abordam os problemas citados e pela importância de se ter soluções para tais, destacamos os três objetivos principais desta tese, cujas metodologias são novas no contexto de geoestatística com amostragem preferencial:

1. Desenvolver um algoritmo SAEM, sob abordagem clássica, com a finalidade de produzir um método de estimação para o problema que considere a informação da amostragem preferencial e que tenha boas propriedades dos estimadores, como convergência para os estimadores de máxima verossimilhança;

2. Desenvolver o modelo t-Student para os dados de geoestatística com amostragem preferencial a fim de acomodar, na modelagem, dados atípicos e realizar inferência correta na presença destes;
3. Desenvolver um modelo exato, sob abordagem Bayesiana, com o objetivo de realizar inferência sem erros de aproximação por discretização e fornecer resultados mais precisos.

Além dos objetivos principais, outros pontos foram abordados e desenvolvidos neste trabalho. Podemos citar o aprimoramento do algoritmo de geração de amostras da distribuição preditiva do processo latente condicionado aos dados, presente em Ferreira and Gamerman (2015), passando da geração de elemento por elemento para a geração de blocos, com o objetivo de torná-lo mais eficiente. Essa etapa foi fundamental para o desenvolvimento prático do nosso algoritmo SAEM para estimação dos parâmetros, pois necessitamos das simulações da distribuição preditiva do processo latente na etapa “E” do algoritmo. Dado que o algoritmo é iterativo, a geração de amostras elemento por elemento dessa distribuição tornaria sua execução extremamente ineficiente. Dessa forma, o desenvolvimento da geração de amostras por blocos da distribuição preditiva do processo latente não apenas reduziu o tempo de obtenção das amostras, como possibilitou boa execução do nosso algoritmo na prática.

Dado que, na inferência clássica, a predição é realizada utilizando os parâmetros estimados como verdadeiros, outro ponto abordado nesta tese é a análise dos resultados fornecidos quando combinamos métodos de estimação com métodos de predição diferentes. Em outras palavras, comparamos a predição obtida plugando o vetor de parâmetros estimados de cada método de estimação abordado em um determinado método de predição, possibilitando uma análise de quais combinações fornecem os melhores resultados.

1.3 Organização da tese

No Capítulo 2 apresentamos uma revisão de resultados que serão a base para construção das metodologias presentes nesta tese. A teoria de simulação via cadeias de Markov é importante para a geração de amostras de uma distribuição que não possui forma fechada. Naturalmente aparece no contexto Bayesiano mas também é importante sob a abordagem clássica, como veremos, na estimação dos parâmetros do modelo pro-

posto. Apresentamos uma revisão sobre o algoritmo EM e suas variações que são a base do algoritmo SAEM que propomos e comparamos a estimação por aproximação de Laplace, método já existente na literatura para o problema de amostragem preferencial. Por fim, introduzimos os conceitos de processos pontuais, cuja realização representará a configuração espacial da amostra obtida. É por essa ligação que a amostragem preferencial é modelada.

No Capítulo 3 apresentamos uma revisão do modelo geoestatístico tradicional, isto é, sem amostragem preferencial, sob as abordagens clássica e Bayesiana. Discutimos como os dados com estrutura espacial em domínio contínuo são modelados, introduzindo os conceitos de processo geoestatístico, correlograma e variograma. Apresentamos os métodos de estimação por máxima verossimilhança, no caso clássico, e por inferência Bayesiana e como realizamos predição para, por exemplo, construir mapas da variável de interesse sob a região de estudo. A partir dessa teoria, introduzimos o problema de amostragem preferencial no contexto da geoestatística e discutimos os principais desafios desse problema e as limitações das metodologias da literatura existente.

Apresentamos nossa metodologia sob a abordagem clássica no Capítulo 4, em que desenvolvemos os algoritmos MCEM e SAEM para o modelo Gaussiano e t-Student. Dessa forma, revisamos as metodologias existentes, como as presentes nos trabalhos de Diggle et al. (2010) e Dinsdale and Salibian-Barrera (2019), para compararmos com a desenvolvida nesta tese. Um estudo de simulação é realizado no Capítulo 5, em que mostramos os efeitos de alguns fatores no processo de estimação e predição: blocagem na geração do processo latente, fixação do parâmetro de correlação e malhas de diferentes tamanhos no processo de estimação. Por fim, comparamos os métodos de predição propostos.

Sob a abordagem Bayesiana, no Capítulo 6 desenvolvemos a metodologia para inferência exata sobre o modelo Gaussiano com amostragem preferencial e discutimos a modelagem considerando a distribuição t-Student. Apresentamos a modelagem e a computação, como obtemos as distribuições a posteriori e como realizamos predição para o modelo exato. No Capítulo 7, realizamos um estudo de simulação para analisar os efeitos da amostragem preferencial, o impacto ao fixar o parâmetro de correlação sob a abordagem Bayesiana e o comportamento do modelo quando há falta de informação. Também verificamos o impacto dos resultados fornecidos pelo modelo proposto quando a função de intensidade real não é a considerada pelo modelo e analisamos o ganho que

a inferência exata possui sobre a inferência com aproximação por discretização. Por fim, apresentamos resultados preliminares considerando o modelo com distribuição t-Student.

No Capítulo 8 apresentamos três aplicações a conjuntos de dados reais para ilustrar a metodologia desenvolvida. A primeira se refere ao dados de biomonitoramento de metais pesados em Galícia, como apresentado anteriormente. A segunda aplicação se refere aos dados de radiação gamma na Alemanha, cujo conjunto de dados foi parte do SIC 2004 - um exercício de comparação de interporlação espacial (*Spatial Interpolation Comparison*). Esses dados não possuem efeito da amostragem preferencial, tendo como objetivo apresentar os resultados do nosso modelo desenvolvido mesmo em situações que não temos preferência na amostragem. A terceira aplicação se refere aos dados de precipitação da cidade do Rio de Janeiro, em que a amostragem é potencialmente preferencial e apresenta dados atípicos, servindo como ilustração da aplicação do modelo t-Student.

Por fim, apresentamos nossas conclusões e direcionamentos de trabalhos futuros no Capítulo 9.

Capítulo 2

Preliminares

O objetivo deste capítulo é introduzir alguns conceitos sobre simulações via cadeias de Markov, métodos de estimação por algoritmo EM e por aproximação de Laplace e resultados sobre processos pontuais, que serão necessários para o desenvolvimento da presente tese.

2.1 Simulação via cadeias de Markov

Uma sequência $\theta_1, \theta_2, \dots$ de variáveis aleatórias é uma cadeia de Markov se a distribuição condicional de θ_{n+1} dado $\theta_1, \dots, \theta_n$ depende apenas de θ_n . Em outras palavras, uma cadeia de Markov é um processo estocástico em que, dado o presente, passado e futuro são independentes (Gamerman, 1997). A distribuição $p(\theta_{n+1}|\theta_n)$ é chamada de distribuição de transição, também denotada por $p(\theta_n, \theta_{n+1})$. Uma cadeia de Markov é aperiódica se todos seus estados são aperiódicos, isto é, há probabilidade positiva de a cadeia continuar no mesmo estado para todos os estados, e é recorrente se seus estados são recorrentes, ou seja, a probabilidade de a cadeia visitar cada um dos estados infinitas vezes é igual a 1. Se a cadeia de Markov é aperiódica e recorrente, então dizemos que é ergódica.

A ideia da simulação via MCMC é construir uma cadeia de Markov cuja distribuição de equilíbrio é igual a distribuição alvo π . Considere uma cadeia de Markov ergódica $(\theta^{(n)})_{n \geq 0}$ com espaço de estados $S \subset \mathbb{R}^d$, com distribuição de transição $p(x, y)$ e distribuição inicial $\pi^{(0)}$. O primeiro valor gerado da cadeia corresponde gerar $\theta^{(0)}$ de $\pi^{(0)}$. O próximo valor, $\theta^{(1)}$, é gerado pela distribuição $p(\theta^{(0)}, \cdot)$. Em seguida, $\theta^{(2)}$ é gerado pela distribuição $p(\theta^{(1)}, \cdot)$, e assim por diante. Após um número finito de simulações, digamos N^* , as amostras geradas se aproximam de amostras que seriam geradas pela distribuição de interesse. A partir desse ponto, as amostras geradas podem ser consideradas provenientes da distribuição π , pois a convergência foi atingida.

Apresentaremos dois métodos para construção de cadeias de Markov: o amostrador de Gibbs e o Metropolis-Hastings.

Amostrador de Gibbs

Suponha que o vetor de parâmetros é $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$ e que a distribuição de interesse é $\pi(\boldsymbol{\theta})$. A distribuição $\pi(\theta_i | \boldsymbol{\theta}_{-i})$, em que $\boldsymbol{\theta}_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)$, é chamada de condicional completa de θ_i . O amostrador de Gibbs (*Gibbs Sampling*) é um método MCMC cuja distribuição de transição é dada pelas distribuições condicionais completas, baseado em gerações sucessivas das distribuições $\pi(\theta_i | \boldsymbol{\theta}_{-i})$, para cada $i = 1, \dots, d$. Dessa forma, considera-se que é possível simular de todas as distribuições condicionais completas em questão.

O processo é realizado da seguinte forma:

1. Inicie as cadeias com valores iniciais $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})$ e o contador $j = 1$.
2. Obtenha um novo valor $\boldsymbol{\theta}^{(j)} = (\theta_1^{(j)}, \dots, \theta_d^{(j)})$ a partir de $\boldsymbol{\theta}^{(j-1)}$ através da geração sucessiva de valores:

$$\begin{aligned}\theta_1^{(j)} &\sim \pi(\theta_1 | \theta_2^{(j-1)}, \dots, \theta_d^{(j-1)}), \\ \theta_2^{(j)} &\sim \pi(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_d^{(j-1)}), \\ &\vdots \\ \theta_d^{(j)} &\sim \pi(\theta_d | \theta_1^{(j)}, \dots, \theta_{d-1}^{(j)}),\end{aligned}$$

3. Altere o contador para $j = j + 1$ e repita o passo 2 até que a convergência seja atingida.

Após a convergência ser atingida, os valores gerados pelo algoritmo são provenientes da distribuição alvo $\pi(\boldsymbol{\theta})$.

Metropolis-Hastings

O método de Metropolis-Hastings (MH) utiliza distribuições auxiliares para gerar amostras compatíveis com a distribuição de alvo $\pi(\boldsymbol{\theta})$. Nesse caso, deve-se construir uma distribuição de transição tal que a distribuição de equilíbrio da cadeia seja $\pi(\boldsymbol{\theta})$. Considere uma cadeia reversível tal que $\pi(\boldsymbol{\theta})p(\boldsymbol{\theta}, \boldsymbol{\theta}') = \pi(\boldsymbol{\theta}')\pi(\boldsymbol{\theta}', \boldsymbol{\theta})$, para todo $\boldsymbol{\theta}$ e $\boldsymbol{\theta}'$.

A condição de reversibilidade é suficiente para que π seja a distribuição de equilíbrio da cadeia (Gamerman, 1997).

A distribuição $p(\boldsymbol{\theta}, \boldsymbol{\theta}')$ é composta por dois elementos: uma distribuição arbitrária $q(\boldsymbol{\theta}, \boldsymbol{\theta}')$, que chamamos de distribuição de proposta, e uma probabilidade $\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}')$ tal que

$$p(\boldsymbol{\theta}, \boldsymbol{\theta}') = q(\boldsymbol{\theta}, \boldsymbol{\theta}')\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}'), \quad \boldsymbol{\theta} \neq \boldsymbol{\theta}'.$$

Para $\boldsymbol{\theta} = \boldsymbol{\theta}'$, temos que

$$p(\boldsymbol{\theta}, \boldsymbol{\theta}) = 1 - \int q(\boldsymbol{\theta}, \boldsymbol{\theta}')\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}')d\boldsymbol{\theta}'.$$

A probabilidade $\alpha(\cdot)$ é chamada de probabilidade de aceitação, dada por

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min \left\{ 1, \frac{\pi(\boldsymbol{\theta}')q(\boldsymbol{\theta}', \boldsymbol{\theta})}{\pi(\boldsymbol{\theta})q(\boldsymbol{\theta}, \boldsymbol{\theta}')} \right\}.$$

Basicamente, dado o estado atual da cadeia em $\boldsymbol{\theta}$, geramos um valor proposto $\boldsymbol{\theta}'$ para o próximo passo da cadeia através da distribuição de proposta $q(\boldsymbol{\theta}, \boldsymbol{\theta}')$ e aceitamos esse valor com probabilidade $\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}')$. Se o valor é aceito, a cadeia se move de $\boldsymbol{\theta}$ para $\boldsymbol{\theta}'$. Caso contrário, ela se mantém no estado $\boldsymbol{\theta}$.

O processo é realizado da seguinte forma:

1. Inicie a cadeia com valor inicial $\boldsymbol{\theta}^{(0)}$ e o contador $j = 1$.
2. Gere um valor proposto $\boldsymbol{\theta}'$ da distribuição de proposta $q(\boldsymbol{\theta}^{(j-1)}, \boldsymbol{\theta}')$.
3. Calcule a probabilidade de aceitar o movimento da cadeia $\alpha(\boldsymbol{\theta}^{(j-1)}, \boldsymbol{\theta}')$.
4. Gere uma variável aleatória uniforme U . Se $u \leq \alpha$, então o movimento proposto é aceito e $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}'$. Caso contrário, $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^{(j-1)}$.
5. Altere o contador para $j = j+1$ e repita os passos 2-4 até que a convergência seja atingida.

Após a convergência ser atingida, os valores gerados pelo algoritmo são provenientes da distribuição alvo π .

No método MH não é necessário conhecer a constante normalizadora da distribuição alvo, permitindo a geração de amostras de uma distribuição que apenas se conhece o núcleo. A probabilidade de aceitação não pode ser muito baixa, pois haveria poucos movimentos da cadeia e o algoritmo levaria muito tempo para convergir. Taxas de aceitação entre 20% e 50% indicam taxas razoáveis, que permitem uma convergência mais rápida do algoritmo garantindo que a cadeia passeie por grande parte do espaço da distribuição alvo (Gamerman, 1997).

Quando trabalhamos com modelos de alta dimensão, pode ser difícil encontrar distribuições de propostas apropriadas e, normalmente, a probabilidade de aceitação é pequena. Nesse caso, uma variação do algoritmo propõe gerar de cada componente do vetor $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$, a partir das distribuições de propostas $q(\theta_i, \theta'_i)$ com probabilidades de aceitação $\alpha(\theta_i, \theta'_i)$, $i = 1, \dots, d$. Uma outra variação é escolher blocos entre os componentes de $\boldsymbol{\theta}$, $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_{d^*}\}$, de acordo com a facilidade de amostragem de cada bloco.

Os algoritmos amostrador de Gibbs e MH podem ser combinados. Quando não há como gerar amostras de alguma condicional completa no amostrador de Gibbs, podemos utilizar um passo de MH. De fato, o amostrador de Gibbs é um caso particular do método MH, quando a distribuição de proposta $q(\theta_i, \cdot)$ é igual a distribuição condicional completa $\pi(\theta_i | \cdot)$. Nesse caso, o valor proposto é aceito com probabilidade 1.

2.2 Algoritmo EM e variações

O algoritmo EM ficou bastante conhecido após publicação do artigo Dempster et al. (1977). Apesar de ter sido proposto várias vezes por autores anteriores para situações específicas, foi nesse trabalho que o algoritmo foi consolidado para uma forma mais geral. O nome do método é dado pelas etapas do algoritmo: a etapa E, de esperança (*expectation*), e a etapa M, de maximização (*maximization*), resultando no algoritmo *Expectation-Maximization*.

O algoritmo EM é um processo iterativo para obtenção dos estimadores de máxima verossimilhança quando trabalhamos com dados incompletos. O contexto de dados incompletos pode ser definido por algumas situações:

- quando há dados faltantes, devido a limitações de observação ou censuras, por exemplo;

- quando há no modelo processos latentes não observados;
- quando introduzimos variáveis latentes no modelo apenas para simplificar a função de verossimilhança. Nesse caso, a variável latente não tem interpretação do modelo, servindo apenas para facilitar a estimação dos parâmetros do modelo.

Seja \mathbf{X} um vetor aleatório com função de densidade $g(\mathbf{x}|\boldsymbol{\theta})$ indexada por um vetor de parâmetros $\boldsymbol{\theta}$, $\boldsymbol{\theta} \in \Theta$, que corresponde ao vetor de dados completos. No contexto de dados incompletos, apenas uma parte de \mathbf{X} é observada, denotamos por $\mathbf{Y} = h(\mathbf{X})$ que possui função de densidade $g(\mathbf{y}|\boldsymbol{\theta})$ e corresponde ao vetor de dados incompletos. A função de log-verossimilhança dos dados completos é denotada por

$$l_c(\boldsymbol{\theta}; \mathbf{x}) = \log g(\mathbf{x}|\boldsymbol{\theta}). \quad (2.1)$$

Para obter os estimadores de máxima verossimilhança, devemos maximizar a função (2.1). Como o vetor \mathbf{x} não é totalmente observado, utilizamos a função de densidade marginal dos dados observados, isto é,

$$g(\mathbf{y}|\boldsymbol{\theta}) = \int g(\mathbf{x}|\boldsymbol{\theta}) d\mathbf{x}_{inc}, \quad (2.2)$$

em que \mathbf{x}_{inc} denota o vetor de dados não observados de \mathbf{x} . O algoritmo EM maximiza indiretamente $g(\mathbf{y}|\boldsymbol{\theta})$ de forma iterativa utilizando $g(\mathbf{x}|\boldsymbol{\theta})$. Na iteração $(j+1)$ do algoritmo, as etapas são realizadas da seguinte forma:

Passo E: Calcule a função $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)})$ dada por

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}) = \int \log g(\mathbf{x}|\boldsymbol{\theta}) p(\mathbf{x}_{inc}|\mathbf{y}, \boldsymbol{\theta}^{(j)}) d\mathbf{x}_{inc} = E[\log g(\mathbf{x}|\boldsymbol{\theta})|\boldsymbol{\theta}^{(j)}, \mathbf{y}];$$

Passo M: Escolha $\boldsymbol{\theta}^{(j+1)}$ como o valor de $\boldsymbol{\theta} \in \Theta$ que maximiza a função $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)})$, isto é,

$$\boldsymbol{\theta}^{(j+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}).$$

Os passos E e M são repetidos até que haja convergência das estimativas de $\boldsymbol{\theta}$.

Dempster et al. (1977) mostram que cada passo do algoritmo aumenta a função de verossimilhança dos dados observados, $g(\mathbf{y}|\boldsymbol{\theta})$. Dessa forma, mostram que a sequência

$\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots\}$ converge para um ponto estacionário de g . A convergência para um máximo global não é garantida, ou seja, não é certo que a convergência se dê em um EMV. Se a função de verossimilhança é unimodal, então o algoritmo converge para o EMV. Mas se a função é multimodal, o algoritmo converge para um valor de máximo local que dependerá do valor inicial $\boldsymbol{\theta}^{(0)}$.

Em alguns casos, não é possível calcular a função Q de forma analítica no passo E. Algumas variações do algoritmo foram propostas para contornar esse problema. Wei and Tanner (1990) propuseram substituir o passo E por uma integração de Monte Carlo, resultando no algoritmo MCEM (Monte Carlo EM). A ideia é aproximar a função Q por

$$\tilde{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}) = \frac{1}{m} \sum_{i=1}^m \log g(\mathbf{x}_{inc_m}, \mathbf{y}|\boldsymbol{\theta}^{(j)}), \quad (2.3)$$

em que \mathbf{x}_{inc_m} é uma simulação do vetor não observado \mathbf{x}_{inc} com distribuição $p(\mathbf{x}_{inc_m}|\mathbf{y}, \boldsymbol{\theta}^{(j)})$. No passo M, deve-se maximizar a função \tilde{Q} e alternar entre os passos iterativamente. Segundo Wei and Tanner (1990), um valor de m alto é ineficiente quando a cadeia de estimação está longe do valor verdadeiro. É recomendado aumentar m quando a cadeia se aproxima da região do valor verdadeiro de $\boldsymbol{\theta}$. Para monitorar a convergência do algoritmo, uma opção é construir o gráfico de cada componente de $\theta_i^{(j)}$ versus iteração, $i = 1, \dots, d$. Após um certo número de iterações, o gráfico revela uma flutuação em torno de $\theta_i = \hat{\theta}_i$. Nesse ponto, pode-se encerrar o algoritmo ou realizar mais iterações com um valor maior de m , que diminuirá a variabilidade da cadeia. Como a convergência do algoritmo se dá em torno de $\hat{\theta}_i$, deve-se tomar um conjunto dos últimos valores da cadeia e obter, por exemplo, o valor médio.

Outra variação que ficou bastante conhecida foi a aproximação estocástica do algoritmo EM (SAEM - *Stochastic Approximation EM*), proposta por Delyon et al. (1999). Assim como no MCEM, a ideia é separar o passo E em duas etapas: uma de simulação e uma de integração. Geramos um valor da variável não observada \mathbf{x}_{inc} e a integração de Monte Carlo é substituída por uma média estocástica. A função Q é aproximada da seguinte forma:

$$\hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}) = \hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j-1)}) + \gamma_j \left(\frac{1}{m} \sum_{i=1}^m \log g(\mathbf{x}_{inc_m}, \mathbf{y}|\boldsymbol{\theta}^{(j)}) - \hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j-1)}) \right), \quad (2.4)$$

em que $\{\gamma_j\}_{j \geq 1}$ é uma sequência de pesos positivos tal que $\sum_{j=1}^{\infty} \gamma_j = \infty$ e $\sum_{j=1}^{\infty} \gamma_j^2 < \infty$.

Os pesos γ_j 's controlam a importância dos valores simulados na iteração j . Se $\gamma_j = 1, \forall j > 0$, então o algoritmo SAEM não tem memória (não utiliza a informação de nenhuma iteração anterior) e se resume ao MCEM. Caso contrário, o algoritmo utiliza a informação gerada das iterações anteriores para aproximar a função Q . A importância das iterações nessa aproximação depende da escolha da sequência $\{\gamma_j\}_{j \geq 1}$. Assim como proposto em Galarza et al. (2017), consideramos a seguinte sequência:

$$\gamma_j = \begin{cases} 1, & \text{se } 1 \leq j \leq cW; \\ \frac{1}{j-cW} & \text{se } cW + 1 \leq j \leq W; \end{cases}$$

em que W é o número máximo de iterações e c é a porcentagem de iterações iniciais sem memória. Se $c = 0$, então o algoritmo tem memória para todas as iterações e convergirá lentamente. Por outro lado, se $c = 1$ o algoritmo não terá memória e convergirá rapidamente para a região de um valor máximo. Um valor de $c \in (0, 1)$ permite que o algoritmo convirja rapidamente em distribuição para uma solução nas iterações sem memória e, em seguida, convirja quase certamente para essa solução (Galarza et al., 2017).

Se desenvolvermos a recursão em (2.4), obtemos

$$\hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}) = \sum_{r=cW}^j \omega_r \hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}), \quad (2.5)$$

em que $\omega_j = \gamma_j$ e $\omega_r = \gamma_r \prod_{s=r+1}^j (1 - \gamma_s)$ para $r = cW, \dots, j - 1$. Se consideramos a sequência proposta por Galarza et al. (2017), então $\omega_r = 0$ para $r = cW$ e $\omega_r = 1/(j - cW)$ para $r = cW + 1, \dots, j$, ou seja, todos os valores amostrados entre as iterações $cW + 1$ e j possuem a mesma importância na obtenção de $\boldsymbol{\theta}^{(j)}$. Os valores de c e W dependem do modelo e do problema em questão. Para avaliar os melhores valores, recomenda-se monitorar graficamente a cadeia de $\boldsymbol{\theta}^{(j)}$.

Para termos informação sobre a incerteza da estimativa dos parâmetros, precisamos conhecer a matriz de informação de Fisher, que nos permite calcular o erro padrão de suas estimativas. A matriz de informação observada é definida por $I(\boldsymbol{\theta}) = -\partial_{\boldsymbol{\theta}}^2 l(\boldsymbol{\theta}; \mathbf{y})$, em que $\partial_{\boldsymbol{\theta}}$ denota o diferencial em relação a $\boldsymbol{\theta}$.

Segundo o princípio da informação de Louis (1982), o vetor gradiente da função

de log-verossimilhança incompleta $l(\boldsymbol{\theta})$ é igual a esperança condicional do vetor gradiente da função de log-verossimilhança completa,

$$\partial_{\theta} l(\boldsymbol{\theta}; \mathbf{y}) = E \left[\partial_{\theta} l_c(\boldsymbol{\theta}; \mathbf{x}) \middle| \mathbf{y}, \boldsymbol{\theta} \right].$$

Dada uma amostra $\mathbf{x}_{inc} = (\mathbf{x}_1, \dots, \mathbf{x}_L)$ da distribuição preditiva de \mathbf{x}_{inc} na iteração atual do algoritmo, o vetor gradiente pode ser aproximado por

$$\partial_{\theta} l(\boldsymbol{\theta}; \mathbf{y}) = \frac{1}{L} \sum_{l=1}^L \partial_{\theta} l_c(\boldsymbol{\theta}; \mathbf{x}_l). \quad (2.6)$$

em que $\mathbf{x}_l = (\mathbf{y}, \mathbf{x}_{inc_l})'$. Essa aproximação é realizada por Wei and Tanner (1990) no contexto do algoritmo MCEM. Delyon et al. (1999) propõe aproximar o vetor gradiente pela aproximação estocástica

$$\nabla_k = \nabla_{k-1} + \gamma_k \left(\frac{1}{L} \sum_{l=1}^L \partial_{\theta} l_c(\boldsymbol{\theta}; \mathbf{x}_l) - \nabla_{k-1} \right), \quad (2.7)$$

em que $\{\gamma_k\}$ é a sequência de pesos utilizada no algoritmo SAEM. Segundo Louis (1982), a matriz Hessiana pode ser expressa por

$$\begin{aligned} \mathbf{H}(\boldsymbol{\theta}) &= E \left[\partial_{\theta}^2 l_c(\boldsymbol{\theta}; \mathbf{x}) \middle| \mathbf{y}, \boldsymbol{\theta}^0 \right] + E \left[(\partial_{\theta} l_c(\boldsymbol{\theta}; \mathbf{x})) (\partial_{\theta} l_c(\boldsymbol{\theta}; \mathbf{x}))' \middle| \mathbf{y}, \boldsymbol{\theta}^0 \right] - \\ &\quad \left(E \left[\partial_{\theta} l_c(\boldsymbol{\theta}; \mathbf{x}) \middle| \mathbf{y}, \boldsymbol{\theta}^0 \right] \right) \left(E \left[\partial_{\theta} l_c(\boldsymbol{\theta}; \mathbf{x}) \middle| \mathbf{y}, \boldsymbol{\theta}^0 \right] \right)'. \end{aligned} \quad (2.8)$$

O primeiro termo de (2.8) corresponde à esperança condicional da matriz de informação observada para os dados completos e os termos restantes produzem a informação esperada para a distribuição condicional de X dado Y . Esses termos podem ser facilmente aproximados por

$$\begin{aligned} \hat{H}(\boldsymbol{\theta}) &= \frac{1}{L} \sum_{l=1}^L \partial_{\theta}^2 l_c(\boldsymbol{\theta}; \mathbf{x}_l) + \frac{1}{L} \sum_{l=1}^L [(\partial_{\theta} l_c(\boldsymbol{\theta}; \mathbf{x}_l)) (\partial_{\theta} l_c(\boldsymbol{\theta}; \mathbf{x}_l))'] - \\ &\quad \left(\frac{1}{L} \sum_{l=1}^L \partial_{\theta} l_c(\boldsymbol{\theta}; \mathbf{x}_l) \right) \left(\frac{1}{L} \sum_{l=1}^L \partial_{\theta} l_c(\boldsymbol{\theta}; \mathbf{x}_l) \right)', \end{aligned} \quad (2.9)$$

proposto por Wei and Tanner (1990). Note que essa aproximação considera apenas a informação da simulação de \mathbf{x}_{inc} na iteração atual do algoritmo (MCEM). Para incorporar

a informação contida das simulações da variável não observada nas iterações anteriores, no contexto do algoritmo SAEM, Delyon et al. (1999) propuseram utilizar

$$\begin{aligned}\mathbf{G}_k &= \mathbf{G}_{k-1} + \gamma_k \left(\frac{1}{L} \sum_{l=1}^L \left(\partial_{\theta}^2 l_c(\boldsymbol{\theta}; \mathbf{x}_l) + (\partial_{\theta} l_c(\boldsymbol{\theta}; \mathbf{x})) (\partial_{\theta} l_c(\boldsymbol{\theta}; \mathbf{x}))' \right) - \mathbf{G}_{k-1} \right), \\ \mathbf{H}_k &= \mathbf{G}_k - \nabla_k \nabla_k' .\end{aligned}\tag{2.10}$$

Dado que o algoritmo SAEM converge para um vetor $\boldsymbol{\theta}^*$ e que $\boldsymbol{\theta} \rightarrow l(\boldsymbol{\theta})$ é suficientemente suave, \mathbf{H}_k converge para a matriz Hessiana em $\boldsymbol{\theta}^*$. Dessa forma, podemos utilizar o valor limite de \mathbf{H}_k para avaliar o erro padrão do estimador. Note que, se $\gamma_k = 1 \forall k$, a aproximação equivale à do trabalho de Wei and Tanner (1990), pois o algoritmo SAEM corresponde ao MCEM nesse caso.

2.3 Aproximação de Laplace

Suponha que queremos calcular a integral $\int_{\mathbb{R}^n} g(\mathbf{x}) d\mathbf{x}$, em que $\mathbf{x} \in \mathbb{R}^n$, e que esse cálculo seja difícil de realizar. Nesse caso, o uso de aproximação é desejável. Seja $h(\mathbf{x}) = \log g(\mathbf{x})$. A aproximação de Laplace será dada por

$$\int \exp(h(\mathbf{x})) d\mathbf{x} \approx \exp(h(\mathbf{x}_0)) (2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2},\tag{2.11}$$

em que $\boldsymbol{\Sigma} = \left[-\frac{d^2}{d\mathbf{x}d\mathbf{x}'} h(\mathbf{x}_0) \right]^{-1}$ é menos a inversa da matriz Hessiana avaliada na moda da função. Esse resultado é obtido expandindo a função $h(\mathbf{x})$ em série de Taylor em torno da moda da distribuição \mathbf{x}_0 e detalhes podem ser consultados em Gamerman (1997) e Tierney et al. (1989).

Logo, a aproximação de Laplace para a integral em (2.11) é dada pelo valor da função $g(x)$ aplicada em sua moda multiplicado por uma constante que depende da curvatura da função $\log(g(x))$. Essa aproximação funciona bem para funções de quadrado integrável, ou seja, $\int g(x)^2 dx < \infty$, que decrescem rapidamente em ambas as caudas da função.

Note que substituímos o problema de integração por um problema de maximização: necessitamos da moda da função $g(x)$ para poder calcular a aproximação de Laplace. Em geral, esse método é utilizado quando o problema de encontrar o máximo da função é mais fácil do que integrar a função em si.

2.4 Processos pontuais

Processos Pontuais espaciais são utilizados para modelar padrões pontuais em que os pontos usualmente são posições em uma região bi ou tridimensional. Sua aplicação abrange várias áreas, como astronomia, ecologia, silvicultura e epidemiologia. Com o desenvolvimento tecnológico e computacional, o uso de métodos como MCMC se intensificou e permitiu o crescimento no estudo de processos pontuais. Møller and Waagepetersen (2003) apresentam uma revisão teórica abrangendo vários modelos de processos pontuais espaciais e várias aplicações.

Considere um espaço $B \subseteq \mathbb{R}^d$, usualmente $d = 2$ ou 3 . Um processo pontual X é um subconjunto aleatório enumerável do espaço B e o número de eventos em B é uma variável aleatória denotada por $N(B)$. Nessa tese trabalhamos com subconjuntos finitos e, conseqüentemente, $N(B) < \infty$. Uma realização do processo pontual X é denotada por $x = (x_1, \dots, x_n)$, em que n representa o número de pontos de X . Chamamos x de uma configuração pontual e cada x_i de evento.

2.4.1 Processo de Poisson

Para definir um processo de Poisson, vamos primeiramente definir a medida de intensidade em B . Seja $\lambda : B \rightarrow [0, \infty)$ a função de intensidade no espaço B e suponha que seja localmente integrável, isto é, $\int_{B'} \lambda(x) dx < \infty$ para toda região limitada $B' \subseteq B$. A medida de intensidade em uma região B' é dada por

$$m(B') = \int_{B'} \lambda(x) dx.$$

Seja f uma função de densidade em $B' \subseteq B$. Um processo pontual X que consiste em n pontos independentes e identicamente distribuídos (iid.) com densidade f é chamado de processo pontual binomial de n pontos em B com densidade f . Denotamos $X \sim \text{binomial}(B', n, f)$. O caso mais simples ocorre quando os n pontos tem distribuição uniforme em B' , ou seja, $f(x) = 1/|B'|$ é a função de densidade uniforme em B' , sendo $|B'|$ o volume de B' .

A seguir, definimos o processo pontual de Poisson.

Definição 2.1 *Um processo pontual X em B é um processo pontual de Poisson com função de intensidade λ se as seguintes propriedades são satisfeitas:*

- para qualquer subconjunto $B' \subseteq B$, $N(B') \sim \text{Pois}(m(B'))$, a distribuição Poisson com média $m(B') < \infty$,
- para qualquer $n \in \mathbb{N}$ e $B' \subseteq B$ com $0 < m(B) < \infty$, condicional em $N(B) = n$, $\mathbf{X}_{B'} \sim \text{binomial}(B', n, f)$, em que $f(x) = \lambda(x)/m(B)$.

Denotamos $\mathbf{X} \sim \text{Poisson}(B, \lambda)$.

Para qualquer região limitada $B' \subseteq B$, $m(B')$ determina o número de pontos esperado em B' , isto é, $E(N(B')) = m(B')$. Note que $\lambda(x)dx$ é a probabilidade de ocorrência de um ponto em uma bola infinitesimal de centro x e volume dx . Se a função de intensidade é constante, $\lambda(x) = \lambda$, $\forall x \in B$ e $\lambda \geq 0$, então \mathbf{X} é chamado de processo de Poisson homogêneo em B com função de intensidade λ . Caso contrário, é chamado de processo de Poisson não homogêneo. O caso em que $f(x)$ é a função de densidade uniforme em B representa um exemplo de um processo de Poisson homogêneo. A Figura 2.1 apresenta exemplos simulados de processos de Poisson homogêneo e não homogêneo.

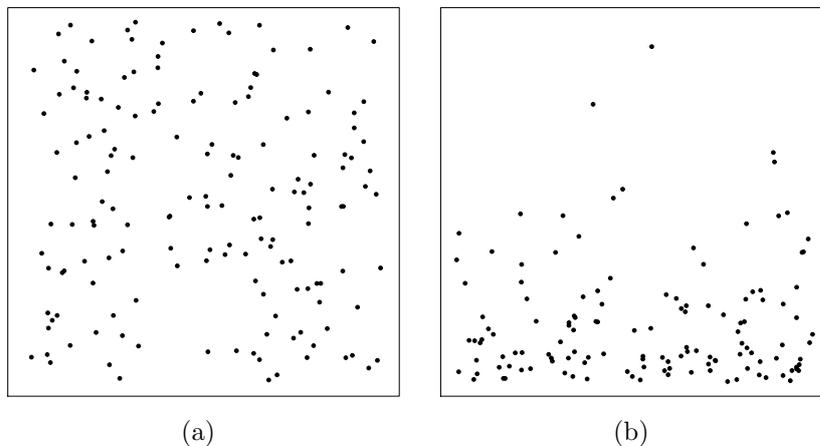


Figura 2.1: Simulação de um processo de Poisson (a) homogêneo e (b) não homogêneo no quadrado unitário. Em ambos os casos, o número esperado de pontos é igual a 150. Para o processo não homogêneo, foi considerada a função de intensidade $\lambda(x) = 750e^{-5c_2}/(1 - e^{-5})$, sendo c_2 a coordenada vertical do ponto x .

Um processo pontual \mathbf{X} em B é estacionário se sua distribuição é invariante sob translações e é isotrópico se sua distribuição é invariante sob rotações em relação a origem em \mathbb{R}^d .

A distribuição de um processo pontual X em uma região B é equivalente à distribuição conjunta do número de pontos $N(B)$ que ocorrem nessa região e a configuração dos $N(B)$ pontos em B . Assim, a função de densidade do processo de Poisson X pode

ser construída a partir da Definição 2.1, tomando a seguinte forma:

$$\begin{aligned}
 f(n, x|\lambda) &= P(X = x|N = n, \lambda) \times P(N = n|\lambda) \\
 &= \left[\prod_{i=1}^n \frac{\lambda(x_i)}{m(B)} \right] \times e^{-m(B)} \frac{m(B)^n}{n!} \\
 &= \frac{e^{-m(B)}}{n!} \left[\prod_{i=1}^n \lambda(x_i) \right].
 \end{aligned} \tag{2.12}$$

Para o processo de Poisson homogêneo, a função de densidade se reduz a

$$f(x|\lambda) = \frac{e^{-\lambda|B|} \lambda^n}{n!}, \tag{2.13}$$

em que $|B|$ é o volume do região B .

No caso em que o número de pontos $N(B)$ não é tratado como variável aleatória, apenas a posição dos pontos no espaço é desconhecida. Dessa forma, a função de densidade se dá apenas pela distribuição dos n pontos no espaço B , ficando

$$f(x|\lambda) = m(B)^{-n} \left[\prod_{i=1}^n \lambda(x_i) \right]. \tag{2.14}$$

Vamos definir duas operações sobre processos pontuais: superposição e *thinning* (Møller and Waagepetersen (2003)).

Definição 2.2 A união disjunta $\bigcup_{i=1}^{\infty} X_i$ de processos pontuais X_1, X_2, \dots é chamada de *superposição*.

Definição 2.3 Seja $p : B \rightarrow [0, 1]$ uma função e X um processo pontual em B . O processo pontual $X_{thin} \subseteq X$ é obtido da seguinte forma:

- cada ponto $b \in X$ é incluído em X_{thin} com probabilidade $p(b)$,
- os pontos são incluídos/excluídos de forma independente.

O processo X_{thin} é chamado de *thinning independente* de X com probabilidades de retenção $p(b)$, $b \in B$.

Da Definição (2.2), a superposição é a união de processos pontuais disjuntos. Por exemplo, o processo resultante da superposição dos processos pontuais disjuntos X_1 e X_2

consiste da união dos pontos de ambos processos. De forma geral, se X é o processo resultante da superposição de processos pontuais disjuntos X_1, X_2, \dots , então $N_X(B) = N_{X_1}(B) + N_{X_2}(B) + \dots$

Da Definição (2.3), a operação de *thinning* se dá ao eliminar pontos de um processo pontual. Por exemplo, seja $Z \sim \text{Bernoulli}(p(b))$, $b \in X$, X um processo pontual em B . Realizar um *thinning* em X significa que, para cada ponto de X , b será retido ou eliminado de acordo com uma realização da variável Z . O ponto é retido se $Z = 1$ (com probabilidade $p(b)$) e eliminado caso contrário.

Os processos pontuais X_{thin} e $X \setminus X_{thin}$ são independentes e disjuntos. Logo, X pode ser visto como a superposição de ambos processos. Nesse sentido, a operação de *thinning* pode ser considerada como a operação inversa à superposição. A Figura 2.2 apresenta um exemplo dessas operações.

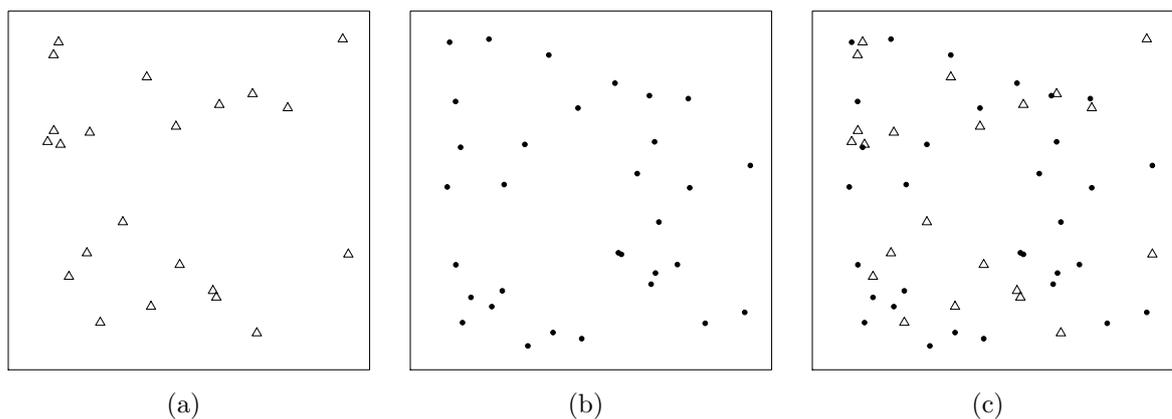


Figura 2.2: Simulação de dois processos pontuais (a) X_1 e (b) X_2 e o processo resultante da superposição (c) X . Podemos observar a operação inversa, em que o processo pontual X_1 é o resultado de se aplicar o *thinning* no processo X e o processo X_2 é composto pelos pontos eliminados de X .

A classe de processos de Poisson é fechada sob as operações de superposição e *thinning*, como apresentam as proposições a seguir (Møller and Waagepetersen, 2003).

Proposição 2.1 *Suponha que $X_i \sim \text{Poisson}(B, \lambda_i)$, $i = 1, 2, \dots$ são mutuamente independentes e que $\lambda = \sum \lambda_i$ seja localmente integrável. Então $X = \bigcup_{i=1}^{\infty} X_i$ é uma superposição de processos disjuntos e $X \sim \text{Poisson}(B, \lambda)$.*

Proposição 2.2 *Suponha que $X \sim \text{Poisson}(B, \lambda)$ é um processo que será aplicado a operação de *thinning* com probabilidades de retenção $p(b)$, $b \in X \subset B$. Seja $\lambda_{thin}(b) =$*

$p(b)\lambda(b)$. Então, X_{thin} e $X \setminus X_{thin}$ são processos de Poisson independentes com funções de intensidade dadas por λ_{thin} e $\lambda - \lambda_{thin}$, respectivamente.

Das Proposições 2.1 e 2.2, temos que ao aplicar qualquer uma das operações de *thinning* e superposição em processos de Poisson, obtemos como resultado um processo de Poisson. Esse é um resultado bastante útil na simulação de processos de Poisson não homogêneos, como veremos a seguir.

2.4.2 Simulação de processo de Poisson

Para simular de um processo de Poisson, vamos considerar primeiramente o caso homogêneo. Da Definição 2.1 e da Equação (2.13), é fácil ver que a simulação de um processo de Poisson homogêneo é trivial e é descrita pelo Algoritmo 1.

Algoritmo 1: Simulação de um processo de Poisson homogêneo

1. Simule $N(B) \sim Pois(\lambda|B|)$ o número de pontos do processo X , em que $|B|$ é o volume de B ,
2. Dado $N(B) = n$, distribua uniformemente os n pontos em B .
3. Compute os locais $\{x_1, \dots, x_n\}$.

Ou seja, para realizar uma simulação de um processo de Poisson homogêneo, basta simular o número de pontos do processo através de uma distribuição de Poisson com média $\lambda|B|$ e distribuir uniformemente os pontos na região B . A Equação (2.13) representa a versão analítica do algoritmo 1.

A simulação não é tão direta para processos de Poisson não homogêneos. Seja $\lambda(x)$ a função de intensidade de um processo de Poisson não homogêneo X e assumamos que $\lambda(x)$ seja limitada superiormente por λ^* . Nesse caso, a simulação do processo X pode ser obtida considerando a realização de *thinning* em um processo de Poisson homogêneo X^* cuja função de intensidade é igual ao $\sup(\lambda(x)) = \lambda^*$. O Algoritmo 2 apresenta o processo de simulação nesse caso.

Algoritmo 2: Simulação de um processo de Poisson não homogêneo com função de intensidade limitada

1. Simule o processo de Poisson homogêneo X^* com função de intensidade constante λ^* na região B :
 - simule $N_{X^*}(B) \sim Pois(\lambda^*|B|)$ o número de pontos do processo X^* , em que $|B|$ é o volume de B ,
 - dado $N_{X^*}(B) = n_{X^*}$, distribua uniformemente os n_{X^*} pontos em B , obtendo os locais $\{x_1, \dots, x_{n_{X^*}}\}$,
2. retenha cada ponto com probabilidades $\lambda(x_i)/\lambda^*$, $i = 1, \dots, n_{X^*}$,
3. compute os locais retidos na etapa 2.

A versão analítica do Algoritmo 2 é representada pela distribuição conjunta dos processos X e X^* , que pode ser escrita como

$$\begin{aligned}
 P(x, x^* | \lambda) &= P(x | x^*, n_{X^*}, \lambda) P(x^* | n_{X^*}, \lambda) P(n_{X^*} | \lambda) \\
 &= P(x | x^*, \lambda) P(x^* | n_{X^*}) P(n_{X^*} | \lambda^*) \\
 &= \left[\prod_{i=1}^{n_X} \frac{\lambda(x_i)}{\lambda^*} \right] \left[\prod_{i=n_X+1}^{n_{X^*}} \left(1 - \frac{\lambda(x_i)}{\lambda^*} \right) \right] \left[\prod_{i=1}^{n_{X^*}} \frac{1}{|B|} \right] e^{-\lambda^*|B|} \frac{(\lambda^*|B|)^{n_{X^*}}}{n_{X^*}!} \\
 &= \left[\prod_{i=1}^{n_X} \frac{\lambda(x_i)}{\lambda^*} \right] \left[\prod_{i=n_X+1}^{n_{X^*}} \left(1 - \frac{\lambda(x_i)}{\lambda^*} \right) \right] e^{-\lambda^*|B|} \frac{(\lambda^*)^{n_{X^*}}}{n_{X^*}!}. \tag{2.15}
 \end{aligned}$$

Dessa forma, a saída do Algoritmo 2 é uma simulação do processo não homogêneo X . Basta notar que, ao final do passo 1 obtemos uma simulação de um processo homogêneo com função de intensidade λ^* . Ao final do passo 3, realizamos um *thinning* em que a probabilidade de retenção dos pontos de X^* é dada por $\lambda(x)/\lambda^*$. Logo, obtemos um processo de Poisson com função de intensidade $\lambda^* \lambda(x)/\lambda^* = \lambda(x)$.

Podemos estar interessados em gerar um número específico de pontos do processo de Poisson. Nesse caso, $N_X(B)$ não é considerado aleatório e é fixado em n . No caso do processo de Poisson homogêneo, basta distribuir uniformemente n pontos sobre a região B . Para o processo de Poisson não homogêneo, basta gerar um número de pontos do processo X^* maior que n e selecionar n pontos com probabilidades $\lambda(x_i)/\lambda^*$.

2.4.3 Processo de Cox

Muitas vezes o processo de Poisson é simples demais para analisar dados reais. Uma extensão do processo de Poisson é o processo de Cox, que apresenta uma função de intensidade estocástica. A seguir, apresentamos a definição do processo de Cox presente em Møller and Waagepetersen (2003).

Definição 2.4 *Suponha que $S = \{S(b) : b \in B\}$ seja um campo aleatório não negativo em B . Se a distribuição condicional do processo pontual X dado S é um processo de Poisson em B com função de intensidade S , então X é chamado de processo de Cox induzido por S .*

Da Definição 2.4, um processo de Cox é um processo pontual cuja função de intensidade é dada por um campo aleatório. Quando S é constante, o processo de Cox se resume a um processo de Poisson com função de intensidade $\lambda = S$. Para o processo de Cox X , utilizaremos a notação $\Lambda(x)$ para denotar sua função de intensidade, diferenciando, assim, da função de intensidade do processo de Poisson $X|S$, denotada por $\lambda(x)$. A função de intensidade do processo de Cox é dada por

$$\Lambda(x) = E[S(x)],$$

e do processo de Poisson $X|S$ é

$$\lambda(x) = S(x).$$

Usualmente o campo aleatório S não é observado e não é possível diferenciar o processo de Cox X do processo de Poisson $X|S$ com apenas uma realização de X . Nesse caso, um conhecimento sobre o fenômeno em estudo é importante para decidir qual dos dois processos utilizar para modelagem. Em outras palavras, considerar que S seja aleatório ou determinístico depende do problema em mãos.

Uma classe particular do processo de Cox é o processo de Cox log-Gaussiano (Møller and Waagepetersen, 2003).

Definição 2.5 *Suponha que $S = \{S(b) : b \in B\}$ seja um processo Gaussiano em B e que X seja um processo de Cox induzido por $\exp\{S\}$. Então, X é chamado de processo de Cox log-Gaussiano (LGCP - log Gaussian Cox process).*

A função de intensidade de um LGCP é dada por

$$\Lambda(x) = E[\exp\{S(x)\}] = \exp\{\mu(x) + \sigma^2(x)/2\}, \quad (2.16)$$

em que $\mu(x)$ e $\sigma^2(x)$ são a média e variância do processo Gaussiano S no ponto x .

A simulação de LGCPs se dá em duas etapas: (a) a simulação do processo Gaussiano S e (b) a do processo de Poisson $X|S$. Como o processo Gaussiano é infinito dimensional, simular S na região B significa simular infinitos pontos $S(b)$, $b \in B$. Computacionalmente não é possível, então é necessário utilizar uma malha refinada o suficiente para aproximar o processo Gaussiano S . Seja b'_i o centróide da i -ésima célula, $i = 1, \dots, n_g$, e considere que o processo S é constante dentro de cada célula. Se assumirmos $\tilde{S} = (S(b'_1), S(b'_2), \dots, S(b'_{n_g}))'$, então, $\tilde{S} \sim NM(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, em que $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ são o vetor de médias e a matriz de covariâncias de \tilde{S} . Dessa forma, simulamos de forma aproximada o processo S através de uma realização de uma distribuição normal multivariada.

Dada a amostra gerada de S , utilizamos o algoritmo 1 para gerar de um processo de Poisson homogêneo em cada célula (pois, internamente, a função de intensidade é constante). A coleção dos pontos gerados será uma realização do LGCP desejado. Note que, se tomarmos como $\sup(\lambda) = \max(\exp(\tilde{S}))$ e utilizarmos o Algoritmo 2 para simular do processo de Poisson não homogêneo, o número de pontos obtidos será pequeno, pois as probabilidades de retenção serão muito pequenas para a maioria dos pontos.

2.4.4 Inferência em processos pontuais

Dada uma realização $x = (x_1, \dots, x_n)'$ de um processo de Poisson, sua função de verossimilhança é dada por

$$L(\lambda|x) \propto \exp\left\{-\int_B \lambda(u)du\right\} \left[\prod_{i=1}^n \lambda(x_i)\right], \quad (2.17)$$

em que a integral é calculada sobre toda a região B . Nota-se, então, que a complexidade da função de verossimilhança depende da escolha da função de intensidade. Para um processo de Poisson homogêneo, a Equação (2.17) se resume a

$$L(\lambda|x) \propto \exp\{-\lambda|B|\} \lambda^n. \quad (2.18)$$

e a inferência sobre λ , nesse caso, é trivial. Considerando a abordagem clássica, o estimador de máxima verossimilhança para λ será $\hat{\lambda} = n/|B|$. Para a abordagem Bayesiana, se consideramos como distribuição a priori para λ uma distribuição Gama(α, β), a distribuição a posteriori será Gama($\alpha+n, \beta + |B|$).

Para funções de intensidade mais complexas, a integral em (2.17) pode não possuir forma fechada e a inferência se torna complicada. A função de verossimilhança para o processo de Poisson $X|S$, em que marginalmente X é um LGCP, é dada por

$$L(\lambda|x) \propto \exp \left\{ - \int_B \exp\{S(u)\} du \right\} \left[\prod_{i=1}^n \exp\{S(x_i)\} \right] \quad (2.19)$$

e a integral acima não possui forma analítica. Porém, se discretizarmos a região B , podemos aproximar tal integral e realizar inferência sobre o modelo. Dessa forma, trabalhamos com a função de verossimilhança aproximada

$$\tilde{L}(\lambda|x) \propto \exp \left\{ - \sum_{b' \in B} \exp\{S(b')\} \right\} \left[\prod_{i=1}^n \exp\{S(x_i)\} \right], \quad (2.20)$$

em que b' representa os centróides das células da malha em B . Vários autores utilizam a discretização para realizar inferência sobre LGCPs, veja como exemplos Diggle et al. (2010), Dinsdale and Salibian-Barrera (2019), Ferreira and Gamerman (2015). Porém, tal aproximação é uma fonte de erros de difícil mensuração, podendo produzir resultados viesados (Simpson et al., 2016). Além disso, é necessário se ter um balanço entre o refinamento da malha e o custo computacional, pois a simulação de processos Gaussianos são da ordem n^3 , tornando-se impraticável já para tamanhos de malha moderados. Como consequência, o nível de discretização necessária para se obter bons resultados pode não ser alcançável.

Capítulo 3

Geoestatística com Amostragem Preferencial

Neste capítulo, revisamos os conceitos básicos de geoestatística, discutindo como incorporar a estrutura de correlação espacial na modelagem desse tipo de dado, como estimar os parâmetro do modelo via inferência clássica e Bayesiana e, mais importante, como realizar predição para locais não observados da região de estudo. Então, introduzimos a abordagem da amostragem preferencial aplicada na geoestatística, apresentando resultados já obtidos na literatura e suas principais limitações e, a partir dessa discussão, abordamos nossas contribuições da tese para esse problema.

3.1 Processo geoestatístico

Um processo geoestatístico $\{S(\mathbf{b}); \mathbf{b} \in B\}$ é um processo estocástico em que \mathbf{b} representa um ponto que se desloca continuamente sobre uma região $B \subset \mathbb{R}^d$, com $d \geq 1$. A região B é denominada domínio do processo geoestatístico e usualmente é uni, bi ou tri-dimensional. O ponto \mathbf{b} é denominado localização e contém a informação das coordenadas em B , isto é, $\mathbf{b} = (b_1, b_2, \dots, b_d)'$. A natureza contínua do processo se dá pela continuidade da região em estudo e não da variável aleatória em si. Por exemplo, se $S(\cdot)$ for uma variável aleatória discreta, há infinitas variáveis aleatórias $S(b)$ no espaço contínuo B .

Um processo $S(\mathbf{b})$ é caracterizado através de sua função de distribuição de probabilidade (*f.d.p.*) conjunta, definida para um número finito de localizações,

$$F_{\mathbf{b}_1, \dots, \mathbf{b}_n}(s_1, \dots, s_n) = P[S(\mathbf{b}_1) \leq s_1, \dots, S(\mathbf{b}_n) \leq s_n],$$

em que $\mathbf{b}_1, \dots, \mathbf{b}_n$ são localizações do domínio B , s_1, \dots, s_n pertencem a \mathbb{R} e n é um número inteiro.

Para duas localizações \mathbf{b}_i e \mathbf{b}_j , as variáveis aleatórias $S(\mathbf{b}_i)$ e $S(\mathbf{b}_j)$ não possuem, necessariamente, a mesma distribuição e normalmente não são independentes. É comum conhecermos apenas uma realização do processo $S(\mathbf{b})$, o que torna impossível realizar inferência sobre o mesmo. Portanto, é necessário admitir hipóteses restritivas: as condições de estacionariedade. Essas garantem a regularidade do processo estocástico em todo o

domínio da região em estudo. Algumas admitem fortes suposições ao passo que outras tem suposições consideradas mais fracas. A seguir apresentamos a definição da condição de estacionariedade forte.

Definição 3.1 *Seja um processo geoestatístico $S(\mathbf{b})$ com domínio $B \subset \mathbb{R}^d$. O processo $S(\mathbf{b})$ apresenta estacionariedade forte se, para qualquer conjunto finito de localizações $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$, para qualquer conjunto finito de número reais $\{s_1, \dots, s_n\}$ e para qualquer vetor $\mathbf{h} \in \mathbb{R}^d$ tal que $\mathbf{b}_i + \mathbf{h} \in B$, $i = 1, \dots, n$, então*

$$F_{\mathbf{b}_1, \dots, \mathbf{b}_n}(s_1, \dots, s_n) = F_{\mathbf{b}_1 + \mathbf{h}, \dots, \mathbf{b}_n + \mathbf{h}}(s_1, \dots, s_n), \quad \forall n \in \mathbb{N}.$$

A Definição 3.1 diz que a lei de distribuição do processo $S(\mathbf{b})$ é invariante por translação e os fatores do seu comportamento agem de forma similar em toda a área de estudo. Em outras palavras, as variáveis aleatórias do processo são identicamente distribuídas. Essa é uma hipótese forte para ser assumida em diversos fenômenos naturais. Por isso, uma hipótese menos restritiva deve ser assumida, a estacionariedade de segunda ordem.

Definição 3.2 *Um processo $S(\mathbf{b})$ com domínio $B \subset \mathbb{R}^d$ apresenta estacionariedade de segunda ordem se possui primeiro e segundo momentos constantes. Para dois pontos \mathbf{b}_i e \mathbf{b}_j , seja \mathbf{h} o vetor de distância de \mathbf{b}_i a \mathbf{b}_j . Então, temos que*

$$(i) \quad E[S(\mathbf{b})] = \mu, \quad \forall \mathbf{b} \in B;$$

$$(ii) \quad Cov[S(\mathbf{b}_i), S(\mathbf{b}_j)] = Cov[S(\mathbf{b}), S(\mathbf{b} + \mathbf{h})] = C(\mathbf{h});$$

$$(iii) \quad Var[S(\mathbf{b})] = E[(S(\mathbf{b}) - \mu)^2] = C(0).$$

A primeira condição da Definição 3.2 garante que o processo possui média constante em todo domínio D . A segunda condição garante que a covariância entre duas observações depende apenas da distância entre elas. A função $C(\cdot)$ é definida como o covariograma do processo ou função de covariância estacionária.

Apesar de ser menos restritiva, a estacionariedade de segunda ordem supõe que a variável aleatória tenha variância finita, o que às vezes é falso. Quando a variável aleatória possui uma capacidade infinita de dispersão, precisamos de outra hipótese mais fraca, a hipótese intrínseca.

Definição 3.3 Um processo $S(\mathbf{b})$ com domínio $B \subset \mathbb{R}^d$ apresenta estacionariedade intrínseca se

$$(i) E[S(\mathbf{b}_i) - S(\mathbf{b}_j)] = 0, \quad \forall \mathbf{b}_i, \mathbf{b}_j \in B;$$

$$(ii) Var[S(\mathbf{b}_i) - S(\mathbf{b}_j)] = 2\gamma(\mathbf{b}_i - \mathbf{b}_j) = 2\gamma(\mathbf{h}), \quad \forall \mathbf{b}_i, \mathbf{b}_j \in B.$$

Sob hipótese intrínseca, são os acréscimos da variável aleatória que obedecem à estacionariedade de segunda ordem e não a própria variável. As variáveis aleatórias $S(\mathbf{b}_i) - S(\mathbf{b}_j)$, $\forall \mathbf{b}_i, \mathbf{b}_j \in B$, são chamadas de incrementos do processo. A primeira condição da Definição 3.3 garante que os incrementos possuem média nula. Caso $S(\mathbf{b})$ possua esperança finita, esta é apenas outra forma de garantir que o processo é estacionário na média, como visto na primeira condição da Definição 3.2. A segunda condição da Definição 3.3 garante que os incrementos tenham variância constante definida apenas pelo vetor de distância \mathbf{h} que separa as duas observações. A função $2\gamma(\mathbf{h})$ é definida como o variograma do processo, e a função $\gamma(\mathbf{h})$ como o semivariograma.

Um processo fortemente estacionário é, logicamente, estacionário de segunda ordem, uma vez que todos os momentos são invariantes por translação. O contrário pode não ser verdade. No caso de um processo Gaussiano (em que a *f.d.p* conjunta segue uma distribuição normal), a estacionariedade de segunda ordem implica em estacionariedade forte, já que os primeiro e segundo momentos definem completamente a distribuição do processo. Se o processo verifica estacionariedade de segunda ordem, verifica a hipótese intrínseca, mas o inverso pode não ser válido. Isso se deve ao fato de que os semivariogramas para a estacionariedade de segunda ordem são sempre limitados, ao contrário que há semivariogramas de processos que se enquadram na hipótese intrínseca e não são limitados.

3.2 Covariograma e variograma

Seja $S(\mathbf{b})$ um processo geoestatístico sob estacionariedade de segunda ordem. Como visto na Definição 3.2, o processo possui uma função de covariância estacionária, $C(\cdot)$, denominado covariograma. Para que a função $C(\cdot)$ seja um covariograma válido, deve satisfazer algumas propriedades, como

$$(i) C(0) = Var[S(\mathbf{b})], \quad \forall \mathbf{b} \in B;$$

$$(ii) \ C(\mathbf{h}) = C(-\mathbf{h}), \ \forall \mathbf{h} \in \mathbb{R}^d;$$

(iii) o covariograma é positivo definido, isto é,

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j C(\mathbf{b}_i - \mathbf{b}_j) \geq 0, \ \forall n \in \mathbb{N}, \forall a_1, \dots, a_n \in \mathbb{R}, \forall \mathbf{b}_1, \dots, \mathbf{b}_n \in B.$$

A condição (iii) é necessária e suficiente para que se defina uma família de covariogramas válidos, porém não é uma condição fácil de ser verificada (Diggle and Ribeiro, 2007). Existem famílias paramétricas de covariogramas que satisfazem as propriedades citadas, como veremos posteriormente. Um covariograma é isotrópico quando depende apenas do vetor \mathbf{h} através de seu módulo, não dependendo de sua direção. Nesse caso, para qualquer localização \mathbf{b} , a covariância entre o processo em \mathbf{b} e em qualquer localização sobre uma circunferência de raio $\|\mathbf{h}\|$ é constante. Caso contrário, temos anisotropia.

O correlograma é definido por $\rho(\mathbf{h}) = C(\mathbf{h})/C(0)$, desde que $C(0)$ seja finito e não nulo. Essencialmente, o correlograma fornece a mesma informação do covariograma. Apesar disso, às vezes é melhor trabalhar com o correlograma na modelagem espacial, como veremos posteriormente.

A função variograma é definida por

$$2\gamma(\mathbf{b}, \mathbf{b}') = Var[S(\mathbf{b}) - S(\mathbf{b}')] = Var[S(\mathbf{b} + \mathbf{h}) - S(\mathbf{b})] = 2\gamma(\mathbf{h}), \quad (3.1)$$

em que $2\gamma(\mathbf{b}, \mathbf{b}') = 2\gamma(\mathbf{h})$ pelo fato do processo ser estacionário. Supondo a existência do covariograma, podemos expressar a função de variograma como

$$\begin{aligned} 2\gamma(\mathbf{h}) &= Var[S(\mathbf{b} + \mathbf{h}) - S(\mathbf{b})], \\ &= Var[S(\mathbf{b} + \mathbf{h})] + Var[S(\mathbf{b})] - 2Cov[S(\mathbf{b} + \mathbf{h}), S(\mathbf{b})], \\ &= 2C(0) - 2C(\mathbf{h}). \end{aligned} \quad (3.2)$$

Note que o variograma pode ser definido pelo covariograma quando este existe. A função de variograma deve satisfazer algumas propriedades, como

$$(i) \ 2\gamma(0) = 0;$$

$$(ii) \ 2\gamma(\mathbf{h}) = 2\gamma(-\mathbf{h}), \ \forall \mathbf{h} \in \mathbb{R}^d;$$

(iii) o variograma é negativo definido, isto é,

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j 2\gamma(\mathbf{b}_i - \mathbf{b}_j) \leq 0, \quad \forall n \in \mathbb{N},$$

em que, a_1, \dots, a_n são tais que $\sum_{i=1}^n a_i = 0$ e $b_i, \dots, b_j \in B$.

(iv) o variograma cresce mais lentamente que $\|\mathbf{h}\|^2$, isto é,

$$\lim_{\|\mathbf{h}\| \rightarrow \infty} \frac{2\gamma(\mathbf{h})}{\|\mathbf{h}\|^2} = 0.$$

Se um variograma satisfizer as propriedades (i) a (iv), é dito um variograma válido. O semivariograma é definido como metade do variograma, $\gamma(\|\mathbf{h}\|) = C(0) - C(\|\mathbf{h}\|)$, e possui as mesmas propriedades que o variograma. Desse modo, é comum utilizar o semivariograma em vez do variograma (Cressie, 1993). Assim como o covariograma, um variograma é isotrópico quando depende apenas do vetor \mathbf{h} através de seu módulo, não dependendo de sua direção. Desse modo, o valor do variograma é o mesmo para quaisquer pares de variáveis aleatórias equidistantes. Uma forma de verificar isotropia é estudar o comportamento do variograma nas direções 0° , 45° , 90° e 135° , em que 0° é a direção Norte-Sul com sentido para o Norte.

Note que o variograma é definido para qualquer processo que obedeça à hipótese intrínseca, já o covariograma somente existe para processos sob estacionariedade de segunda ordem. Assim, se o processo possui covariograma, existe também seu variograma como visto em (3.2). Porém, a existência do variograma não implica a existência do covariograma. Esse é um dos motivos pelo qual a função de variograma é preferida. É usual utilizar a função de semivariância, uma vez que possui as mesmas propriedades que o variograma.

Apesar da propriedade (i) do semivariograma dizer que seu gráfico começa na origem, na prática pode haver descontinuidade nesse ponto. Esse efeito é chamado de efeito pepita (*nugget effect*). Se pudéssemos coletar duas amostras no mesmo local ($\|\mathbf{h}\| = 0$), os valores observados deveriam ser exatamente os mesmos. Porém, haverá uma variação entre essas amostras, seja por erro de medida, de amostragem ou outros. Assim, o gráfico do semivariograma não converge para zero quando $\|\mathbf{h}\|$ se aproxima da origem. Podemos definir o efeito pepita como o valor $\tau^2 = \lim_{\|\mathbf{h}\| \rightarrow 0} \gamma(\mathbf{h})$. Quando não há correlação espa-

cial, a variação da variável aleatória se dá apenas pelo efeito pepita e é chamado de efeito pepita puro.

A Figura 3.1 apresenta um gráfico de semivariograma limitado. No gráfico de

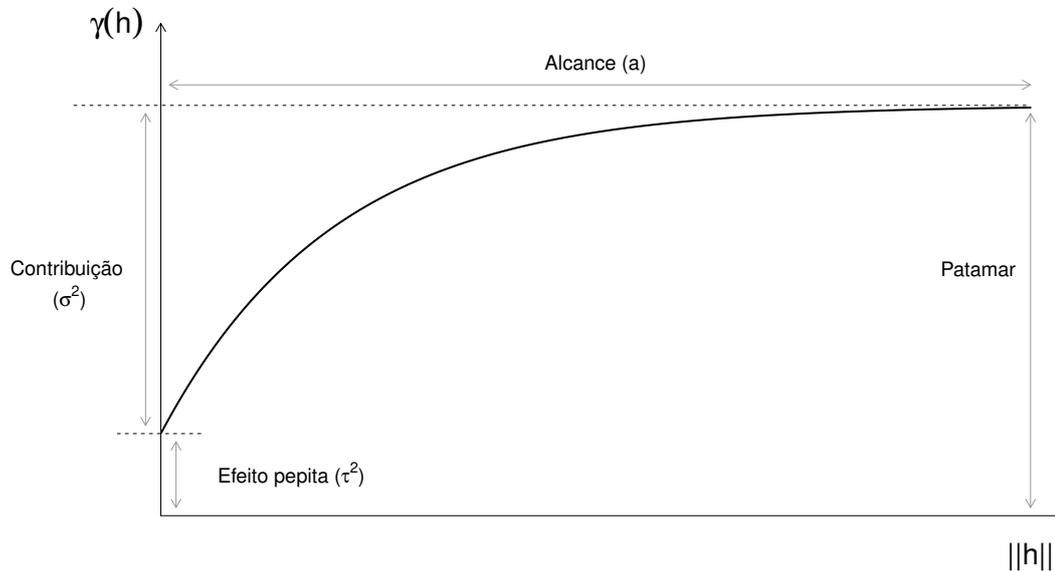


Figura 3.1: Parâmetros do semivariograma.

semivariograma $\gamma(\mathbf{h})$, podemos observar os seguintes parâmetros:

- Efeito pepita ou *nugget effect* (τ^2): representa a variabilidade não espacial.
- Contribuição ou *partial sill* (σ^2): representa a variabilidade espacial.
- Patamar: é o valor do semivariograma correspondente ao seu alcance a partir do qual o semivariograma se estabiliza, dado por $\tau^2 + \sigma^2$. Quando o semivariograma não é limitado, não existe patamar.
- Alcance ou *range* (a): distância a partir da qual as amostras passam a ser consideradas independentes, ou seja, a partir da qual a variação média entre duas observações não é mais função da distância entre elas. Quando essa distância tende a ∞ , considera-se que o semivariograma atinge um patamar apenas assintoticamente. Então, é utilizado um alcance chamado de alcance prático. Diggle and Ribeiro (2007) definem o alcance prático como sendo o valor de $\|\mathbf{h}\|$ tal que o valor da correlação seja 0,05.

O modelo paramétrico Matérn é muito utilizado na literatura e suas funções de

semivariogramas, covariogramas e correlogramas isotrópicos são dadas por

$$\gamma(h) = \begin{cases} 0 & \text{se } \|h\| = 0; \\ \tau^2 + \sigma^2 \left[1 - (2^{\kappa-1}\Gamma(\kappa))^{-1} \left(\frac{\|h\|}{\phi} \right)^\kappa K_\kappa \left(\frac{\|h\|}{\phi} \right) \right] & \text{se } \|h\| > 0; \end{cases}$$

$$C(h) = \begin{cases} \sigma^2 & \text{se } \|h\| = 0; \\ \sigma^2 (2^{\kappa-1}\Gamma(\kappa))^{-1} \left(\frac{\|h\|}{\phi} \right)^\kappa K_\kappa \left(\frac{\|h\|}{\phi} \right) & \text{se } \|h\| > 0; \end{cases}$$

$$\rho(h) = \begin{cases} 1 & \text{se } \|h\| = 0; \\ (2^{\kappa-1}\Gamma(\kappa))^{-1} \left(\frac{\|h\|}{\phi} \right)^\kappa K_\kappa \left(\frac{\|h\|}{\phi} \right) & \text{se } \|h\| > 0; \end{cases}$$

em que $\theta = (\tau^2, \sigma^2, \phi, \kappa)$, $\phi > 0$, $\kappa > 0$, K_κ é a função de Bessel modificada de ordem κ e Γ é a função Gama. Esse modelo foi proposto por Matérn (1960) e é válido em qualquer dimensão $d \geq 1$. Os modelos exponencial e Gaussiano são casos particulares do modelo Matérn, em que $\kappa = 0,5$ no primeiro caso e κ tende a ∞ no segundo caso. Pela dificuldade de identificabilidade de todos os parâmetros desse modelo, normalmente o valor de κ é fixado e avaliado em vários valores. A Figura 3.2 apresenta exemplos de gráficos de semivariogramas.

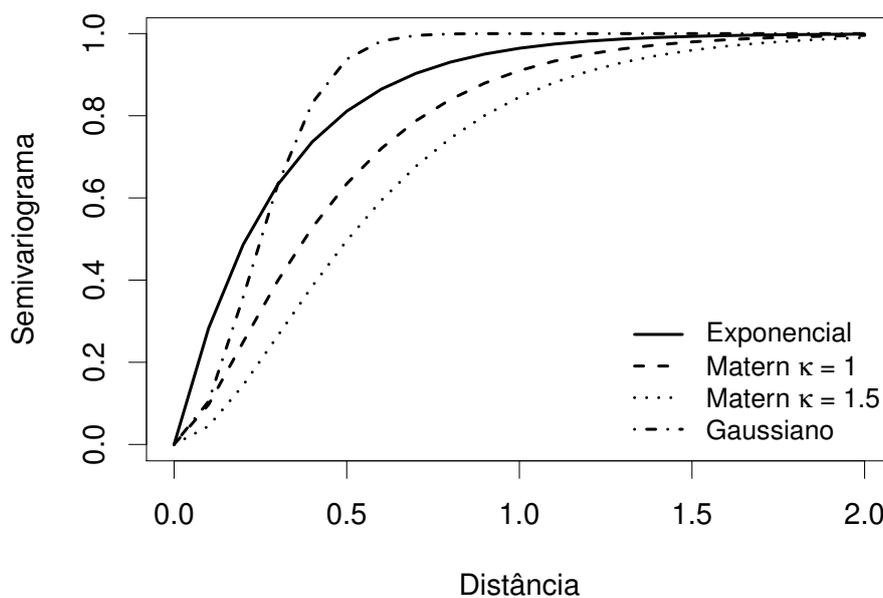


Figura 3.2: Exemplos de semivariograma.

3.3 Modelo geoestatístico

Tendo como base o processo geoestatístico, nosso interesse é modelar o processo e realizar inferências a partir de dados coletados em pontos conhecidos do espaço. Na etapa de coleta de dados, é natural que haja variação entre o verdadeiro valor do processo geoestatístico e o valor medido ocasionada, por exemplo, por erros na medição devido a precisão de instrumento utilizado. Um conjunto de valores observados será denotado por (\mathbf{b}_i, y_i) , $i = 1, \dots, n$, em que \mathbf{b}_i 's são as coordenadas dos locais amostrados e y_i 's são os valores medidos associados aos locais \mathbf{b}_i . Dessa forma, vamos denotar como $S(b)$ (ou simplesmente S , deixando a dependência do espaço B implícita) o processo geoestatístico desconhecido e como $Y(b)$, ou Y , o valor medido na localização b , $b \in B$.

O modelo mais simples considera o processo geoestatístico como um processo Gaussiano, o qual definimos a seguir.

Definição 3.4 *Um processo Gaussiano é um processo estocástico $\{S(\mathbf{b}); \mathbf{b} \in B\}$, $B \subset \mathbb{R}^d$, com $d \geq 1$, no qual qualquer conjunto finito de pontos segue uma distribuição normal.*

Como vimos na Seção 3.1, precisamos considerar hipóteses sobre o processo para podermos realizar inferência. Neste trabalho, vamos considerar processos que obedecem a condição de estacionariedade de segunda ordem, isto é, processos cujos primeiro e segundo momentos são constantes. Se S é um processo Gaussiano estacionário, sua média e variância são constantes em toda região B . É usual trabalhar com a função de correlação ao invés da função de covariância. Dessa forma, dizemos que S é um processo Gaussiano estacionário com variância σ^2 e função de correlação $\rho(S(\mathbf{b}), S(\mathbf{b}'))$.

A seguir introduzimos a definição do modelo geoestatístico.

Definição 3.5 *Definimos o seguinte modelo geoestatístico*

$$Y_i = \mu(\mathbf{b}_i) + S(\mathbf{b}_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (3.3)$$

em que

- Y_i representa o valor observado no local \mathbf{b}_i ,
- $S(\cdot)$ é um processo Gaussiano estacionário com domínio $B \subset \mathbb{R}^m$, $m \geq 1$, média zero, variância σ^2 e função de correlação $\rho(S(\mathbf{b}), S(\mathbf{b}'))$,

- $\mu(\mathbf{b}_i) = E[Y_i]$ é um componente determinístico que representa a média de Y no local \mathbf{b}_i ,
- ϵ_i é um componente de erro. Assume-se que as variáveis ϵ_i 's são independentes com distribuição $N(0, \tau^2)$. A variância desse termo representa o efeito pepita.

Observe que a variável aleatória $Y_i - \mu(\mathbf{b}_i)$ representa o processo Gaussiano S em \mathbf{b}_i acrescido de um ruído aleatório. Do modelo geoestatístico, temos que o processo S é responsável por modelar a associação espacial da variável de interesse e a incerteza não espacial está relacionada ao componente ϵ . Se não houvesse estrutura espacial, o modelo em (3.3) se reduz a um modelo de regressão normal. A função de correlação deve ser positiva definida.

Observe que $\mathbf{S} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{R}(\phi))$, em que $\mathbf{R}(\phi) = [(r_{ij})]$ representa a matriz de correlação de dimensão $n \times n$, com elementos $r_{ij} = \rho(S(\mathbf{b}_i), S(\mathbf{b}_j))$. Dessa forma, \mathbf{Y} possui distribuição normal com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\tau^2, \sigma^2, \phi) = \tau^2 \mathbf{I}_n + \sigma^2 \mathbf{R}(\phi). \quad (3.4)$$

Os parâmetros τ^2 , σ^2 e ϕ são os parâmetros do semivariograma discutidos na Seção 3.2.

3.4 Estimação via método de máxima verossimilhança

Denotamos o vetor de parâmetros do modelo como $\boldsymbol{\theta} = (\mu, \tau^2, \sigma^2, \phi)$ e $\boldsymbol{\mu} = \mu \mathbf{1}_n$. A função de log-verossimilhança do modelo (3.3) é dada por

$$l(\boldsymbol{\theta}; \mathbf{y}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log|\boldsymbol{\Sigma}| - \frac{1}{2} (\mathbf{y} - \boldsymbol{\mu} \mathbf{1}_n)' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu} \mathbf{1}_n), \quad (3.5)$$

em que $|\boldsymbol{\Sigma}|$ corresponde ao determinante da matriz $\boldsymbol{\Sigma}$.

A estimação dos parâmetros se dá maximizando a função de log-verossimilhança em (3.5). A função score é definida como o vetor gradiente da função de log-verossimilhança, isto é,

$$U(\boldsymbol{\theta}) = \left(\frac{\partial l(\boldsymbol{\theta}, \mathbf{y})}{\partial \mu}, \frac{\partial l(\boldsymbol{\theta}, \mathbf{y})}{\partial \tau^2}, \frac{\partial l(\boldsymbol{\theta}, \mathbf{y})}{\partial \sigma^2}, \frac{\partial l(\boldsymbol{\theta}, \mathbf{y})}{\partial \phi} \right)'. \quad (3.6)$$

Quando igualamos a função score a zero, podemos encontrar o estimador do parâmetros de interesse caso esse tenha forma analítica. O estimador de μ possui forma

fechada e pode ser expresso como

$$\hat{\boldsymbol{\mu}} = (\mathbf{1}'_n \boldsymbol{\Sigma}^{-1} \mathbf{1}_n)^{-1} \mathbf{1}'_n \boldsymbol{\Sigma}^{-1} \mathbf{y}. \quad (3.7)$$

Não é possível calcular o estimador de forma fechada para os demais parâmetros e suas estimações devem ser realizadas numericamente. Desse modo, o vetor de parâmetros $(\tau^2, \sigma^2, \phi)'$ deve ser estimado por

$$(\hat{\tau}^2, \hat{\sigma}^2, \hat{\phi})' = \underset{(\tau^2, \sigma^2, \phi)}{\operatorname{argmax}} \left(l(\tau^2, \sigma^2, \phi; \hat{\boldsymbol{\mu}}, \mathbf{y}) \right).$$

3.5 Estimação via inferência Bayesiana

Na abordagem Bayesiana, o modelo (3.3) toma a seguinte forma

$$\mathbf{Y} | \mathbf{S}_n, \mu, \tau^2 \sim N_n(\mu \mathbf{1}_n + \mathbf{S}_n, \tau^2 \mathbf{I}_n), \quad (3.8)$$

$$\mathbf{S}_n | \sigma^2, \phi \sim N_n(\mathbf{0}, \sigma^2 \mathbf{R}(\phi)) \quad (3.9)$$

$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}), \quad (3.10)$$

em que $\mathbf{S}_n = (S(\mathbf{b}_1), \dots, S(\mathbf{b}_n))'$, $\mathbf{1}_n$ é um vetor de 1's de tamanho n , \mathbf{I}_n é a matriz identidade de dimensão n e $\boldsymbol{\theta} = (\mu, \tau^2, \sigma^2, \phi)'$.

As quantidades desconhecidas do modelo são \mathbf{S}_n e $\boldsymbol{\theta}$. Sob o paradigma Bayesiano, a inferência sobre os parâmetros do modelo é realizada através da distribuição a posteriori. Desse modo,

$$\begin{aligned} \pi(\boldsymbol{\theta}, \mathbf{S}_n | \mathbf{y}) &\propto \pi(\mathbf{y}, \mathbf{S}_n, \boldsymbol{\theta}), \\ &\propto L(\mathbf{y} | \mathbf{S}_n, \boldsymbol{\theta}) \pi(\mathbf{S}_n | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}), \\ &\propto \left(\frac{1}{\tau^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\tau^2} \sum_{i=1}^n (y_i - S_i - \mu)^2 \right\} \times \\ &\quad \left(\frac{1}{\sigma^2} \right)^{n/2} \frac{1}{|\mathbf{R}(\phi)|^{1/2}} \exp \left\{ -\frac{1}{2\sigma^2} \mathbf{S}'_n \mathbf{R}(\phi)^{-1} \mathbf{S}_n \right\} \times \pi(\boldsymbol{\theta}) \end{aligned} \quad (3.11)$$

em que $L(\mathbf{y} | \mathbf{S}, \boldsymbol{\theta})$ é a função de verossimilhança do modelo e $\pi(\boldsymbol{\theta})$ é a distribuição a priori de $\boldsymbol{\theta}$.

Como a distribuição a posteriori (3.11) não possui forma fechada, vamos aproximá-la utilizando MCMC. As distribuições condicionais completas são dadas por

$$\pi(\mu|\cdot) \propto \exp \left\{ -\frac{1}{2\tau^2} \sum_{i=1}^n (y_i - S_i - \mu)^2 \right\} \pi(\mu) \quad (3.12)$$

$$\pi(\tau^2|\cdot) \propto \left(\frac{1}{\tau^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\tau^2} \sum_{i=1}^n (y_i - S_i - \mu)^2 \right\} \pi(\tau^2) \quad (3.13)$$

$$\pi(\sigma^2|\cdot) \propto \left(\frac{1}{\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \mathbf{S}'_n \mathbf{R}(\phi)^{-1} \mathbf{S}_n \right\} \pi(\sigma^2) \quad (3.14)$$

$$\pi(\phi|\cdot) \propto \frac{1}{|\mathbf{R}(\phi)|^{1/2}} \exp \left\{ -\frac{1}{2\sigma^2} \mathbf{S}'_n \mathbf{R}(\phi)^{-1} \mathbf{S}_n \right\} \pi(\phi) \quad (3.15)$$

$$\pi(\mathbf{S}_n|\cdot) \propto \exp \left\{ -\frac{1}{2\tau^2} \sum_{i=1}^n (y_i - S_i - \mu)^2 \right\} \exp \left\{ -\frac{1}{2\sigma^2} \mathbf{S}'_n \mathbf{R}(\phi)^{-1} \mathbf{S}_n \right\} \quad (3.16)$$

Dessa forma, a geração de amostras foi realizada pelo seguinte amostrador de Gibbs.

Passo 1: μ . Considerando como distribuição a priori para μ uma distribuição $N(\mu_0, \sigma_\mu^2)$, a distribuição condicional completa de μ é dada por

$$(\mu|\cdot) \sim N \left(\left(\frac{\sum_{i=1}^n (y_i - S_i)}{\tau^2} + \frac{\mu_0}{\sigma_\mu^2} \right) \left(\frac{n}{\tau^2} + \frac{1}{\sigma_\mu^2} \right)^{-1}; \left(\frac{n}{\tau^2} + \frac{1}{\sigma_\mu^2} \right)^{-1} \right). \quad (3.17)$$

Passo 2: τ^2 . Considerando como distribuição a priori $\tau^2 \sim GI(a_\tau, b_\tau)$, em que $GI(a, b)$ representa a distribuição Gama Inversa com função de densidade

$$f(x) = \frac{b^a}{\Gamma(a)} e^{-b/x} (1/x)^{a+1},$$

a distribuição condicional completa de τ^2 é dada por

$$(\tau^2|\cdot) \sim GI \left(\frac{n}{2} + a_\tau; \frac{\sum_{i=1}^n (y_i - S_i - \mu)^2}{2} + b_\tau \right). \quad (3.18)$$

Passo 3: σ^2 . Considerando como distribuição a priori $\sigma^2 \sim GI(a_\sigma, b_\sigma)$, a distribuição condicional completa de σ^2 é dada por

$$(\sigma^2|\cdot) \sim GI \left(\frac{n}{2} + a_\sigma; \frac{\mathbf{S}'_n \mathbf{R}(\phi)^{-1} \mathbf{S}_n}{2} + b_\sigma \right). \quad (3.19)$$

Passo 4: ϕ . Considerando como distribuição a priori $\phi \sim Gama(a_\phi, b_\phi)$, em que

$Gama(a, b)$ representa a distribuição Gama com função de densidade

$$f(x) = \frac{b^a}{\Gamma(a)} e^{-bx} x^{a-1},$$

a distribuição condicional completa de ϕ é dada por

$$(\phi|.) \sim |\mathbf{R}(\phi)|^{-1/2} \phi^{a_\phi-1} \exp \left\{ -\frac{1}{2\sigma^2} \mathbf{S}_n \mathbf{R}(\phi)^{-1} \mathbf{S}_n - b_\phi \phi \right\}. \quad (3.20)$$

Para amostrar dessa distribuição será implementado um passo de Metropolis-Hastings (MH). Assumimos a distribuição Lognormal($\log(\phi_c); \delta_\phi$) como distribuição de proposta $q(\phi_p|\phi_c)$, em que ϕ_p é o valor proposto de ϕ e ϕ_c é o valor de ϕ no estado corrente da cadeia. A função de densidade de probabilidade de $q(\phi_p|\phi_c)$ é expressa por:

$$q(\phi_p|\phi_c) = \frac{1}{\phi_p \sqrt{2\pi\delta_\phi}} \exp \left\{ -\frac{(\log(\phi_p) - \log(\phi_c))^2}{2\delta_\phi} \right\}.$$

De forma equivalente, $\log(\phi_p) \sim N(\log(\phi_c), \delta_\phi)$. A probabilidade de aceitação é $A_\phi = \min\{1, p_\phi\}$, em que

$$p_\phi = \left(\frac{|R(\phi_p)|}{|R(\phi_c)|} \right)^{-1/2} \left(\frac{\phi_p}{\phi_c} \right)^{a_\phi} \exp \left\{ -\frac{1}{2\sigma^2} \left(\mathbf{S}'_n (\mathbf{R}(\phi_p)^{-1} - \mathbf{R}(\phi_c)^{-1}) \mathbf{S}_n \right) - b_\phi (\phi_p - \phi_c) \right\}.$$

Passo 5: \mathbf{S}_n . Remanejando os termos referentes a \mathbf{S}_n em (3.16), obtemos que $(\mathbf{S}_n|.) \sim N_n(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, em que

$$\begin{aligned} \boldsymbol{\Sigma}^* &= \left(\frac{1}{\tau^2} \mathbf{I}_n + \frac{1}{\sigma^2} \mathbf{R}(\phi)^{-1} \right)^{-1} \\ \boldsymbol{\mu}^* &= \boldsymbol{\Sigma}^* \left(\frac{1}{\tau^2} \mathbf{I}_n \right) (\mathbf{y} - \boldsymbol{\mu}). \end{aligned}$$

Os passos 1-5 fornecerão, após convergência da cadeia, amostras da distribuição a posteriori de $(\boldsymbol{\theta}, \mathbf{S})$. A partir dessas amostras é possível realizar a inferência sobre os parâmetros do modelo.

3.6 Predição

Na análise geoestatística, a confecção de mapas temáticos é importante para descrição da variável resposta em todo o espaço de domínio. Para isso, os valores observados

são utilizados para prever o valor da variável resposta em pontos que não foram amostrados e que são internos à região de estudo. Nesse caso, é realizado um processo de interpolação espacial. Esse processo, em geoestatística, é chamado de krigagem, desenvolvido por Matheron em homenagem à Daniel Krige pelo seu trabalho de minas de ouro na África do Sul (Cressie, 1993).

Considerando o modelo (3.3), assumindo um processo Gaussiano, sejam $\mathbf{Y}_o = (Y_{o_1}, \dots, Y_{o_{n_o}})'$ o vetor de Y nos locais não observados $\{\mathbf{b}_{o_1}, \dots, \mathbf{b}_{o_{n_o}}\}$ e \mathbf{Y} o vetor de Y nos locais observados $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$. A distribuição conjunta de \mathbf{Y}_o e \mathbf{Y} será

$$(\mathbf{Y}_o, \mathbf{Y}) \sim NM_{(n_o+n)} \left(\begin{pmatrix} \boldsymbol{\mu}_o \\ \boldsymbol{\mu} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right),$$

em que $\boldsymbol{\mu}_o$ é o vetor de médias de \mathbf{Y}_o , $\boldsymbol{\mu}$ é o vetor de médias do vetor \mathbf{Y} , $\boldsymbol{\Sigma}_{11}$ corresponde à matriz de covariâncias de \mathbf{Y}_o , $\boldsymbol{\Sigma}_{22}$ corresponde à matriz de covariâncias de \mathbf{Y} , $\boldsymbol{\Sigma}_{12}$ corresponde à matriz de covariâncias de \mathbf{Y}_o e \mathbf{Y} e $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}'_{12}$.

Através de resultados relacionados à distribuição normal multivariada, é possível mostrar que a distribuição de $\mathbf{Y}_o|\mathbf{Y}$ é normal com

$$E[\mathbf{Y}_o|\mathbf{Y}] = \boldsymbol{\mu}_o + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{Y} - \boldsymbol{\mu}), \quad (3.21)$$

$$Var[\mathbf{Y}_o|\mathbf{Y}] = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \quad (3.22)$$

Dessa forma, para realizar predição em locais não observados para o modelo Gaussiano, utilizamos os parâmetros estimados como verdadeiros e substituímos na Equação (3.21).

Sob o enfoque Bayesiano, necessitamos da distribuição preditiva a posteriori de \mathbf{Y}_o , dada por

$$\begin{aligned} \pi(\mathbf{Y}_o|\mathbf{y}) &= \int \pi(\mathbf{Y}_o, \mathbf{S}_o, \mathbf{S}_n, \boldsymbol{\theta}|\mathbf{y})d\mathbf{S}_od\mathbf{S}_nd\boldsymbol{\theta}, \\ &= \int \pi(\mathbf{Y}_o|\mathbf{S}_o, \boldsymbol{\theta})\pi(\mathbf{S}_o|\mathbf{S}_n, \boldsymbol{\theta})\pi(\mathbf{S}_n, \boldsymbol{\theta})d\mathbf{S}_od\mathbf{S}_nd\boldsymbol{\theta}, \end{aligned}$$

em que $\pi(\mathbf{S}_n, \boldsymbol{\theta})$ é a distribuição a posteriori do modelo, $\pi(\mathbf{S}_o|\mathbf{S}_n, \boldsymbol{\theta})$ é uma distribuição

normal com vetor de médias e matriz de covariâncias dadas por

$$\begin{aligned}\boldsymbol{\mu}_{S_o|S_n} &= \mathbf{R}_o(\phi)\mathbf{R}_n(\phi)^{-1}\mathbf{S}_n, \\ \boldsymbol{\Sigma}_{S_o|S_n} &= \sigma^2(\mathbf{R}_o(\phi) - \mathbf{R}_{on}(\phi)\mathbf{R}_n(\phi)^{-1}(\mathbf{R}_{on}(\phi))'),\end{aligned}$$

sendo $\mathbf{R}_o(\phi)$ a matriz de correlação de \mathbf{S}_o , $\mathbf{R}_n(\phi)$ a matriz de correlação de \mathbf{S}_n e $\mathbf{R}_{on}(\phi)$ a matriz de correlação entre os elementos de \mathbf{S}_o e \mathbf{S}_n ; e $\pi(\mathbf{Y}_o|\mathbf{S}_o, \boldsymbol{\theta})$ uma distribuição normal com vetor de médias $\boldsymbol{\mu}_{Y_o|S_o} = \mu\mathbf{1}_o + \mathbf{S}_o$ e matriz de covariâncias $\boldsymbol{\Sigma}_{Y_o|S_o} = \tau^2\mathbf{I}_o$.

A partir das amostras geradas de $(\mathbf{S}_n, \boldsymbol{\theta})$ no processo de estimação dos parâmetros do modelo, geramos amostras de $\mathbf{S}_o|\mathbf{S}_n, \boldsymbol{\theta}$ para realizar previsão do processo Gaussiano nos locais não amostrados e, em seguida, geramos amostras de $\mathbf{Y}_o|\mathbf{S}_o, \boldsymbol{\theta}$ para realizar previsão sobre o processo observado Y . Note que, se trabalhamos com a distribuição marginal de \mathbf{Y} no modelo Bayesiano, a distribuição preditiva $(\mathbf{Y}_o|\mathbf{y})$ coincide com a distribuição preditiva no caso clássico.

3.7 Modelo geoestatístico com amostragem preferencial

Em estudos geoestatísticos, é comum considerar que nenhuma informação sobre o fenômeno é utilizada na escolha dos locais amostrados. Esse tipo de amostragem é denominado não preferencial e o modelo com essa característica foi abordado na Seção 3.1. No entanto, há casos em que se opta por escolher locais na amostragem considerando o fenômeno em questão, como exemplo podemos considerar alocar estações de monitoramento de poluentes atmosféricos próximas a possíveis fontes de poluição. Nesse caso, o desenho amostral é construído com base em informações sobre o problema e esse tipo de amostragem é denominada preferencial. Desse modo, os métodos de análise discutidos na Seção 3.1 não são mais adequados e outro tipo de modelagem se torna necessária. Neste seção, vamos abordar como incluímos a informação de dependência do desenho amostral com o fenômeno em estudo na modelagem do problema.

Quando falamos de desenho amostral em geoestatística, nos referimos a configuração dos locais da amostra na região de estudo. Logo, a realização de um processo pontual X na região B representa o desenho amostral nessa região. Quando dizemos que a amostragem é realizada de modo não preferencial, significa que o processo pontual X independe do processo Gaussiano S (que representa o fenômeno em estudo). Porém, quando

estamos no contexto de amostragem preferencial, $f(S, X) \neq f(S)f(X)$. Vale lembrar que S é um processo latente (não observado) e que observamos o processo de medida Y que representa os valores observados nos pontos amostrados de X . Desse modo, o modelo é especificado pela distribuição conjunta de S , X e Y , dada por

$$f(S, X, Y) = f(S)f(X|S)f(Y|X, S) = f(S)f(X|S)f(Y|S(X)). \quad (3.23)$$

No caso não preferencial, a Equação (3.23) se resume a $f(S)f(X)f(Y|S(X))$ e a dependência de S e X no termo $f(Y|S(X))$ pode ser ignorada. No caso preferencial, desconsiderar a característica estocástica do processo X pode resultar em uma inferência incorreta. No pioneiro artigo de Diggle et al. (2010), os autores definem uma classe de modelos para o presente problema que apresentamos a seguir.

Definição 3.6 *O modelo geoestatístico Gaussiano com amostragem preferencial é definido pela seguinte forma hierárquica:*

- 1 : *S é um processo Gaussiano estacionário e isotrópico em uma região B com média 0, variância σ^2 e função de correlação $\rho(h, \phi)$, em que $h = \|\mathbf{x} - \mathbf{x}'\|$ é a distância euclidiana entre os pontos \mathbf{x} e \mathbf{x}' , ϕ é o parâmetro da função de correlação e $\mathbf{x}, \mathbf{x}' \in B$,*
- 2 : *Condicional em S , X é um processo de Poisson não homogêneo com taxa $\lambda(\mathbf{x}) = \exp(\alpha + \beta S(\mathbf{x}))$, em que α e β são números reais.*
- 3 : *Condicional em S e X , Y é um conjunto de v.a. Gaussianas independentes com $Y(\mathbf{x}_i) \sim N(\mu + S(\mathbf{x}_i), \tau^2)$.*

Das suposições (1) e (2) da Definição 3.6, segue que, marginalmente, X é um processo de Cox log-Gaussiano (Møller and Waagepetersen, 2003). O parâmetro β define a intensidade de preferência do modelo. Se $\beta = 0$, a função de intensidade $\lambda(x)$ se torna constante e o processo de Poisson X se torna homogêneo significando que não há preferência na amostragem. Além disso, marginalmente Y terá distribuição Gaussiana multivariada com vetor de média $\mu \mathbf{1}_n$ e matriz de covariância $\tau^2 \mathbf{I}_n + \sigma^2 \mathbf{R}(\phi)$, em que $\mathbf{1}_n$ é o vetor de 1's de tamanho n , \mathbf{I}_n é a matriz identidade de ordem n e $\mathbf{R}(\phi)$ é a matriz de correlação $n \times n$ com elementos $r_{ij} = \rho(\|x_i - x_j\|, \phi)$. Caso $\beta \neq 0$, a distribuição marginal

de Y não possui forma fechada. Um valor positivo de β indica uma associação positiva entre S e X e os pontos de maior valor são preferencialmente amostrados. Um valor negativo de β indica uma associação negativa entre S e X no sentido de que os pontos de menor valor de S possuem preferência na amostragem. A Figura 3.3 exemplifica o efeito do parâmetro β no desenho amostral no espaço bidimensional. O parâmetro α está associado ao número médio de pontos do processo X . Da Seção 2.4, o valor esperado do número de pontos de X na região B é dado por

$$E[N(B)] = \int_B \lambda(u) du = \int_B \exp\{\alpha + \beta S(u)\} du = e^\alpha \int_B \exp\{\beta S(u)\} du.$$

Se $\beta = 0$, então $E[N(B)] = e^\alpha |B|$, em que $|B|$ é o volume da região B . Se β é não nulo, podemos dizer que o número esperado de pontos de X flutuará em torno de e^α , dependendo do processo S .

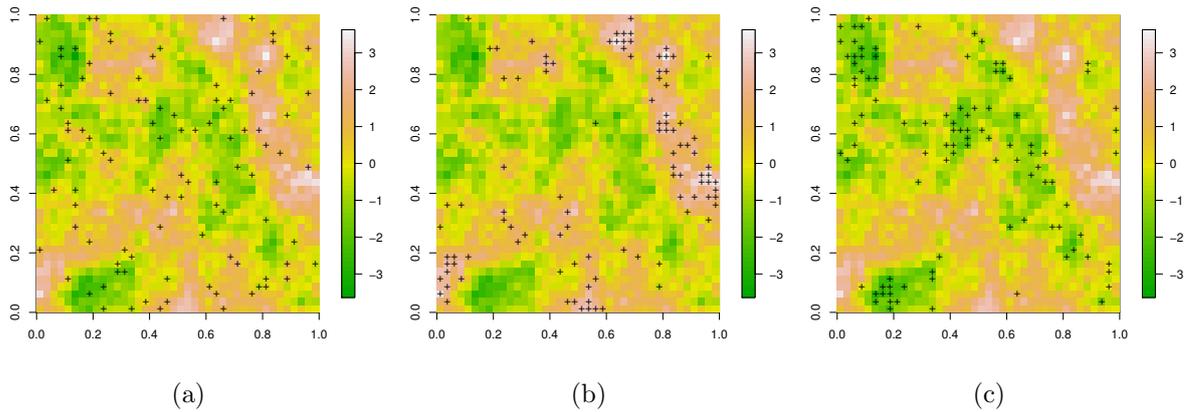


Figura 3.3: Exemplos de amostragem considerando (a) $\beta = 0$, (b) $\beta = 1, 5$ e (c) $\beta = -1, 5$. Os pontos marcados indicam os locais pertencentes a amostra.

Note que a distribuição do processo S é infinito dimensional. Para realizar inferência, Diggle et al. (2010) consideram uma malha regular na região B composta de N células. Desse modo, o processo Gaussiano S é aproximado por $\mathbf{S} \sim N_N(\mathbf{0}\mathbf{1}_N, \sigma^2 \mathbf{R}_N(\phi))$. Quanto maior o refinamento da malha, melhor será a aproximação do processo real S . Porém, o tempo computacional cresce de forma não linear no processo de estimação (como veremos adiante) e um balanço entre refinamento e custo deverá ser levado em conta.

A distribuição do processo de Poisson $X|S$ depende da abordagem do problema, isto é, se consideramos o número de pontos da amostra n como fixo ou não. Diggle et al.

(2010) consideram o tamanho da amostra como fixo e, então,

$$f(X|S, n) = \prod_{i=1}^n \frac{\exp\{\alpha + \beta S(x_i)\}}{\int_B \exp\{\alpha + \beta S(u)\} du} = \left(\int_B \exp\{\beta S(u)\} du \right)^{-n} \prod_{i=1}^n \exp\{\beta S(x_i)\}. \quad (3.24)$$

Por conveniência, a dependência sobre o tamanho de amostra será omitida, denotando a distribuição em (3.24) como $f(X|S)$. Note que o termo $\int_B \exp\{\beta S(u)\} du$ é infinito-dimensional, pois depende de S em todos os pontos da região B . Utilizando uma malha fina na região B , esse termo será aproximado por

$$\int_B \exp\{\beta S(u)\} du \simeq \sum_{i=1}^N \Delta_i \exp\{\beta S(x_i)\} \quad (3.25)$$

em que Δ_i corresponde ao volume da i -ésima célula e a função de intensidade é considerada constante em toda volume da célula. Quando consideramos uma malha regular, então $\Delta_i = \Delta, \forall i$. A distribuição de $X|S$ fica aproximada por

$$f(X|S) \simeq \left(\sum_{i=1}^N \Delta_i \exp\{\beta S(x_i)\} \right)^{-n} \left[\prod_{i=1}^N \exp\{\beta S(x_i)\}^{n_i} \right], \quad (3.26)$$

em que n_i denota o número de pontos de X da i -ésima célula da malha e $\sum_{i=1}^N n_i = n$. Note que, ao considerar o tamanho de amostra n fixo, o parâmetro α não se torna necessário na análise.

Se há mais de um valor observado em uma determinada célula, haverá mais de um valor de Y_i para um determinado S_j . Se a área Δ_i for reduzida de tal forma que seja esperado no máximo um ponto em cada célula, a aproximação corresponde a utilizada por Diggle et al. (2010).

A função de verossimilhança do modelo é dada por

$$L(\boldsymbol{\theta}; \mathbf{y}, \mathbf{x}) = f(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta}) = \int_B f(\mathbf{y}|\mathbf{x}, S, \boldsymbol{\theta}) f(\mathbf{x}|S, \boldsymbol{\theta}) f(S|\boldsymbol{\theta}) dS,$$

em que $\boldsymbol{\theta} = (\mu, \tau^2, \sigma^2, \phi, \beta)$.

Sob a abordagem clássica, precisamos encontrar os estimadores de máxima verossimilhança para realizar inferência. A integral acima é intratável e não é possível obter os estimadores de máxima verossimilhança de forma analítica. Assim, formas alternati-

vas de estimação devem ser consideradas. Diggle et al. (2010) realizam a estimação dos parâmetros do modelo utilizando aproximação da verossimilhança por Monte Carlo. A principal crítica sobre essa abordagem é sobre os autores utilizarem a distribuição de $S|Y$ como distribuição preditiva no processo de estimação, ignorando a informação do processo pontual X . Dinsdale and Salibian-Barrera (2019) utilizam aproximação de Laplace para a função de verossimilhança no intuito de considerar a informação do processo X corretamente na estimação dos parâmetros do modelo. A aproximação de Laplace funciona bem para distribuições semelhantes à normal, porém não conhecemos a forma da função de verossimilhança devido a intratabilidade da integral, o que se torna a principal crítica sobre essa abordagem. Dessa forma, nesta tese, propomos e desenvolvemos um algoritmo SAEM para estimação dos parâmetros do modelo geoestatístico com amostragem preferencial, gerando amostras da distribuição preditiva correta de $S|Y, X$.

Sob o enfoque Bayesiano, a inferência acerca do vetor de parâmetros é realizada através da distribuição a posteriori. Assumimos o seguinte modelo:

$$\mathbf{Y}|\mathbf{S}_n, \mu, \tau^2 \sim N_n(\mu\mathbf{1}_n + \mathbf{S}_n, \tau^2\mathbf{I}_n), \quad (3.27)$$

$$\mathbf{X}|S, \lambda \sim PP(\lambda); \quad (3.28)$$

$$\lambda(x) = \exp\{\alpha + \beta S(x)\}; \quad (3.29)$$

$$S|\sigma^2, \phi \sim PG(0, \sigma^2\rho(\phi)); \quad (3.30)$$

$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}), \quad \boldsymbol{\theta} = (\mu, \tau^2, \sigma^2, \phi, \alpha, \beta), \quad (3.31)$$

em que $\mathbf{S}_n = (S(\mathbf{x}_1), \dots, S(\mathbf{x}_n))'$. A distribuição a posteriori é dada por

$$\begin{aligned} \pi(\boldsymbol{\theta}, \mathbf{S}|\mathbf{y}, \mathbf{x}) &\propto \pi(\mathbf{y}, \mathbf{x}, \mathbf{S}, \boldsymbol{\theta}), \\ &\propto L(\mathbf{y}|\mathbf{S}_n, \mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{x}|\mathbf{S}, \boldsymbol{\theta})\pi(\mathbf{S}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}). \end{aligned}$$

Ferreira and Gamerman (2015) utilizam a distribuição do processo de Poisson $X|S$ considerando o tamanho da amostra como aleatório. Temos então que

$$f(X|S) \propto \exp\left\{-\int_B \exp\{\alpha + \beta S(u)\}du\right\} \prod_{i=1}^n \exp\{\alpha + \beta S(x_i)\}. \quad (3.32)$$

Novamente temos o problema da intratabilidade da integral em (3.32) e recorremos a

aproximação em (3.25) para tornar a inferência possível.

Desde o artigo de Diggle et al. (2010), muito tem sido desenvolvido sobre o modelo geoestatístico Gaussiano com amostragem preferencial, porém pouco há na literatura sobre o modelo considerando um processo latente com caudas pesadas. Isso significa que os modelos, até o presente momento, não comportam dados atípicos e não há tratamento para *outliers*, havendo duas alternativas possíveis: a remoção do dado atípico e a perda de informação relevante ou o ajuste do modelo com a possibilidade de viesar os resultados. Motivados pela falta de estudo nessa área, propomos, no Capítulo 4, o modelo geoestatístico t-Student com amostragem preferencial, considerando o processo S como um processo t-Student para acomodar dados atípicos e fornecer uma inferência correta sobre o problema.

Note que, sobre ambas as abordagens clássica e Bayesiana, utilizamos aproximação do processo S . Apesar do uso de discretização ser uma forma de contornar o problema da intratabilidade da integral do processo de Poisson não homogêneo, tal aproximação é uma fonte de erros difícil de mensurar, podendo produzir resultados viesados (Simpson et al., 2016). Outro problema está no custo computacional, pois a simulação de processos Gaussianos são da ordem de n^3 e rapidamente se tornam ineficientes. Dessa forma, é necessário ter um balanço entre custo computacional e a qualidade da aproximação do processo S . No artigo de Gonçalves and Gamerman (2018), os autores, sob a abordagem Bayesiana, realizam inferência exata sobre processos de Cox, considerando que a função de intensidade do processo pontual é limitada, contornando os possíveis erros que a aproximação possa acarretar. Motivados pela ideia dos autores, no Capítulo 6, propomos realizar a inferência exata sobre o modelo geoestatístico com amostragem preferencial, sob abordagem Bayesiana, evitando a discretização do processo latente S e seus possíveis erros.

Capítulo 4

Inferência clássica - Modelagem e computação

Neste capítulo apresentaremos a estimação e predição do modelo geoestatístico com amostragem preferencial via estatística clássica. Desse modo, na Seção 4.1 apresentamos o modelo geoestatístico Gaussiano com amostragem preferencial, na qual revisamos os métodos de estimação existentes e propomos um algoritmo SAEM para estimação dos parâmetros do modelo. Na Seção 4.2, desenvolvemos nossa proposta do modelo considerando a distribuição t-Student para o processo latente S , elaborando a estimação e a predição.

4.1 Modelo Gaussiano

O modelo geoestatístico Gaussiano com amostragem preferencial foi elaborado por Diggle et al. (2010) e apresentado no Capítulo 3. Dessa forma, a estimação será desenvolvida sobre o modelo referido na Definição (3.6). Descrevemos, a seguir, a estimação por aproximação da verossimilhança por Monte Carlo (MCLA - *Monte Carlo likelihood approximation*), presente em Diggle et al. (2010), por aproximação de Laplace, presente em Dinsdale and Salibian-Barrera (2019), e por SAEM, nossa proposta nesse trabalho.

Vamos considerar as seguintes notações: $\mathbf{y} = (y_1, \dots, y_n)'$ o vetor do processo observado, $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$ o vetor de localizações da amostra, \mathbf{S} o processo aproximado por discretização, $S|\mathbf{x}, \mathbf{y}$ significando $S|\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}$.

4.1.1 Estimação via MCLA

A função de verossimilhança do modelo da Definição (3.6) é dada por

$$\begin{aligned}
 L(\boldsymbol{\theta}) &= f(\mathbf{x}, \mathbf{y}) = \int_B f(\mathbf{x}, \mathbf{y}, s) ds \\
 &= \int_B f(\mathbf{y}|\mathbf{x}, s) f(\mathbf{x}|s) f(s) ds \\
 &= E_S[f(\mathbf{y}|\mathbf{x}, s) f(\mathbf{x}|s)]
 \end{aligned} \tag{4.1}$$

Como a integral em (4.1) não possui forma fechada, é necessário o uso de técnicas de integração. Uma alternativa seria aproximar a função de verossimilhança por integral

de Monte Carlo, ficando

$$L_{MC}(\theta) = \frac{1}{m} \sum_{j=1}^m f(\mathbf{y}|\mathbf{x}, s_j) f(\mathbf{x}|s_j), \quad (4.2)$$

em que s_j é a j -ésima simulação de S .

Para estimar o vetor de parâmetros θ , são simuladas m amostras de S para então maximizar a função em (4.2). Segundo Diggle et al. (2010), essa estratégia falha quando τ^2 é igual a zero, pois as simulações de S nos pontos de X serão incompatíveis com os valores observados Y . Essa estratégia também falha quando o erro de medida é pequeno em relação a variância de S , que é o caso da maioria dos problemas reais. Para corrigir isso, os autores utilizam amostragem por importância, utilizando a distribuição de $S|\mathbf{y}$ como distribuição de importância. Desse modo, as simulações de S serão compatíveis com os valores observados de Y . Nesse ponto, os autores utilizam a distribuição de $S|\mathbf{y}$ no contexto não preferencial, que é um problema que será discutido adiante.

Seja $S = \{\mathbf{S}_0, S_1\}$, em que \mathbf{S}_0 denota os valores de S nos pontos de X e S_1 denota os valores de S nos pontos restantes da região B . A função de verossimilhança em (4.1) é reescrita como

$$\begin{aligned} L(\theta) &= \int f(\mathbf{y}|\mathbf{x}, s) f(\mathbf{x}|s) \frac{f(s|\mathbf{y})}{f(s|\mathbf{y})} f(s) ds \\ &= \int f(\mathbf{x}|s) \frac{f(\mathbf{y}|\mathbf{s}_0)}{f(\mathbf{s}_0|\mathbf{y})} f(\mathbf{s}_0) f(s|\mathbf{y}) ds \\ &= E_{S|\mathbf{y}} \left[f(\mathbf{x}|s) \frac{f(\mathbf{y}|\mathbf{s}_0)}{f(\mathbf{s}_0|\mathbf{y})} f(\mathbf{s}_0) \right] \end{aligned} \quad (4.3)$$

Observe que

$$\frac{f(\mathbf{y}|\mathbf{s}_0)}{f(\mathbf{s}_0|\mathbf{y})} f(\mathbf{s}_0) = \frac{f(\mathbf{y}, \mathbf{s}_0)}{f(\mathbf{s}_0|\mathbf{y})} = f(\mathbf{y}),$$

e (4.3) fica

$$L(\theta) = E_{S|\mathbf{y}} [f(\mathbf{x}|s) f(\mathbf{y})] \quad (4.4)$$

em que $f(\mathbf{y})$ é a função de densidade marginal de \mathbf{Y} (no contexto não preferencial). A

aproximação da função de verossimilhança por Monte Carlo fica dada por

$$L_{MC}(\boldsymbol{\theta}) = \frac{1}{m} f(\mathbf{y}) \sum_{j=1}^m f(\mathbf{x}|s_j). \quad (4.5)$$

Para gerar amostras da distribuição de $S|\mathbf{y}$, Diggle et al. (2010) utilizam a seguinte construção considerando o processo aproximado \mathbf{S} . Os locais de $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ são um subconjunto dos $N \geq n$ locais de $X^* = \{\mathbf{x}_1^*, \dots, \mathbf{x}_N^*\}$ que compõe a malha da região B . Defina \mathbf{C} como a matriz $n \times N$ com i -ésima linha consistindo em $N - 1$ 0's e um único 1 identificando a posição de \mathbf{x}_i em X^* . Marginalmente, $\mathbf{S} \sim N_N(0, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} = \sigma^2 \mathbf{R}(\phi)$, e $\mathbf{Y} \sim N_n(\mu \mathbf{1}_n, \boldsymbol{\Sigma}_Y)$, $\boldsymbol{\Sigma}_Y = \mathbf{C} \boldsymbol{\Sigma} \mathbf{C}' + \tau^2 \mathbf{I}_n$. Rue and Held (2005) mostram que a variável $\mathbf{S} + \boldsymbol{\Sigma} \mathbf{C}' \boldsymbol{\Sigma}_Y^{-1} (\mathbf{y} - \mu \mathbf{1}_n + \mathbf{Z} - \mathbf{C} \mathbf{S})$ tem distribuição $\mathbf{S}|\mathbf{y}$, em que $\mathbf{Z} \sim N_n(0, \tau^2 \mathbf{I}_n)$. Então, uma amostra da distribuição $\mathbf{S}|\mathbf{y}$ pode ser obtida da seguinte forma:

Algoritmo: Simulação de $\mathbf{S}|\mathbf{y}$

1. Simule uma variável $\mathbf{Z} \sim N_n(0, \tau^2 \mathbf{I}_n)$,
2. Simule uma amostra de $\mathbf{S} \sim N_N(0, \boldsymbol{\Sigma})$.
3. Calcule $\mathbf{S} + \boldsymbol{\Sigma} \mathbf{C}' \boldsymbol{\Sigma}_Y^{-1} (\mathbf{y} - \mu \mathbf{1}_n + \mathbf{Z} - \mathbf{C} \mathbf{S})$.

A partir de m amostras de $\mathbf{S}|\mathbf{y}$, a estimativa de $\boldsymbol{\theta}$ é obtida pela maximização da função de verossimilhança (4.5), isto é,

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} (L_{MC}(\boldsymbol{\theta})).$$

Dinsdale and Salibian-Barrera (2019) apontam o problema dessa metodologia: Diggle et al. (2010) não consideram o processo pontual X na distribuição de importância, utilizando a distribuição de $S|\mathbf{y}$ do modelo tradicional de geoestatística. Segundo Dinsdale and Salibian-Barrera (2019), essa abordagem fornece boas estimativas apenas no contexto não preferencial. Assim, os autores apresentam a versão corrigida da função de

verossimilhança, utilizando $f(s|\mathbf{x}, \mathbf{y})$ como distribuição de importância, dada por

$$\begin{aligned} L(\boldsymbol{\theta}) &= \int f(\mathbf{y}|\mathbf{x}, s)f(\mathbf{x}|s)f(s)\frac{f(s|\mathbf{x}, \mathbf{y})}{f(s|\mathbf{x}, \mathbf{y})}ds \\ &= E_{S|\mathbf{x}, \mathbf{y}} \left[f(\mathbf{y}|\mathbf{x}, s)\frac{f(\mathbf{x}|s)f(s)}{f(s|\mathbf{x}, \mathbf{y})} \right]. \end{aligned} \quad (4.6)$$

Apesar da distribuição $S|\mathbf{x}, \mathbf{y}$ ser geralmente intratável, é possível gerar amostras utilizando métodos de Monte Carlo, como será discutido adiante. Porém, se observarmos o termo dentro da esperança da Equação (4.6), percebemos que este não depende de S , pois

$$f(\mathbf{y}|\mathbf{x}, s)\frac{f(\mathbf{x}|s)f(s)}{f(s|\mathbf{x}, \mathbf{y})} = \frac{f(\mathbf{y}, \mathbf{x}, s)}{f(s|\mathbf{x}, \mathbf{y})} = \frac{f(s|\mathbf{x}, \mathbf{y})f(\mathbf{x}, \mathbf{y})}{f(s|\mathbf{x}, \mathbf{y})} = f(\mathbf{x}, \mathbf{y}).$$

Ou seja, voltamos ao início do problema. Dessa forma, o desenvolvimento da estimação por esse caminho se torna inviável.

4.1.2 Estimação via aproximação de Laplace

Alternativamente à proposta de estimação por MCLA, Dinsdale and Salibian-Barrera (2019) propõem utilizar aproximação de Laplace para a estimação dos parâmetros do modelo. Podemos reescrever a função de verossimilhança em (4.1) como

$$L(\boldsymbol{\theta}) = \int_B f(\mathbf{y}, \mathbf{x}, s)ds = \int_B \exp\{f(s, \boldsymbol{\theta})\}ds,$$

em que $f(s, \boldsymbol{\theta}) = \log(f(\mathbf{y}, \mathbf{x}, s))$. Novamente vamos recorrer a aproximação do processo infinito dimensional S por uma malha na região B . Assim, escrevemos o processo aproximado por \mathbf{S} e a aproximação de Laplace será dada por

$$\hat{L}(\boldsymbol{\theta}) = (2\pi)^{N/2}|\mathbf{H}(\boldsymbol{\theta})|^{1/2} \exp\{f(\hat{\mathbf{s}}(\boldsymbol{\theta}), \boldsymbol{\theta})\}, \quad (4.7)$$

em que N é o tamanho da malha considerada, $\hat{\mathbf{s}}(\boldsymbol{\theta})$ é o valor de \mathbf{S} que maximiza a função $f(\mathbf{s}, \boldsymbol{\theta})$, isto é,

$$\hat{\mathbf{s}}(\boldsymbol{\theta}) = \underset{\mathbf{s}}{\operatorname{argmax}} f(\mathbf{s}, \boldsymbol{\theta}),$$

e $\mathbf{H}(\boldsymbol{\theta})$ é a matriz Hessiana de $f(\mathbf{s}, \boldsymbol{\theta})$ com respeito a \mathbf{S} aplicada em $\hat{\mathbf{s}}(\boldsymbol{\theta})$, ou seja,

$$\mathbf{H}(\boldsymbol{\theta}) = \left[-\frac{d}{d\mathbf{S}d\mathbf{S}'} f(\mathbf{s}, \boldsymbol{\theta}) \right]^{-1} \Bigg|_{\mathbf{S}=\hat{\mathbf{s}}(\boldsymbol{\theta})}.$$

A estimação do vetor de parâmetros $\boldsymbol{\theta}$ se dá maximizando a função de verossimilhança aproximada \hat{L} em (4.7). O método é iterativo e, a partir de um valor inicial para o vetor de parâmetros $\boldsymbol{\theta}^{(0)}$, podemos esquematizar o processo de estimação com o seguinte algoritmo:

Algoritmo: Estimação de parâmetros via aproximação de Laplace.

1. Encontre o valor que maximiza a função $f(\mathbf{s}, \boldsymbol{\theta}^{(k-1)})$, seja $\hat{\mathbf{s}}(\boldsymbol{\theta})^{(k)}$.
2. Calcule a Hessiana $\mathbf{H}(\boldsymbol{\theta})$ como função de $\boldsymbol{\theta}$ aplicada em $\hat{\mathbf{s}}(\boldsymbol{\theta})^{(k)}$.
3. Maximize a função \hat{L} em relação a $\boldsymbol{\theta}$ e obtenha $\boldsymbol{\theta}^{(k)}$.
4. Se foi observada convergência, compute $\boldsymbol{\theta}^{(k)}$. Caso contrário, repita os passos de (1) a (3).

Para encontrar a matriz Hessiana $\mathbf{H}(\boldsymbol{\theta})$, necessitamos das derivadas de segunda ordem de $f(\mathbf{s}, \boldsymbol{\theta})$ em relação à \mathbf{S} . Apesar de conseguirmos calcular essas derivadas de forma analítica, podemos utilizar métodos automáticos de diferenciação para obter a matriz Hessiana. Diferenciação automática (*Automatic Differentiation - AD*) é a área em que se estuda e desenvolve métodos capazes de calcular e avaliar derivadas de funções especificadas por um programa de computador (Griewank and Walther, 2008). Para fazer uso da AD, utilizamos o pacote TMB (Kristensen et al., 2016), que utiliza AD para computar a aproximação de Laplace de uma função de verossimilhança e a maximiza com respeito ao vetor de parâmetros de interesse. Dessa forma, o pacote realiza a integral em relação ao processo latente S na função de verossimilhança em (4.1) e obtém a estimativa de $\boldsymbol{\theta}$.

Foi comparado o tempo de estimação para obtenção das estimativas dos parâmetros via pacote TMB e via construção explícita da matriz Hessiana e, em nossas simulações, obtemos resultados mais rápidos quando utilizamos o primeiro. Uma implementação mais eficiente da matriz Hessiana exata talvez inverta esse resultado, porém esse não foi

o foco do presente trabalho. Dessa forma, os resultados de estimação via aproximação de Laplace apresentados nessa tese foram obtidos utilizando o pacote TMB.

Note que $\hat{\mathbf{s}}(\boldsymbol{\theta})$ é a moda da distribuição de $S|\mathbf{x}, \mathbf{y}$ para um determinado valor do vetor de parâmetros $\boldsymbol{\theta}$ (pois essa distribuição é proporcional a $f(\mathbf{y}, \mathbf{x}, \mathbf{s})$). Dessa forma, se $\hat{\boldsymbol{\theta}}$ é a estimativa do vetor de parâmetros do modelo, então $\hat{\mathbf{s}}(\hat{\boldsymbol{\theta}})$ fornece um preditor para o processo S dados (\mathbf{x}, \mathbf{y}) . Esse método de predição considera a informação dos dados observados e do processo pontual, corrigindo a predição realizada por Diggle et al. (2010).

Como visto na Seção 2.3, a aproximação de Laplace funciona bem para distribuições que se assemelham a distribuição Gaussiana na vizinhança de sua moda. Além disso, não sabemos se a distribuição de $(S, \mathbf{X}, \mathbf{Y})$ é unimodal. Como não conhecemos a distribuição conjunta de $(S, \mathbf{X}, \mathbf{Y})$, considerar essa aproximação pode não ser uma boa escolha. Dessa forma, apresentamos a seguir nossa proposta de estimação para contornar esse possível problema.

4.1.3 Estimação via SAEM

Como outra alternativa para o método de estimação, podemos utilizar o algoritmo EM se pensarmos os dados do processo S como dados faltantes. Dessa forma, a função de log-verossimilhança completa é dada por

$$l_c(\boldsymbol{\theta}) = \log f(\mathbf{y}, \mathbf{x}, s) = \log [f(\mathbf{y}|\mathbf{x}, s)f(\mathbf{x}|s)f(s)] = \log f(\mathbf{y}|\mathbf{x}, s) + \log f(\mathbf{x}|s) + \log f(s).$$

Pelo modelo apresentado na Definição (3.6), a função de log-verossimilhança completa toma a forma

$$\begin{aligned} l_c(\boldsymbol{\theta}) = & -\frac{n}{2}\log(\tau^2) - \frac{1}{2\tau^2} \sum_{i=1}^n (y_i - \mu - s(\mathbf{x}_i))^2 + \beta \sum_{i=1}^n s(\mathbf{x}_i) \\ & -n\log\left(\int_B \exp\{\beta s(\mathbf{u})\} d\mathbf{u}\right) + \log f(s) + c^*, \end{aligned} \quad (4.8)$$

em que c^* é uma constante em relação aos parâmetros. Considerando \mathbf{S} o processo aproximado para S em uma malha sobre a região B e substituindo a aproximação do

termo da integral em (3.26) na Equação (4.8), obtemos

$$\begin{aligned}
l_c(\boldsymbol{\theta}) &= -\frac{n}{2}\log(\tau^2) - \frac{1}{2\tau^2} \sum_{i=1}^n (y_i - \mu - s(\mathbf{x}_i))^2 + \beta \sum_{j=1}^N n_j s(\mathbf{x}_j) \\
&\quad - n \log \left(\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\} \right) - \frac{N}{2} \log(\sigma^2) - \frac{1}{2} \log(|\mathbf{R}|) \\
&\quad - \frac{1}{2\sigma^2} \mathbf{s}' \mathbf{R}^{-1} \mathbf{s} + c^*,
\end{aligned} \tag{4.9}$$

em que n_j é o número de locais amostrados que se encontram na j -ésima célula da malha e $(\mathbf{x}_1, \dots, \mathbf{x}_N)$ são os centróides das células da malha.

No primeiro passo do algoritmo EM, calculamos o valor esperado da função de log-verossimilhança completa condicionado aos dados e a estimativa do vetor de parâmetros $\boldsymbol{\theta}$ na k -ésima iteração. Esse passo é dado pela função $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = E[l_c(\boldsymbol{\theta})|\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}^{(k)}]$, e será

$$\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) &= -\frac{n}{2}\log(\tau^2) - \frac{1}{2\tau^2} \sum_{i=1}^n E[(y_i - \mu - S(\mathbf{x}_i))^2|\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}^{(k)}] + \\
&\quad + \beta \sum_{j=1}^N n_j E[S(\mathbf{x}_j)|\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}^{(k)}] - n E \left[\log \left(\sum_{j=1}^N \Delta_j \exp\{\beta S(\mathbf{x}_j)\} \right) \middle| \mathbf{y}, \mathbf{x}, \boldsymbol{\theta}^{(k)} \right] \\
&\quad - \frac{N}{2} \log \sigma^2 - \frac{1}{2} \log(|\mathbf{R}|) - \frac{1}{2\sigma^2} E[\mathbf{S}' \mathbf{R}^{-1} \mathbf{S}|\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}^{(k)}] + c^*.
\end{aligned} \tag{4.10}$$

Os valores esperados necessários não possuem forma fechada, inviabilizando essa etapa do algoritmo. Métodos de Monte Carlo para aproximar essas esperanças se tornam úteis e recorreremos ao algoritmo SAEM, proposto por Delyon et al. (1999). No algoritmo SAEM, a função $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$ é substituída pela aproximação

$$\begin{aligned}
\hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) &= \hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)}) + \gamma_k \left(\frac{1}{L} \sum_{i=1}^L l_c(\boldsymbol{\theta}; \mathbf{s}_i, \mathbf{x}, \mathbf{y}) - \hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)}) \right), \\
&= \gamma_k \left(\frac{1}{L} \sum_{i=1}^L l_c(\boldsymbol{\theta}; \mathbf{s}_i, \mathbf{x}, \mathbf{y}) \right) + (1 - \gamma_k) \hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)})
\end{aligned} \tag{4.11}$$

em que $\{\gamma_j\}_{j \geq 1}$ é uma sequência de pesos positivos tal que $\sum_{j=1}^{\infty} \gamma_j = \infty$ e $\sum_{j=1}^{\infty} \gamma_j^2 < \infty$ e \mathbf{s}_l é uma simulação de $\mathbf{S}|\mathbf{x}, \mathbf{y}$, $l = 1, \dots, L$.

Para o modelo Gaussiano, a função \hat{Q} é dada por

$$\begin{aligned} \hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) &= -\frac{n}{2}\log(\tau^2) - \frac{1}{2\tau^2} \sum_{i=1}^n \left((y_i - \mu - s(\mathbf{x}_i))_{SA}^{(k)} \right)^2 + \\ &+ \beta \sum_{j=1}^N n_j (s(\mathbf{x}_j))_{SA}^{(k)} - n \left(\log \left(\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\} \right) \right)_{SA}^{(k)} \\ &- \frac{N}{2}\log\sigma^2 - \frac{1}{2}\log(|\mathbf{R}|) - \frac{1}{2\sigma^2} \left(\mathbf{s}'\mathbf{R}^{-1}\mathbf{s} \right)_{SA}^{(k)} + c^*. \end{aligned} \quad (4.12)$$

em que

$$\begin{aligned} (s(\mathbf{x}_j))_{SA}^{(k)} &= \gamma_k \frac{1}{L} \sum_{l=1}^L s_l(\mathbf{x}_j) + (1 - \gamma_k) (s(\mathbf{x}_j))_{SA}^{(k-1)}, \\ \left((y_i - \mu - s(\mathbf{x}_i))^2 \right)_{SA}^{(k)} &= \gamma_k \frac{1}{L} \sum_{l=1}^L (y_i - \mu - s_l(\mathbf{x}_i))^2 + \\ &(1 - \gamma_k) \left((y_i - \mu - s(\mathbf{x}_i))^2 \right)_{SA}^{(k-1)}, \\ \left(\log \left(\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\} \right) \right)_{SA}^{(k)} &= \gamma_k \frac{1}{L} \sum_{l=1}^L \log \left(\sum_{j=1}^N \Delta_j \exp\{\beta s_l(\mathbf{x}_j)\} \right) + \\ &(1 - \gamma_k) \left(\log \left(\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\} \right) \right)_{SA}^{(k-1)}, \\ \left(\mathbf{s}'\mathbf{R}^{-1}\mathbf{s} \right)_{SA}^{(k)} &= \gamma_k \frac{1}{L} \sum_{l=1}^L \mathbf{s}'_l \mathbf{R}^{-1} \mathbf{s}_l + (1 - \gamma_k) \left(\mathbf{s}'\mathbf{R}^{-1}\mathbf{s} \right)_{SA}^{(k-1)}. \end{aligned}$$

O segundo passo do algoritmo - Maximização - é constituído em encontrar o valor de parâmetros $\boldsymbol{\theta}$ que maximiza a função em (4.12). Derivando a função \hat{Q} em relação a cada parâmetro de interesse, obtemos

$$\begin{aligned} \partial_{\mu} Q_{MC}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) &= \frac{1}{\tau^2} \sum_i (y_i - \mu - (s(\mathbf{x}_i))_{SA}^{(k)}), \\ \partial_{\tau^2} Q_{MC}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) &= -\frac{1}{2\tau^2} \left(n - \frac{1}{\tau^2} \sum_i \left((y_i - \mu - s(\mathbf{x}_i))_{SA}^{(k)} \right)^2 \right), \\ \partial_{\sigma^2} Q_{MC}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) &= -\frac{1}{2\sigma^2} \left(N - \frac{1}{\sigma^2} \left(\mathbf{s}'\mathbf{R}^{-1}\mathbf{s} \right)_{SA}^{(k)} \right), \\ \partial_{\phi} Q_{MC}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) &= -\frac{1}{2} \text{tr} \left(\mathbf{R}^{-1} \frac{d}{d\phi} \mathbf{R} \right) - \frac{1}{2\sigma^2} \left(\frac{d}{d\phi} \left(\mathbf{s}'\mathbf{R}^{-1}\mathbf{s} \right)_{SA}^{(k)} \right), \\ \partial_{\beta} Q_{MC}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) &= \sum_{j=1}^N n_j (s(\mathbf{x}_j))_{SA}^{(k)} - n \frac{d}{d\phi} \left(\log \left(\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\} \right) \right)_{SA}^{(k)}, \end{aligned}$$

em que ∂_θ denota a derivada em relação a θ .

Os estimadores dos parâmetros μ , τ^2 e σ^2 possuem forma fechada e são dados por

$$\begin{aligned}\hat{\mu}^{(k)} &= \frac{\sum_i \left(y_i - (s(\mathbf{x}_i))_{SA}^{(k)} \right)}{n}, \\ \hat{\tau}^2{}^{(k)} &= \frac{\sum_i \left((y_i - \mu - s(\mathbf{x}_i))^2 \right)_{SA}^{(k)}}{n}, \\ \hat{\sigma}^2{}^{(k)} &= \frac{(\mathbf{s}'\mathbf{R}^{-1}\mathbf{s})_{SA}^{(k)}}{N}.\end{aligned}$$

Os parâmetros ϕ e β não possuem estimadores de forma fechada e devem ser estimados numericamente, isto é,

$$(\hat{\phi}, \hat{\beta}) = \underset{(\phi, \beta)}{\operatorname{argmax}} \left(\hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) \right).$$

A estimação é realizada recursivamente até que se observe convergência. Note que a função \hat{Q} possui termos que englobam parâmetros e que dependem de iterações anteriores, como o termo $(\mathbf{s}'\mathbf{R}(\phi)^{-1}\mathbf{s})_{SA}^{(k)}$. Dessa forma, a estimação numérica deve levar em conta os termos que contêm os parâmetros nas iterações anteriores. Por isso, escrevemos (4.12) de forma recursiva

$$\hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = \sum_{r=k_0}^k \omega_r \hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}), \quad (4.13)$$

em que $\omega_k = \gamma_k$, $\omega_r = \gamma_r \prod_{s=r+1}^k (1 - \gamma_s)$ para $r = k_0, \dots, k-1$ e k_0 é o número de iterações sem memória do algoritmo. Consideramos a sequência proposta por Galarza et al. (2017), em que $\omega_r = 0$ para $r = k_0$ e $\omega_r = 1/(k - k_0)$ para $r = k_0 + 1, \dots, k$.

Para termos informação sobre a incerteza da estimativa dos parâmetros, precisamos conhecer a matriz de informação de Fisher, que nos permite calcular o erro padrão das estimativas dos parâmetros. A matriz de informação observada é definida por $I(\boldsymbol{\theta}) = -\partial_\theta^2 l(\boldsymbol{\theta}; \mathbf{y})$.

No Capítulo 2, vimos como aproximar a matriz de informação observada quando utilizamos os algoritmos MCEM e SAEM. Para isso, devemos conhecer o vetor gradiente da função de log-verossimilhança dos dados completos e a matriz de Hessiana. As

derivadas de primeira ordem para $l_c(\boldsymbol{\theta})$ são dadas por:

$$\begin{aligned}\partial_\mu l_c(\boldsymbol{\theta}) &= \frac{1}{\tau^2} \sum_i (y_i - \mu - s(\mathbf{x}_i)), \\ \partial_{\tau^2} l_c(\boldsymbol{\theta}) &= -\frac{n}{2\tau^2} + \frac{1}{2\tau^4} \sum_i (y_i - \mu - s(\mathbf{x}_i))^2, \\ \partial_{\sigma^2} l_c(\boldsymbol{\theta}) &= -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \mathbf{s}' \mathbf{R}^{-1} \mathbf{s}, \\ \partial_\phi l_c(\boldsymbol{\theta}) &= -\frac{1}{2} \text{tr}(\mathbf{R}^{-1} \partial_\phi(\mathbf{R})) + \frac{1}{2\sigma^2} \mathbf{s}' \mathbf{R}^{-1} \partial_\phi(\mathbf{R}) \mathbf{R}^{-1} \mathbf{s}, \\ \partial_\beta l_c(\boldsymbol{\theta}) &= \sum_{j=1}^N n_j s(\mathbf{x}_j) - n \left(\frac{\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\} s(\mathbf{x}_j)}{\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\}} \right).\end{aligned}$$

Considerando a função de correlação exponencial, temos que $\partial_\phi \mathbf{R} = \mathbf{R} \mathbf{D} (1/\phi)$, em que \mathbf{D} é a matriz de distâncias dos N centróides da malha utilizada. As derivadas de segunda ordem são dadas por

$$\begin{aligned}\partial_\mu^2 l_c(\boldsymbol{\theta}) &= -\frac{n}{\tau^2}, \\ \partial_{\tau^2}^2 l_c(\boldsymbol{\theta}) &= \frac{n}{2\tau^4} - \frac{1}{\tau^6} \sum_i (y_i - \mu - s(\mathbf{x}_i))^2, \\ \partial_{\sigma^2}^2 l_c(\boldsymbol{\theta}) &= \frac{N}{2\sigma^4} - \frac{1}{\sigma^6} \mathbf{s}' \mathbf{R}^{-1} \mathbf{s}, \\ \partial_\phi^2 l_c(\boldsymbol{\theta}) &= \frac{1}{2} \text{tr}(\mathbf{R}^{-1} (\partial_\phi(\mathbf{R}) \mathbf{R}^{-1} \partial_\phi(\mathbf{R}) - \partial_\phi^2(\mathbf{R}))) + \\ &\quad \frac{1}{2\sigma^2} \mathbf{s}' \mathbf{R}^{-1} (-2\partial_\phi(\mathbf{R}) \mathbf{R}^{-1} \partial_\phi(\mathbf{R}) + \partial_\phi^2(\mathbf{R})) \mathbf{R}^{-1} \mathbf{s}, \\ \partial_\beta^2 l_c(\boldsymbol{\theta}) &= n \left[\left(\frac{\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\} s(\mathbf{x}_j)}{\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\}} \right)^2 - \frac{\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\} s(\mathbf{x}_j)^2}{\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\}} \right], \\ \partial_\mu \partial_{\tau^2} l_c(\boldsymbol{\theta}) &= -\frac{1}{\tau^4} \sum_i (y_i - \mu - s(\mathbf{x}_i)), \\ \partial_{\sigma^2} \partial_\phi l_c(\boldsymbol{\theta}) &= -\frac{1}{2\sigma^4} \mathbf{s}' \mathbf{R}^{-1} \partial_\phi(\mathbf{R}) \mathbf{R}^{-1} \mathbf{s}, \\ \partial_{\tau^2} \partial_\mu &= \partial_\mu \partial_{\tau^2}, \\ \partial_\phi \partial_{\sigma^2} &= \partial_{\sigma^2} \partial_\phi.\end{aligned}$$

As demais derivadas cruzadas de segunda ordem são nulas. Dessa forma, obtemos a matriz Hessiana aproximada pelas equações (2.8) ou (2.10) e conseguimos estimar o erro padrão dos estimadores por $\text{diag}((\hat{I}(\boldsymbol{\theta}))^{-1})^{1/2}$, em que $\hat{I}(\boldsymbol{\theta})$ é a matriz de informação observada aproximada.

4.1.4 Predição

Como visto anteriormente, Dinsdale and Salibian-Barrera (2019) corrigem a função de verossimilhança apresentada em Diggle et al. (2010) substituindo a distribuição de importância $S|\mathbf{y}$ por $S|\mathbf{x}, \mathbf{y}$. Utilizar a distribuição $S|\mathbf{x}, \mathbf{y}$ corrigiria a estimação no caso preferencial, porém esta não possui forma fechada, o que dificulta sua utilização. Ferreira and Gamerman (2015) realizam a amostragem dessa distribuição utilizando o algoritmo de Metropolis Hastings (MH) e esse será o método abordado nesse trabalho.

Considerando o modelo Gaussiano da Definição (3.6), a distribuição preditiva $S|\mathbf{x}, \mathbf{y}$ é dada por

$$\begin{aligned} f(s|\mathbf{x}, \mathbf{y}) &\propto f(\mathbf{y}|\mathbf{x}, s)f(\mathbf{x}|s)f(s) \\ &\propto \exp\left\{-\frac{1}{2}(\mathbf{y} - \mu\mathbf{1}_n - \mathbf{s}(\mathbf{x}))'(\tau^2\mathbf{I}_n)^{-1}(\mathbf{y} - \mu\mathbf{1}_n - \mathbf{s}(\mathbf{x}))\right\} \times \\ &\quad \left[\prod_{i=1}^n \frac{\exp\{\beta s(\mathbf{x}_i)\}}{\int_B \exp\{\beta s(\mathbf{u})\}d\mathbf{u}}\right] \times f(s). \end{aligned} \quad (4.14)$$

Utilizando o processo aproximado \mathbf{S} , obtemos

$$\begin{aligned} f(s|\mathbf{x}, \mathbf{y}) &\propto \exp\left\{-\frac{1}{2}(\mathbf{y} - \mu\mathbf{1}_n - \mathbf{s}(\mathbf{x}))'(\tau^2\mathbf{I}_n)^{-1}(\mathbf{y} - \mu\mathbf{1}_n - \mathbf{s}(\mathbf{x}))\right\} \times \\ &\quad \frac{\prod_{j=1}^N \exp\{\beta s(\mathbf{x}_j)\}^{n_j}}{\left(\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\}\right)^n} \times \exp\left\{-\frac{1}{2}\mathbf{s}'(\sigma^2\mathbf{R}(\phi))^{-1}\mathbf{s}\right\}. \end{aligned} \quad (4.15)$$

A distribuição em (4.15) não possui forma fechada e, assim, não obtemos seus valores esperados de forma analítica. Técnicas de integração de Monte Carlo se tornam necessárias para avaliar $E[S|\mathbf{x}, \mathbf{y}]$, que corresponde aos valores preditos de S . Como alternativa, utilizamos o algoritmo MH. Note que a dimensão de \mathbf{S} depende do tamanho da malha e, para boa aproximação, será uma dimensão grande. Como consequência, a probabilidade de aceitação do algoritmo é muito baixa e o torna extremamente ineficiente. Ferreira and Gamerman (2015) propõe gerar cada observação de \mathbf{S} por sua condicional completa para contornar esse problema. No nosso trabalho, vamos realizar a amostragem por blocos para reduzir o tempo computacional do algoritmo da abordagem de Ferreira and Gamerman (2015).

Sejam n_g o número de blocos do vetor \mathbf{S} e G o tamanho do bloco e denote \mathbf{S}_g como o g -ésimo bloco, $g = 1, \dots, n_g$. A distribuição de $\mathbf{S}_g|\mathbf{s}_{-g}, \mathbf{x}, \mathbf{y}$, em que $\mathbf{S}_{-g} = \mathbf{S} \setminus \mathbf{S}_g$,

é dada por

$$\begin{aligned}
f(\mathbf{s}_g | \mathbf{s}_{-g}, \mathbf{x}, \mathbf{y}) &\propto \exp \left\{ -\frac{1}{2\tau^2} \sum_{i=1}^n [(y_i - \mu - s(\mathbf{x}_i))^2 \mathbb{1}\{s(\mathbf{x}_i) \in \mathbf{S}_g\}] \right\} \times \\
&\quad \left[\prod_{j=1}^N (\exp\{\beta s(\mathbf{x}_j) \mathbb{1}\{s(\mathbf{x}_j) \in \mathbf{S}_g\}\})^{n_j} \right] \left[\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\} \right]^{-n} \times \\
&\quad \exp \left\{ -\frac{1}{2} \mathbf{s}' (\sigma^2 \mathbf{R}(\phi))^{-1} \mathbf{s} \right\}. \tag{4.16}
\end{aligned}$$

Dessa forma, é gerado um valor proposto para cada elemento de \mathbf{S}_g com distribuição de proposta Gaussiana univariada centrada no seu valor na iteração anterior e variância δ_s . A probabilidade de aceitação do bloco \mathbf{S}_g é dada por

$$\begin{aligned}
p_{\mathbf{S}_g} &\propto \exp \left\{ -\frac{1}{2\tau^2} \sum_{i=1}^n \left[((y_i - \mu - s_p(\mathbf{x}_i))^2 - (y_i - \mu - s_c(\mathbf{x}_i))^2) \mathbb{1}\{s(\mathbf{x}_i) \in \mathbf{S}_g\} \right] \right\} \times \\
&\quad \left[\prod_{j=1}^N (\exp\{\beta(s_p(\mathbf{x}_j) - s_c(\mathbf{x}_j)) \mathbb{1}\{s(\mathbf{x}_j) \in \mathbf{S}_g\}\})^{n_j} \right] \left[\frac{\sum_{j=1}^N \Delta_j \exp\{\beta s_p(\mathbf{x}_j)\}}{\sum_{j=1}^N \Delta_j \exp\{\beta s_c(\mathbf{x}_j)\}} \right]^{-n} \times \\
&\quad \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{s}'_p R^{-1} \mathbf{s}_p - \mathbf{s}'_c R^{-1} \mathbf{s}_c) \right\}.
\end{aligned}$$

Como alternativa e discutido na Seção 4.1.2, dado o vetor de parâmetros estimados $\hat{\boldsymbol{\theta}}$, podemos realizar a predição do processo Gaussiano S através da moda da distribuição de $S | \mathbf{x}, \mathbf{y}$. Logo, $\hat{\mathbf{s}}(\boldsymbol{\theta})$ fornece valores de predição dados os parâmetros estimados e os dados.

4.2 Modelo t-Student

Há pelo menos duas maneiras de induzir uma distribuição de t-Student no modelo Gaussiano da Definição (3.6). A primeira é considerar o processo latente S como um processo t-Student. Dessa forma, estamos definindo que a densidade dos valores do processo S nas caudas da distribuição é maior quando comparada ao processo Gaussiano, isto é, observações que eram consideradas atípicas passam a ser consideradas plausíveis pelo modelo. A segunda forma é considerar que, dados o processo S e o processo pontual X, Y possui distribuição t-Student. Nesse sentido, dizemos que o processo de medida possui caudas mais pesadas e os valores considerados atípicos pelo modelo Gaussiano são

interpretados como valores em que houve um erro maior na mensuração. Também podemos incluir a distribuição t-Student no processo S e no processo Y , mas o modelo fica mais complexo.

Nessa tese, trabalhamos com a distribuição t-Student no processo latente S , significando que os valores reais do processo tem distribuição com caudas pesadas. Dessa forma, vamos definir o processo t-Student e o modelo geoestatístico t-Student com amostragem preferencial.

Definição 4.1 *Um vetor aleatório $\mathbf{S} \in \mathbb{R}^d$ possui distribuição t-Student com ν graus de liberdade (gl), vetor de locação $\boldsymbol{\mu}$ e matriz de escala $\boldsymbol{\Sigma}$ se sua função de densidade de probabilidade é dada por*

$$f(\mathbf{s}) = \frac{\Gamma((\nu + d)/2)}{\Gamma(\nu/2)(\nu\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \left[1 + \frac{(\mathbf{s} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{s} - \boldsymbol{\mu})}{\nu} \right]^{-(\nu+d)/2},$$

em que $\Gamma(\cdot)$ é a função gama. Se $\nu > 1$, então a distribuição possui média $\boldsymbol{\mu}$ e se $\nu > 2$, $\nu(\nu - 2)^{-1}\boldsymbol{\Sigma}$ é a matriz de covariâncias.

Definição 4.2 *Um processo t-Student com ν gl é um processo estocástico $\{S(\mathbf{b}); \mathbf{b} \in B\}$, $B \subset \mathbb{R}^d$, com $d \geq 1$, no qual qualquer conjunto finito de pontos segue uma distribuição t-Student com ν gl.*

Definição 4.3 *O modelo geoestatístico t-Student com amostragem preferencial é definido pela seguinte forma hierárquica:*

- 1 : S é um processo estacionário e isotrópico t-Student com ν gl em uma região B com média 0, parâmetro de escala σ^2 e função de correlação $\rho(h, \phi)$, em que $h = \|\mathbf{x} - \mathbf{x}'\|$ é a distância euclidiana entre os pontos \mathbf{x} e \mathbf{x}' , ϕ é o parâmetro da função de correlação e $\mathbf{x}, \mathbf{x}' \in B$,
- 2 : Condicional em S , X é um processo de Poisson não homogêneo com taxa $\lambda(\mathbf{x}) = \exp(\alpha + \beta S(\mathbf{x}))$, em que α e β são números reais.
- 3 : Condicional em S e X , Y é um conjunto de v.a. Gaussianas independentes com $Y(\mathbf{x}_i) \sim N(\mu + S(\mathbf{x}_i), \tau^2)$.

Vale ressaltar que os graus de liberdade (ν) não serão estimados, sendo considerados fixos durante toda a tese. A estimação de ν foi considerada para trabalhos futuros, como será discutido posteriormente.

4.2.1 Estimação via aproximação de Laplace

Considerando que o processo latente S seja um processo t-Student com ν gl, vamos aproximá-lo por uma malha na região B . O processo \mathbf{S} aproximado possui distribuição t-Student com ν gl com função de densidade dada por

$$f(\mathbf{s}) = \frac{\Gamma((\nu + N)/2)}{\Gamma(\nu/2)(\nu\pi)^{N/2}(\sigma^2)^{N/2}|\mathbf{R}|^{1/2}} \left[1 + \frac{\mathbf{s}'\mathbf{R}^{-1}\mathbf{s}}{\nu\sigma^2} \right]^{-(\nu+N)/2},$$

em que $\mathbf{R} = \mathbf{R}(\phi)$ é a matriz de correlação do processo aproximado \mathbf{S} .

A estimação via aproximação de Laplace para o modelo t-Student é equivalente a estimação para o modelo Gaussiano, alterando a função de densidade do processo aproximado \mathbf{S} para a distribuição t-Student. A alteração também se dá na matriz Hessiana $\mathbf{H}(\theta)$, cujas derivadas dependem da distribuição de \mathbf{S} . O pacote TMB constrói essas derivadas utilizando AD, não sendo necessárias calculá-las analiticamente. Porém, para o caso do modelo t-Student com amostragem preferencial, o uso de memória RAM para armazenar as derivadas e a matriz Hessiana é alto, sendo essa uma limitação computacional do método. Em nossas simulações, o tamanho de malha máximo utilizado foi de 20×20 , utilizando cerca de 26 Gb de memória RAM.

Dessa forma, foi utilizado o pacote TMB para estimação dos parâmetros do modelo t-Student, recorrendo-se ao algoritmo apresentado na Seção 4.1.2.

4.2.2 Estimação via SAEM

Desenvolvemos a seguir a estimação dos parâmetros do modelo geoestatístico t-Student com amostragem preferencial. Construindo uma malha sobre a região B , a função de log-verossimilhança completa do modelo é dada por

$$\begin{aligned} l_c(\theta) &= \log [f(\mathbf{y}|\mathbf{x}, \mathbf{s})f(\mathbf{x}|\mathbf{s})f(\mathbf{s})] = \log f(\mathbf{y}|\mathbf{x}, \mathbf{s}) + \log f(\mathbf{x}|\mathbf{s}) + \log f(\mathbf{s}) \\ &= -\frac{n}{2}\log(\tau^2) - \frac{1}{2\tau^2} \sum_{i=1}^n (y_i - \mu - s(\mathbf{x}_i))^2 + \beta \sum_{j=1}^N n_j s(\mathbf{x}_j) \\ &\quad - n \log \left(\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\} \right) - \frac{N}{2}\log(\sigma^2) - \frac{1}{2}\log(|\mathbf{R}|) \\ &\quad - \left(\frac{\nu + N}{2} \right) \log \left[1 + \frac{\mathbf{s}'\mathbf{R}^{-1}\mathbf{s}}{\nu\sigma^2} \right] + c_t^*, \end{aligned} \tag{4.17}$$

em que Δ_j é o volume da j -ésima célula da malha, n_j é o número de observações na j -ésima célula e c_t^* é uma constante em relação aos parâmetros. A função \hat{Q} é dada por

$$\begin{aligned} \hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) &= -\frac{n}{2}\log(\tau^2) - \frac{1}{2\tau^2} \sum_{i=1}^n \left((y_i - \mu - s(\mathbf{x}_i))^2 \right)_{SA}^{(k)} + \\ &+ \beta \sum_{j=1}^N n_j (S(\mathbf{x}_j))_{SA}^{(k)} - n \left(\log \left(\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\} \right) \right)_{SA}^{(k)} \\ &- \frac{N}{2}\log\sigma^2 - \frac{1}{2}\log(|\mathbf{R}|) - \left(\frac{\nu + N}{2} \right) \left(\log \left[1 + \frac{\mathbf{s}'\mathbf{R}^{-1}\mathbf{s}}{\nu\sigma^2} \right] \right)_{SA}^{(k)} + c^*, \end{aligned} \quad (4.18)$$

em que

$$\begin{aligned} (s(\mathbf{x}_j))_{SA}^{(k)} &= \gamma_k \frac{1}{L} \sum_{l=1}^L s_l(\mathbf{x}_j) + (1 - \gamma_k) (s(\mathbf{x}_j))_{SA}^{(k-1)}, \\ \left((y_i - \mu - s(\mathbf{x}_i))^2 \right)_{SA}^{(k)} &= \gamma_k \frac{1}{L} \sum_{l=1}^L (y_i - \mu - s_l(\mathbf{x}_i))^2 \\ &+ (1 - \gamma_k) \left((y_i - \mu - s(\mathbf{x}_i))^2 \right)_{SA}^{(k-1)}, \\ \left(\log \left(\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\} \right) \right)_{SA}^{(k)} &= \gamma_k \frac{1}{L} \sum_{l=1}^L \log \left(\sum_{j=1}^N \Delta_j \exp\{\beta s_l(\mathbf{x}_j)\} \right) + \\ &(1 - \gamma_k) \left(\log \left(\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\} \right) \right)_{SA}^{(k-1)}, \\ \left(\log \left[1 + \frac{\mathbf{s}'\mathbf{R}^{-1}\mathbf{s}}{\nu\sigma^2} \right] \right)_{SA}^{(k)} &= \gamma_k \frac{1}{L} \sum_{l=1}^L \log \left[1 + \frac{\mathbf{s}'_l \mathbf{R}^{-1} \mathbf{s}_l}{\nu\sigma^2} \right] + \\ &(1 - \gamma_k) \left(\log \left[1 + \frac{\mathbf{s}'\mathbf{R}^{-1}\mathbf{s}}{\nu\sigma^2} \right] \right)_{SA}^{(k-1)}. \end{aligned}$$

Os estimadores dos parâmetros μ e τ^2 possuem forma fechada e são dados por

$$\begin{aligned} \hat{\mu}^{(k)} &= \frac{\sum_i (y_i - (s(\mathbf{x}_i))_{SA}^{(k)})}{n}, \\ \hat{\tau}^2{}^{(k)} &= \frac{\sum_i \left((y_i - \mu - s(\mathbf{x}_i))^2 \right)_{SA}^{(k)}}{n}. \end{aligned}$$

Diferente do modelo Gaussiano, o estimador do parâmetro σ^2 para o modelo t não possui forma analítica. Assim, os parâmetros σ^2 , ϕ e β devem ser estimados numérica-

mente, isto é,

$$(\hat{\sigma}^2, \hat{\phi}, \hat{\beta}) = \underset{(\sigma^2, \phi, \beta)}{\operatorname{argmax}} (\hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})).$$

A estimação é realizada recursivamente até que se observe convergência. Assim como no modelo Gaussiano, a função \hat{Q} possui termos que englobam parâmetros e dependem de valores das iterações anteriores. A forma recursiva da função \hat{Q} em (4.13) é adotada.

Para avaliar a incerteza sobre a estimação dos parâmetros, vamos calcular a matriz de informação observada, que será aproximada como visto no Capítulo 2. Para isso, necessitamos das derivadas de primeira e segunda ordem da função de verossimilhança completa dos dados.

Para os parâmetros μ , τ^2 e β , as derivadas de primeira e segunda ordem (inclusive as derivadas cruzadas) são idênticas às do modelo Gaussiano. Em relação aos parâmetros σ^2 e ϕ , obtemos

$$\begin{aligned} \partial_{\sigma^2} l_c(\boldsymbol{\theta}) &= -\frac{N}{2\sigma^2} + \frac{\nu + N}{2\sigma^2} \frac{\mathbf{s}'\mathbf{R}^{-1}\mathbf{s}}{\nu\sigma^2 + \mathbf{s}'\mathbf{R}^{-1}\mathbf{s}}, \\ \partial_{\phi} l_c(\boldsymbol{\theta}) &= -\frac{1}{2}\operatorname{tr}(\mathbf{R}^{-1}\partial_{\phi}\mathbf{R}) + \frac{\nu + N}{2(\nu\sigma^2 + \mathbf{s}'\mathbf{R}^{-1}\mathbf{s})} \mathbf{s}'\mathbf{R}^{-1}(\partial_{\phi}\mathbf{R})\mathbf{R}^{-1}\mathbf{s}, \\ \partial_{\sigma^2}^2 l_c(\boldsymbol{\theta}) &= \frac{1}{2\sigma^2} \left[\frac{N}{\sigma^2} + \left(\frac{(\nu + N)\mathbf{s}'\mathbf{R}^{-1}\mathbf{s}}{\nu\sigma^2 + \mathbf{s}'\mathbf{R}^{-1}\mathbf{s}} \right) \left(\frac{1}{\sigma^2} + \frac{\nu}{\nu\sigma^2 + \mathbf{s}'\mathbf{R}^{-1}\mathbf{s}} \right) \right], \\ \partial_{\phi}^2 l_c(\boldsymbol{\theta}) &= \frac{1}{2}\operatorname{tr}(\mathbf{R}^{-1}(\partial_{\phi}(\mathbf{R})\mathbf{R}^{-1}\partial_{\phi}(\mathbf{R}) - \partial_{\phi}^2(\mathbf{R}))) + \\ &\quad -\frac{(\nu + N)}{2} \left[(\nu\sigma^2 + \mathbf{s}'\mathbf{R}^{-1}\mathbf{s})^{-2} (\mathbf{s}'\mathbf{R}^{-1}\partial_{\phi}(\mathbf{R})\mathbf{R}^{-1}\mathbf{s})^2 + \right. \\ &\quad \left. (\nu\sigma^2 + \mathbf{s}'\mathbf{R}^{-1}\mathbf{s})^{-1} \mathbf{s}'\mathbf{R}^{-1} (-2\partial_{\phi}(\mathbf{R})\mathbf{R}^{-1}\partial_{\phi}(\mathbf{R}) + \partial_{\phi}(\mathbf{R})) \mathbf{R}^{-1}\mathbf{s} \right] \\ \partial_{\sigma^2}\partial_{\phi} l_c(\boldsymbol{\theta}) &= -\frac{\nu(\nu + N)}{2(\nu\sigma^2 + \mathbf{s}'\mathbf{R}^{-1}\mathbf{s})^2} \mathbf{s}'\mathbf{R}^{-1}(\partial_{\phi}\mathbf{R})\mathbf{R}^{-1}\mathbf{s}, \\ \partial_{\sigma^2}\partial_{\phi} &= \partial_{\phi}\partial_{\sigma^2}. \end{aligned}$$

Dessa forma, obtemos a matriz Hessiana aproximada pelas equações (2.8) ou (2.10) e conseguimos estimar o erro padrão dos estimadores por $\operatorname{diag}(\hat{I}(\boldsymbol{\theta})^{-1})^{1/2}$, em que $\hat{I}(\boldsymbol{\theta})$ é a matriz de informação observada aproximada.

4.2.3 Predição

Considerando o processo aproximado \mathbf{S} , a distribuição preditiva de $\mathbf{S}|\mathbf{x}, \mathbf{y}$ para o modelo t-Student com amostragem preferencial é dada por

$$\begin{aligned}
 f(\mathbf{s}|\mathbf{x}, \mathbf{y}) &\propto f(\mathbf{y}|\mathbf{x}, \mathbf{s})f(\mathbf{x}|\mathbf{s})f(\mathbf{s}) \\
 &\propto \exp\left\{-\frac{1}{2}(\mathbf{y} - \mu\mathbf{1}_n - \mathbf{s}(\mathbf{x}))'(\tau^2\mathbf{I}_n)^{-1}(\mathbf{y} - \mu\mathbf{1}_n - \mathbf{s}(\mathbf{x}))\right\} \times \\
 &\quad \frac{\prod_{j=1}^N \exp\{\beta s(\mathbf{x}_j)\}^{n_j}}{\left(\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\}\right)^n} \times \left[1 + \frac{\mathbf{s}'\mathbf{R}^{-1}\mathbf{s}}{\nu\sigma^2}\right]^{-\left(\frac{\nu+N}{2}\right)}. \quad (4.19)
 \end{aligned}$$

Para gerar amostras da distribuição em (4.19), será utilizado o método MH em blocos. Novamente, sejam n_g o número de blocos do vetor \mathbf{S} e G o tamanho do bloco e denote \mathbf{S}_g como o g -ésimo bloco, $g = 1, \dots, n_g$. A distribuição de $\mathbf{S}_g|\mathbf{s}_{-g}, \mathbf{x}, \mathbf{y}$, em que $\mathbf{S}_{-g} = \mathbf{S} \setminus \mathbf{S}_g$, é dada por

$$\begin{aligned}
 f(\mathbf{S}_g|\mathbf{s}_{-g}, \mathbf{x}, \mathbf{y}) &\propto \exp\left\{-\frac{1}{2\tau^2} \sum_{i=1}^n [(y_i - \mu - s(\mathbf{x}_i))^2 \mathbb{1}\{s(\mathbf{x}_i) \in \mathbf{S}_g\}]\right\} \times \\
 &\quad \left[\prod_{j=1}^N (\exp\{\beta s(\mathbf{x}_j) \mathbb{1}\{s(\mathbf{x}_j) \in \mathbf{S}_g\}\})^{n_j}\right] \left[\sum_{j=1}^N \Delta_j \exp\{\beta s(\mathbf{x}_j)\}\right]^{-n} \times \\
 &\quad \left[1 + \frac{\mathbf{s}'\mathbf{R}^{-1}\mathbf{s}}{\nu\sigma^2}\right]^{-\left(\frac{\nu+N}{2}\right)}. \quad (4.20)
 \end{aligned}$$

Dessa forma, é gerado um valor proposto para cada elemento de \mathbf{S}_g com distribuição de proposta Gaussiana univariada centrada no seu valor na iteração anterior e variância δ_s . A probabilidade de aceitação do bloco \mathbf{S}_g é dada por

$$\begin{aligned}
 p_{\mathbf{S}_g} &\propto \exp\left\{-\frac{1}{2\tau^2} \sum_{i=1}^n \left[((y_i - \mu - s_p(\mathbf{x}_i))^2 - (y_i - \mu - s_c(\mathbf{x}_i))^2) \mathbb{1}\{s(\mathbf{x}_i) \in \mathbf{S}_g\} \right]\right\} \times \\
 &\quad \left[\prod_{j=1}^N (\exp\{\beta(s_p(\mathbf{x}_j) - s_c(\mathbf{x}_j)) \mathbb{1}\{s(\mathbf{x}_j) \in \mathbf{S}_g\}\})^{n_j}\right] \left[\frac{\sum_{j=1}^N \Delta_j \exp\{\beta s_p(\mathbf{x}_j)\}}{\sum_{j=1}^N \Delta_j \exp\{\beta s_c(\mathbf{x}_j)\}}\right]^{-n} \times \\
 &\quad \left[\frac{\nu\sigma^2 + \mathbf{s}'_p\mathbf{R}^{-1}\mathbf{s}_p}{\nu\sigma^2 + \mathbf{s}'_c\mathbf{R}^{-1}\mathbf{s}_c}\right]^{-\left(\frac{\nu+N}{2}\right)}.
 \end{aligned}$$

Capítulo 5

Inferência clássica - Estudo de simulação

Apresentamos a seguir resultados de simulações para avaliar a metodologia apresentada. Discutimos o comportamento das cadeias de Markov da distribuição preditiva $S|\mathbf{x}, \mathbf{y}$ obtidas via Metropolis Hastings. Além disso, estudamos o efeito da blocagem no algoritmo MH para gerar amostras da distribuição preditiva na estimação dos parâmetros pelos métodos MCEM e SAEM e verificamos também o impacto de se fixar o parâmetro de correlação nas estimativas dos parâmetros do modelo. Para comparação dos métodos de estimação, realizamos um estudo comparativo entre os métodos MCLA, TMB, MCEM e SAEM. Por fim, comparamos os métodos de predição a partir dos parâmetros estimados.

5.1 Distribuição preditiva de $S|\mathbf{x}, \mathbf{y}$

A distribuição preditiva de $S|\mathbf{x}, \mathbf{y}$ é obtida através de simulações, pois não possui forma analítica, e foi apresentado na Seção 4.1.4 um algoritmo MCMC com passo de Metropolis-Hastings para gerar amostras de tal distribuição. Dessa forma, torna-se necessária a avaliação da convergência das cadeias de Markov para identificarmos se a distribuição alvo foi atingida. É a partir da convergência que se obtém as amostras da distribuição de interesse, sendo desconsideradas as amostras geradas antes da convergência, o que chamamos de *burn in*. No caso, temos a cadeia de Markov de cada S_i gerado, $i = 1, \dots, N$, N o tamanho da malha de predição. Como são muitas cadeias, verificar cada uma se torna impraticável e uma alternativa é verificar se a cadeia de Markov de uma função de todas as variáveis converge. Dessa forma, verificamos se a cadeia da função de log-densidade de \mathbf{S} converge e o tamanho de *burn in* necessário para alcançar a convergência.

Para gerar uma amostra do processo Gaussiano, foi utilizada uma malha 50×50 no quadrado unitário, totalizando 2500 células. Nessa malha, foi gerada uma realização do processo Gaussiano S . A partir da amostra de S , uma realização do processo pontual X de tamanho 100 foi gerada através do Algoritmo 2 da Seção 2.4.2 considerando n fixo. Dados S e X , uma amostra do processo Y foi simulada. Dessa forma, obtemos uma amostra do modelo geoestatístico Gaussiano. Os parâmetros considerados foram $\mu = 4$,

$\tau^2 = 0,10$, $\sigma^2 = 1,5$, $\phi = 0,15$ e $\beta = 2$. Para o processo t-Student, a simulação foi realizada de forma análoga, considerando $\nu = (1, 3, 5, 7, 10)$ graus de liberdade na geração do processo S .

Para realizar a predição, foram consideradas malhas regulares de tamanhos N iguais a 225, 400, 625 e 900 e considerados tamanhos de blocos 1, 5 e 10. Dessa forma, podemos verificar o efeito do tamanho da malha e do tamanho do bloco nas cadeias de predição. Como valor inicial para o vetor \mathbf{S} , foi gerado um valor de uma distribuição Normal padrão para cada S_i , $i = 1, \dots, N$ e foram realizadas 1000 iterações do algoritmo MH.

A Figura 5.1 apresenta o mapa do processo \mathbf{S} simulado para $N = 625$ e os pontos do processo pontual X para os modelos Gaussiano e t-Student com 3 gl. Como esperado, os pontos da amostra de X são compostos, em sua maioria, por valores altos (pois escolhemos um valor positivo para o parâmetro de preferência β).

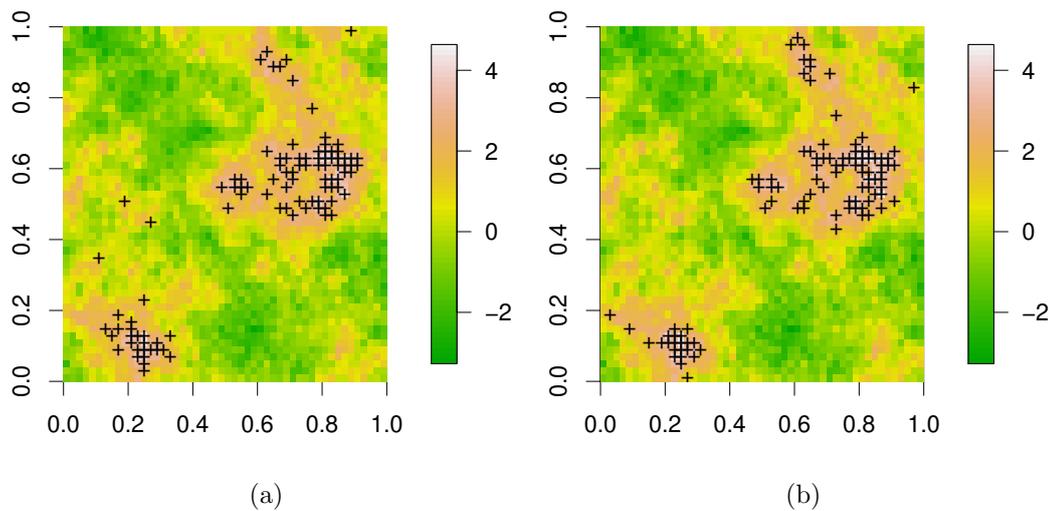


Figura 5.1: Mapas de S para $N = 625$ e o processo pontual X (pontos marcados) para os modelos (a) Gaussiano e (b) t-Student com 3 gl.

A Figura 5.2 apresenta as cadeias da função de log-densidade de \mathbf{S} para ambos modelos, considerando tamanho de bloco igual a 1 e diferentes tamanhos de malha N . Observamos rápida convergência para todos os tamanhos de malha considerados para ambos modelos, sendo que as cadeias de S do modelo Gaussiano apresentaram menor variabilidade. Notamos também um *burn in* necessário de 100 iterações para visualizar convergência da cadeia de log-densidade de S para o modelo Gaussiano e 200 iterações para o modelo t-Student. Considerando o modelo t-Student, observamos maior variabilidade

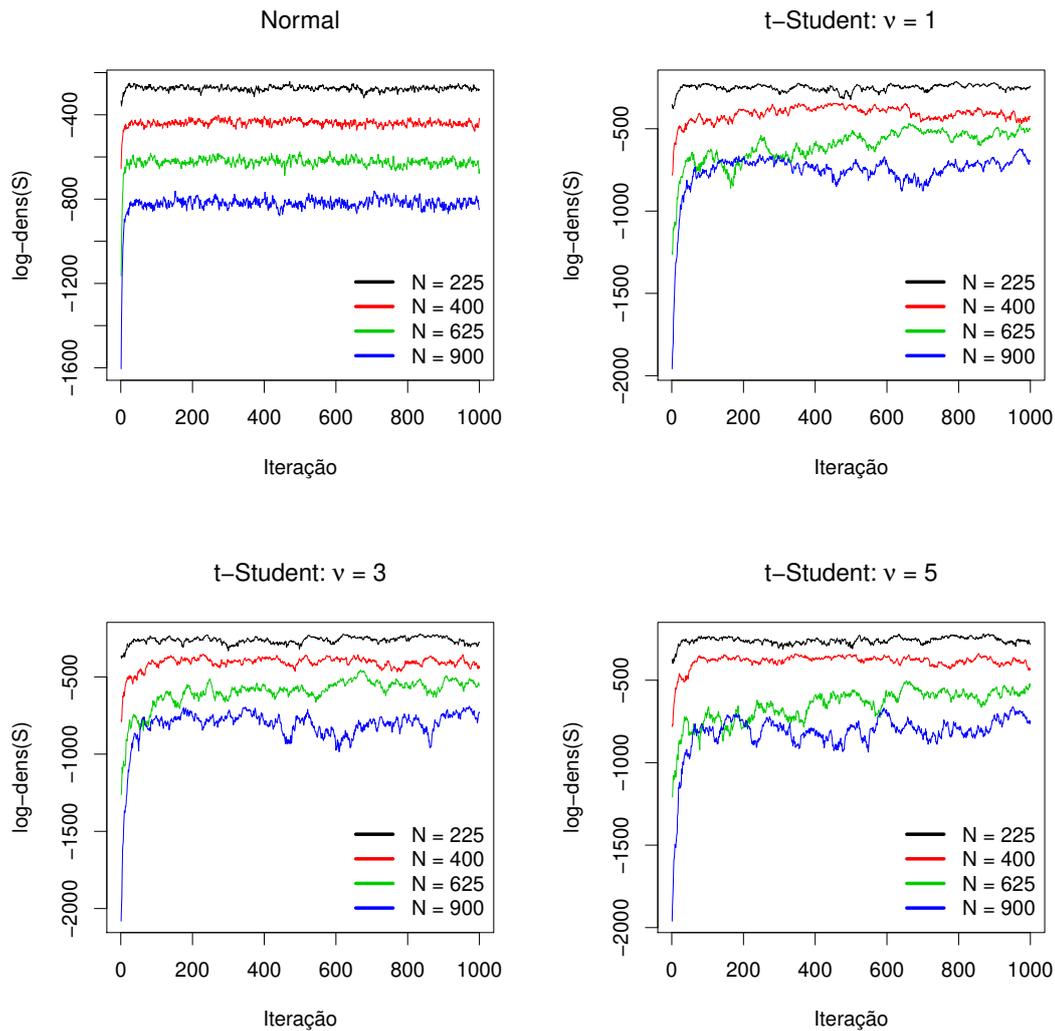


Figura 5.2: Gráfico das cadeias de Markov da log-densidade de S para os modelos Gaussiano e t-Student considerando vários valores de N e graus de liberdade.

das cadeias para maiores valores de N considerados.

A Tabela 5.1 apresenta o tempo computacional que foi necessário para se realizar as 1000 iterações do algoritmo MH para ambos modelos. Para um N fixo, o tempo é semelhante para os modelos Gaussiano e t-Student. Considerando os valores de N , percebemos um aumento não linear no tempo computacional, chegando a cerca de 20 minutos para $N = 900$.

A Figura 5.3 apresenta as cadeias da função de log-densidade de S para ambos modelos, considerando $N = 225$ e tamanhos de bloco iguais a 1, 5 e 10. Observamos que a convergência é mais rápida para a geração de elemento por elemento. Quanto maior o tamanho do bloco, mais lenta é a convergência, porém menor o tempo computacional como

Tabela 5.1: Tempo computacional (em segundos) para realizar 1000 iterações do algoritmo MH para os modelos Gaussiano e t-Student considerando simulação de elemento por elemento de \mathbf{S} .

N	Normal	t_1	t_3	t_5	t_7	t_{10}
225	21,25	21,34	20,83	20,79	20,78	20,78
400	110,48	107,35	107,59	106,91	107,47	107,20
625	398,62	391,26	394,79	391,46	394,74	391,44
900	1176,48	1159,14	1167,35	1159,44	1167,43	1159,05

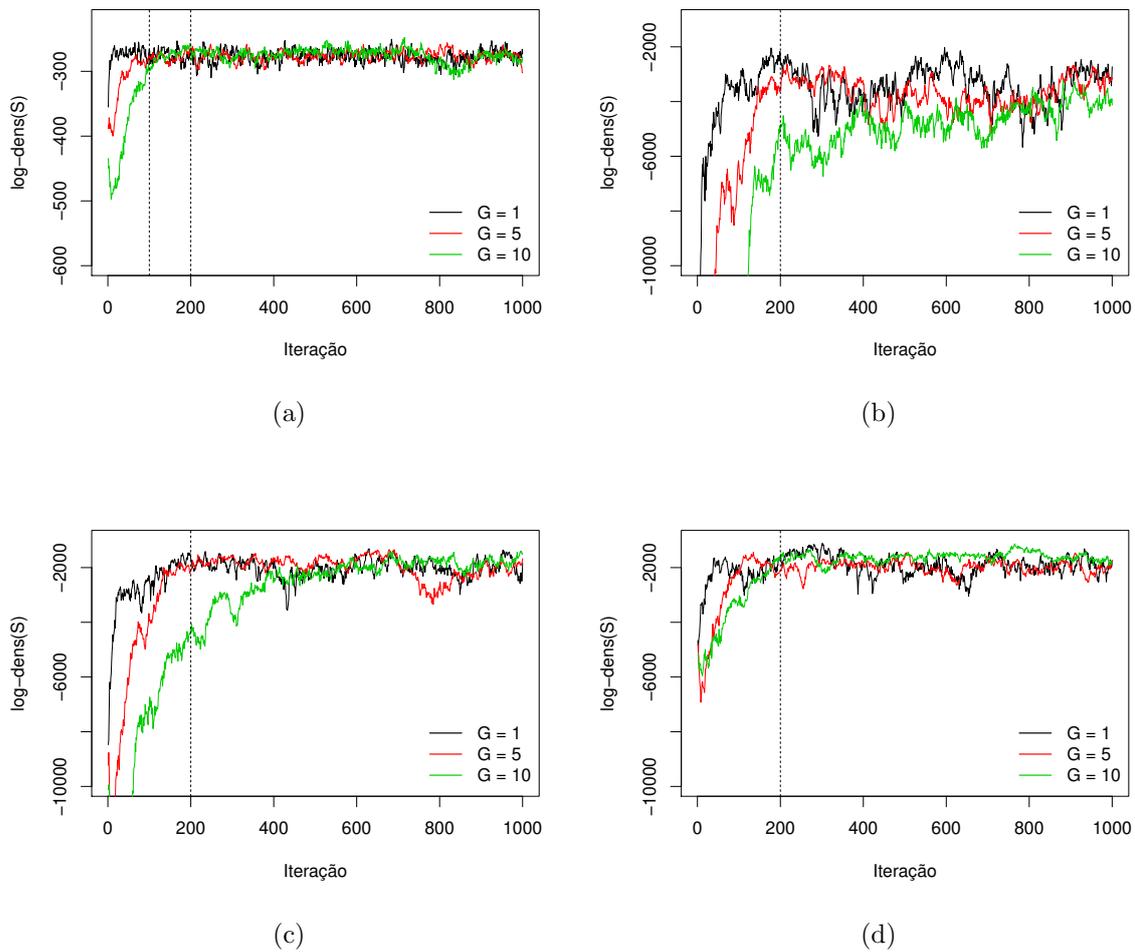


Figura 5.3: Gráfico das cadeias de Markov da log-densidade de S para os modelos (a) Gaussiano e t-Student com (b) 3, (c) 5 e (d) 10 gl, considerando vários valores de G .

mostra a Tabela 5.2. Logo, um balanço entre o tempo de *burn in* e tempo computacional deve ser levado em consideração para realizar predições.

A importância em escolher a melhor combinação de tamanho de malha e tamanho de blocos se dá, principalmente, na estimação dos parâmetros do modelo. Pelo algoritmo SAEM, necessitamos gerar amostras da distribuição preditiva $S|\mathbf{x}, \mathbf{y}$ para toda iteração do

Tabela 5.2: Tempo computacional (em segundos) para realizar 1000 iterações do algoritmo MH para os modelos Gaussiano e t-Student com 3 e 5 gl, considerando tamanhos de malha 225 e 900.

G	$N = 225$			$N = 900$		
	Normal	t_3	t_5	Normal	t_3	t_5
1	21,41	21,48	21,49	1331,92	1369,32	1370,46
5	4,28	4,41	4,39	271,15	269,89	274,93
10	2,17	2,16	2,21	135,18	136,44	135,65

algoritmo e, se escolhermos uma malha refinada e tamanho de bloco igual a 1, por exemplo, o algoritmo será bastante ineficiente em relação a tempo computacional. Por outro lado, se admitimos um tamanho de malha e de bloco tal que não comprometa a estimação dos parâmetros e aumente a velocidade para obtenção das estimativas, tornamos o algoritmo mais eficiente. Por exemplo, considere realizar 300 iterações do algoritmo SAEM para o modelo Gaussiano com $N = 225$ e $G = 5$. Com essa combinação, o tempo para gerar uma amostra da distribuição preditiva com *burn in* de 100 iterações é 2,4 segundos, o que significa que seriam necessários 12 minutos do algoritmo para realizar predição. Por outro lado, se escolhermos $N = 900$ e $G = 1$ e mesmo tamanho de *burn in*, o algoritmo levaria quase 12 horas somente na etapa de predição.

Após a convergência da cadeia da log-densidade de S , obtemos amostras da distribuição preditiva $S|\mathbf{x}, \mathbf{y}$ que nos permite realizar a predição do processo S na região de estudo e construir mapas de predição. Para avaliar a qualidade da predição obtida, foram consideradas as amostras do processo Gaussiano e t-Student com 3 gl, tamanho de malha 900 e calculadas duas medidas de erro, sendo elas:

- Erro Médio Absoluto:

$$EMA = \frac{1}{N} \sum_{i=1}^N |\hat{S}_i - S_i|,$$

em que \hat{S}_i é a estimativa de S na i -ésima localização da malha.

- Raiz quadrada do Erro Quadrático Médio de Predição:

$$REQMP = \sqrt{\frac{1}{N} (\hat{\mathbf{S}} - \mathbf{S})' (\hat{\mathbf{S}} - \mathbf{S})},$$

sendo $\hat{\mathbf{S}}$ a estimativa do vetor \mathbf{S} .

As amostras foram simuladas considerando valores de β iguais a zero (amostragem

não preferencial), 1 e 2. Dessa forma, podemos comparar a predição realizada pela distribuição preditiva e pelo método de krigagem para diferentes forças de preferenciabilidade. Foram realizadas 300 iterações na predição pela distribuição preditiva e descartadas as 100 primeiras.

A Figura 5.4 apresenta o mapa do processo S original, considerando distribuição normal, e os mapas preditos pela distribuição preditiva e por krigagem, considerando os vários valores de β . Para o processo com amostragem não preferencial ($\beta = 0$), observamos que as predições obtidas pelos métodos se diferem em várias regiões do mapa, como no quadrante inferior esquerdo, em que a predição realizada pela distribuição preditiva apresenta valores menores comparado aos valores obtidos pela krigagem. Para os processos com amostragem preferencial, notamos que as estimativas pela krigagem são superestimadas, não recuperando os menores valores do processo S . Já para as predições obtidas pela distribuição preditiva, obtemos estimativas mais próximas das verdadeiras. Observe que, a medida que o valor de β aumenta, a amostra é composta cada vez mais por valores altos do processo, reduzindo a informação sobre os menores valores do mesmo. Desse modo, não possuímos informação suficiente para estimar os valores na cauda inferior do processo S . Apesar disso, o método de predição pela distribuição preditiva consegue obter um alcance da cauda inferior do processo maior que o método de krigagem, tornando-o mais adequado e menos viesado na presença de amostragem preferencial.

A Tabela 5.3 apresenta as medidas de qualidade de predição para o modelo Gaussiano considerando amostras com os diferentes valores de β . Notamos que o método de krigagem resultou em menores EAM e REQMP na situação de não preferenciabilidade, indicando melhor predição que a obtida pela distribuição preditiva. A justificativa para esse resultado é que a krigagem fornece a predição exata (de forma analítica) no caso não preferencial, fornecendo melhores resultados comparados aos de simulação. Na presença de amostragem preferencial, o método que apresenta menores valores de EMA e REQMP é o da distribuição preditiva. Nesses casos, o método de krigagem não considera a informação do processo pontual, resultando em estimativas com maiores medidas de erro. Observe que, quanto maior o grau de preferenciabilidade, maior será o viés da predição, pois não possuímos informação suficiente para recuperar completamente o processo S original. Porém, notamos que as medidas de erro de predição aumentaram mais para o método de krigagem do que para o método da distribuição preditiva, quando β aumentou

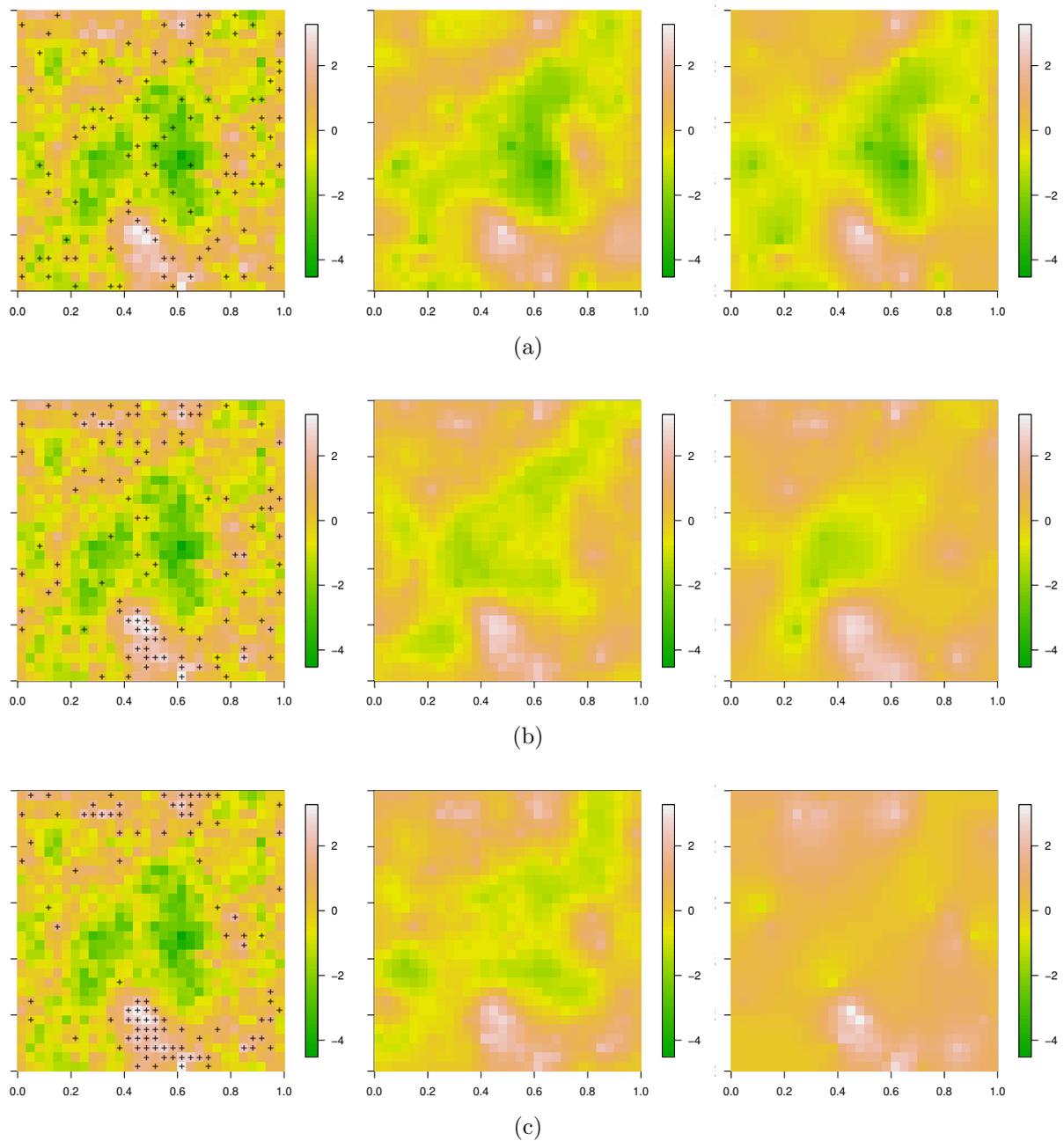


Figura 5.4: Mapas do processo S Gaussiano simulado (primeira coluna) e preditos pela distribuição de $S|\mathbf{x}, \mathbf{y}$ (segunda coluna) e por krigagem (terceira coluna), considerando amostras com (a) $\beta = 0$, (b) $\beta = 1$ e (c) $\beta = 2$.

de 1 para 2, indicando que esse é mais adequado no caso de amostragem preferencial que o primeiro.

A Figura 5.5 apresenta os gráficos de valores preditos versus valores verdadeiros do processo Gaussiano S considerando os dois métodos de predição. O método de krigagem apresentou melhores estimativas no caso não preferencial, apresentando pontos que se concentram mais ao redor da reta de 45° comparado ao método da distribuição preditiva,

Tabela 5.3: Medidas de qualidade de predição para o modelo Gaussiano.

β	Método	EMA	REQMP
0	Dist. Preditiva	0,635	0,826
	Krigagem	0,583	0,754
1	Dist. Preditiva	0,613	0,793
	Krigagem	0,744	0,971
2	Dist. Preditiva	0,623	0,817
	Krigagem	1,007	1,279

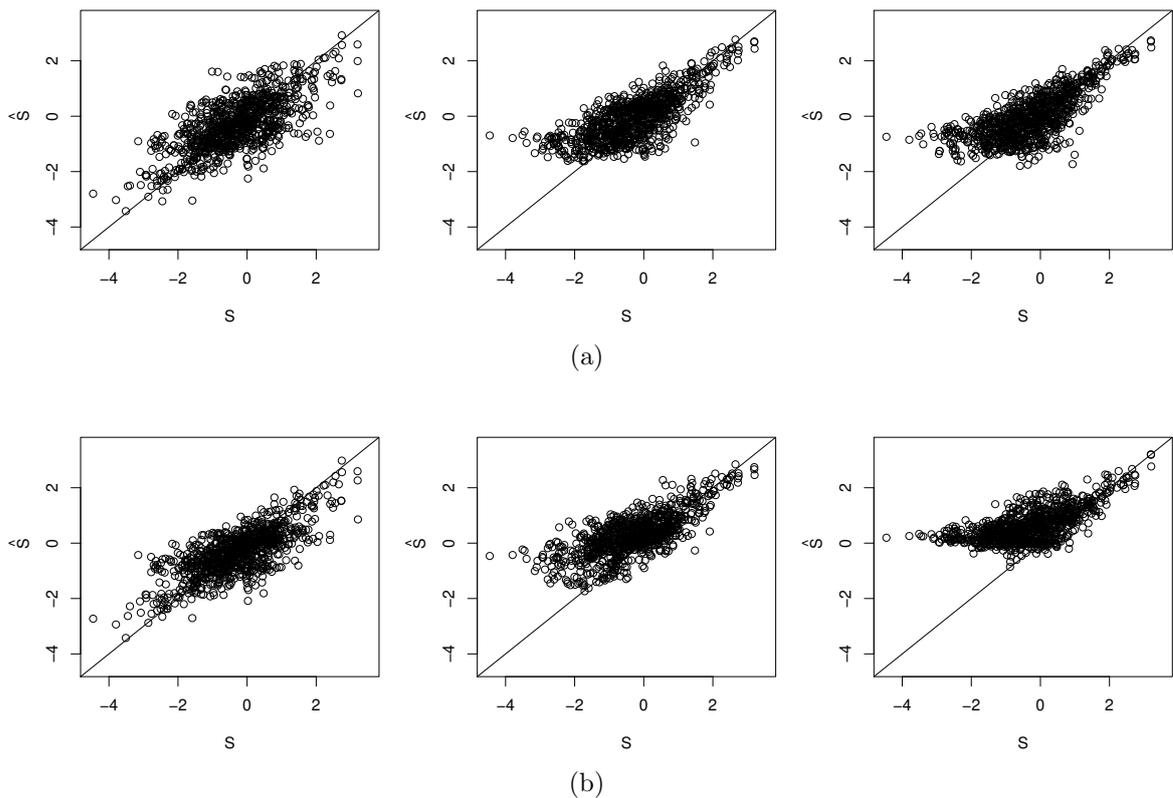


Figura 5.5: Gráfico de valores preditos versus valores verdadeiros do processo Gaussiano S considerando (a) a distribuição preditiva $S|\mathbf{x}, \mathbf{y}$ e (b) o método de krigagem para valores de β iguais a 0 (primeira coluna), 1 (segunda coluna) e 2 (terceira coluna).

corroborando com o mapa e os valores de erro de medida apresentados. Para o caso preferencial, observamos claramente que o método de krigagem não consegue recuperar a cauda inferior do processo S , tornando as estimativas viesadas. O método da distribuição preditiva conseguiu recuperar parte da cauda inferior do processo, proporcionando estimativas melhores.

Para o modelo t-Student com 3 gl, foram realizadas 400 iterações do método de distribuição preditiva e descartadas as 200 primeiras. A Figura 5.6 apresenta o mapa de S simulado e os mapas de predição obtidos pela distribuição preditiva $S|\mathbf{x}, \mathbf{y}$ e pelo

método de krigagem. Assim como no modelo Gaussiano, a predição realizada por krigagem apresenta melhores resultados na situação não preferencial e a distribuição preditiva nas situações de presença de amostragem preferencial. A Tabela 5.4 apresenta as medidas de qualidade de predição para o modelo t-Student com 3 gl e a Figura 5.7 mostra os gráficos de valores preditos versus valores verdadeiros do processo S considerando os dois métodos de predição. Novamente notamos que o método de krigagem não recupera a cauda inferior do processo S , apresentando maiores medidas de erro quando comparadas às da distribuição preditiva. Vale ressaltar que o método de distribuição preditiva recupera apenas uma parte da cauda inferior do processo S devido a falta de informação dos menores valores que a amostra provê.

Tabela 5.4: Medidas de qualidade de predição para o modelo t-Student com 3 gl.

β	Método	EMA	REQMP
0	Dist. Preditiva	0,631	0,817
	Krigagem	0,563	0,736
1	Dist. Preditiva	0,615	0,795
	Krigagem	0,743	0,968
2	Dist. Preditiva	0,763	1,015
	Krigagem	1,045	1,323

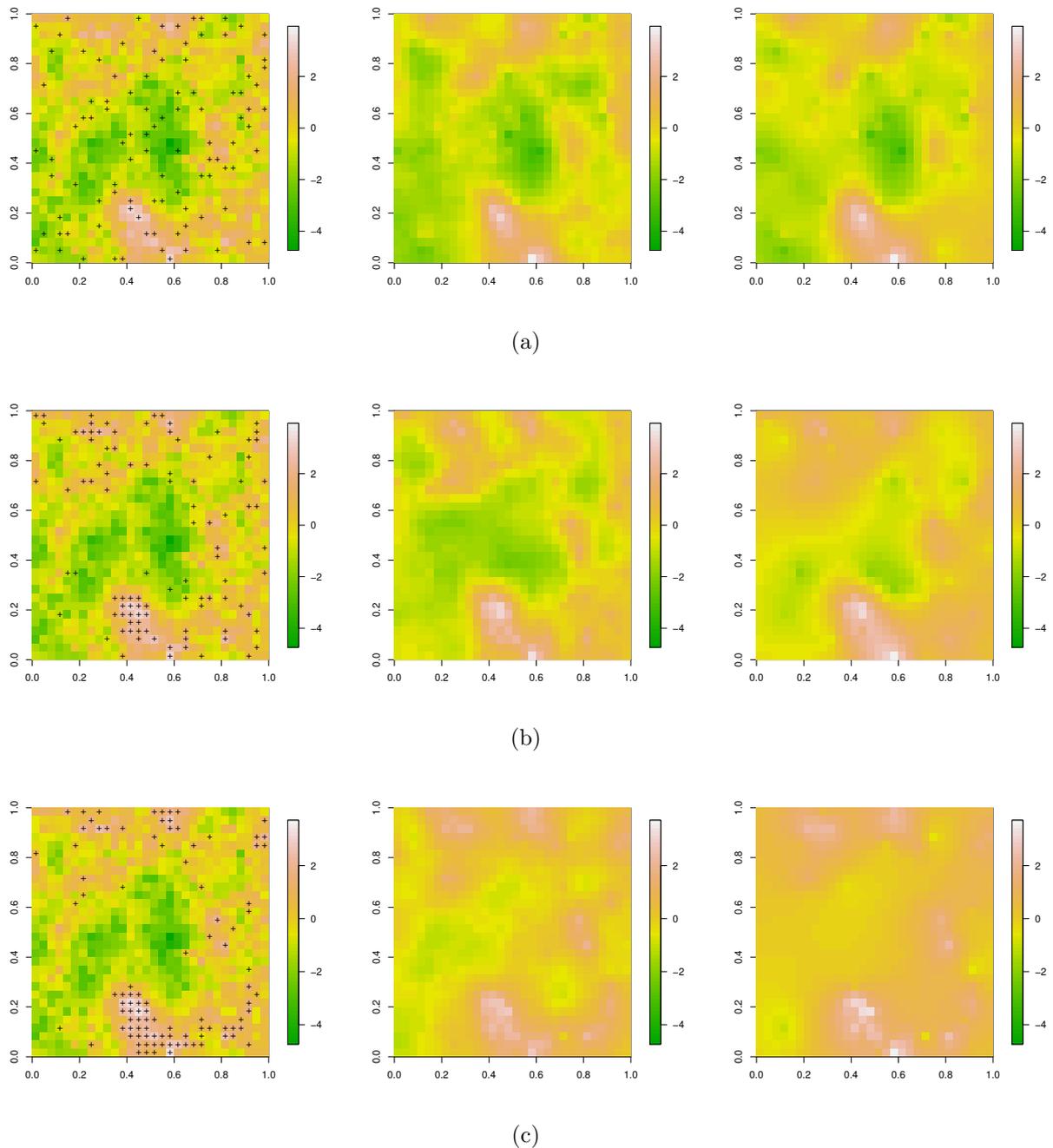


Figura 5.6: Mapas do processo S t -Student com 3 gl simulado e preditos pela distribuição de $S|\mathbf{x}, \mathbf{y}$ e por krigagem considerando amostras com (a) $\beta = 0$, (b) $\beta = 1$ e (c) $\beta = 2$.

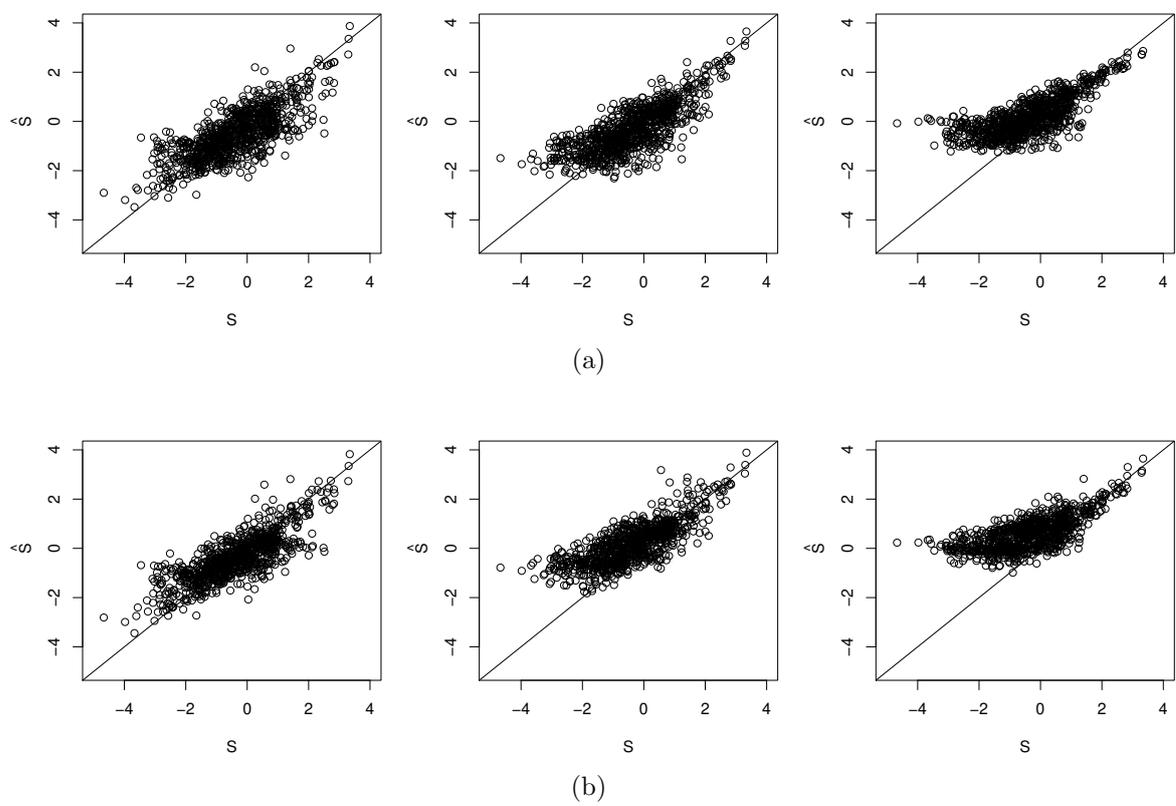


Figura 5.7: Gráfico de valores preditos versus valores verdadeiros do processo t-Student S com 3 gl considerando (a) a distribuição preditiva $S|\mathbf{x}, \mathbf{y}$ e (b) o método de krigagem.

5.2 Efeito da blocagem na estimação dos parâmetros do modelo

Ferreira and Gamerman (2015) utilizaram blocos de tamanho 1 na geração de amostras de $S|\mathbf{x}, \mathbf{y}$, ou seja, geraram elemento por elemento do vetor \mathbf{S} do processo Gaussiano. Em nosso trabalho, consideramos blocos de tamanho maior a fim de melhorar o desempenho computacional e, dessa forma, uma análise do efeito do tamanho dos blocos na estimação dos parâmetros do modelo deve ser considerada.

Foi realizado um estudo Monte Carlo no qual amostras do processo foram simuladas no quadrado unitário utilizando uma malha 50×50 , totalizando 2500 células. Os parâmetros considerados foram $\mu = 4$, $\tau^2 = 0,10$, $\sigma^2 = 1,5$, $\phi = 0,15$ e $\beta = 2$. Para a estimação dos parâmetros, foi considerada malha de tamanho 15×15 totalizando 225 células. Foram simuladas 50 amostras e, apesar de pequeno, essa quantidade foi escolhida devido ao tempo computacional para realizar todas as estimações.

A Figura 5.8 apresenta os gráficos de boxplot das estimativas dos parâmetros do modelo Gaussiano por MCEM para diferentes tamanhos de simulações (L) da distribuição preditiva de $S|\mathbf{x}, \mathbf{y}$ e diferentes tamanhos de bloco considerando tamanho de malha igual a 225. Observamos resultados semelhantes entre o número de simulações da distribuição preditiva e também entre os blocos. A diferença na estimação dos parâmetros para blocos diferentes da distribuição preditiva está no tamanho do *burn in* necessário para a convergência da cadeia de \mathbf{S} , como visto na seção anterior. Para os blocos 1 e 5, foi utilizado um *burn in* de tamanho 150, ao passo que para o tamanho de bloco 25, esse tamanho foi de 500. Por isso, notamos através da Tabela 5.5 que a redução do tempo computacional não foi tão grande entre os blocos 5 e 25.

Por proporcionar resultados semelhantes entre os blocos e pela redução do tempo computacional, escolhemos trabalhar com tamanhos de bloco iguais a 25 para realizar a estimação.

Tabela 5.5: Tempo médio (em minutos) para estimação dos parâmetros do modelo Gaussiano considerando tamanho de malha 225.

$L \setminus G$	1	5	25
1	5,9	2,5	1,7
10	6,4	2,1	1,8
20	8,2	2,8	1,8
50	7,9	2,7	2,1

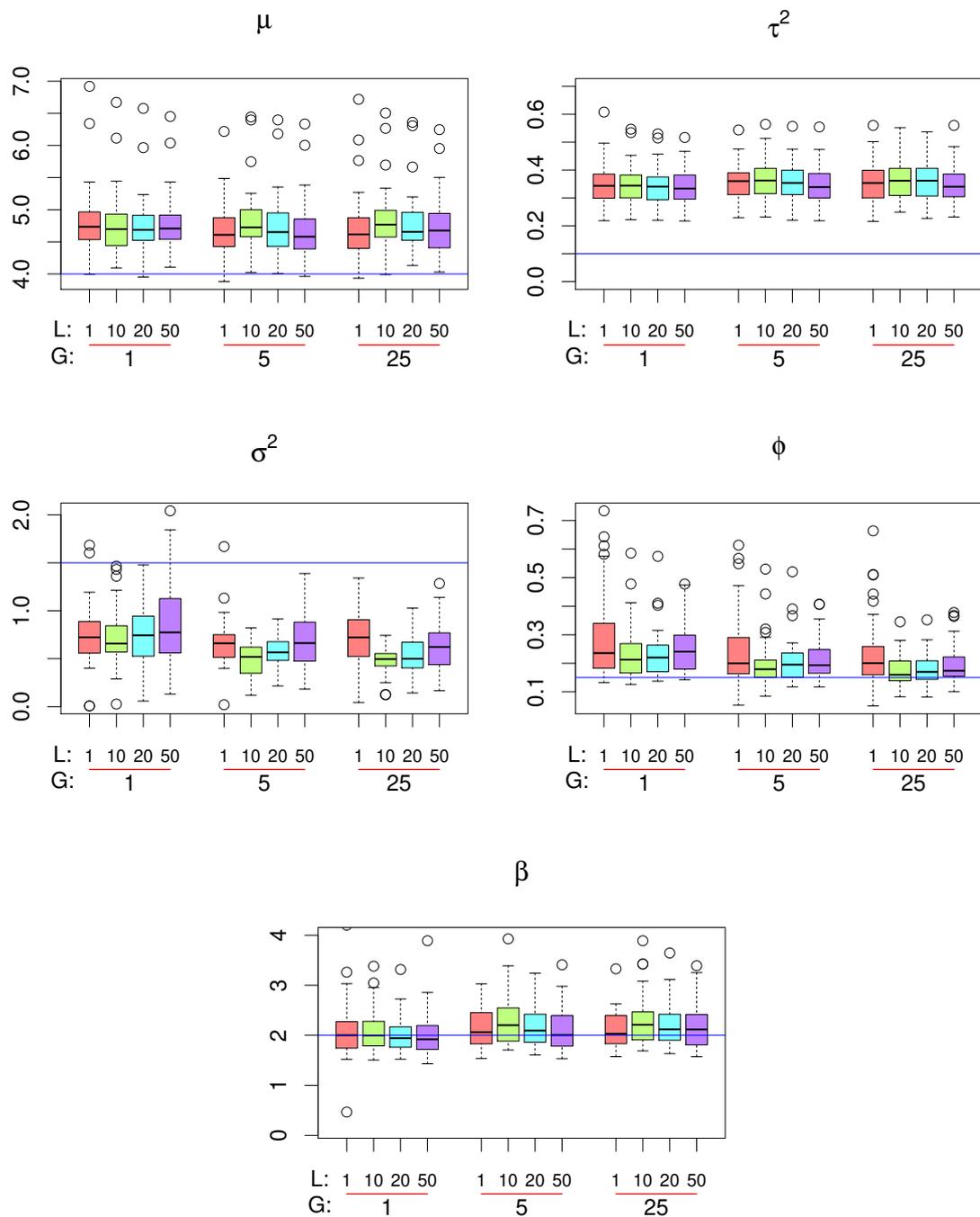


Figura 5.8: Gráficos de boxplot de 50 estimativas dos parâmetros do modelo pelo MCEM considerando uma malha regular de tamanho 225 e o modelo Gaussiano. Alguns pontos foram omitidos do gráfico do parâmetro β cujos valores estavam abaixo de 0 ou maior que 4, correspondendo a, no máximo, 4% em relação às 50 estimativas.

5.3 Efeito de fixar o parâmetro de correlação

Zhang and Wang (2010) mostram empiricamente que não se perde qualidade nas estimativas dos parâmetros do modelo tradicional Gaussiano ao fixar ϕ , ganhando tempo

computacional e evitando possíveis problemas numéricos, uma vez que ϕ seria estimado numericamente. Nesta seção, investigamos o efeito de fixar o parâmetro de correlação no modelo geoestatístico com amostragem preferencial.

A estimação foi realizada de duas formas: a primeira estimando o parâmetro ϕ e a segunda considerando um valor fixo para ϕ . Para a segunda forma de estimação, assumimos os valores 0,05, 0,15 (verdadeiro), 0,5 e 1,0. Dessa forma, analisamos as estimativas dos parâmetros quando subestimamos e superestimamos o valor do parâmetro de correlação.

Realizamos 100 iterações do algoritmo MCEM, considerando uma malha regular de tamanho 15×15 , um tamanho de *burn in* 500 para simulação da distribuição preditiva mais 20 simulações. Foram realizadas 50 estimativas, sendo que as estimativas dos parâmetros foram calculadas pela média das últimas 50 iterações do algoritmo e a Figura 5.9 apresenta o gráfico de boxplot das estimativas. Observamos que fixar o valor de ϕ não altera muito

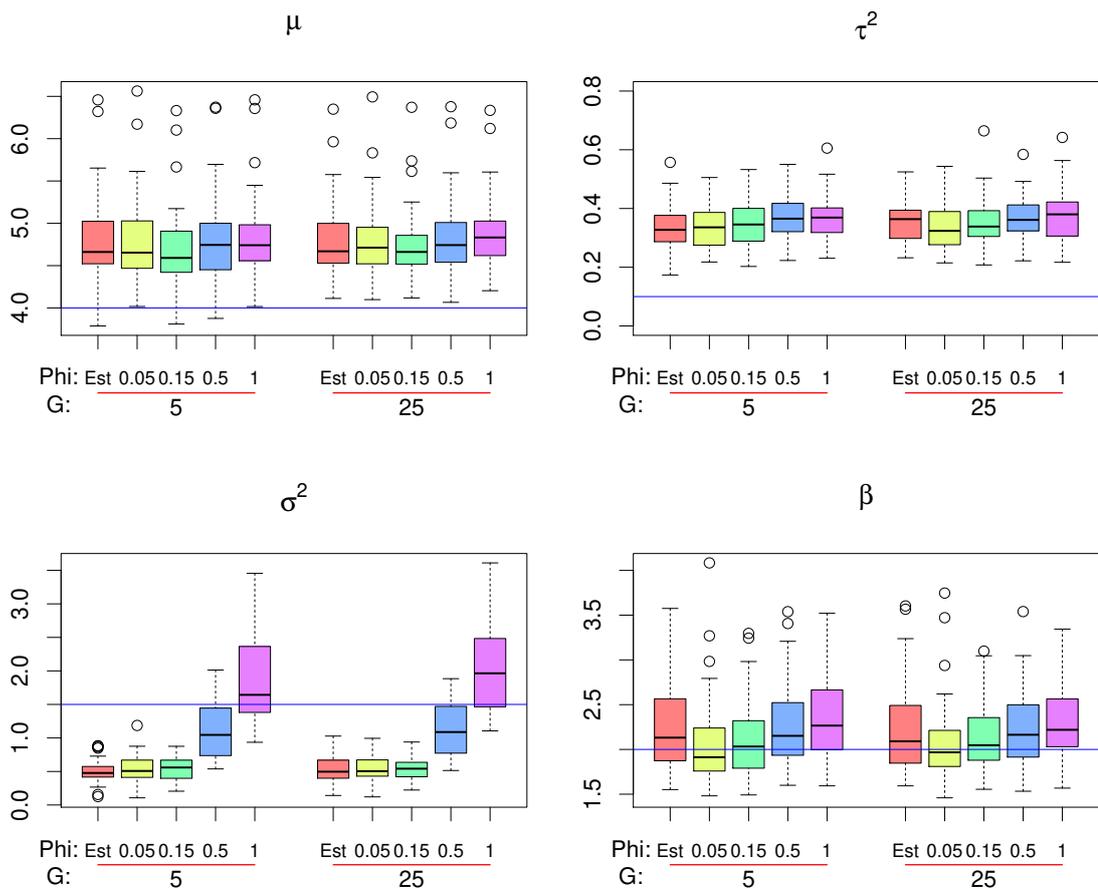


Figura 5.9: Gráficos de boxplot de 50 estimativas dos parâmetros do modelo pelo MCEM considerando uma malha regular de tamanho 225 e o modelo Gaussiano. O parâmetro ϕ foi fixado nos valores 0,05, 0,15 (verdadeiro), 0,5 e 1.

as estimativas dos parâmetros considerando ambos tamanhos de bloco. Analisando o gráfico para σ^2 , percebemos, empiricamente, que nosso modelo não estima os parâmetros σ^2 e ϕ de forma consistente, corroborando com o resultado de Zhang (2004) para processos Gaussianos: a estimativa de σ^2 depende do valor de ϕ .

A Figura 5.10 apresenta o gráfico de boxplot das estimativas da razão σ^2/ϕ e, exceto para o valor fixo subestimado (0,05), os resultados são semelhantes ao resultado obtido quando estimamos ϕ . Em ambos os casos de estimar ou fixar o parâmetro de

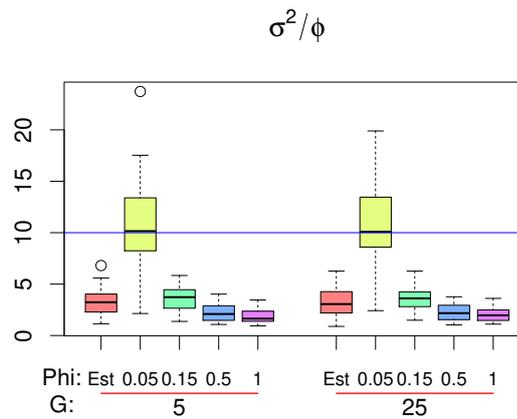


Figura 5.10: Gráficos de boxplot de 50 estimativas de σ^2/ϕ pelo MCEM considerando uma malha regular de tamanho 225 e o modelo Gaussiano. O parâmetro ϕ foi fixado nos valores 0,05, 0,15 (verdadeiro), 0,5 e 1.

correlação, a razão σ^2/ϕ é subestimada e, para o caso do parâmetro ϕ fixado, isso ocorre porque o modelo não consegue estimar σ^2 no ponto que corresponde ao valor verdadeiro da razão de σ^2/ϕ , que no caso é igual a 10. Por exemplo, para $\phi = 1$, as estimativas de σ^2 deveriam estar em torno de 10, porém os resultados estão bem abaixo desse valor. A justificativa se dá pelo fato de que, no modelo, temos outras relações de σ^2 além de ϕ . Comparar os dois casos (parâmetro de correlação fixo ou não) nos dá uma noção de como o modelo se comporta, pois a razão σ^2/ϕ foi subestimada em ambos.

Considerando o método SAEM com as últimas 50 iterações com memória, a Figura 5.11 apresenta a mesma análise e observamos que os resultados obtidos são semelhantes aos do método MCEM.

A Tabela 5.6 apresenta o tempo médio em minutos da estimação dos parâmetros do modelo.

Em relação ao MCEM, observamos que há uma redução considerável de mais de 90% quando consideramos tamanho de bloco igual a 5 e fixamos o parâmetro de

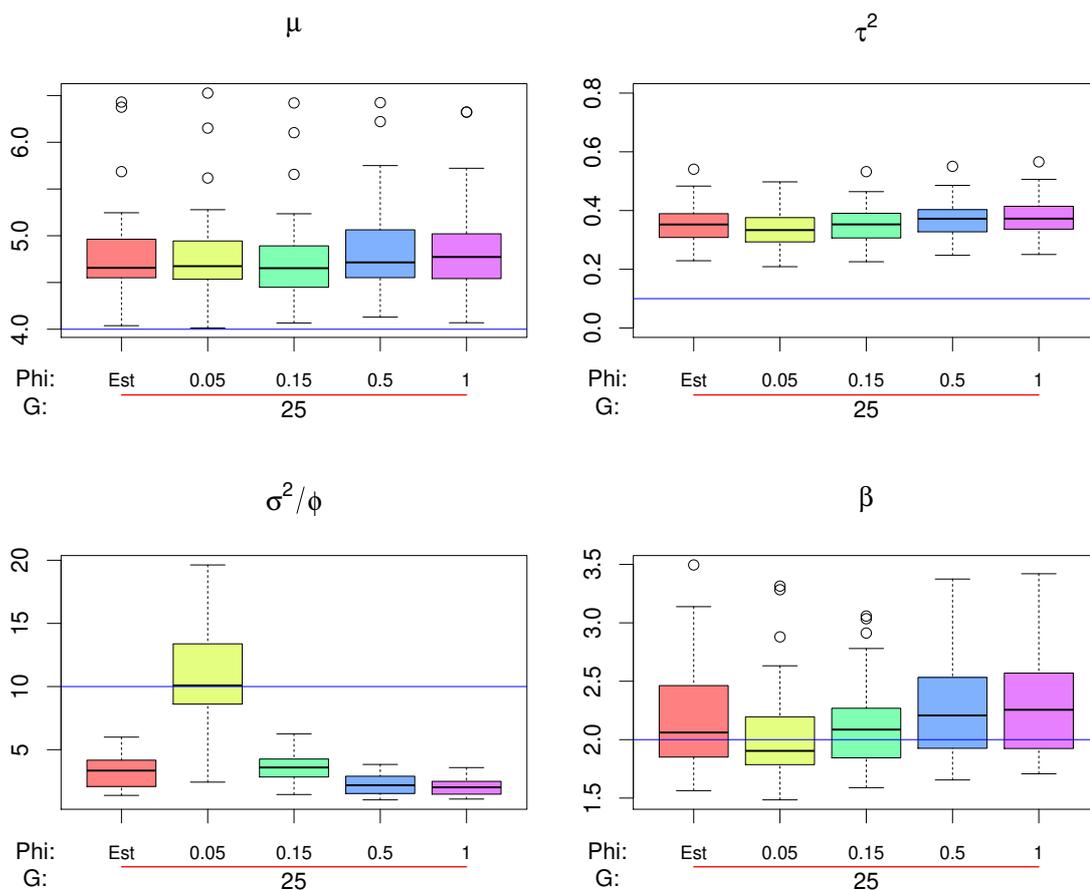


Figura 5.11: Gráficos de boxplot de 50 estimativas dos parâmetros do modelo pelo SAEM considerando uma malha regular de tamanho 225 e o modelo Gaussiano. O parâmetro ϕ foi fixado nos valores 0,05, 0,15 (verdadeiro), 0,5 e 1.

correlação e de cerca de 60% para tamanho de bloco igual a 25. Analisando o tempo para o método SAEM, notamos uma redução de 97% do tempo necessário para a estimação dos parâmetros. Vale ressaltar que, no caso SAEM, a forma recursiva da função Q é levada ao processo de otimização, o que significa que o tempo cresce de forma não linear quanto maior o número de iterações com memória.

Tabela 5.6: Tempo médio (em minutos) para estimação dos parâmetros do modelo Gaussiano.

Método	Bloco	ϕ				
		Estimado	0,05	0,15	0,5	1
MCEM	5	13,89	1,16	1,16	1,16	1,17
	25	1,82	0,75	0,75	0,75	0,79
SAEM	25	32,20	0,85	0,84	0,84	0,84

A Figura 5.12 apresenta a mesma análise para os dados gerados sem amostragem

preferencial e observamos resultados equivalentes.

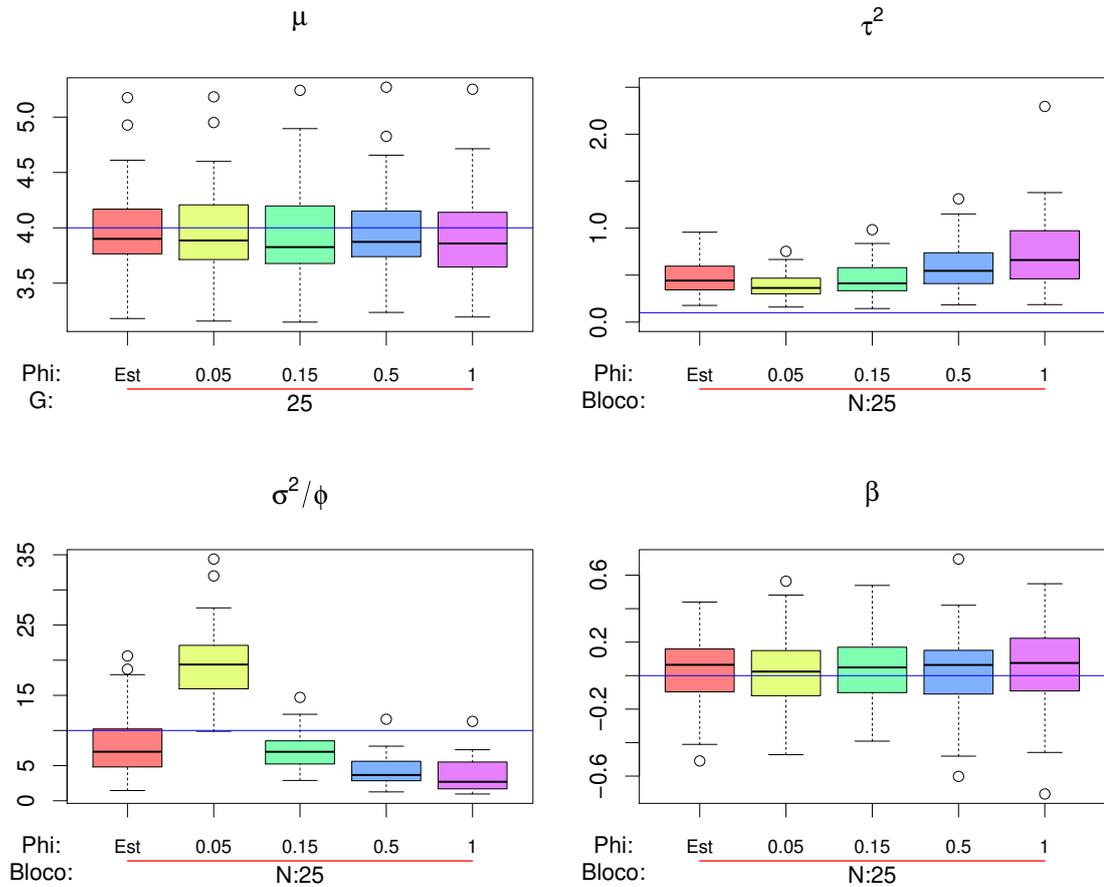


Figura 5.12: Gráficos de boxplot de 50 estimativas dos parâmetros do modelo pelo MCEM considerando uma malha regular de tamanho 225 e o modelo Gaussiano sem preferencialidade. O parâmetro ϕ foi fixado nos valores 0,05, 0,15 (verdadeiro), 0,5 e 1.

5.4 Estimação dos parâmetros do modelo

Apresentaremos agora a análise da qualidade de estimação dos parâmetros do modelo geoestatístico com amostragem preferencial pelo pacote TMB e através dos algoritmos MCEM e SAEM. Comparamos as estimativas obtidas por esses métodos com as obtidas pelo método de aproximação da função de verossimilhança por Monte Carlo (MCLA) presente em Diggle et al. (2010).

Para gerar uma amostra do processo Gaussiano, foi utilizada uma malha 50×50 no quadrado unitário, totalizando 2500 células. Nessa malha, foi gerada uma realização do processo Gaussiano S . A partir da amostra de S , uma realização do processo pontual X de tamanho 100 foi gerada e, dados S e X , uma amostra do processo Y foi simulada. Os parâmetros considerados para foram $\mu = 4$, $\tau^2 = 0,10$, $\sigma^2 = 1,5$, $\phi = 0,15$ e $\beta = 2$. Para o processo t-Student, a simulação foi realizada de forma análoga, considerando 3 graus de liberdade na geração do processo S . A partir da amostra gerada, foram construídas malhas regulares de tamanho 225 e 400 para estimação. As localizações dos valores observados foram substituídas pelas coordenadas do centróide mais próximo da malha utilizada, como pode ser observado na Figura 5.13. Foram, então, realizadas 200 estimações através dos métodos considerados.

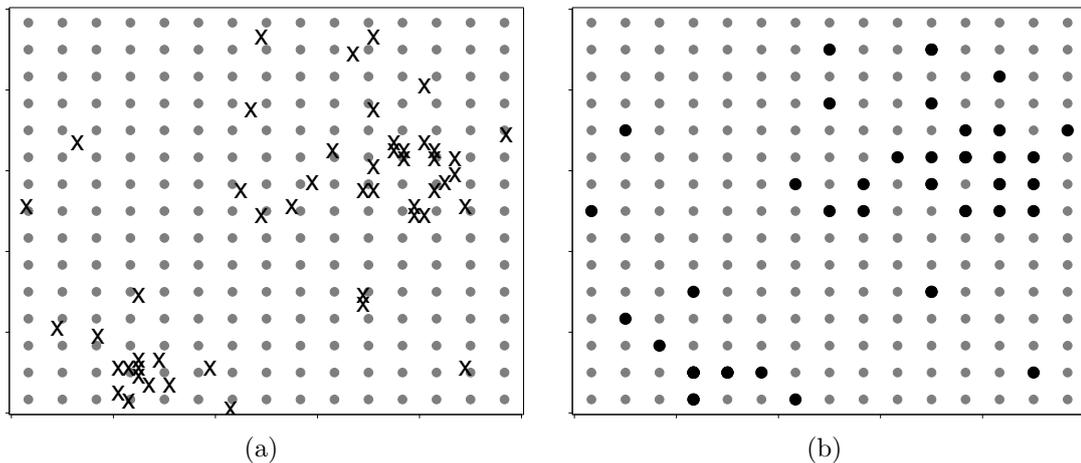


Figura 5.13: Exemplo da etapa de alocar os valores observados na malha considerada para estimação. (a) malha e amostra (225 pontos da malha mais 50 locais observados) e (b) 225 pontos da malha na qual os pontos destacados contêm informação de Y .

Modelo Gaussiano

Para comparar os métodos de estimação, a Figura 5.14 apresenta o gráfico de boxplot das estimativas dos parâmetros do modelo Gaussiano para os métodos descritos.

Notamos que considerar a informação do processo pontual na distribuição preditiva de $S|\mathbf{x}, \mathbf{y}$ melhora a estimativa do parâmetro de média, apresentando valores mais próximos do verdadeiro em comparação com o método MCLA. O efeito pepita foi superestimado nos métodos TMB e MCEM, mas apresentam melhores estimativas de σ^2 . O parâmetro de correlação e o parâmetro de preferenciabilidade foram melhores estimados pelo nosso método MCEM, evidenciando a qualidade do método.

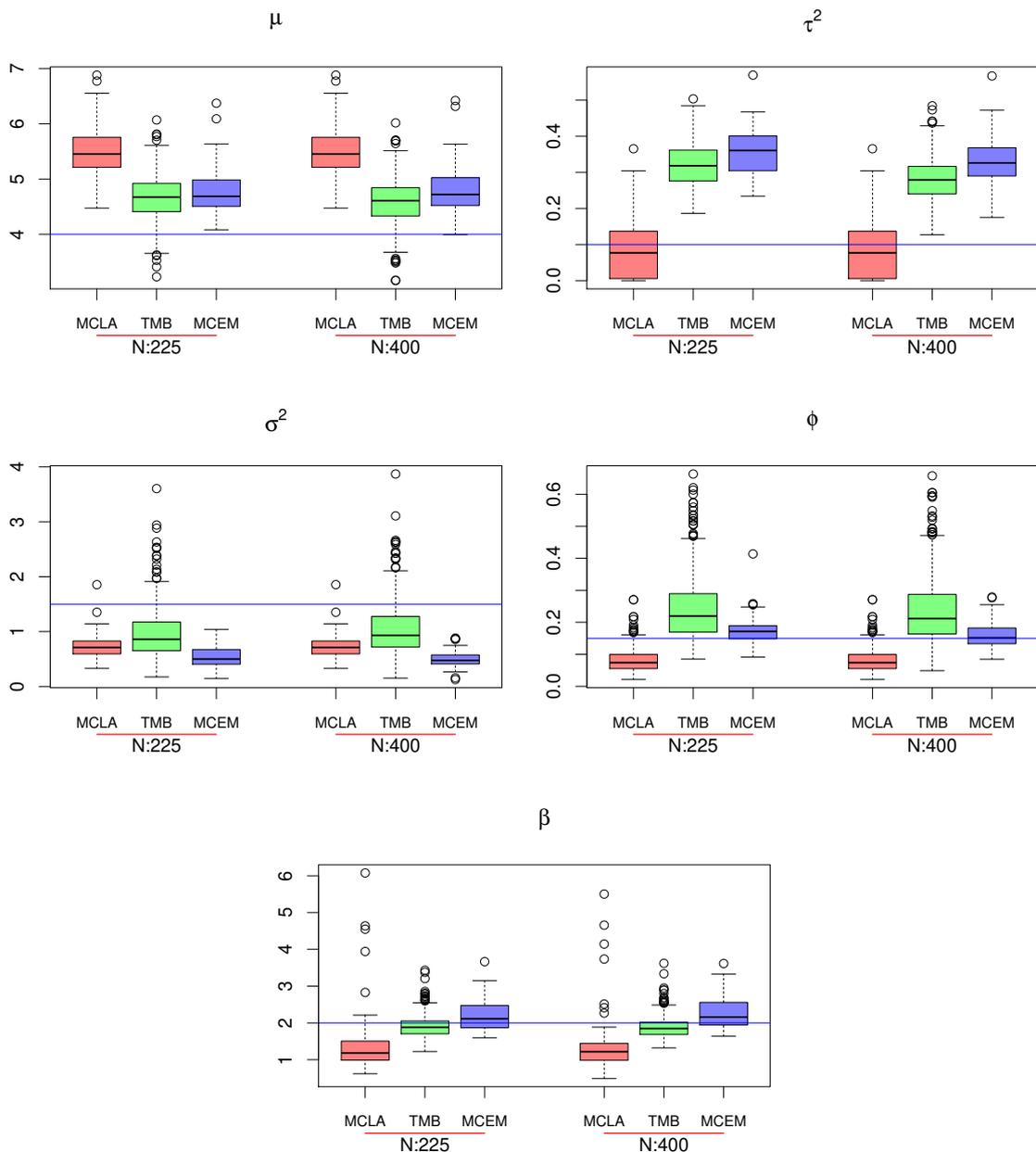


Figura 5.14: Gráficos de boxplot das estimativas dos parâmetros do modelo Gaussiano pelos métodos MCLA, TMB e MCEM considerando malhas de tamanho 225 e 400.

A Tabela 5.7 apresenta o tempo médio para estimação dos parâmetros do modelo

para cada método utilizado. Observamos que o método MCEM fornece resultados em tempo muito superior aos demais métodos para os dois tamanhos de malha, tornando-o menos atrativo em relação ao método TMB no quesito computacional.

Tabela 5.7: Tempo médio (em minutos) para estimação dos parâmetros do modelo Gaussiano.

N\Método	MCLA	TMB	MCEM
225	0,02	0,14	4,06
400	0,03	0,63	13,96

Como utilizamos a forma recursiva da função \hat{Q} na etapa de maximização numérica do método SAEM, esse apresenta tempo computacional muito superior aos demais métodos quando estimamos o parâmetro de correlação. Por exemplo, o tempo médio de execução do SAEM com 100 iterações, sendo as últimas 50 com memória, em uma grid de tamanho 225 foi de 32 minutos. Como devemos inverter a matriz de correlação quando estimamos ϕ e carregamos a forma recursiva de \hat{Q} na etapa de otimização, isso causa um grande aumento no tempo computacional, tornando o método pouco atrativo. Dessa forma, para comparar as estimativas obtidas por MCEM e SAEM, realizamos o estudo considerando o parâmetro ϕ fixo em seu valor verdadeiro (0,15). Consideramos 150 iterações, sendo as últimas 50 com memória para o SAEM. Como pode ser observado na Figura 5.15, os resultados são similares para ambos os modelos. O tempo médio de execução foi de 2,69 minutos para MCEM e 2,92 minutos para SAEM, evidenciando o ganho computacional que obtemos ao fixar ϕ . Segundo Zhang (2004), não há grandes impactos ao fixar o

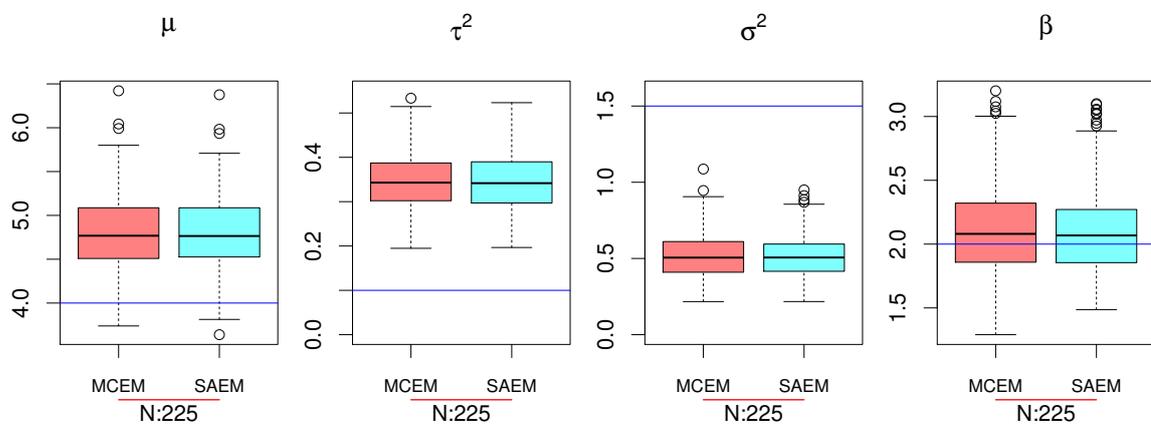


Figura 5.15: Gráficos de boxplot das estimativas dos parâmetros do modelo Gaussiano pelos métodos MCEM e SAEM, considerando uma grid de tamanho 225. O parâmetro de correlação foi fixado em seu valor verdadeiro (0,15).

parâmetro de correlação em um valor razoável (e maior que o valor verdadeiro) e, nesse cenário, o método SAEM se torna uma boa opção.

Modelo t-Student

Para o modelo t-Student, não é possível estimar os parâmetros pelo método MCLA. Em relação à estimação pelo pacote TMB, a diferenciação automática utiliza a memória RAM para armazenar o processo de derivação. No caso do modelo t-Student com amostragem preferencial, esse processo demanda muita memória, o que não ocorreu com o modelo Gaussiano. Em um computador com 28 GB de memória RAM, foi possível estimar os parâmetros do modelo com uma malha de tamanho máximo 400, sendo uma limitação desse método no nosso contexto.

Pela Figura 5.16, observamos que o efeito pepita é superestimado para todos os métodos. As estimativas do parâmetro μ foram melhores para o método MCEM, apresentando valores mais próximos do valor verdadeiro para ambos tamanhos de malha, ao passo que o método TMB apresentou estimativas subestimadas. Em relação ao parâmetro σ^2 , o método MCEM apresentou resultados próximos do real, ao passo que o método TMB superestima σ^2 e também apresenta maior variabilidade nas estimações, indicando maior incerteza ao estimar esse parâmetro. Assim como no modelo Gaussiano, o método MCEM se mostrou melhor para estimar os parâmetros de correlação e de preferenciabilidade no modelo t-Student.

A Tabela 5.8 apresenta o tempo médio para estimação dos parâmetros do modelo para cada método utilizado. Observamos que o método TMB apresentou, novamente, menor tempo médio para obtenção das estimativas dos parâmetros do modelo. Também é o que apresenta o menor aumento no tempo quando passamos de uma malha de tamanho 225 para uma de tamanho 400. Porém, a qualidade da estimação é melhor para os método MCEM, sendo mais atrativo mesmo apresentando tempo computacional maior.

Tabela 5.8: Tempo médio (em minutos) para estimação dos parâmetros do modelo t-Student com 3 gl.

N\Método	TMB	MCEM
225	0,35	3,40
400	2,18	15,94

Para comparar as estimativas obtidas por MCEM e SAEM, realizamos novamente o

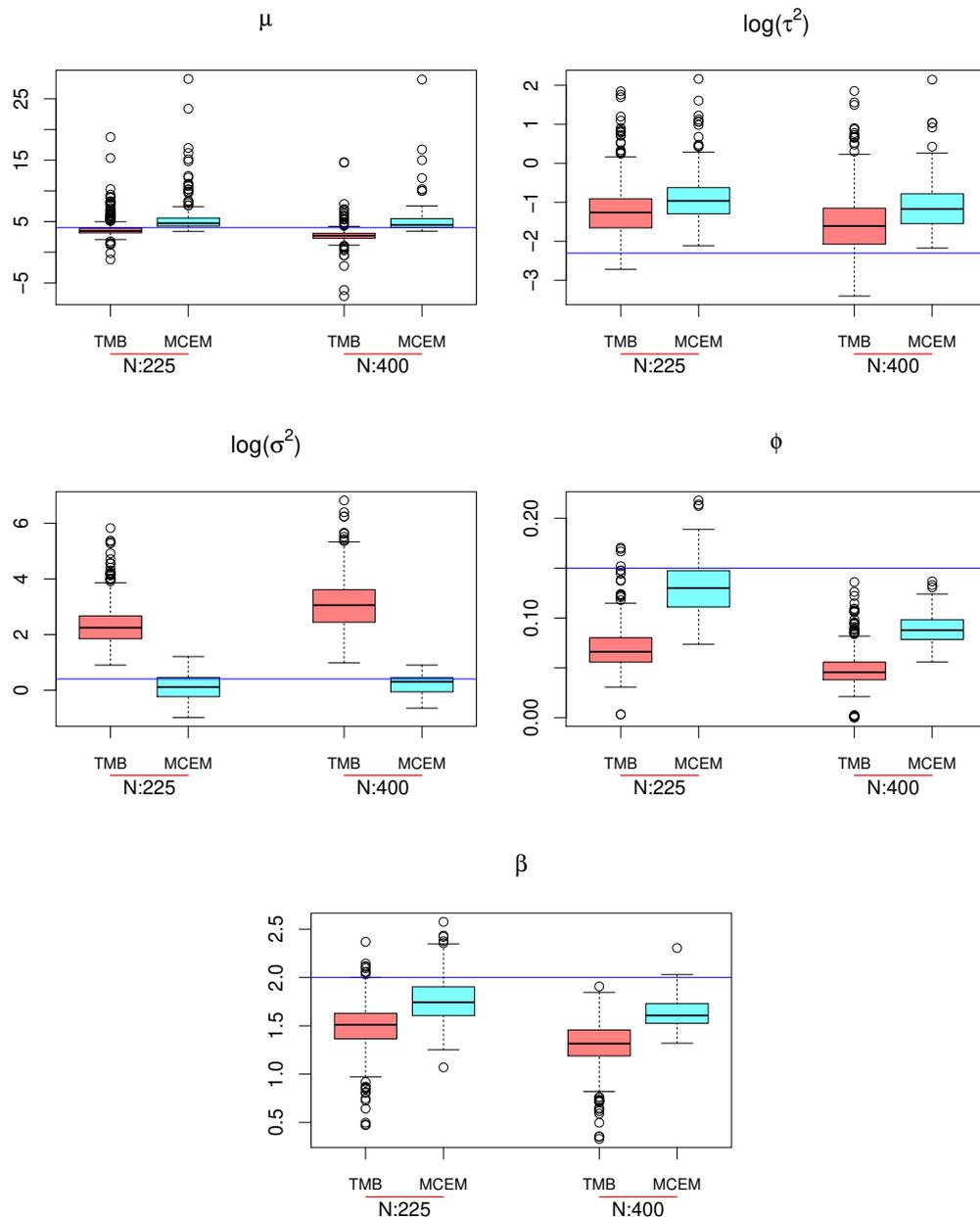


Figura 5.16: Gráficos de boxplot das estimativas dos parâmetros do modelo t-Student com 3 gl pelos métodos TMB e MCEM considerando malhas de tamanho 225 e 400.

estudo considerando o parâmetro ϕ fixo em seu valor verdadeiro. Como podemos observar na Figura 5.17, os resultados obtidos são similares para ambos métodos. O tempo médio de estimação para o MCEM foi de 0,88 minutos e para o SAEM foi de 2,09 minutos. Diferente do modelo Gaussiano, para o modelo t-Student o método SAEM apresenta tempo de execução muito maior que o MCEM (aproximadamente 2,5 vezes maior), o que é justificado pela estimação numérica de σ^2 . Além disso, notamos a grande redução de tempo quando fixamos o parâmetro de correlação, tornando o algoritmo mais eficiente.

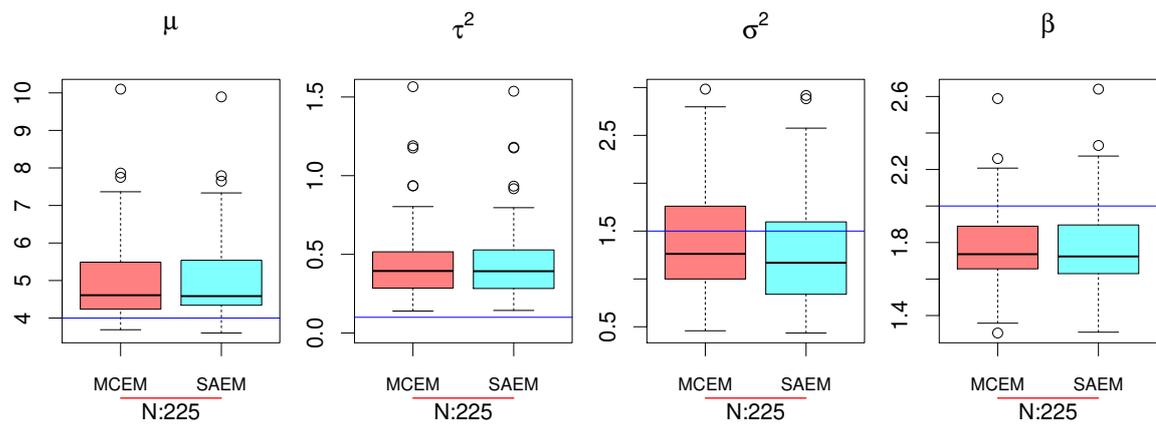


Figura 5.17: Gráficos de boxplot das estimativas dos parâmetros do modelo t-Student com 3 gl pelos métodos MCEM e SAEM, considerando uma malha de tamanho 225. O parâmetro de correlação foi fixado em seu valor verdadeiro (0,15).

5.5 Comparação dos métodos de predição

Após a estimação dos parâmetros, a próxima etapa de uma análise geoestatística é realizar a predição do processo de interesse em locais não amostrados, no caso o processo Y . Dessa forma, se faz necessário um estudo comparativo entre os métodos de predição utilizados nesse trabalho. O primeiro método é a krigagem. O segundo método é maximizar a distribuição conjunta $f(s, \mathbf{x}, \mathbf{y})$ em relação a S , ou seja, encontrar o valor correspondente à sua moda. Note que esse processo é equivalente a maximizar a distribuição preditiva de $S|\mathbf{x}, \mathbf{y}$, pois $f(s|\mathbf{x}, \mathbf{y}) \propto f(s, \mathbf{x}, \mathbf{y})$. O terceiro método é o algoritmo MH para gerar amostras da distribuição de S condicionado aos dados, em que o valor predito é calculado pela média dos valores gerados.

Os métodos utilizam os valores dos parâmetros estimados como os verdadeiros das distribuições e podemos realizar combinações de método de estimação e método de predição para verificar qual par fornecerá a melhor predição. Por exemplo, podemos estimar os parâmetros por aproximação de Laplace e utilizar o algoritmo MH para realizar a predição. Dessa forma, estudamos, para dez conjuntos de dados, as predições obtidas combinando os métodos de estimação MCLA, TMB e MCEM e os métodos de predição considerados, totalizando nove pares. Vale ressaltar que não realizamos o estudo para SAEM devido ao maior tempo computacional do algoritmo e pela similaridade de estimação com o MCEM. A Figura 5.18 apresenta o mapa simulado de Y para as amostras 1 e 2 considerando o modelo normal.

Os métodos de predição geram valores para o processo latente S e a predição para

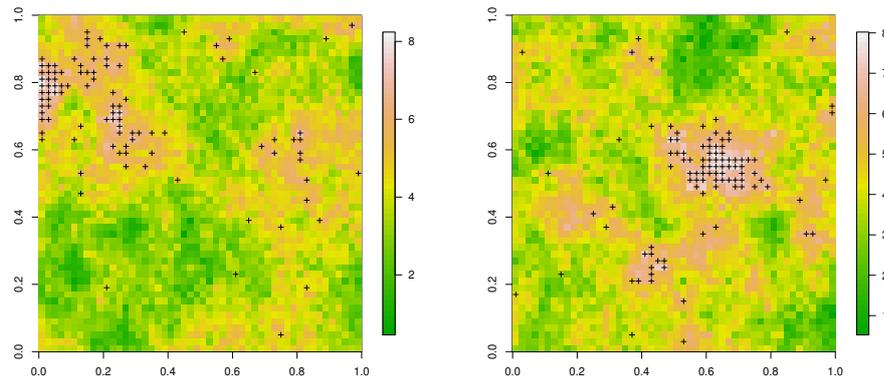


Figura 5.18: Mapas do processo Y simulado para as amostras 1 (esquerda) e 2 (direita) dos dados simulados com distribuição normal.

o processo Y é realizada por $Y_p|\mathbf{x}, \mathbf{s} \sim N(\hat{\mu} + \hat{S}(\mathbf{x}_p), \hat{\tau}^2)$, em que \mathbf{x}_p é um local não observado. As Figuras 5.19 e 5.20 apresentam os valores de EMA e REQMP para as predições realizadas por cada combinação para dados gerados com distribuição normal sobre uma malha 20×20 . Podemos observar que o método de krigagem fornece piores

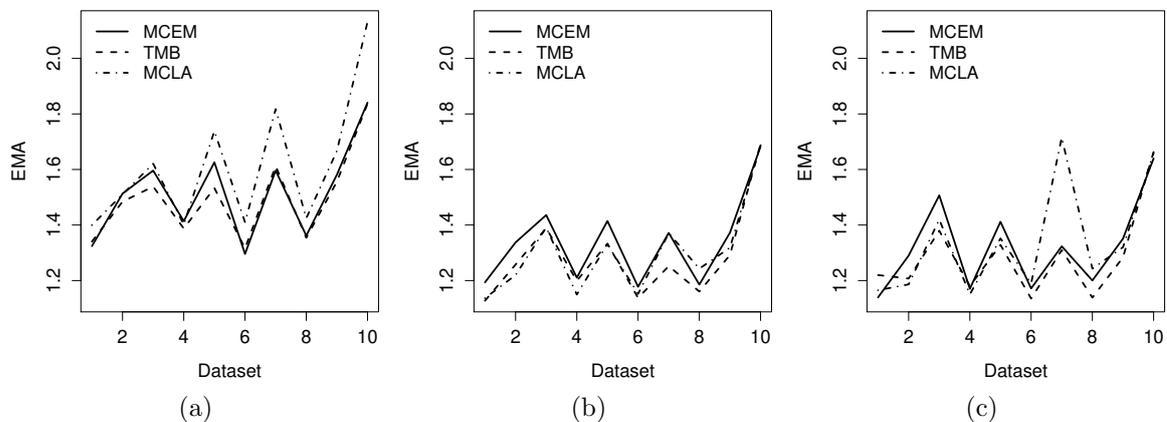


Figura 5.19: Gráficos de EMA para 10 conjuntos de dados simulados com distribuição normal para predições realizadas por (a) krigagem, (b) moda da distribuição preditiva e (c) algoritmo MH.

medidas de qualidade de predição, mesmo considerando os parâmetros estimados por MCEM e TMB. A predição pela moda da distribuição preditiva e pelo algoritmo MH produzem medidas de qualidade semelhantes. Para os dez conjuntos de dados, 7 foram melhor preditos considerando os valores estimados dos parâmetros por TMB. Desses, 4 foram preditos pela moda de S e 3 por MH.

Os mapas de predição de Y para a amostra 1 podem ser observados na Figura

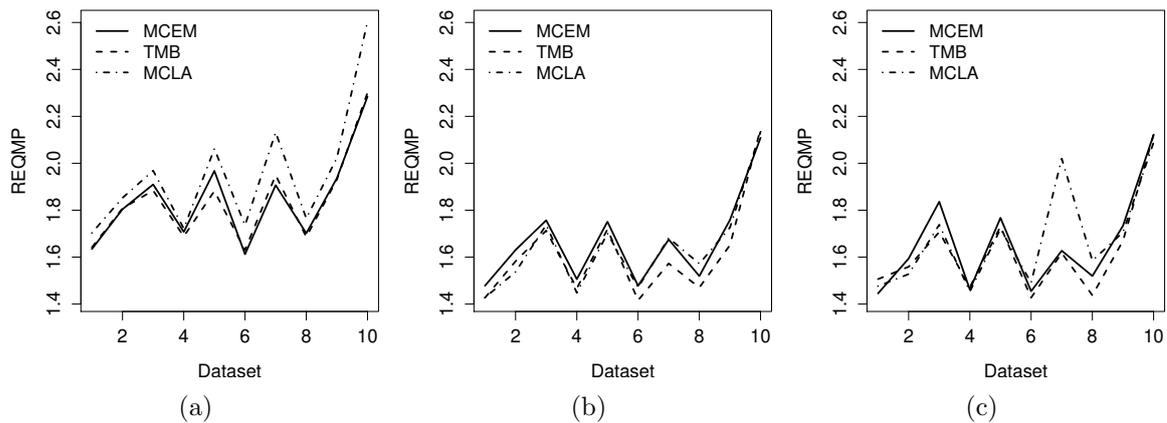


Figura 5.20: Gráficos de REQMP para 10 conjuntos de dados simulados com distribuição normal para predições realizadas por (a) krigagem, (b) moda da distribuição preditiva e (c) algoritmo MH.

5.21. Como indica as medidas de qualidade de predição, os valores do processo não são bem estimados quando utilizamos a krigagem, independente da estimativa do vetor de parâmetros utilizado. Quando consideramos realizar predições que consideram a informação do processo pontual, obtemos melhores estimativas, principalmente em áreas com pouca informação. Analisando a Figura 5.22, observamos que a predição realizada por MH apresenta menores erros de predição, indicando que é um método mais preciso.

Em relação ao modelo t-Student, os mapas simulados para as amostras 1 e 2 podem ser observados na Figura 5.23. As medidas de qualidade de predição são apresentadas nas Figuras 5.24 e 5.25. Notamos que as medidas com menores vieses são associadas, em maioria, à combinação de estimação de parâmetros por TMB e moda de S para predição, com 5 e 4 das melhores medidas por EMA e REQMP, respectivamente. Porém, essas são medidas que não levam em conta a incerteza dos valores preditos. Como visto na seção anterior, o parâmetro σ^2 é superestimado pelo TMB, o que aumenta a incerteza das predições.

As Figuras 5.26 e 5.27 apresentam os mapas dos valores preditos e erro padrão de predição para o processo Y . Podemos observar o efeito da superestimação de σ^2 na incerteza dos valores preditos nos mapas associados ao método de estimação TMB. Os baixos valores preditos por esse método em regiões de pouca informação estão associados a subestimação da média e não, necessariamente, em uma predição mais precisa. Os valores dos parâmetros estimados para a amostra 1 são apresentados na Tabela 5.9, na qual podemos observar a magnitude do erro padrão da estimativa de σ^2 . Dessa forma, uma

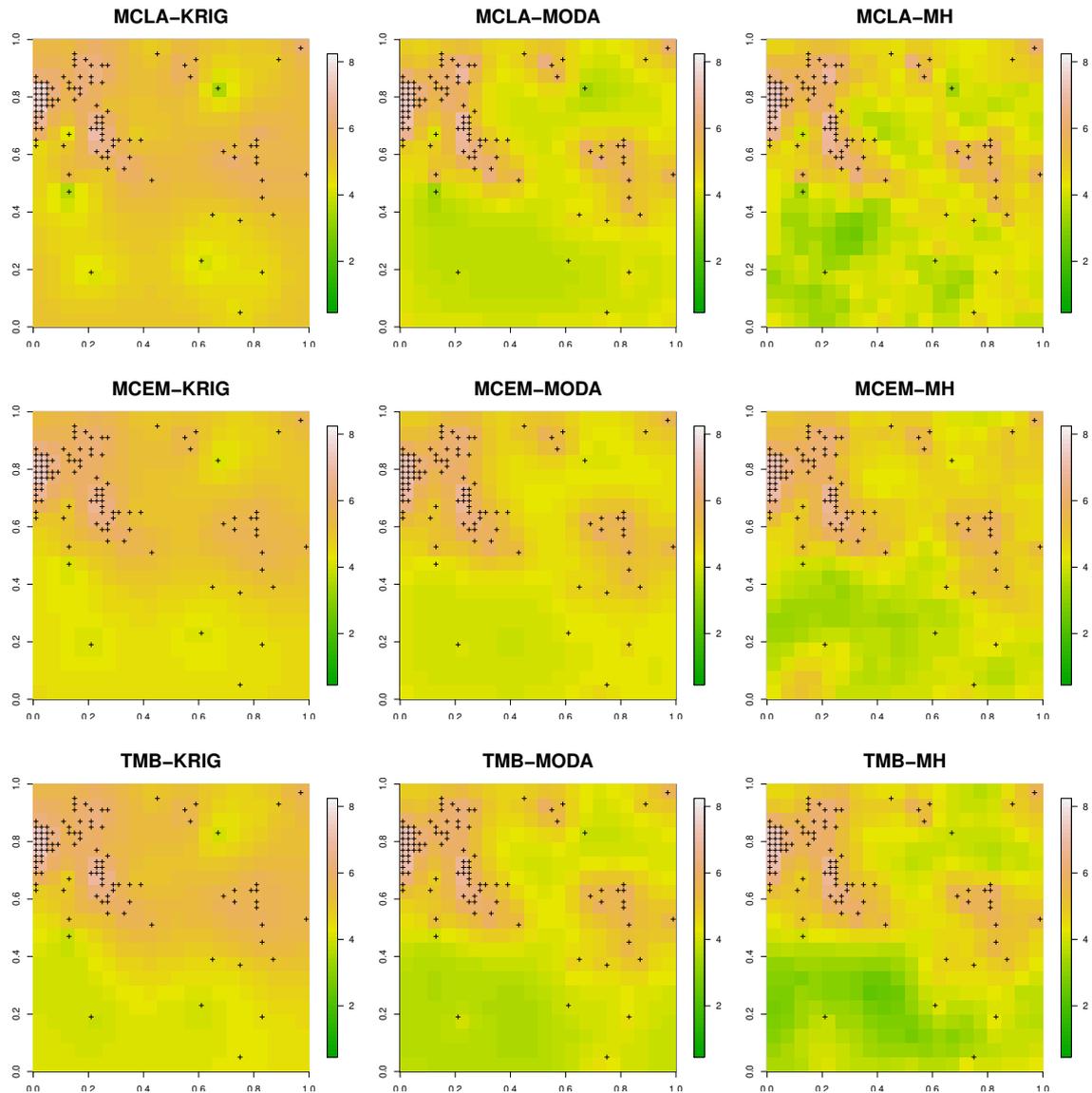


Figura 5.21: Mapas de predição do processo Y para a amostra 1 dos dados simulados com distribuição normal considerando as combinações de métodos de estimação e de predição.

Tabela 5.9: Parâmetros estimados do modelo t-Student com 3 gl para a amostra 1.

Parâmetro	Real	MCEM	TMB
μ	4,00	4,632 (0,135)	1,911 (0,896)
τ^2	0,10	0,525 (0,043)	0,371 (0,058)
σ^2	1,50	1,984 (0,283)	37,314 (22,908)
ϕ	0,15	0,087 (0,013)	0,053 (0,012)
β	2,00	1,631 (0,087)	1,164 (0,158)

predição mais confiável está associada ao método de estimação MCEM. Observe também que não há informação de valores baixos do processo na amostra, o que, naturalmente,

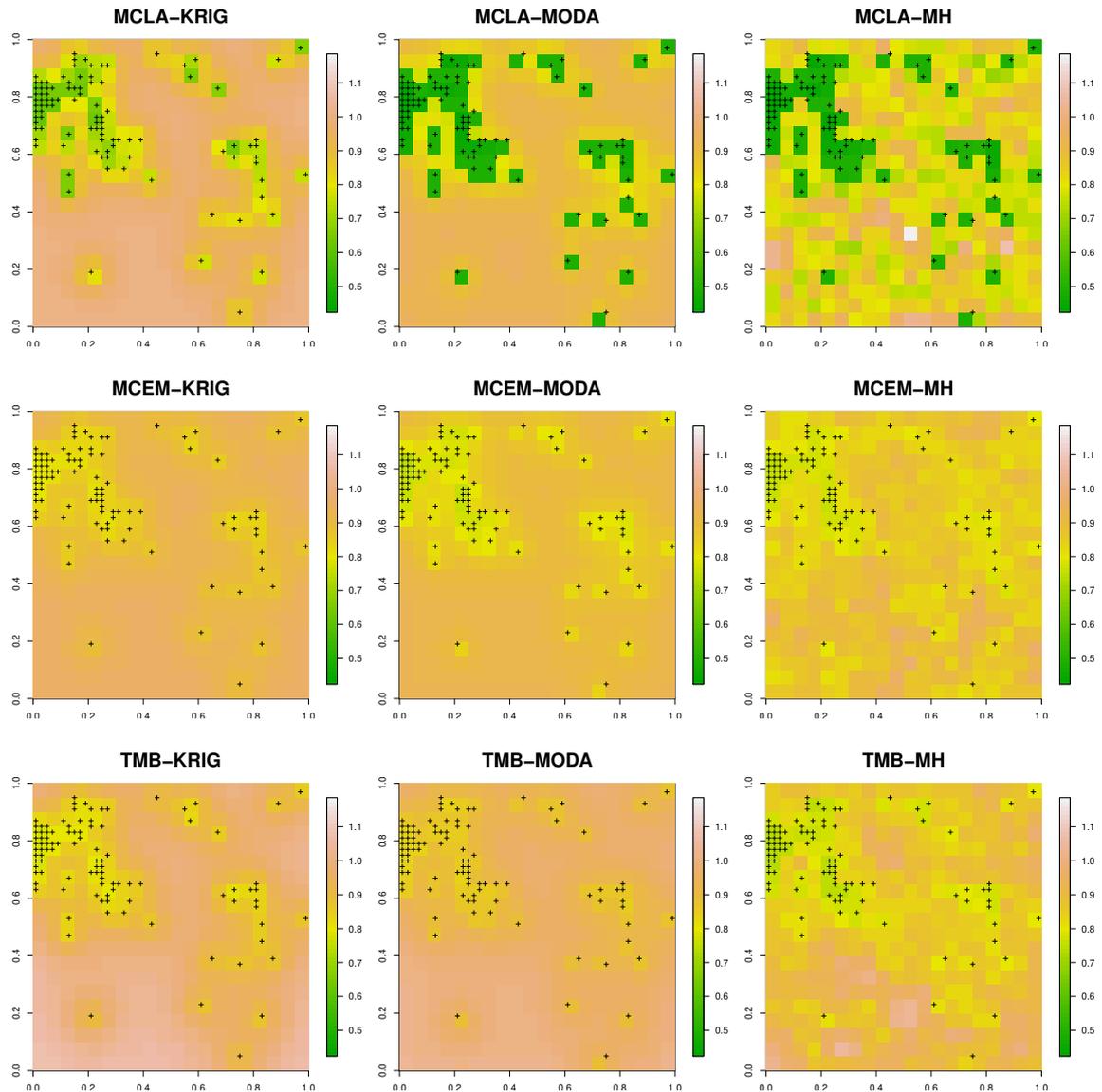


Figura 5.22: Mapas de erro de predição do processo Y para a amostra 1 dos dados simulados com distribuição normal considerando as combinações de métodos de estimação e de predição.

reflete nos valores preditos do processo, em que não conseguimos recuperar os menores valores de Y .

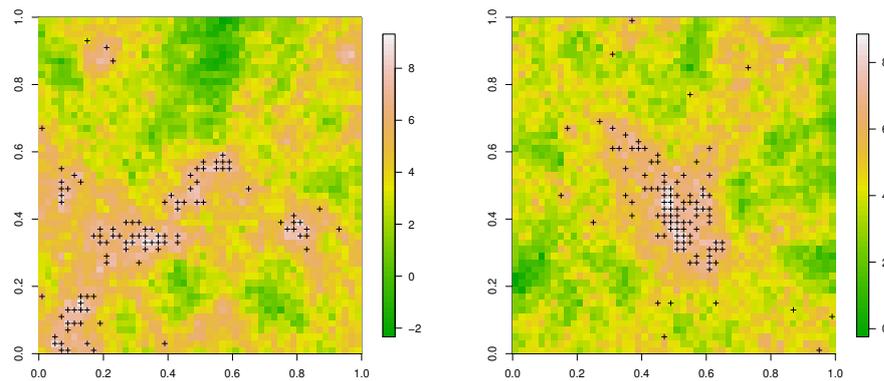


Figura 5.23: Mapas do processo Y simulado para as amostras 1 (esquerda) e 2 (direita) dos dados simulados com distribuição t com 3 gl.

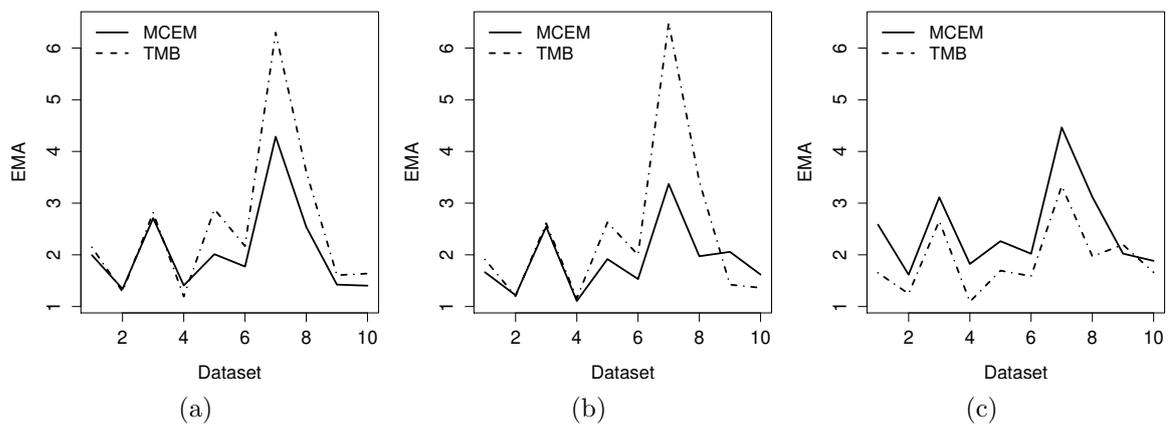


Figura 5.24: Gráficos de EMA para 10 conjuntos de dados para previsões realizadas por (a) krigagem, (b) moda da distribuição preditiva e (c) algoritmo MH.

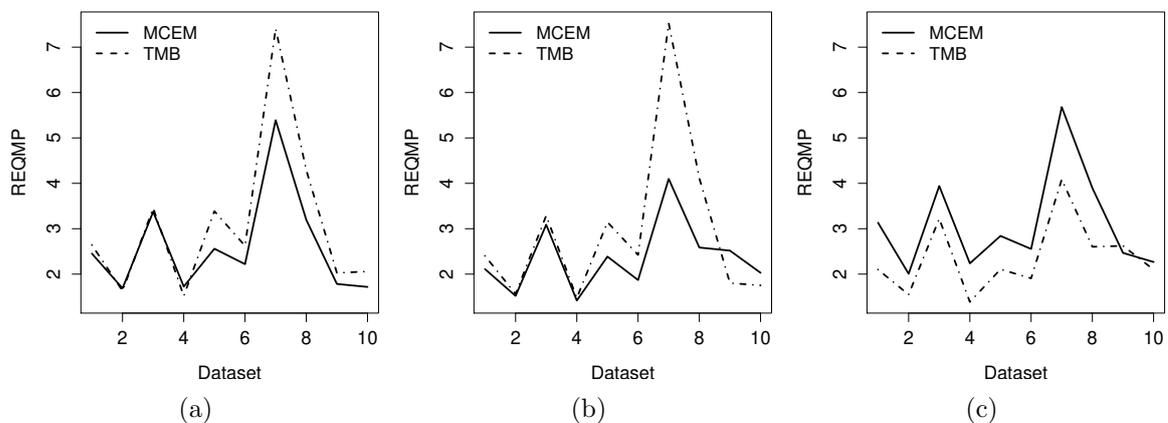


Figura 5.25: Gráficos de REQMP para 10 conjuntos de dados para previsões realizadas por (a) krigagem, (b) moda da distribuição preditiva e (c) algoritmo MH.

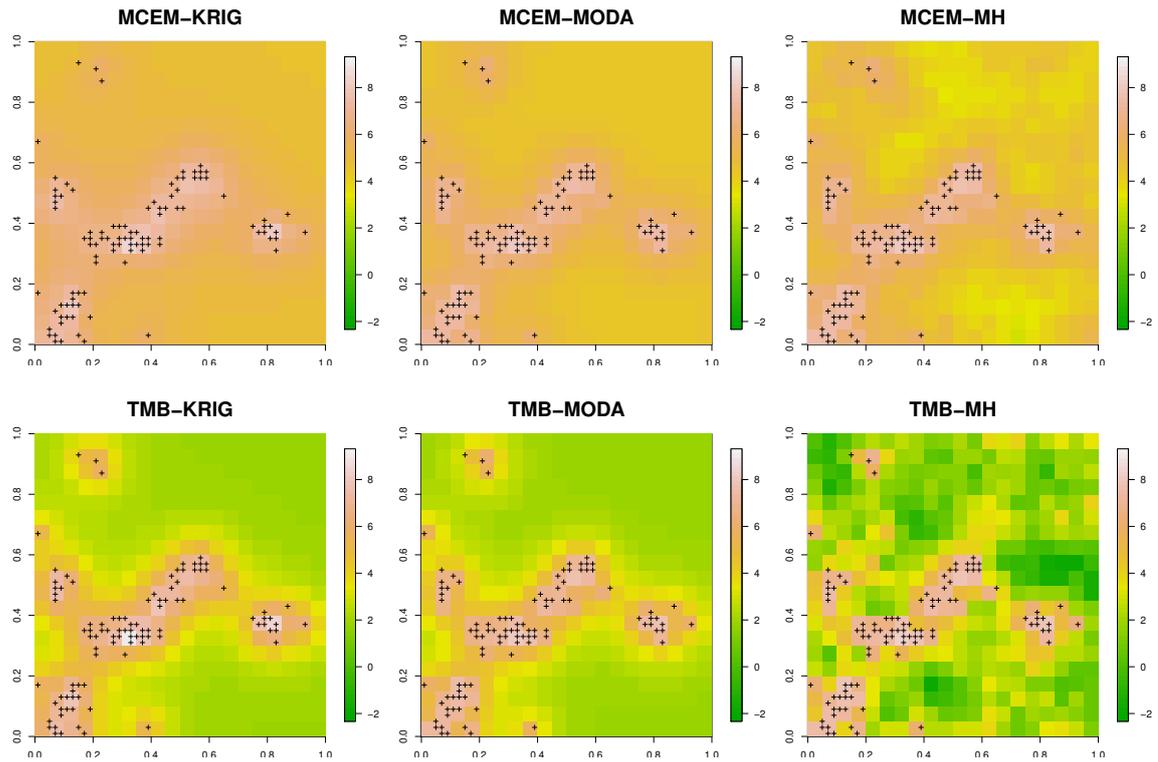


Figura 5.26: Mapas de predição do processo Y para a amostra 1 dos dados simulados com distribuição t com 3 gl considerando as combinações de métodos de estimação e de predição.

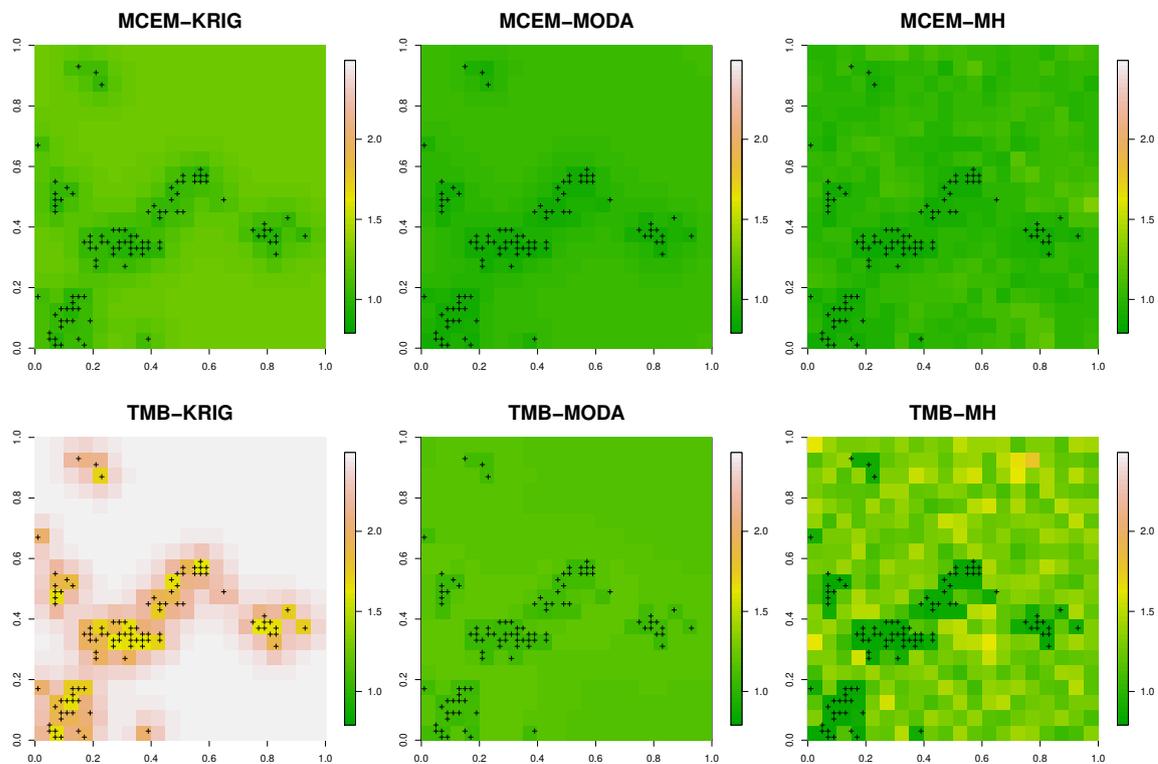


Figura 5.27: Mapas de erro de predição do processo Y para a amostra 1 dos dados simulados com distribuição t com 3 gl considerando as combinações de métodos de estimação e de predição.

Capítulo 6

Inferência Bayesiana - Modelagem e computação

Sob o enfoque Bayesiano, a inferência é realizada a partir da distribuição a posteriori dos parâmetros do modelo. Ferreira and Gamerman (2015) realizam a inferência do modelo geoestatístico com amostragem preferencial recorrendo à discretização da região S , pois a função de verossimilhança não possui forma analítica. Gonçalves and Gamerman (2018) apresentam uma inferência exata para o processo de Cox, utilizando como função de intensidade a função probito, o que possibilita também a geração exata do processo de pontual.

Na Seção 6.1 apresentamos o modelo geoestatístico Bayesiano com amostragem preferencial e como realizamos inferência exata, desenvolvemos um algoritmo MCMC para a estimação dos parâmetros do modelo e mostramos como realizar previsões nesse contexto. Apresentamos, na Seção 6.2, uma extensão da função de intensidade utilizada. A extensão do modelo para t-Student é discutida na Seção 6.3 .

6.1 Modelo Bayesiano Exato

O modelo geoestatístico com amostragem preferencial foi apresentado no Capítulo 3. Alguns problemas persistem quando passamos da abordagem clássica para a abordagem Bayesiana, sendo o principal a dificuldade de trabalhar com a função de densidade do processo pontual X , dada por

$$\pi(x|S, \theta) \propto \exp\left\{-\int_B \lambda(u) du\right\} \prod_{i=1}^N \lambda(x_i).$$

Para realizar inferência sobre o modelo geoestatístico com amostragem preferencial, foi necessário utilizar aproximação do processo S devido a intratabilidade da integral acima. Apesar do uso de discretização ser uma forma de contornar o problema, veja por exemplo Møller and Waagepetersen (2003), Diggle et al. (2010) e Dinsdale and Salibian-Barrera (2019), tal aproximação é uma fonte de erros difícil de mensurar, podendo produzir resultados viesados (Simpson et al., 2016). Além disso, é necessário se ter um balanço entre o refinamento da malha e o custo computacional, pois a simulação de processos

Gaussianos são da ordem de n^3 , tornando-se impraticável já para tamanhos de malha moderados. Como consequência, o nível de discretização necessária para se obter bons resultados pode não ser alcançável.

Gonçalves and Gamerman (2018) realizam simulação e inferência exata sobre processos de Cox, admitindo que a função de intensidade do processo pontual é limitada superiormente (condicionada nos parâmetros). Dessa forma, o processo pontual não homogêneo X é considerado como o resultado da aplicação de um *Poisson thinning* em um processo pontual homogêneo W . No caso da função de intensidade exponencial, esse método não pode ser aplicado basicamente por dois motivos: (1) a função não é limitada superiormente e (2) se truncarmos o valor da função em um valor alto, as probabilidades de retenção são baixas devido ao comportamento da função exponencial. Logo, considerar que $\lambda(x)$ possui um limite superior (condicionada nos parâmetros) foi a estratégia que permitiu aos autores contornar o problema da aproximação, obtendo uma inferência mais precisa.

Dessa forma, propomos aplicar a metodologia presente em Gonçalves and Gamerman (2018) no modelo geoestatístico com amostragem preferencial. Assumimos o seguinte modelo:

$$\mathbf{Y}|X, S \sim N_n(\mu \mathbf{1}_n + S(\mathbf{x}), \tau^2 \mathbf{I}_n), \quad (6.1)$$

$$X|S, \lambda \sim PP(\lambda), \quad (6.2)$$

$$\lambda(\mathbf{x}) = \lambda^* \Phi(\beta S(\mathbf{x})/\sigma), \quad (6.3)$$

$$S|\sigma^2, \phi \sim PG(0, \sigma^2 \rho(\phi)), \quad (6.4)$$

$$\theta \sim \pi(\theta), \quad \theta = (\lambda^*, \mu, \tau^2, \sigma^2, \phi, \beta), \quad (6.5)$$

em que PG = Processo Gaussiano, PP = Processo de Poisson, $\mathbf{1}_n$ é um vetor de 1's de tamanhos n , \mathbf{I}_n é a matriz identidade de tamanho n , $\rho(\phi)$ é a função de correlação de S indexada pelo parâmetro ϕ e $\lambda^* = \sup\{\lambda(\mathbf{x})\}$. Qualquer função contínua limitada poderia ser utilizada ao invés de Φ na função de intensidade como, por exemplo, a função logística. A escolha particular da função probito auxilia na construção do algoritmo MCMC, como será visto posteriormente.

Seja W um processo de Poisson homogêneo na região B com função de intensidade λ^* e suponha que o processo X é obtido através do processo W pelo método *Poisson*

thinning. Isto é, dividimos o processo W nos processos X e \tilde{X} tais que

$$\begin{aligned} X &= \{(x, z_x) : x \in W, z_x \sim \text{Bernoulli}(\Phi(\beta S(x))), z_x = 1\}, \\ \tilde{X} &= \{(\tilde{x}, z_{\tilde{x}}) : \tilde{x} \in W, z_{\tilde{x}} \sim \text{Bernoulli}(\Phi(\beta S(\tilde{x}))), z_{\tilde{x}} = 0\}. \end{aligned}$$

Denotamos o número de pontos da realização do processo W por k , do processo X por n e do processo \tilde{X} por $k-n$. \tilde{X} representa o processo composto pelos pontos descartados de W após aplicar o método *Poisson thinning* e os processos X e \tilde{X} são independentes condicionados aos parâmetros do modelo com $\tilde{X}|\tilde{\lambda} \sim PP(\tilde{\lambda})$, $\tilde{\lambda}(x) = \lambda^* \Phi(-\beta S(x)/\sigma)$.

A função de verossimilhança do modelo aumentado é dada por

$$l(S, \boldsymbol{\theta}; \mathbf{y}, \mathbf{x}) = \pi(\mathbf{y}, \mathbf{w}, \mathbf{x}|S, \boldsymbol{\theta}) = \pi(\mathbf{y}|S, \mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{x}, \mathbf{w}|S, \boldsymbol{\theta}), \quad (6.6)$$

em que \mathbf{w} , \mathbf{x} e \mathbf{y} são os valores observados de W , X e Y , respectivamente.

O primeiro termo do lado direito da Equação (6.6) é dado por

$$\pi(\mathbf{y}|S, \mathbf{x}, \boldsymbol{\theta}) = \left(\frac{1}{2\pi\tau^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\tau^2} \sum_{i=1}^n (y_i - \mu - S(\mathbf{x}_i))^2\right\}. \quad (6.7)$$

Ou seja, o primeiro termo da função de verossimilhança depende de $S = (\mathbf{S}_n, S_{-n})$ apenas através de \mathbf{S}_n , que representa S nos n valores observados.

O segundo termo de (6.6) é dado por $\pi(\mathbf{x}, \mathbf{w}|S, \boldsymbol{\theta})$, que é equivalente a $P(\mathbf{x}, \mathbf{w}, k|S, \boldsymbol{\theta})$, como visto no Algoritmo (1) supondo uma função de intensidade genérica. Considerando a função de intensidade em (6.3), podemos escrever

$$\begin{aligned} P(\mathbf{x}, \mathbf{w}, k|S, \boldsymbol{\theta}) &= P(\mathbf{x}|\mathbf{w}, k, S, \boldsymbol{\theta})P(\mathbf{w}|k, S, \boldsymbol{\theta})P(k|\boldsymbol{\theta}), \\ &= P(\mathbf{x}|\mathbf{w}, S, \boldsymbol{\theta})P(\mathbf{w}|k)P(k|\lambda^*), \\ &= \left[\prod_{i=1}^n \frac{\lambda(\mathbf{x}_i)}{\lambda^*}\right] \left[\prod_{i=n+1}^k 1 - \frac{\lambda(\mathbf{x}_i)}{\lambda^*}\right] \left[\prod_{i=1}^k \frac{1}{|B|}\right] e^{-\lambda^*|B|} \frac{(\lambda^*|B|)^k}{k!}, \\ &= \left[\prod_{i=1}^n \Phi\left(\beta \frac{S(\mathbf{x}_i)}{\sigma}\right)\right] \left[\prod_{i=n+1}^k \Phi\left(-\beta \frac{S(\mathbf{x}_i)}{\sigma}\right)\right] e^{-\lambda^*|B|} \frac{(\lambda^*)^k}{k!}, \\ &= \Phi_n\left(\frac{\beta}{\sigma} S_n; \mathbf{I}_n\right) \Phi_{k-n}\left(-\frac{\beta}{\sigma} S_{k-n}; \mathbf{I}_{k-n}\right) e^{-\lambda^*|B|} \frac{(\lambda^*)^k}{k!}, \end{aligned} \quad (6.8)$$

em que $\Phi_k(\cdot; \mathbf{I}_k)$ é a função de distribuição acumulada de uma distribuição Gaussiana

k -dimensional com vetor de média 0 e matriz de covariância \mathbf{I}_k .

Note que a função de verossimilhança depende de $S = (\mathbf{S}_k, S_{-k})$ apenas através de $\mathbf{S}_k = (\mathbf{S}_n, \mathbf{S}_{k-n})'$, que representa o processo Gaussiano S nos k pontos do processo pontual W . Dessa forma, podemos construir um algoritmo MCMC apenas com \mathbf{S}_k aplicando amostragem retrospectiva. A inferência para pontos de S_{-k} pode ser realizada através da posteriori. Dessa forma, a computação para nosso problema se torna finito-dimensional e evitamos o uso de discretização, sendo $\mathbf{S}_k \sim N_k(\mathbf{0}, \sigma^2 \mathbf{R}(\phi))$, em que $\mathbf{R}(\phi)$ é a matriz de correlação de \mathbf{S}_k . Para simplificar a notação, denotaremos $\mathbf{R} = \mathbf{R}(\phi)$.

A distribuição a posteriori do modelo será

$$\begin{aligned}
\pi(\mathbf{S}_k, \boldsymbol{\theta} | \mathbf{y}, \mathbf{x}) &\propto l(\mathbf{S}_k, \boldsymbol{\theta}; \mathbf{y}, \mathbf{x}) \times \pi(\mathbf{S}_k, \boldsymbol{\theta}), \\
&\propto \pi(\mathbf{y} | \mathbf{S}_n, \mathbf{x}, \boldsymbol{\theta}) \pi(\mathbf{x}, \mathbf{w}, k | \mathbf{S}_k, \boldsymbol{\theta}) \pi(\mathbf{S}_k | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \\
&\propto (\tau^2)^{-n/2} \exp \left\{ -\frac{1}{2\tau^2} \sum_{i=1}^n (y_i - \mu - S(\mathbf{x}_i))^2 \right\} \times \\
&\quad \left[\prod_{i=1}^n \Phi \left(\beta \frac{S(\mathbf{x}_i)}{\sigma} \right) \right] \left[\prod_{i=n+1}^k \Phi \left(-\beta \frac{S(\mathbf{x}_i)}{\sigma} \right) \right] e^{-\lambda^* |B|} (\lambda^*)^k \times \\
&\quad (\sigma^2)^{-k/2} |\mathbf{R}|^{-1/2} \exp \left\{ -\frac{1}{2\sigma^2} \mathbf{S}'_k \mathbf{R}^{-1} \mathbf{S}_k \right\} \pi(\boldsymbol{\theta}), \tag{6.9}
\end{aligned}$$

em que $\pi(\boldsymbol{\theta})$ é a distribuição a priori de $\boldsymbol{\theta}$.

Note que a distribuição a posteriori (6.9) não possui forma fechada e não identificamos nenhuma distribuição conhecida. Dessa forma, é necessário o uso de aproximação da distribuição a posteriori que será realizada via MCMC, como será visto na próxima seção.

6.1.1 Computação

Como não conhecemos a distribuição a posteriori de forma analítica, devemos realizar uma inferência baseada em aproximação por amostragem. Para isso, construímos um amostrador de Gibbs para poder gerar amostras de cada distribuição condicional. As

distribuições condicionais completas do modelo são dadas por:

$$\pi(\tilde{X}|\cdot) \sim PP(\tilde{\lambda}), \quad (6.10)$$

$$\pi(\mathbf{S}_k|\cdot) \propto \Phi_n\left(\frac{\beta}{\sigma}\mathbf{S}_n; \mathbf{I}_n\right) \Phi_{k-n}\left(-\frac{\beta}{\sigma}\mathbf{S}_{k-n}; \mathbf{I}_{k-n}\right) \pi(\mathbf{y}|\mathbf{S}_n, \mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{S}_k|\mathbf{w}, \boldsymbol{\theta}), \quad (6.11)$$

$$\pi(\lambda^*|\cdot) \propto e^{-\lambda^*|B|}(\lambda^*)^k\pi(\lambda^*), \quad (6.12)$$

$$\pi(\beta|\cdot) \propto \Phi_n\left(\frac{\beta}{\sigma}\mathbf{S}_n; \mathbf{I}_n\right) \Phi_{k-n}\left(-\frac{\beta}{\sigma}\mathbf{S}_{k-n}; \mathbf{I}_{k-n}\right) \pi(\beta), \quad (6.13)$$

$$\pi(\mu, \tau^2|\cdot) \propto \pi(\mathbf{y}|\mathbf{S}_n, \mathbf{x}, \boldsymbol{\theta})\pi(\mu, \tau^2), \quad (6.14)$$

$$\pi(\sigma^2|\cdot) \propto \Phi_n\left(\frac{\beta}{\sigma}\mathbf{S}_n; \mathbf{I}_n\right) \Phi_{k-n}\left(-\frac{\beta}{\sigma}\mathbf{S}_{k-n}; \mathbf{I}_{k-n}\right) \pi(\mathbf{S}_k|\mathbf{w}, \boldsymbol{\theta})\pi(\sigma^2), \quad (6.15)$$

$$\pi(\phi|\cdot) \propto \pi(\mathbf{S}_k|\mathbf{w}, \boldsymbol{\theta})\pi(\phi). \quad (6.16)$$

A geração de λ^* fornece o valor do número médio de pontos que serão gerados para o processo \tilde{X} . Note que os $k - n$ pontos do processo \tilde{X} são gerados aleatoriamente sobre a região contínua B , ou seja, a cadeia do processo dos pontos descartados passeia sobre o domínio contínuo da região. Após a geração dos pontos do processo descartado, o processo Gaussiano infinito-dimensional é gerado apenas nos k pontos do processo W . Em outras palavras, a abordagem continua sendo infinito-dimensional, porém necessitamos apenas de um subconjunto de k pontos para construir o algoritmo MCMC, o que é a vantagem da amostragem retrospectiva. A partir da amostra de \mathbf{S}_k e dos locais \mathbf{x}_k , os demais parâmetros são amostrados. Inferência para S_{-k} é realizada através da posteriori (condicionada nos \mathbf{S}_k pontos).

Dessa forma, a geração de amostras das distribuições condicionais completas foram realizadas considerando os seguintes passos.

Passo 1: λ^* . Se escolhermos como distribuição a priori para λ^* uma distribuição $Gama(a_{\lambda^*}, b_{\lambda^*})$, (6.12) será uma distribuição $Gama(a_{\lambda^*} + k, b_{\lambda^*} + |B|)$.

Passo 2: \tilde{X} . Essa amostragem foi realizada utilizando o algoritmo 2 (Seção 2.4.2):

1. simulou-se $K^* \sim Pois(\lambda^*|B)$,
2. os $K^* = k^*$ pontos foram distribuídos uniformemente em B , obtendo os locais $\{\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_{k^*}\}$,
3. simulou-se \mathbf{S}_{k^*} retrospectivamente de $\pi(\mathbf{S}_{k^*}|\mathbf{S}_k, \boldsymbol{\theta})$, em que \mathbf{S}_{k^*} é o valor de S nos k^* locais gerados,
4. aplicou-se o *Poisson thinning* com probabilidades de retenção $\Phi(-\beta S(\tilde{\mathbf{x}}_j)/\sigma)$, $j = 1, \dots, k^*$,
5. foram computados os locais retidos na etapa 4.

Note que simulamos da distribuição de \mathbf{S}_{k^*} condicional a $\mathbf{S}_k = \{\mathbf{S}_n, \mathbf{S}_{k-n}\}$, em que \mathbf{S}_{k-n} é o processo Gaussiano S simulado nos $k - n$ pontos de \tilde{X} do passo anterior. Dessa forma, agregamos a informação dos pontos de \tilde{X} do passo $i - 1$ para o passo i .

Passo 3: \mathbf{S}_k . Mostramos a seguir que a distribuição condicional completa de \mathbf{S}_k pode ser escrita como o núcleo de uma distribuição Skew normal (SN). Utilizaremos a definição da distribuição SN presente no artigo de Gonçalves and Gamerman, 2018. Sejam um vector m -dimensional $\boldsymbol{\xi}$, uma matrix \mathbf{W} $m' \times m$ e uma matriz \mathbf{E} $m \times m$, definimos $\mathbf{U} = (\mathbf{U}_0, \mathbf{U}_1)' \sim N_{m'+m}(\mathbf{0}, \mathbf{E}^*)$,

$$\mathbf{E}^* = \begin{pmatrix} \boldsymbol{\Gamma} & \mathbf{B}' \\ \mathbf{B} & \mathbf{E} \end{pmatrix},$$

em que $\boldsymbol{\Gamma} = \mathbf{I}_{m'} + \mathbf{W}\mathbf{E}\mathbf{W}'$ e $\mathbf{B}' = \mathbf{W}\mathbf{E}$. Seja $\mathbf{a} = (a_1, \dots, a_r) > \mathbf{b} = (b_1, \dots, b_r)$ no sentido de que $a_i > b_i, \forall i$, e defina $\boldsymbol{\gamma} = \mathbf{W}\boldsymbol{\xi}$. Dizemos que $(\mathbf{U}_1 + \boldsymbol{\xi}|\mathbf{U}_0 > -\boldsymbol{\gamma})$ possui distribuição $SN(\boldsymbol{\xi}, \mathbf{E}, \mathbf{W})$ cuja densidade é dada por

$$f(\mathbf{u}) = \frac{1}{\Phi_{m'}(\boldsymbol{\gamma}; \boldsymbol{\Gamma})} \phi_m(\mathbf{u} - \boldsymbol{\xi}; \mathbf{E}) \Phi_{m'}(\mathbf{W}\mathbf{u}; \mathbf{I}_m),$$

em que $\phi_m(\cdot; \boldsymbol{\Omega})$ e $\Phi_m(\cdot; \boldsymbol{\Omega})$ são as funções de densidade e de probabilidade, respectivamente, de uma distribuição Gaussiana m -dimensional com vetor de médias $\mathbf{0}$ e matriz de covariâncias $\boldsymbol{\Omega}$.

O termo $\pi(\mathbf{y}|\mathbf{S}_n, \mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{S}_k|\mathbf{w}, \boldsymbol{\theta})$ na distribuição em (6.11) pode ser escrito como o núcleo de uma função de densidade de probabilidade em \mathbf{S}_k . Considerando $\boldsymbol{\Sigma}_y = \tau^2 \mathbf{I}_N$, temos que

$$\begin{aligned}\pi(\mathbf{y}|\mathbf{S}_n, \mathbf{x}, \boldsymbol{\theta}) &\propto \exp\left\{-\frac{1}{2}\left[(\mathbf{y} - \mu\mathbf{1}_n - \mathbf{S}_n)' \boldsymbol{\Sigma}_y^{-1}(\mathbf{y} - \mu\mathbf{1}_n - \mathbf{S}_n)\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[2(\mathbf{y} - \mu\mathbf{1}_n)' \boldsymbol{\Sigma}_y^{-1} \mathbf{S}_n + \mathbf{S}_n' \boldsymbol{\Sigma}_y^{-1} \mathbf{S}_n\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[2(\mathbf{y} - \mu\mathbf{1}_n)' \boldsymbol{\Sigma}_y^{-1} \mathbf{C} \mathbf{S}_k + \mathbf{S}_k' \mathbf{C}' \boldsymbol{\Sigma}_y^{-1} \mathbf{C} \mathbf{S}_k\right]\right\},\end{aligned}$$

em que $\mathbf{S}_n = \mathbf{C} \mathbf{S}_k$, \mathbf{C} é uma matriz $n \times k$ com i -ésima linha consistindo em $n - 1$ 0's e um único 1 identificando a posição de S_i em \mathbf{S}_k , $i = 1, \dots, N$. Dessa forma, podemos escrever

$$\begin{aligned}\pi(\mathbf{y}|\mathbf{S}_n, \mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{S}_k|\mathbf{w}, \boldsymbol{\theta}) &\propto \exp\left\{-\frac{1}{2}\left[\mathbf{S}_k' \mathbf{C}' \boldsymbol{\Sigma}_y^{-1} \mathbf{C} \mathbf{S}_k - 2(\mathbf{y} - \mu\mathbf{1}_n)' \boldsymbol{\Sigma}_y^{-1} \mathbf{C} \mathbf{S}_k\right.\right. \\ &\quad \left.\left.+ \mathbf{S}_k' (\sigma^2 \mathbf{R})^{-1} \mathbf{S}_k\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\mathbf{S}_k' (\mathbf{C}' \boldsymbol{\Sigma}_y^{-1} \mathbf{C} + (\sigma^2 \mathbf{R})^{-1}) \mathbf{S}_k\right.\right. \\ &\quad \left.\left.- 2((\mathbf{y} - \mu\mathbf{1}_n)' \boldsymbol{\Sigma}_y^{-1} \mathbf{C}) \mathbf{S}_k\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}(\mathbf{S}_k - \boldsymbol{\mu}^*)' \mathbf{D}^{-1}(\mathbf{S}_k - \boldsymbol{\mu}^*)\right\},\end{aligned}\quad (6.17)$$

em que $\mathbf{D} = (\mathbf{C}' \boldsymbol{\Sigma}_y^{-1} \mathbf{C} + (\sigma^2 \mathbf{R})^{-1})^{-1}$, $\boldsymbol{\mu}^* = \mathbf{D} \mathbf{C}' \boldsymbol{\Sigma}_y^{-1} (\mathbf{y} - \mu\mathbf{1}_n)$ e $\mathbf{R} = \mathbf{R}(\phi)$. Dessa forma, podemos reescrever a distribuição condicional completa em (6.11) como

$$\pi(\mathbf{S}_k|\cdot) \propto \phi_k(\mathbf{S}_k; \boldsymbol{\mu}^*, \mathbf{D}) \Phi_k(\mathbf{G} \mathbf{S}_k; \mathbf{I}_K), \quad (6.18)$$

em que $\mathbf{G} = \beta/\sigma * \text{diag}(\mathbf{I}_n, -\mathbf{I}_{k-n})$.

A distribuição em (6.18) corresponde ao núcleo de uma distribuição $SN(\boldsymbol{\mu}^*, \mathbf{D}, \mathbf{G})$ e pode ser amostrado pelo Algoritmo 4 apresentado em Gonçalves and Gamerman (2018).

Passo 4: β . Foi implementado um passo de MH para gerar amostras de β . Assumindo como distribuição a priori $\beta \sim N(\mu_\beta, \sigma_\beta^2)$, a distribuição condicional completa

é dada por

$$\pi(\beta|\cdot) \propto \Phi_n\left(\frac{\beta}{\sigma}\mathbf{S}_n; \mathbf{I}_n\right) \Phi_{k-n}\left(-\frac{\beta}{\sigma}\mathbf{S}_{k-n}; \mathbf{I}_{k-n}\right) \phi(\beta; \mu_\beta, \sigma_\beta^2).$$

Considerando como distribuição de proposta $\beta_p|\beta_c \sim N(\beta_c, \delta_\beta)$, em que β_p é o valor proposto para β e β_c é o valor corrente da cadeia, a probabilidade de aceitação é $A_\beta = \min\{1, p_\beta\}$, em que

$$p_\beta = \frac{\Phi_n\left(\frac{\beta_p}{\sigma}\mathbf{S}_n; \mathbf{I}_n\right) \Phi_{k-n}\left(-\frac{\beta_p}{\sigma}\mathbf{S}_{k-n}; \mathbf{I}_{k-n}\right) \phi(\beta_p; \mu_\beta, \sigma_\beta^2)}{\Phi_n\left(\frac{\beta_c}{\sigma}\mathbf{S}_n; \mathbf{I}_n\right) \Phi_{k-n}\left(-\frac{\beta_c}{\sigma}\mathbf{S}_{k-n}; \mathbf{I}_{k-n}\right) \phi(\beta_c; \mu_\beta, \sigma_\beta^2)}.$$

Passo 5: μ e τ^2 . Vamos considerar que, a priori, μ é independente de τ^2 e $\pi(\mu, \tau^2) = \pi(\mu)\pi(\tau^2)$. Assumimos que $\mu \sim N(\mu_0, \sigma_\mu^2)$ e $\tau^2 \sim GI(a_\tau, b_\tau)$. Dessa forma, as distribuições condicionais completas de μ e τ^2 são dadas por

$$\begin{aligned} (\mu|\cdot) &\sim N\left(\left(\frac{\sum_{i=1}^n (y_i - S(\mathbf{x}_i))}{\tau^2} + \frac{\mu_0}{\sigma_\mu^2}\right)\sigma^{2*}, \sigma^{2*}\right); \sigma^{2*} = \left(\frac{n}{\tau^2} + \frac{1}{\sigma_\mu^2}\right)^{-1}, \\ (\tau^2|\cdot) &\sim GI\left(\frac{n}{2} + a_\tau, \frac{\sum_{i=1}^n (y_i - S(\mathbf{x}_i) - \mu)^2}{2} + b_\tau\right). \end{aligned}$$

Passo 6: σ^2 . Foi implementado um passo de MH para gerar amostras de σ^2 . Se escolhermos como distribuição a priori para σ^2 uma distribuição $GI(a_\sigma, b_\sigma)$, a distribuição condicional completa é dada por:

$$\begin{aligned} \pi(\sigma^2|\cdot) &\propto \Phi_n\left(\frac{\beta}{\sigma}\mathbf{S}_n; \mathbf{I}_n\right) \Phi_{k-n}\left(-\frac{\beta}{\sigma}\mathbf{S}_{k-n}; \mathbf{I}_{k-n}\right) \times \\ &\quad \left(\frac{1}{\sigma^2}\right)^{k/2+a_\sigma+1} \exp\left\{-\left(\frac{\mathbf{S}'_k \mathbf{R}^{-1} \mathbf{S}_k}{2} + b_\sigma\right) \frac{1}{\sigma^2}\right\}. \end{aligned}$$

Considerando a distribuição de proposta $q(\sigma_p^2|\sigma_c^2)$ como Lognormal($\log(\sigma_c^2)$; δ_{σ^2}), em que σ_p^2 é o valor proposto para σ^2 e σ_c^2 é o valor corrente da cadeia, a probabilidade de aceitação é $A_\sigma = \min\{1, p_\sigma\}$, em que

$$p_\sigma = \frac{\Phi_k\left((\beta/\sigma_p)\mathbf{S}_k \mathbf{I}_k^*; \mathbf{I}_k\right) \left(\frac{\sigma_p^2}{\sigma_c^2}\right)^{-\frac{k}{2}-a_\sigma}}{\Phi_k\left((\beta/\sigma_c)\mathbf{S}_k \mathbf{I}_k^*; \mathbf{I}_k\right) \left(\frac{\sigma_c^2}{\sigma_c^2}\right)^{-\frac{k}{2}-a_\sigma}} \exp\left\{-\left(\frac{\mathbf{S}'_k \mathbf{R}^{-1} \mathbf{S}_k}{2} + b_\sigma\right) \left(\frac{1}{\sigma_p^2} - \frac{1}{\sigma_c^2}\right)\right\}.$$

Passo 7: ϕ . Assumindo uma distribuição a priori $Gama(a_\phi, b_\phi)$, a distribuição

condicional completa fica dada por

$$\pi(\phi|.) \propto |\mathbf{R}|^{-1/2} \phi^{a_\phi - 1} \exp \left\{ -\frac{1}{2\sigma^2} \mathbf{S}'_k \mathbf{R}^{-1} \mathbf{S}_k - b_\phi \phi \right\}.$$

Para amostrar dessa distribuição será implementado um passo de MH. Assumimos a distribuição de proposta $q(\phi_p|\phi_c)$ como distribuição Lognormal($\log(\phi_c)$; δ_ϕ), em que ϕ_p é o valor proposto de ϕ e ϕ_c é o valor de ϕ no estado corrente da cadeia. De forma equivalente, $\log(\phi_p) \sim N(\log(\phi_c), \delta_\phi)$. A probabilidade de aceitação é $A_\phi = \min\{1, p_\phi\}$, em que

$$p_\phi = \left(\frac{|\mathbf{R}(\phi_p)|}{|\mathbf{R}(\phi_c)|} \right)^{-1/2} \left(\frac{\phi_p}{\phi_c} \right)^{a_\phi} \exp \left\{ -\frac{1}{2\sigma^2} \left(\mathbf{S}'_k (\mathbf{R}(\phi_p)^{-1} - \mathbf{R}(\phi_c)^{-1}) \mathbf{S}_k \right) - b_\phi (\phi_p - \phi_c) \right\}.$$

Os passos 1-7 fornecerão, após convergência das cadeias, amostras da distribuição a posteriori de $\boldsymbol{\theta}$. A partir dessas amostras é possível realizar a inferência sobre os parâmetros do modelo.

O parâmetro λ^* controla o número esperado de pontos na simulação de \tilde{X} , isto é, $E[N(B)] = \lambda^*|B|$. Isso significa que valores altos de λ^* impactam na eficiência do algoritmo, pois precisamos realizar simulações do processo Gaussiano nos pontos gerados. Logo, truncar a região dos possíveis valores desse parâmetro é uma alternativa para se evitar o problema de simular S de uma dimensão alta. Neste trabalho, utilizamos como distribuição a priori para λ^* uma distribuição gama truncada $GamaT(a_{\lambda^*}, b_{\lambda^*}, \lambda_{max})$, em que λ_{max} é o valor máximo que λ^* pode assumir.

Podemos acrescentar uma estrutura de regressão na média, tomando

$$\mu(\mathbf{b}_i) = \eta_0 + \sum_{j=1}^p \eta_j d_j(\mathbf{b}_i), \quad (6.19)$$

em que η_0, \dots, η_p são parâmetros de regressão a serem estimados e $d_j(\cdot)$ são funções conhecidas para $j = 1, \dots, p$. A Equação (6.1) do modelo toma a forma

$$\begin{aligned} \mathbf{Y}|X, S &\sim N_n(\mathbf{D}\boldsymbol{\eta} + S(\mathbf{x}), \tau^2 \mathbf{I}_n), \\ \boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}), \quad \boldsymbol{\theta} = (\lambda^*, \boldsymbol{\eta}, \tau^2, \sigma^2, \phi, \beta), \end{aligned}$$

em que \mathbf{D} é a matriz de delineamento $N \times (p+1)$ e $\boldsymbol{\eta} = (\eta_0, \eta_1, \dots, \eta_p)$. As distribuições

condicionais completas de \mathbf{S}_k e τ^2 são equivalentes ao caso anterior, alterando $\boldsymbol{\mu}$ pela estrutura de regressão. A distribuição condicional completa de $\boldsymbol{\eta}$ é dada por

$$\pi(\boldsymbol{\eta}|\cdot) \propto \pi(y|\mathbf{S}_n, \mathbf{x}, \boldsymbol{\theta})\pi(\boldsymbol{\eta}).$$

Se consideramos como distribuição a priori para $\boldsymbol{\eta}$ uma distribuição normal $(p+1)$ -variada com vetor de médias $\boldsymbol{\eta}_0$ e matriz de covariâncias $\sigma_\eta^2 \mathbf{I}_{(p+1)}$, sua distribuição condicional completa será normal $(p+1)$ -variada com vetor de médias e matriz de covariâncias dados por

$$\begin{aligned} \boldsymbol{\mu}_\eta^* &= \boldsymbol{\Sigma}_\eta^* \left(\frac{1}{\tau^2} \mathbf{D}'(\mathbf{y} - \mathbf{S}_n) + \frac{1}{\sigma_\eta^2} \mathbf{I}_{(p+1)} \right), \\ \boldsymbol{\Sigma}_\eta^* &= \left(\frac{1}{\tau^2} \mathbf{D}'\mathbf{D} + \frac{1}{\sigma_\eta^2} \mathbf{I}_{(p+1)} \right)^{-1}. \end{aligned}$$

6.1.2 Predição

O principal objetivo em geoestatística é realizar predição da variável de interesse em locais não observados. No nosso modelo, estamos interessados em prever o valor de Y em locais distintos dos da amostra. Geralmente é considerada uma malha regular na área de estudo para a construção de mapas. Seja $\mathbf{x}_p = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n_p})$ o vetor de localizações que queremos prever Y , em que cada $\mathbf{x}_i = (b_1, b_2, \dots, b_d) \in B \subset \mathbb{R}^d$, $i = 1, \dots, n_p$. Assuma que \mathbf{S}_p e \mathbf{Y}_p sejam os valores de S e Y nas localizações \mathbf{x}_p .

A distribuição preditiva a posteriori de \mathbf{Y}_p será dada por

$$\begin{aligned} \pi(\mathbf{Y}_p|\mathbf{x}, \mathbf{y}) &= \int \pi(\mathbf{Y}_p, \mathbf{S}_p, \mathbf{S}_k, \boldsymbol{\theta}|\mathbf{x}, \mathbf{y}) d\mathbf{S}_p d\mathbf{S}_k d\boldsymbol{\theta} \\ &= \int \pi(\mathbf{Y}_p|\mathbf{S}_p, \boldsymbol{\theta})\pi(\mathbf{S}_p|\mathbf{S}_k, \boldsymbol{\theta})\pi(\mathbf{S}_k, \boldsymbol{\theta}|\mathbf{x}, \mathbf{y}) d\mathbf{S}_p d\mathbf{S}_k d\boldsymbol{\theta}, \end{aligned} \quad (6.20)$$

em que $\pi(\mathbf{S}_k, \boldsymbol{\theta}|\mathbf{x}, \mathbf{y})$ é a distribuição a posteriori de $(\mathbf{S}_k, \boldsymbol{\theta})$. Como S é um processo Gaussiano, $(\mathbf{S}_p, \mathbf{S}_k|\boldsymbol{\theta})$ possui distribuição Normal multivariada, ou seja,

$$(\mathbf{S}_p, \mathbf{S}_k|\boldsymbol{\theta}) \sim N_{(n_p+k)} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \sigma^2 \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix} \right),$$

em que \mathbf{R}_{11} corresponde à matriz de correlação de \mathbf{S}_p , \mathbf{R}_{22} corresponde à matriz de correlação de \mathbf{S}_k , \mathbf{R}_{12} corresponde à matriz de correlação de \mathbf{S}_p e \mathbf{S}_k e $\mathbf{R}_{21} = \mathbf{R}'_{12}$. Dessa

forma, $(\mathbf{S}_p | \mathbf{S}_k, \boldsymbol{\theta}) \sim N_{n_p}(\boldsymbol{\mu}_{\mathbf{S}_p | \mathbf{S}_k}, \sigma^2 \mathbf{R}_{\mathbf{S}_p | \mathbf{S}_k})$, em que $\boldsymbol{\mu}_{\mathbf{S}_p | \mathbf{S}_k} = \mathbf{R}_{12} \mathbf{R}_{22}^{-1} \mathbf{S}_k$, $\mathbf{R}_{\mathbf{S}_p | \mathbf{S}_k} = \mathbf{R}_{11} - \mathbf{R}_{12} \mathbf{R}_{22}^{-1} \mathbf{R}_{21}$. Dados \mathbf{S}_p e $\boldsymbol{\theta}$, \mathbf{Y}_p possui distribuição normal com média $\mathbf{S}_p + \mu \mathbf{1}_{n_p}$ e matriz de covariâncias $\tau^2 \mathbf{I}_{n_p}$. Logo, obter uma amostra da distribuição preditiva de Y equivale a, dada uma amostra de $(\mathbf{S}_k, \boldsymbol{\theta} | \mathbf{x}, \mathbf{y})$, gerar uma amostra de $(\mathbf{S}_p | \boldsymbol{\theta})$ e, em seguida, obter uma amostra de $(\mathbf{Y}_p | \mathbf{S}_p, \boldsymbol{\theta})$.

Em resumo, a informação do processo pontual X é utilizada na distribuição a posteriori de $(\mathbf{S}_k, \boldsymbol{\theta})$ para obtermos estimativas corretas dos parâmetros e dos valores do processo Gaussiano nos locais observados. A partir dessas estimativas “corrigidas” pela informação de X , a predição do processo Gaussiano é realizada de forma independente de X , ou seja, de forma convencional (krigagem). E dados os valores de \mathbf{S}_p , a predição para a variável de interesse Y também é dada de forma convencional.

Uma outra predição importante no contexto de processos pontuais é estimar a função de intensidade em locais não observados. Denotando o vetor dos valores da função de intensidade nos locais \mathbf{x}_p como $\lambda(\mathbf{x}_p) = (\lambda(\mathbf{x}_{p1}), \dots, \lambda(\mathbf{x}_{pn_p}))'$, precisamos de amostras a posteriori de $\lambda(\mathbf{x}_p)$. Para isso, necessitamos de amostras de S nos locais de interesse (\mathbf{S}_p) , β , σ^2 e λ^* . Para cada amostra a posteriori de $(\mathbf{S}_k, \boldsymbol{\theta})$, simulamos um valor de \mathbf{S}_p e, em seguida, calculamos o valor de $\lambda(\mathbf{x})$ nos locais não observados. Assim, obtemos uma amostra da distribuição a posteriori de $\lambda(\mathbf{x}_p)$ e podemos realizar inferência sobre a função de intensidade.

Podemos também estimar a função de intensidade integrada $\Lambda(\cdot)$ em uma subregião de B , dada por $\Lambda(B') = \int_{B'} \lambda(\boldsymbol{\zeta}) d\boldsymbol{\zeta}$, $B' \subset B$ e $\boldsymbol{\zeta} \in B'$, que representa o número médio de pontos em B' . Essa estimação pode ser realizada via Monte Carlo, introduzindo uma variável auxiliar A com distribuição uniforme em B' . Note que $E_A[\lambda(B')] = |B'|^{-1} \int_{B'} \lambda(\boldsymbol{\zeta}) d\boldsymbol{\zeta}$ (Beskos et al., 2006) e um estimador não viciado para $\Lambda(B')$ é dado por

$$\hat{\Lambda}(B') = |B'| \frac{1}{L} \sum_{l=1}^L \lambda_l(A_l),$$

em que $\lambda_l(A_l)$ é a l -ésima amostra de $\lambda(\cdot)$ no local A_l . Essas amostras são obtidas pela distribuição a posteriori.

6.2 Modelo com intercepto na função de intensidade

No modelo proposto, a função de intensidade em (6.3) não possui intercepto e, como consequência, para o valor do processo Gaussiano igual a zero a taxa será $\lambda^*/2$. Essa característica impacta na estimação da função de intensidade quando a taxa real não segue o modelo proposto, isto é, se a função de intensidade para o valor nulo do processo Gaussiano possui valor diferente de $\lambda^*/2$, nosso modelo não é capaz de captar esse comportamento.

De fato, a função de intensidade proposta em (6.3) é um caso particular da seguinte função de intensidade,

$$\lambda(\mathbf{x}) = \lambda^* \Phi(\alpha + \beta S(\mathbf{x})/\sigma), \quad (6.21)$$

em que α representa o intercepto da curva. No nosso caso, α é nulo. A Figura 6.1 apresenta o efeito do intercepto na função de intensidade e percebemos que, a partir de α , podemos deslocar a curva da taxa ao passo que, para α nulo, a curva passa necessariamente no ponto $(0, \lambda^*/2)$.

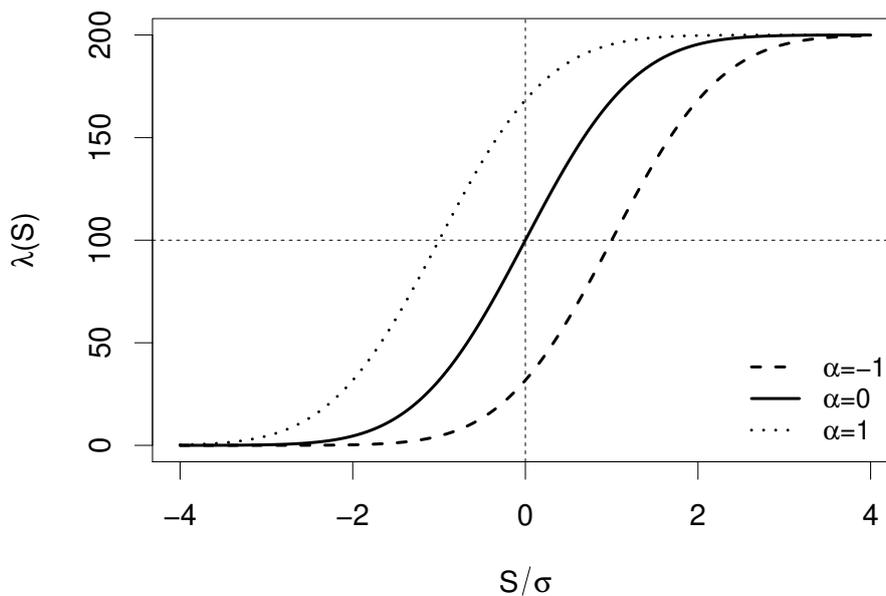


Figura 6.1: Efeito do parâmetro α na função de intensidade ($\beta = 1$ e $\lambda^* = 200$).

A computação para o modelo com intercepto na função de intensidade foi de-

sempre e mostraremos as mudanças em relação ao modelo anterior. As distribuições condicionais completas de λ^* , μ , τ^2 e ϕ permanecem iguais às apresentadas na Seção 6.1.1. Em relação aos demais parâmetros, suas distribuições condicionais completas serão dadas por

$$\pi(\tilde{X}|\cdot) \sim PP(\tilde{\lambda}), \quad (6.22)$$

$$\pi(\mathbf{S}_k|\cdot) \propto \Phi_n\left(\alpha + \frac{\beta}{\sigma}\mathbf{S}_n; \mathbf{I}_n\right) \Phi_{k-n}\left(-\alpha - \frac{\beta}{\sigma}\mathbf{S}_{k-n}; \mathbf{I}_{k-n}\right) \times \pi(\mathbf{y}|\mathbf{S}_n, \mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{S}_k|\mathbf{w}, \boldsymbol{\theta}), \quad (6.23)$$

$$\pi(\alpha, \beta|\cdot) \propto \Phi_n\left(\alpha + \frac{\beta}{\sigma}\mathbf{S}_n; \mathbf{I}_n\right) \Phi_{k-n}\left(-\alpha - \frac{\beta}{\sigma}\mathbf{S}_{k-n}; \mathbf{I}_{k-n}\right) \pi(\alpha, \beta), \quad (6.24)$$

$$\pi(\sigma^2|\cdot) \propto \Phi_n\left(\alpha + \frac{\beta}{\sigma}\mathbf{S}_n; \mathbf{I}_n\right) \Phi_{k-n}\left(-\alpha - \frac{\beta}{\sigma}\mathbf{S}_{k-n}; \mathbf{I}_{k-n}\right) \pi(\mathbf{S}_k|\mathbf{w}, \boldsymbol{\theta})\pi(\sigma^2), \quad (6.25)$$

em que $\tilde{\lambda}(\mathbf{x}) = \lambda^*\Phi(-\alpha - \beta\mathbf{S}(\mathbf{x})/\sigma)$.

A geração de amostras de σ^2 é realizada por passo de MH, da mesma forma do modelo anterior (Passo 6), apenas alterando o argumento da função probito.

Em relação aos parâmetros α e β , assumimos independência a priori e distribuições a priori $\alpha \sim N(\alpha_0, \sigma_\alpha^2)$ e $\beta \sim N(\beta_0, \sigma_\beta^2)$. A distribuição condicional completa é dada por

$$\pi(\alpha, \beta|\cdot) \propto \Phi_n\left(\alpha + \frac{\beta}{\sigma}\mathbf{S}_n; \mathbf{I}_n\right) \Phi_{k-n}\left(-\alpha - \frac{\beta}{\sigma}\mathbf{S}_{k-n}; \mathbf{I}_{k-n}\right) \phi(\alpha; \alpha_0, \sigma_\alpha^2)\phi(\beta; \beta_0, \sigma_\beta^2).$$

A geração de amostras desses parâmetros é realizada conjuntamente por passo de MH, em que consideramos como distribuição de proposta $(\alpha_p, \beta_p|\alpha_c, \beta_c)' \sim N_2((\alpha_c, \beta_c), \boldsymbol{\Sigma}_{\alpha,\beta})$, em que $\boldsymbol{\Sigma}_{\alpha,\beta}$ é uma matriz diagonal com $\text{diag}(\boldsymbol{\Sigma}_{\alpha,\beta}) = (\delta_\alpha, \delta_\beta)$ e α_p , α_c , β_p e β_c são os valores propostos e correntes das cadeias de α e β , respectivamente. A probabilidade de aceitação é dada por $A_{\alpha,\beta} = \min\{1, p_{\alpha,\beta}\}$, em que

$$p_{\alpha,\beta} = \frac{\Phi_n\left(\alpha_p + \frac{\beta_p}{\sigma}\mathbf{S}_n; \mathbf{I}_n\right) \Phi_{k-n}\left(-\alpha_p - \frac{\beta_p}{\sigma}\mathbf{S}_{k-n}; \mathbf{I}_{k-n}\right) \phi(\alpha_p; \alpha_0, \sigma_\alpha^2)\phi(\beta_p; \beta_0, \sigma_\beta^2)}{\Phi_n\left(\alpha_c + \frac{\beta_c}{\sigma}\mathbf{S}_n; \mathbf{I}_n\right) \Phi_{k-n}\left(-\alpha_c - \frac{\beta_c}{\sigma}\mathbf{S}_{k-n}; \mathbf{I}_{k-n}\right) \phi(\alpha_c; \alpha_0, \sigma_\alpha^2)\phi(\beta_c; \beta_0, \sigma_\beta^2)}.$$

Em relação à \mathbf{S}_k , poderíamos gerar amostras a partir de um passo de MH. Porém, se quisermos utilizar a distribuição Skew Normal, precisamos realizar a seguinte trans-

formação

$$\mathbf{T}_k = \alpha + \frac{\beta}{\sigma} \mathbf{S}_k \quad (6.26)$$

A distribuição condicional completa de \mathbf{T}_k será

$$\begin{aligned} \pi(\mathbf{T}_k | \cdot) &= \pi(\mathbf{S}_k | \cdot) |_{h^{-1}(\mathbf{S}_k)} \times |\mathbf{J}| \\ &\propto \Phi_n(\mathbf{T}_n; \mathbf{I}_n) \Phi_{k-n}(-\mathbf{T}_{k-n}; \mathbf{I}_{k-n}) \pi(\mathbf{y} | \mathbf{T}_n, \mathbf{x}, \boldsymbol{\theta}) \pi(\mathbf{T}_k | \mathbf{w}, \boldsymbol{\theta}) \times \left(\frac{\sigma}{\beta}\right)^k, \end{aligned}$$

em que \mathbf{J} é o Jacobiano da transformação. De forma análoga ao modelo anterior, a distribuição condicional completa de \mathbf{T}_k pode ser escrita como

$$\pi(\mathbf{T}_k | \cdot) \propto \phi_k(\mathbf{T}_k; \boldsymbol{\mu}^*, \mathbf{D}) \Phi_k(\mathbf{G}\mathbf{T}_k; \mathbf{I}_k), \quad (6.27)$$

em que

$$\begin{aligned} \mathbf{D} &= \left(\frac{\sigma^2}{\beta^2} \mathbf{C}' \boldsymbol{\Sigma}_y^{-1} \mathbf{C} + \frac{1}{\beta^2} \mathbf{R}^{-1} \right)^{-1} \\ \boldsymbol{\mu}^* &= \mathbf{D} \left(\frac{\sigma}{\beta} \mathbf{C}' \boldsymbol{\Sigma}_y^{-1} (\mathbf{y} - \mu \mathbf{1}_n + \alpha \frac{\sigma}{\beta} \mathbf{1}_n) + \frac{\alpha}{\beta^2} \mathbf{R}^{-1} \mathbf{1}_k \right), \end{aligned}$$

$\mathbf{G} = \text{diag}(\mathbf{I}_n, -\mathbf{I}_{k-n})$, $\boldsymbol{\Sigma}_y = \tau^2 \mathbf{I}_n$ e \mathbf{C} é uma matriz $n \times k$ com i -ésima linha consistindo em $n-1$ 0's e um único 1 identificando a posição de S_i em \mathbf{S}_k , $i = 1, \dots, n$. A distribuição em (6.27) corresponde ao núcleo de uma distribuição Skew Normal $SN(\boldsymbol{\mu}^*, \mathbf{D}, \mathbf{G})$ e pode ser amostrado pelo Algoritmo 4 apresentado no artigo de Gonçalves and Gamerman (2018).

Algumas dificuldades foram encontradas na estimação do intercepto desse modelo e serão discutidas na seção de simulações.

6.3 Modelo t-Student

Como visto na Seção 4.2, podemos induzir a distribuição t-Student no modelo de duas maneiras: considerando que o processo latente S é o um processo t ou considerando que o processo de medida Y é um processo t. Enquanto a primeira significa que é plausível termos dados que estão localizados na cauda da distribuição, a segunda sugere que podemos obter alguns erros na mensuração maiores que o normal. Dessa forma, conhecer o

contexto para o qual se considera o modelo t-Student se torna importante.

Vamos construir o modelo t-Student para os dois casos, apontando suas diferenças com o modelo Gaussiano, apresentado pelas equações (6.1) a (6.5). Vale ressaltar que os graus de liberdade (ν) não serão estimados, sendo considerados fixos durante toda a tese. A estimação de ν foi considerada para trabalhos futuros, como será discutido posteriormente.

Primeiro caso: processo S t-Student

Nesse caso, o modelo Gaussiano sofre alteração na Equação (6.4) referente ao processo S e na Equação (6.3) referente à função de intensidade, tomando a seguinte forma:

$$\begin{aligned}\lambda(\mathbf{x}) &= \lambda^* \Phi(\beta \sqrt{U} S(\mathbf{x}) / \sigma), \\ S | \sigma^2, \phi, U &\sim PG(0, \sigma^2 U^{-1} \rho(\phi)), \\ U | \nu &\sim Gama(\nu/2, \nu/2), \quad \nu > 0,\end{aligned}$$

em que ν é o número de graus de liberdade (gl) da distribuição t-Student. Note que, marginalmente, $S | \sigma^2, \phi$ é um processo t. A função de verossimilhança do modelo não se altera e depende apenas de S_k . Dessa forma,

$$\mathbf{S}_k | \sigma^2, \phi, U \sim N_k(0, \sigma^2 U^{-1} \mathbf{R}(\phi)).$$

A distribuição a posteriori do modelo será

$$\begin{aligned}\pi(\mathbf{S}_k, U, \boldsymbol{\theta} | \mathbf{y}, \mathbf{x}) &\propto l(\mathbf{S}_k, U, \boldsymbol{\theta}; \mathbf{y}, \mathbf{x}) \times \pi(\mathbf{S}_k, U, \boldsymbol{\theta}), \\ &\propto \pi(\mathbf{y} | \mathbf{S}_n, \mathbf{x}, \boldsymbol{\theta}) \pi(\mathbf{x}, \mathbf{w}, k | \mathbf{S}_k, \boldsymbol{\theta}) \pi(\mathbf{S}_k | U, \boldsymbol{\theta}) \pi(U) \pi(\boldsymbol{\theta}) \\ &\propto (\tau^2)^{-n/2} \exp \left\{ -\frac{1}{2\tau^2} \sum_{i=1}^n (y_i - \mu - S(\mathbf{x}_i))^2 \right\} \times \\ &\quad \left[\prod_{i=1}^n \Phi \left(\frac{\beta \sqrt{U} S(\mathbf{x}_i)}{\sigma} \right) \right] \left[\prod_{i=n+1}^k \Phi \left(-\beta \sqrt{U} \frac{S(\mathbf{x}_i)}{\sigma} \right) \right] e^{-\lambda^* |B| (\lambda^*)^k} \times \\ &\quad U^{k/2} (\sigma^2)^{-k/2} |\mathbf{R}|^{-1/2} \exp \left\{ -\frac{U}{2\sigma^2} \mathbf{S}'_k \mathbf{R}^{-1} \mathbf{S}_k \right\} \\ &\quad \exp \left\{ -\frac{\nu}{2} U \right\} U^{\nu/2-1} \pi(\boldsymbol{\theta}),\end{aligned}$$

em que $\pi(\boldsymbol{\theta})$ é a distribuição a priori de $\boldsymbol{\theta}$.

Novamente vamos aproximar a distribuição a posteriori por métodos de amostragem, construindo um amostrador de Gibbs para gerar de cada distribuição condicional. As distribuições condicionais de λ^* , μ e τ^2 permanecem as mesmas do modelo Gaussiano. Para os demais parâmetros, as distribuições condicionais completas são dadas por

$$\begin{aligned}\pi(\mathbf{S}_k|\cdot) &\propto \Phi_n\left(\frac{\beta\sqrt{U}}{\sigma}\mathbf{S}_n;\mathbf{I}_n\right)\Phi_{k-n}\left(-\frac{\beta\sqrt{U}}{\sigma}\mathbf{S}_{k-n};\mathbf{I}_{k-n}\right)\pi(\mathbf{y}|\mathbf{S}_n,\mathbf{x},\boldsymbol{\theta})\pi(\mathbf{S}_k|\mathbf{w},U,\boldsymbol{\theta}), \\ \pi(\beta|\cdot) &\propto \Phi_n\left(\frac{\beta\sqrt{U}}{\sigma}\mathbf{S}_n;\mathbf{I}_n\right)\Phi_{k-n}\left(-\frac{\beta\sqrt{U}}{\sigma}\mathbf{S}_{k-n};\mathbf{I}_{k-n}\right)\pi(\beta), \\ \pi(\sigma^2|\cdot) &\propto \Phi_n\left(\frac{\beta\sqrt{U}}{\sigma}\mathbf{S}_n;\mathbf{I}_n\right)\Phi_{k-n}\left(-\frac{\beta\sqrt{U}}{\sigma}\mathbf{S}_{k-n};\mathbf{I}_{k-n}\right)\pi(\mathbf{S}_k|\mathbf{w},U,\boldsymbol{\theta})\pi(\sigma^2), \\ \pi(\phi|\cdot) &\propto \pi(\mathbf{S}_k|\mathbf{w},U,\boldsymbol{\theta})\pi(\phi), \\ \pi(U|\cdot) &\propto \pi(\mathbf{S}_k|\mathbf{w},U,\boldsymbol{\theta})\pi(U).\end{aligned}$$

Para geração de amostras de \mathbf{S}_k , β , σ^2 e ϕ , basta utilizar os Passos 2, 4, 6 e 7, considerando σ^2U^{-1} como variância de \mathbf{S}_k em suas respectivas distribuições condicionais. Em relação ao parâmetro U , sua distribuição condicional completa é dada por

$$\pi(U|\cdot) \propto \exp\left\{-\left(\frac{\mathbf{S}'_k\mathbf{R}^{-1}\mathbf{S}_k}{2\sigma^2} + \frac{\nu}{2}\right)U\right\}U^{\frac{(k+\nu)}{2}-1},$$

que é o núcleo de uma distribuição *Gama* $\left(\frac{(k+\nu)}{2}, \frac{\mathbf{S}'_k\mathbf{R}^{-1}\mathbf{S}_k}{2\sigma^2} + \frac{\nu}{2}\right)$ e é fácil de amostrar.

Considerando \mathbf{Y}_p e \mathbf{S}_p os processos Y e S nos \mathbf{x}_p locais não observados que queremos realizar predição, a distribuição preditiva de \mathbf{Y}_p será

$$\begin{aligned}\pi(\mathbf{Y}_p|\mathbf{x},\mathbf{y}) &= \int \pi(\mathbf{Y}_p,\mathbf{S}_p,\mathbf{S}_k,U,\boldsymbol{\theta}|\mathbf{x},\mathbf{y})d\mathbf{S}_pd\mathbf{S}_kdUd\boldsymbol{\theta} \\ &= \int \pi(\mathbf{Y}_p|\mathbf{S}_p,\boldsymbol{\theta})\pi(\mathbf{S}_p|\mathbf{S}_k,U,\boldsymbol{\theta})\pi(\mathbf{S}_k,U,\boldsymbol{\theta}|\mathbf{x},\mathbf{y})d\mathbf{S}_pd\mathbf{S}_kdUd\boldsymbol{\theta},\end{aligned}$$

em que $\pi(\mathbf{S}_k,U,\boldsymbol{\theta}|\mathbf{x},\mathbf{y})$ é a distribuição a posteriori de $(\mathbf{S}_k,U,\boldsymbol{\theta})$. Como $S|U$ é um processo Gaussiano, $(\mathbf{S}_p|\mathbf{S}_k,U,\boldsymbol{\theta}) \sim N_{n_p}(\boldsymbol{\mu}_{\mathbf{S}_p|\mathbf{S}_K},\sigma^2U^{-1}\mathbf{R}_{\mathbf{S}_p|\mathbf{S}_K})$, em que $\boldsymbol{\mu}_{\mathbf{S}_p|\mathbf{S}_K} = \mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{S}_k$, $\mathbf{R}_{\mathbf{S}_p|\mathbf{S}_K} = \mathbf{R}_{11} - \mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}$. Dados \mathbf{S}_p e $\boldsymbol{\theta}$, \mathbf{Y}_p possui distribuição normal com média $\mathbf{S}_p + \mu\mathbf{1}_{n_p}$ e matriz de covariâncias $\tau^2\mathbf{I}_{n_p}$. Logo, obter uma amostra da distribuição preditiva de Y equivale a, dada uma amostra de $(\mathbf{S}_k,\boldsymbol{\theta}|\mathbf{x},\mathbf{y})$, gerar uma amostra de $(\mathbf{S}_p|\boldsymbol{\theta})$

e, em seguida, obter uma amostra de $(\mathbf{Y}_p | \mathbf{S}_p, \boldsymbol{\theta})$.

Note que acomodamos observações atípicas tanto no processo de estimação quanto no de predição considerando caudas pesadas na distribuição finita de \mathbf{S} , mas mantemos a variância do erro de medida com distribuição normal. Isto implica que, mesmo para valores preditos, um valor atípico representa, provavelmente, um valor extremo do dado real e não de uma medida grosseiramente errada.

Segundo caso: processo Y t-Student

Nesse caso, o modelo Gaussiano sofre alteração na Equação (6.1) referente ao processo Y , tomando a seguinte forma:

$$\begin{aligned} \mathbf{Y} | X, S, U &\sim N_n(\mu \mathbf{1}_n + S(\mathbf{x}), \tau^2 U^{-1} \mathbf{I}_n), \\ U | \nu &\sim \text{Gama}(\nu/2, \nu/2), \quad \nu > 0. \end{aligned}$$

Note que marginalmente \mathbf{Y} possui distribuição t n -variada com ν gl. A distribuição a posteriori do modelo será

$$\begin{aligned} \pi(\mathbf{S}_k, U, \boldsymbol{\theta} | \mathbf{y}, \mathbf{x}) &\propto l(\mathbf{S}_k, U, \boldsymbol{\theta}; \mathbf{y}, \mathbf{x}) \times \pi(\mathbf{S}_k, U, \boldsymbol{\theta}), \\ &\propto \pi(\mathbf{y} | \mathbf{S}_n, \mathbf{x}, U, \boldsymbol{\theta}) \pi(\mathbf{x}, \mathbf{w}, k | \mathbf{S}_k, \boldsymbol{\theta}) \pi(\mathbf{S}_k | \boldsymbol{\theta}) \pi(U) \pi(\boldsymbol{\theta}) \\ &\propto U^{n/2} (\tau^2)^{-n/2} \exp \left\{ -\frac{U}{2\tau^2} \sum_{i=1}^n (y_i - \mu - S(\mathbf{x}_i))^2 \right\} \times \\ &\quad \left[\prod_{i=1}^n \Phi \left(\beta \frac{S(\mathbf{x}_i)}{\sigma} \right) \right] \left[\prod_{i=n+1}^k \Phi \left(-\beta \frac{S(\mathbf{x}_i)}{\sigma} \right) \right] e^{-\lambda^* |B|} (\lambda^*)^k \times \\ &\quad (\sigma^2)^{-k/2} |\mathbf{R}|^{-1/2} \exp \left\{ -\frac{1}{2\sigma^2} \mathbf{S}'_k \mathbf{R}^{-1} \mathbf{S}_k \right\} \\ &\quad \exp \left\{ -\frac{\nu}{2} U \right\} U^{\nu/2-1} \pi(\boldsymbol{\theta}), \end{aligned}$$

em que $\pi(\boldsymbol{\theta})$ é a distribuição a priori de $\boldsymbol{\theta}$.

Novamente vamos aproximar a distribuição a posteriori por métodos de amostragem, construindo um amostrador de Gibbs para gerar de cada distribuição condicional. As distribuições condicionais de λ^* , σ^2 , ϕ e β permanecem as mesmas do modelo Gaussiano.

Para os demais parâmetros, as distribuições condicionais completas são dados por

$$\begin{aligned}\pi(\mathbf{S}_k|\cdot) &\propto \Phi_n\left(\frac{\beta}{\sigma}\mathbf{S}_n; \mathbf{I}_n\right)\Phi_{k-n}\left(-\frac{\beta}{\sigma}\mathbf{S}_{k-n}; \mathbf{I}_{k-n}\right)\pi(\mathbf{y}|\mathbf{S}_n, \mathbf{x}, U, \boldsymbol{\theta})\pi(\mathbf{S}_k|\mathbf{w}, \boldsymbol{\theta}), \\ \pi(\mu, \tau^2|\cdot) &\propto \pi(\mathbf{y}|\mathbf{S}_n, \mathbf{x}, U, \boldsymbol{\theta})\pi(\mu, \tau^2), \\ \pi(U|\cdot) &\propto \pi(\mathbf{y}|\mathbf{S}_n, \mathbf{x}, U, \boldsymbol{\theta})\pi(U).\end{aligned}$$

Para geração de amostras de \mathbf{S}_k , μ e τ^2 , basta utilizar os Passos 2 e 5, considerando τ^2U^{-1} como variância de \mathbf{y} em suas respectivas distribuições condicionais. Em relação ao parâmetro U , sua distribuição condicional completa é dada por

$$\pi(U|\cdot) \propto \exp\left\{-\left(\frac{\sum_{i=1}^n(y_i - S(\mathbf{x}_i) - \mu)^2}{2\tau^2} + \frac{\nu}{2}\right)U\right\}U^{\frac{(n+\nu)}{2}-1},$$

que é o núcleo de uma distribuição *Gama* $\left(\frac{(k+\nu)}{2}, \frac{\sum_{i=1}^n(y_i - S(\mathbf{x}_i) - \mu)^2}{2\tau^2} + \frac{\nu}{2}\right)$ e é fácil de amostrar.

Considerando \mathbf{Y}_p e \mathbf{S}_p os processos Y e S nos \mathbf{x}_p locais não observados que queremos realizar predição, a distribuição preditiva de \mathbf{Y}_p será

$$\begin{aligned}\pi(\mathbf{Y}_p|\mathbf{x}, \mathbf{y}) &= \int \pi(\mathbf{Y}_p, \mathbf{S}_p, \mathbf{S}_k, U, \boldsymbol{\theta}|\mathbf{x}, \mathbf{y})d\mathbf{S}_pd\mathbf{S}_kdUd\boldsymbol{\theta} \\ &= \int \pi(\mathbf{Y}_p|\mathbf{S}_p, U, \boldsymbol{\theta})\pi(\mathbf{S}_p|\mathbf{S}_k, \boldsymbol{\theta})\pi(\mathbf{S}_k, U, \boldsymbol{\theta}|\mathbf{x}, \mathbf{y})d\mathbf{S}_pd\mathbf{S}_kdUd\boldsymbol{\theta},\end{aligned}$$

em que $\pi(\mathbf{S}_k, U, \boldsymbol{\theta}|\mathbf{x}, \mathbf{y})$ é a distribuição a posteriori de $(\mathbf{S}_k, U, \boldsymbol{\theta})$. Como S é um processo Gaussiano, a distribuição de $\mathbf{S}_p|\mathbf{S}_k, \boldsymbol{\theta}$ será normal com mesmos parâmetros do caso Gaussiano. Dados \mathbf{S}_p, U e $\boldsymbol{\theta}$, \mathbf{Y}_p possui distribuição normal com média $\mathbf{S}_p + \mu\mathbf{1}_{n_p}$ e matriz de covariâncias $\tau^2U^{-1}\mathbf{I}_{n_p}$. Logo, obter uma amostra da distribuição preditiva de Y equivale a, dada uma amostra de $(\mathbf{S}_k, \boldsymbol{\theta}|\mathbf{x}, \mathbf{y})$, gerar uma amostra de $(\mathbf{S}_p|\boldsymbol{\theta})$ e, em seguida, obter uma amostra de $(\mathbf{Y}_p|\mathbf{S}_p, \boldsymbol{\theta})$.

Observe que observações atípicas são acomodadas tanto no processo de estimação quanto no de predição considerando caudas pesadas na distribuição de \mathbf{Y} . Isto implica que, mesmo para valores preditos, um valor atípico representa, provavelmente, uma medida grosseiramente errada e não um valor extremo do dado real. Obviamente, não conhecemos a magnitude desse erro, pois não possuímos observações do processo S . Dessa forma, vale discutir com o pesquisador responsável da área para analisar se o *outlier* é

um dado real ou um dado viesado, devendo incorporá-lo no modelo ou retirá-lo da amostra.

Capítulo 7

Inferência Bayesiana - Estudo de simulação

Apresentamos a seguir estudos de simulações para avaliar a metodologia apresentada. O estudo foi dividido em etapas, cada uma com o objetivo de estudar o comportamento do modelo em determinadas situações. Essas etapas são:

- 1: Comparar os resultados de estimação e predição obtidos pelos modelos com amostragem preferencial e sem amostragem preferencial a partir de dados simulados provenientes de amostragem preferencial e não preferencial;
- 2: Analisar o efeito de fixar o parâmetro de correlação do modelo;
- 3: Estudar o comportamento do modelo quando há pouca informação;
- 4: Avaliar o modelo quando os dados são gerados a partir da função de intensidade exponencial;
- 5: Analisar o processo de estimação quando consideramos a função de intensidade probito com intercepto;
- 6: Comparar os resultados obtidos pelo modelo exato com o modelo aproximado por discretização;
- 7: Apresentar alguns resultados sobre a extensão para o modelo t-Student.

Para realizar as simulações, foi considerado o quadrado unitário $[0, 1] \times [0, 1]$ como região de estudo. A menos que se mencione o contrário, assumimos que S é um processo Gaussiano com função de correlação exponencial, $\sigma^2 = 3$ e $\phi = 0,15$, consideramos $\beta = 2$ e $\lambda^* = 150$ para a função de intensidade e $\mu = 4$ e $\tau^2 = 0,10$ para o processo Y . No caso da simulação de dados sem amostragem preferencial, foi assumido $\beta = 0$.

As distribuições a priori consideradas foram: $\lambda^* \sim \text{Gama}T(0.001, 0.001, 250)$, $\mu \sim N(0, 10^6)$, $\tau^2 \sim GI(0.001, 0.001)$, $\sigma^2 \sim GI(0.001, 0.001)$, $\phi \sim \text{Gama}(2, 4)$ e $\beta \sim N(0, 1)$. A computação foi realizada utilizando a linguagem R (R Core Team, 2018) integrada a linguagem C++ (Eddelbuettel and François, 2011).

7.1 Amostragem preferencial e não preferencial

Para realizar esse estudo de simulação, foram considerados 10 conjuntos de dados simulados. Desse modo, evitamos obter resultados apenas para uma amostra específica, que pode ser ruim ou boa para o modelo. A Tabela 7.1 apresenta o número de observações obtidas para cada conjunto de dados.

Tabela 7.1: Número de observações obtidas em cada conjunto de dados.

Conjunto de dados	1	2	3	4	5	6	7	8	9	10
$\beta = 2$	58	63	89	72	85	80	75	108	55	87
$\beta = 0$	67	70	77	87	73	70	78	80	66	70

A Figura 7.1 apresenta o mapa do processo Gaussiano simulado e os locais amostrados considerando amostragem preferencial e não preferencial do conjunto de dados 1. Fica claro que, na amostragem não preferencial, os locais escolhidos estão espalhados de forma homogênea enquanto que, na amostragem preferencial (no caso com $\beta = 2$ positivo), os locais selecionados tendem a ter maiores valores do processo Gaussiano.

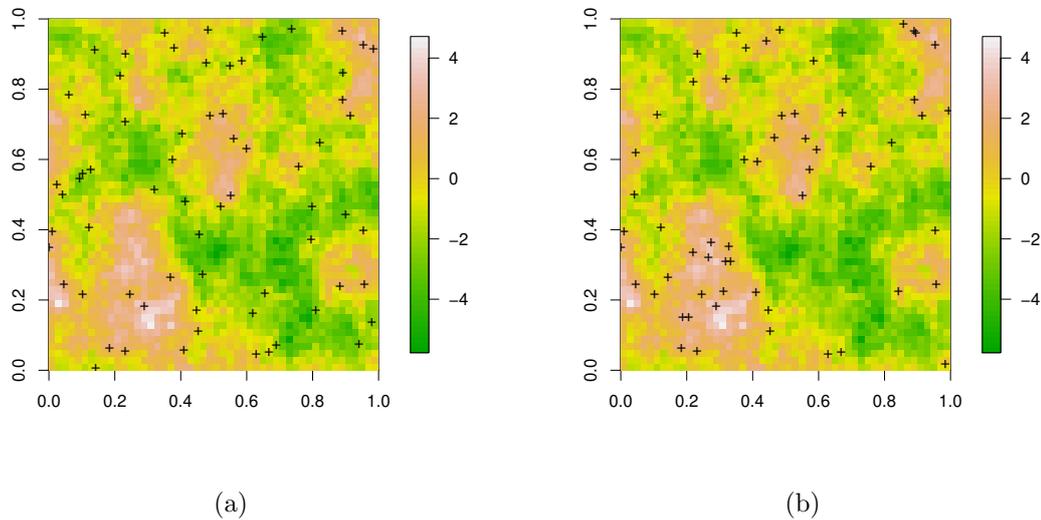


Figura 7.1: Mapa do processo Gaussiano simulado do modelo (a) não preferencial ($\beta = 0$) e (b) preferencial ($\beta = 2$) do conjunto de dados 1.

A Figura 7.2 apresenta o mapa da função de intensidade no quadrado unitário, considerando amostragem não preferencial e preferencial para o conjunto de dados 1. No caso da amostragem não preferencial temos um processo de Poisson homogêneo, por isso todos os pontos possuem o mesmo valor de $\lambda(x)$. No caso da amostragem preferencial, é

interessante notar que, para valores de S acima de 3 ou abaixo de -3, o valor da função de intensidade fica próximo dos extremos (λ^* e 0, respectivamente). Isso se deve pela característica da função probito, que tem valores próximos de zero para argumentos menores que -3 e próximos de 1 para argumentos maiores que 3.

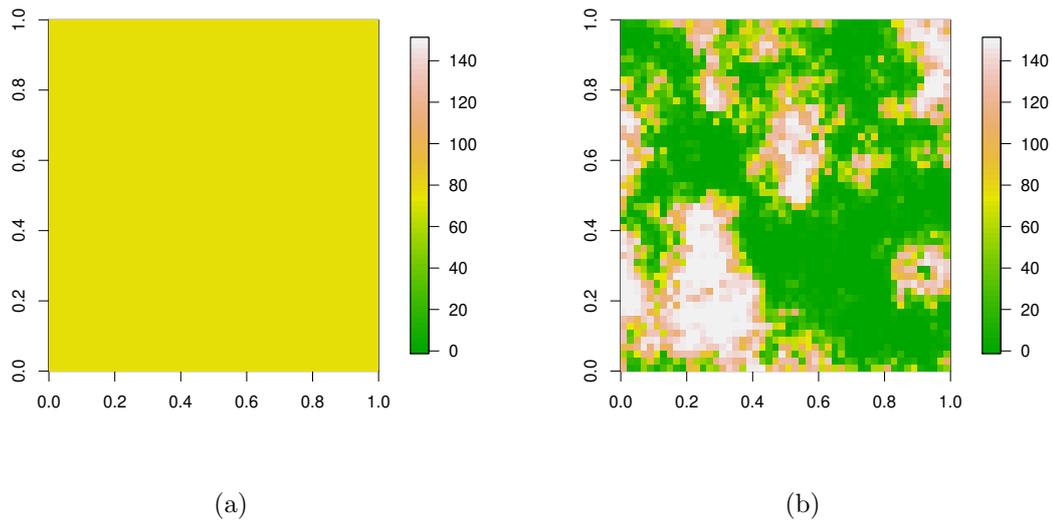


Figura 7.2: Mapa da função de intensidade $\lambda(x)$ simulada do modelo (a) não preferencial e (b) preferencial do conjunto de dados 1.

A partir dos conjuntos de dados simulados, procedeu-se à etapa de inferência.

Dados simulados com amostragem preferencial

Para a estimação dos parâmetros, foram monitoradas duas cadeias com 200 mil iterações, das quais descartamos as 20 mil iniciais e a amostra a posteriori foi constituída pelas iterações restantes com *lag* de 60, resultando em uma amostra de tamanho 6 mil. A convergência das cadeias de Markov foi verificada por análise gráfica e os gráficos das trajetórias dos parâmetros para o conjunto de dados 1 podem ser observados no Apêndice B.

A Figura 7.3 apresenta as densidades das distribuições a posteriori dos parâmetros estimados pelo modelo com amostragem preferencial (AP). Podemos observar que a estimação do parâmetro λ^* é satisfatória, com grande densidade a posteriori em torno de seu valor verdadeiro. Percebemos também que não há grande densidade a posteriori na região de truncamento, ou seja, o truncamento em 250 não afetou negativamente a es-

timação do parâmetro, aumentando a eficiência do MCMC. Para os demais parâmetros, obtemos estimação satisfatória, com altas densidades a posteriori em torno de seus valores verdadeiros. Vale notar que o parâmetro de preferenciabilidade β foi bem recuperado em metade dos conjuntos de dados considerados, sendo que nos demais obtemos estimativas subestimadas porém maiores que zero (na maior densidade da distribuição a posteriori), indicando a presença de amostragem preferencial.

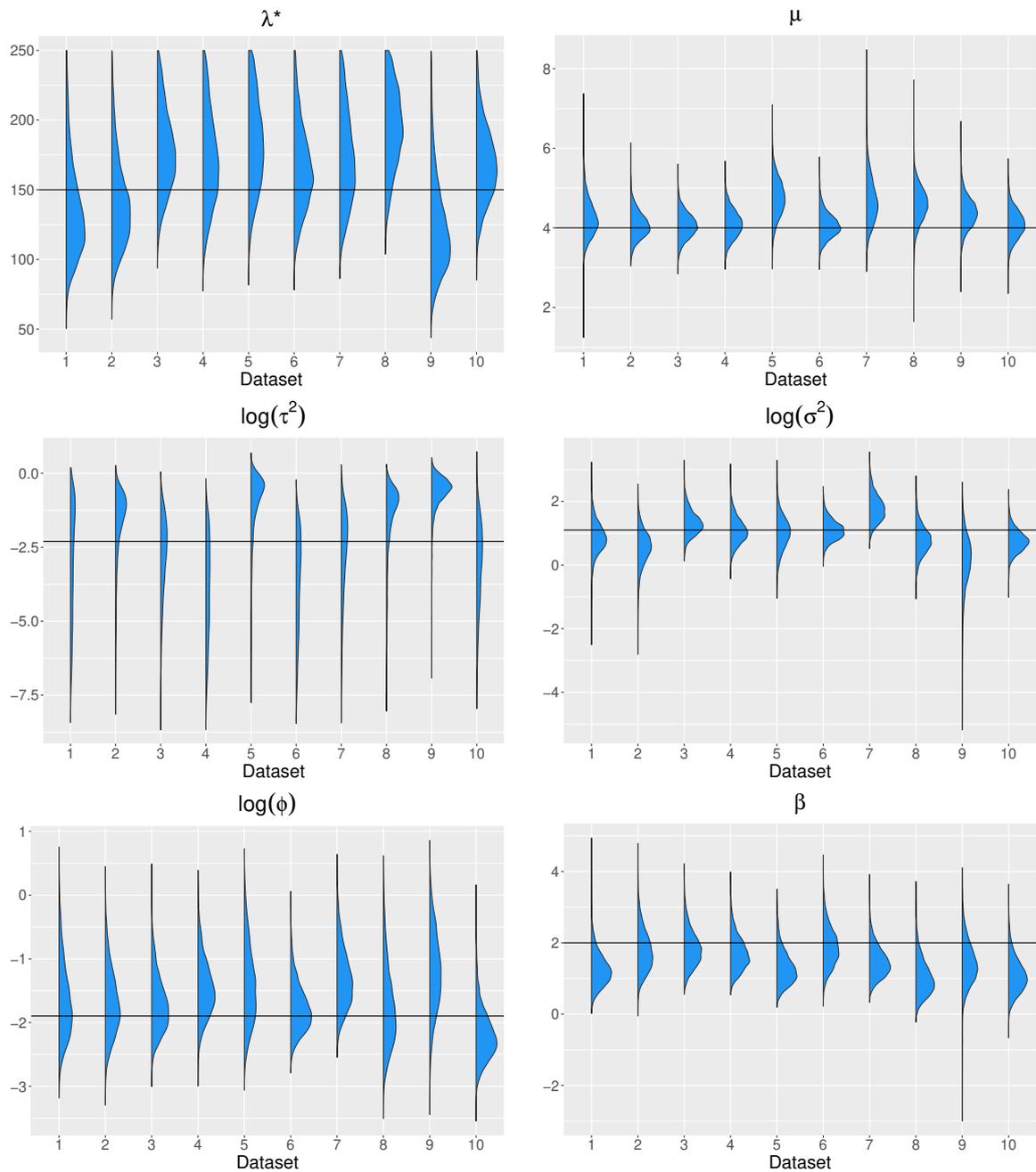


Figura 7.3: Densidades das distribuições a posteriori dos parâmetros do modelo com amostragem preferencial para os dados simulados com amostragem preferencial. A linha horizontal representa o valor verdadeiro do parâmetro.

Através do gráfico de dispersão bidimensional dos parâmetros, podemos analisar como estes se relacionam no modelo. A Figura 7.4 apresenta o gráfico para alguns pares de parâmetros. Observamos uma relação positiva entre σ^2 e ϕ (que já é conhecida na literatura) e entre σ^2 e β , que estão relacionados pela Equação 6.3 do modelo. Considerando que os dados possuem amostragem preferencial, a relação entre μ e β é explicada pelo fato de que, menores valores de β (e próximos de zero) indicam menor força de preferência, o que acarreta na superestimação de μ (em módulo). Valores mais altos de β vão proporcionar uma melhor correção nas estimativas de μ . Notamos também uma relação positiva entre μ e λ^* . Essa relação precisa ser melhor investigada em estudos posteriores.

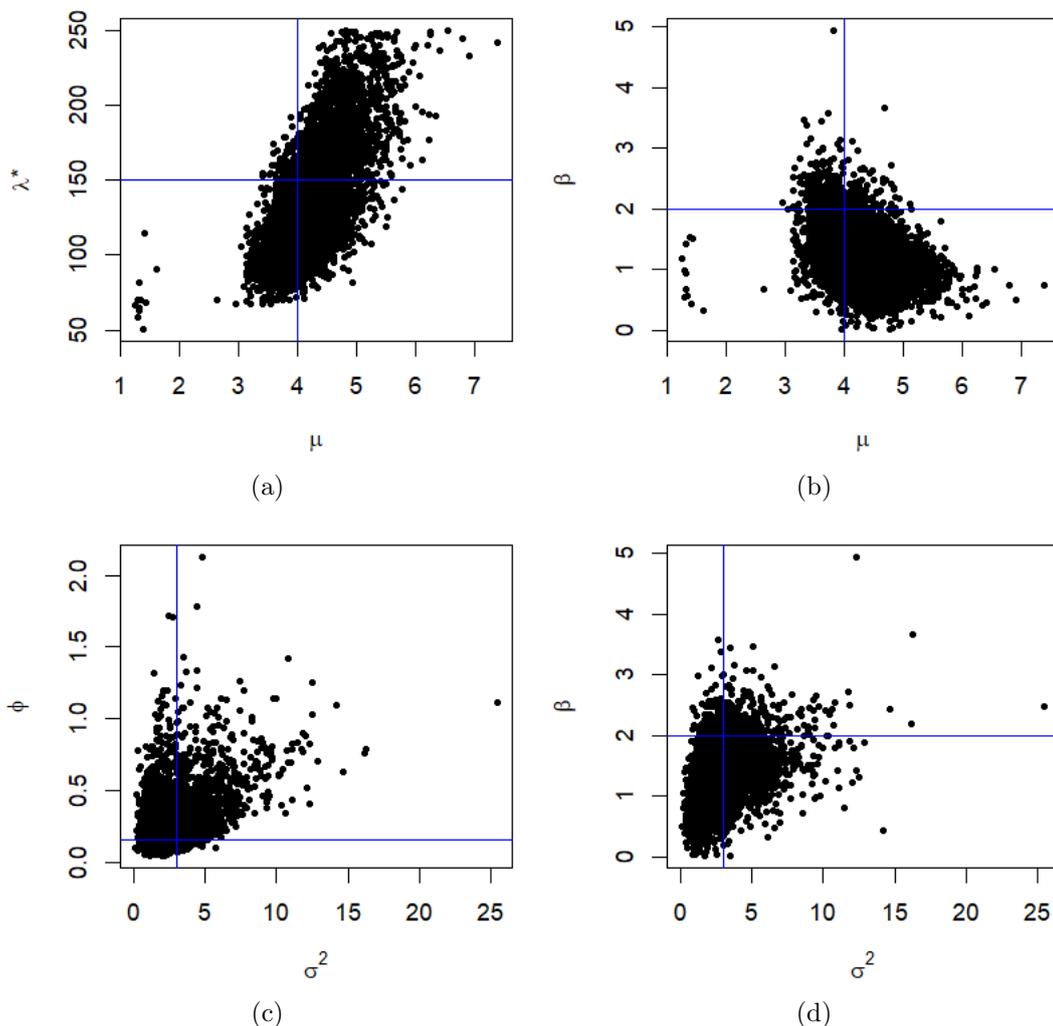


Figura 7.4: Gráficos de dispersão de alguns pares de parâmetros da amostra a posteriori para os dados simulados do conjunto de dados 1.

A Figura 7.5 apresenta as densidades das distribuições a posteriori dos parâmetros estimados pelo modelo sem amostragem preferencial (SAP). Podemos observar que o

parâmetro μ é superestimado na maioria dos conjuntos de dados. Isso ocorre porque a amostra possui, em sua maioria, valores altos do processo e o modelo não consegue captar essa característica. Apesar disso, os demais parâmetros tem estimação satisfatória.

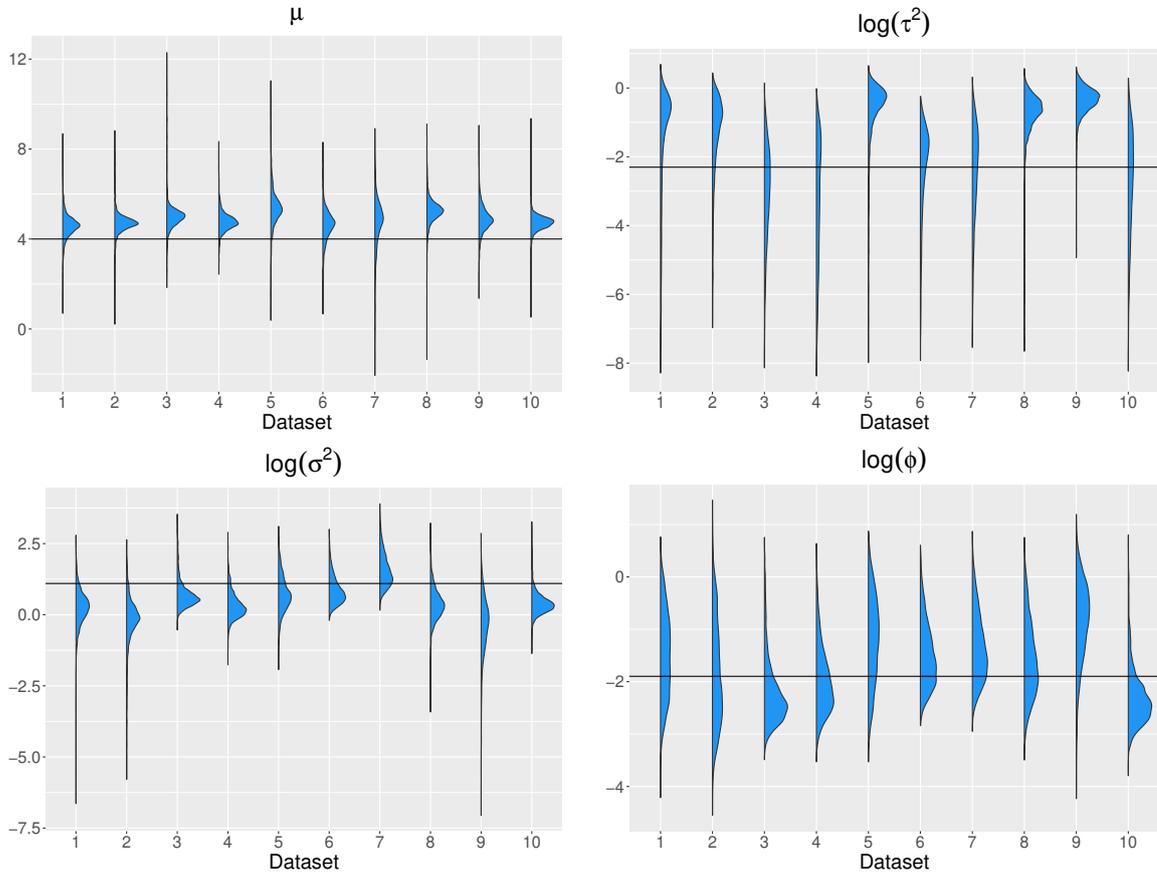


Figura 7.5: Densidades das distribuições a posteriori dos parâmetros do modelo sem amostragem preferencial para os dados simulados com amostragem preferencial. A linha horizontal representa o valor verdadeiro do parâmetro.

A Figura 7.6 apresenta a função de intensidade estimada pelo modelo em função do processo S . Observamos que a curva estimada acompanha a curva real, porém subestima os maiores valores da taxa tanto para a média quanto para a mediana a posteriori. Notamos grande incerteza através da banda de 95% de credibilidade, porém a curva real se encontra dentro da banda em toda trajetória. Através do mapa estimado da função de intensidade (Figura 7.7), podemos observar valores subestimados nas regiões de alta intensidade, comparado com a Figura 7.2 (b), e valores superestimados em regiões de baixa intensidade. Apesar disso, identificamos um padrão semelhante dos mapas. Observamos também que o desvio padrão a posteriori da função de intensidade é menor nos locais amostrados, como esperado.

Para analisar a qualidade preditiva dos modelos, foi considerado um local com valor

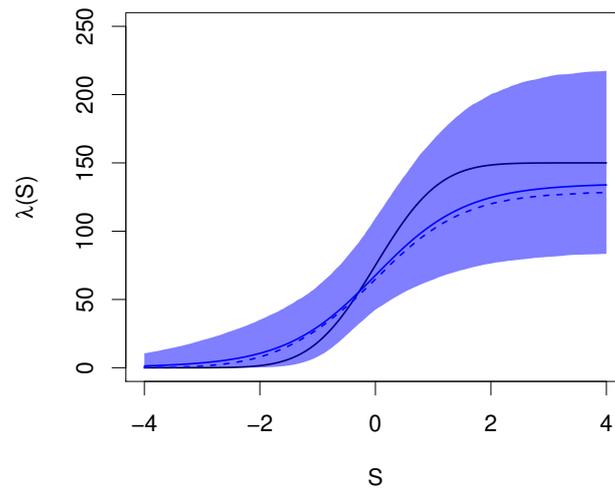


Figura 7.6: Gráfico da função de intensidade estimada pela média (linha contínua) e mediana (linha tracejada) a posteriori e banda de 95% de credibilidade para os dados simulados do conjunto de dados 1.

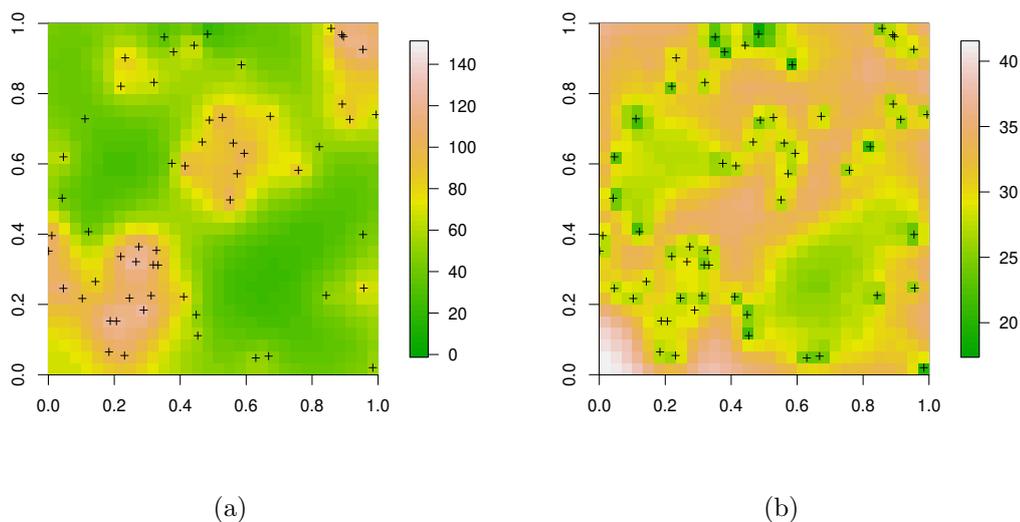


Figura 7.7: Mapas (a) da média a posteriori e (b) do desvio padrão a posteriori da função de intensidade estimada para os dados simulados do conjunto de dados 1.

baixo do processo S que não possui locais amostrados próximos. Esse ponto foi escolhido para cada conjunto de dados simulados, denotado por x_p , e a Figura 7.8 apresenta o local do ponto escolhido para os conjuntos de dados 1 e 2. Foram considerados também dois locais fixos para todos os conjuntos de dados: $x_{p1} = (0,49; 0,49)$ e $x_{p2} = (0,91; 0,09)$. Foram realizadas as previsões para o processo Gaussiano S e para o processo Y para cada um dos conjuntos de dados e calculado o viés de previsão, dado pela subtração do valor predito pelo valor verdadeiro do processo. A Figura 7.9 apresenta o intervalo de

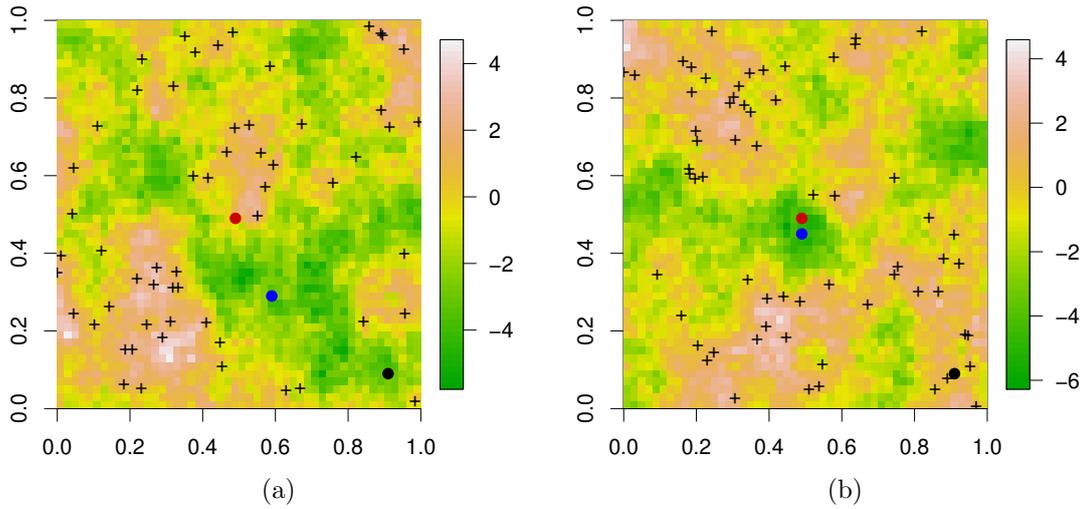


Figura 7.8: Mapa do processo S dos conjuntos de dados (a) 1 e (b) 2 para os dados simulados com amostragem preferencial. Os pontos em azul, vermelho e preto indicam os locais de x_p , x_{p_1} e x_{p_2} , respectivamente.

credibilidade de 95% para o viés de predição do processo S e do processo Y nos três locais de predição, considerando os modelos AP e SAP. Podemos observar que a predição realizada pelo modelo AP é mais satisfatória que a realizada pelo modelo SAP, pois o IC cobre o viés nulo em mais conjuntos de dados, principalmente para o local x_p . Nesse local, notamos que, dos dez conjuntos de dados, apenas um IC do viés de predição sob o modelo sem amostragem preferencial cobre o valor nulo, ao passo que, sob nosso modelo com amostragem preferencial, temos 6 intervalos que contém o zero. Além disso, notamos que o viés médio está mais próximo de zero no modelo AP, indicando melhor predição.

A partir de uma malha regular de tamanho 30×30 , foi realizada a predição do processo Gaussiano S e do processo Y para os modelos com amostragem preferencial e sem amostragem preferencial. A partir dos valores preditos para cada local da malha, foram calculados a raiz quadrada do Erro Quadrático Médio de Predição (REQMP)

$$REQMP = \sqrt{\frac{1}{n_p} \sum_{i=1}^{n_p} (\hat{S}(x_i) - S(x_i))^2}$$

e a Proporção de Cobertura dos Intervalos de Predição (PCIP) de 95%

$$PCIP = \frac{1}{n_p} \sum_{i=1}^{n_p} \mathbb{1}\{S(x_i) \in IC(S(x_i), 95\%)\}.$$

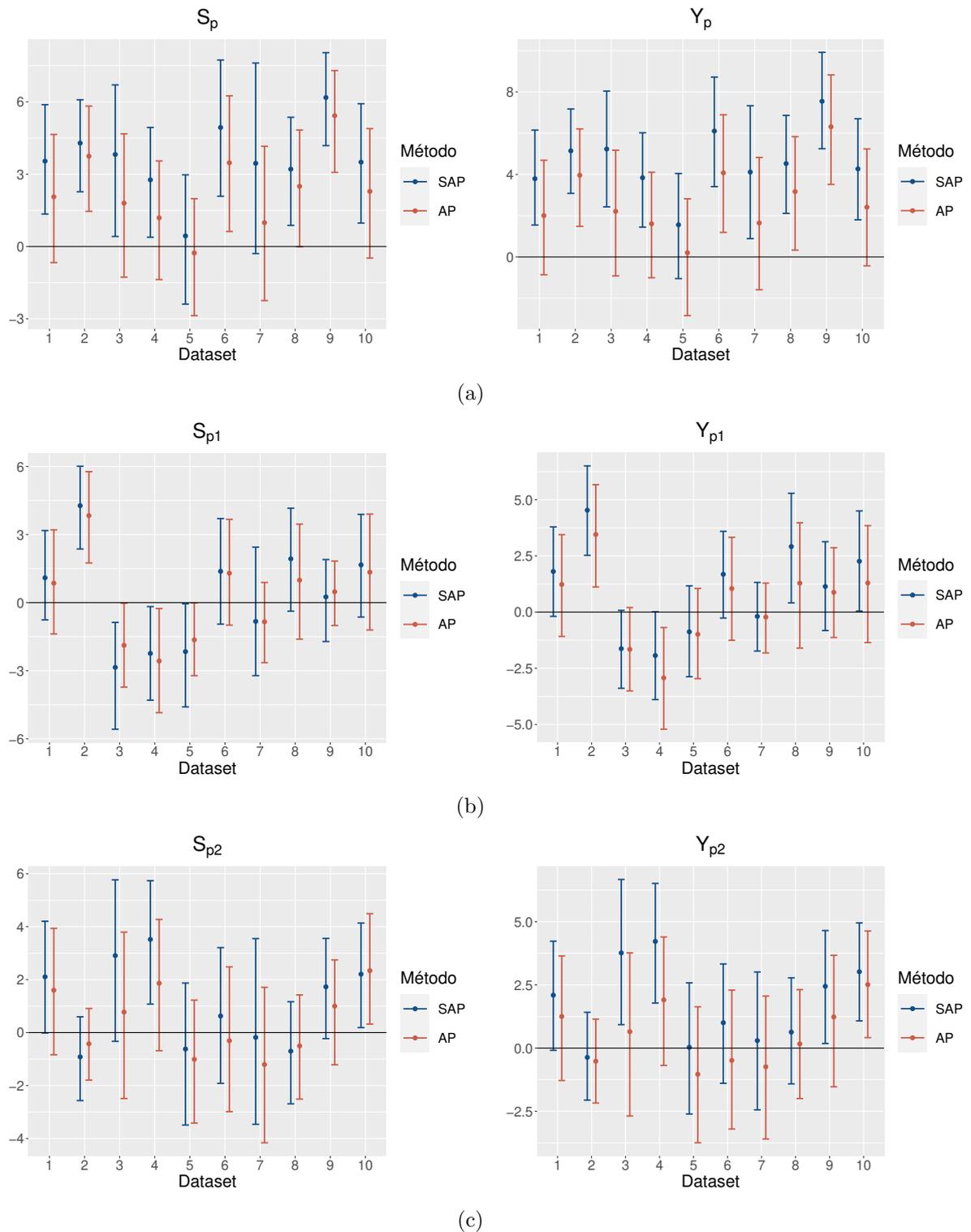


Figura 7.9: Intervalos de credibilidade de 95% do viés de predição dos processos S e Y nos locais (a) x_p , (b) $x_{p1} = (0, 49; 0, 49)$ e (c) $x_{p2} = (0, 91; 0, 09)$, considerando os modelos AP e SAP para os dados simulados com amostragem preferencial.

Para medidas globais, integramos as medidas locais sobre a região de interesse, no caso, dadas pela média das medidas locais, resultando em REQMP Global (REQMPG) e PCIP

Global (PCIPG). As mesmas quantidades são utilizadas para analisar a qualidade de predição do processo Y . Os resultados são apresentados nas Figuras 7.10 e 7.11. Em

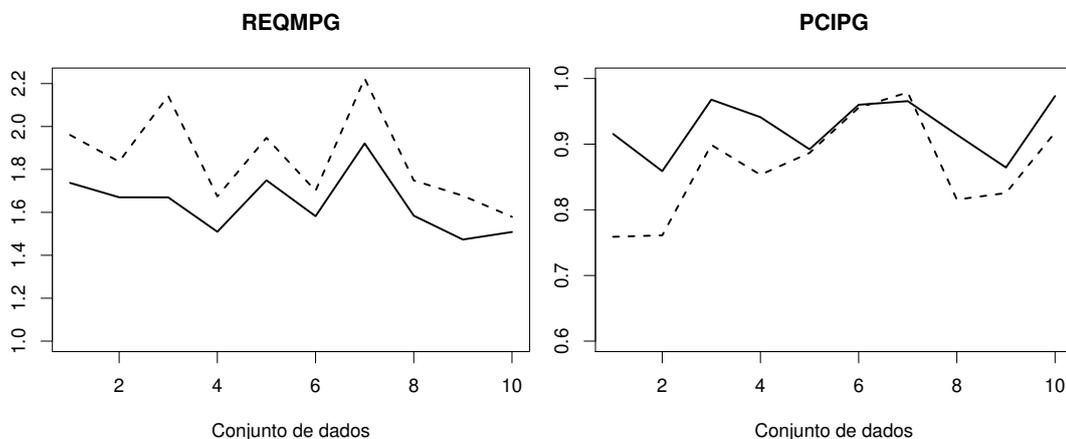


Figura 7.10: Medidas de qualidade de predição para o processo S dos modelos AP (linha contínua) e SAP (linha tracejada) considerando os conjuntos de dados simulados com amostragem preferencial.

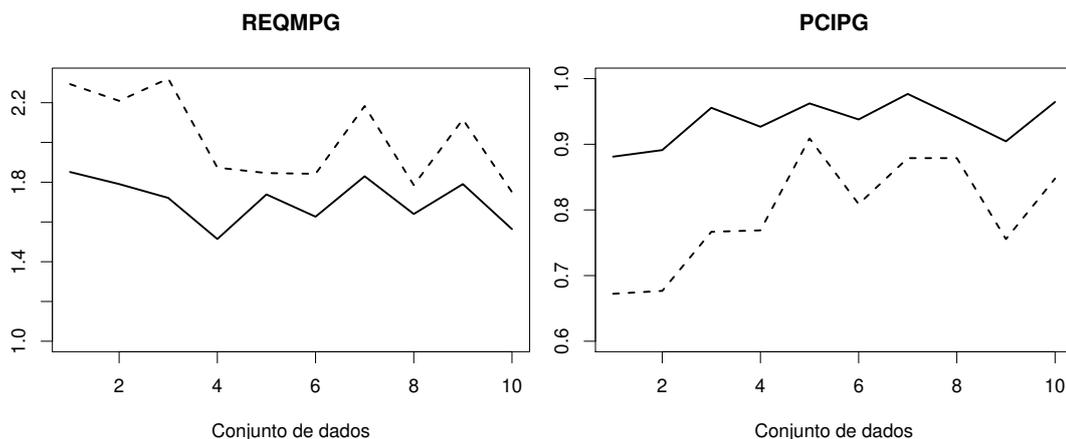


Figura 7.11: Medidas de qualidade de predição para o processo Y dos modelos AP (linha contínua) e SAP (linha tracejada) considerando os conjuntos de dados simulados com amostragem preferencial.

relação à predição do processo latente S , podemos observar que o modelo AP possui menor EQMPG para todos os conjuntos de dados e maior PCIPG para nove dos dez conjuntos de dados (em torno do valor nominal), indicando melhor predição. Quando observamos a predição realizada para o processo Y , o modelo AP se sobressai em relação ao modelo SAP em todos os conjuntos de dados, apresentando menor viés e melhor cobertura de intervalos de predição. Isso indica que o modelo AP é melhor que o modelo SAP na presença de amostragem preferencial em relação à predição, que é o principal objetivo da geoestatística.

As Figuras 7.12 e 7.13 apresentam os mapas estimados dos processos S e Y do conjunto de dados 1 para ambos os modelos. Como o modelo SAP não considera a

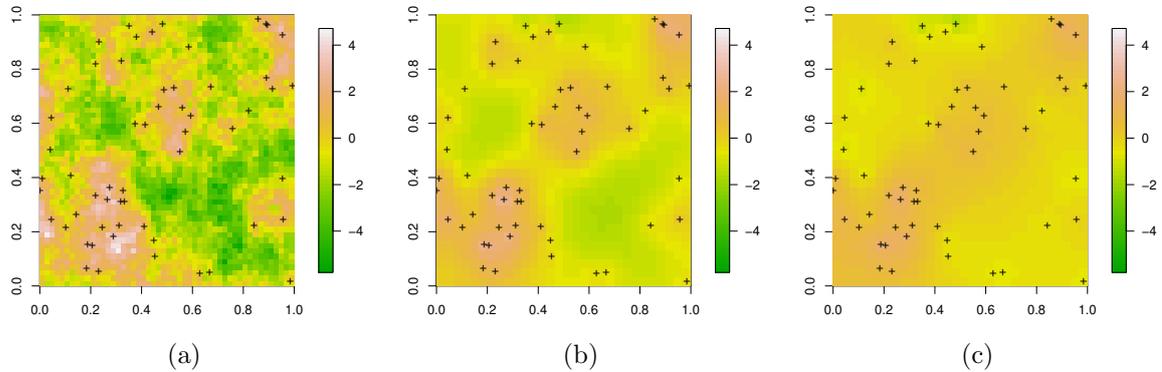


Figura 7.12: Mapa do processo S (a) original, (b) estimado pelo modelo com amostragem preferencial e (c) estimado pelo modelo sem amostragem preferencial, ambos utilizando a mediana da distribuição preditiva a posteriori para o conjunto de dados 1.

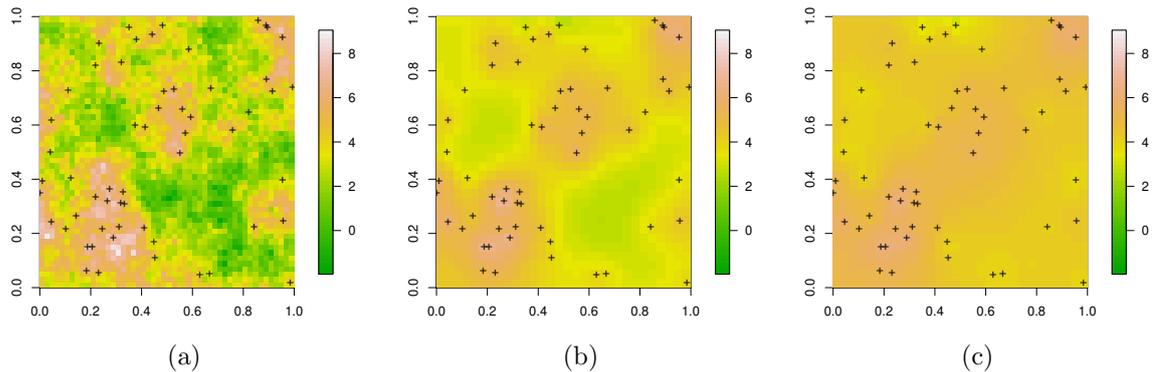


Figura 7.13: Mapa do processo Y (a) original, (b) estimado pelo modelo com amostragem preferencial e (c) estimado pelo modelo sem amostragem preferencial, ambos utilizando a mediana da distribuição preditiva a posteriori para o conjunto de dados 1.

informação do processo pontual na predição, sua predição não é capaz de recuperar os menores valores do processo latente S . Esse efeito é acentuado quando consideramos o processo Y , em que o mapa possui apenas cores que correspondem a valores altos. Já para o modelo AP, observamos no mapa regiões de menores valores dos processos preditos, tanto para S quanto para Y , se aproximando melhor dos processos originais principalmente em áreas de pouca informação.

Dados simulados sem amostragem preferencial

Para a estimação dos parâmetros, foram monitoradas duas cadeias com 200 mil iterações, das quais descartamos as 20 mil iniciais e a amostra a posteriori foi constituída pelas iterações restantes com *lag* de 60, resultando em uma amostra de tamanho 6 mil.

A Figura 7.14 apresenta as densidades das distribuições a posteriori dos parâmetros estimados pelo modelo AP. Podemos observar que a estimação do parâmetro λ^* é satisfatória, com grande densidade a posteriori em torno de seu valor verdadeiro. Para os demais parâmetros, obtemos estimação satisfatória, com altas densidades a posteriori em torno de seus valores verdadeiros. Vale notar que o parâmetro de preferenciabilidade β foi bem recuperado em todos os conjuntos de dados, isto é, o modelo foi capaz de identificar que os dados não provém de uma amostragem preferencial (β igual a zero).

A Figura 7.15 apresenta as densidades das distribuições a posteriori dos parâmetros estimados pelo modelo SAP. Podemos observar que o parâmetro μ é estimado de forma satisfatória, porém possui maior variabilidade se comparamos com a estimação obtida pelo modelo AP. Os demais parâmetros tem estimação satisfatória.

A Figura 7.16 apresenta a função de intensidade estimada pelo modelo em função do processo S . Observamos que a banda de 95% de credibilidade contém a curva real em toda trajetória. Como estamos lidando com incertezas, era esperado que nossa função de intensidade estimada apresentasse inclinação no caso de amostragem não preferencial.

Para analisar a qualidade preditiva dos modelos, foi considerado um local com valor baixo do processo S que não possui locais amostrados próximos. Esse ponto foi escolhido para cada conjunto de dados simulados, denotado por x_p , e a Figura 7.17 apresenta o local do ponto escolhido para os conjuntos de dados 1 e 2. Foram considerados também dois locais fixos para todos os conjuntos de dados: $x_{p1} = (0,49; 0,49)$ e $x_{p2} = (0,91; 0,09)$. A predição foi realizada e calculou-se o viés de predição para o processo S e o processo Y , dado pela subtração do valor predito pelo valor verdadeiro do processo. A Figura 7.18 apresenta o intervalo de credibilidade de 95% para o viés de predição do processo S e do processo Y nos três locais de predição, considerando os modelos AP e SAP. Podemos observar que a predição é bem semelhante para ambos modelos, com a maioria dos intervalos contendo o valor de viés igual a zero.

A partir de uma malha regular 30×30 , foram calculadas as medidas de qualidade

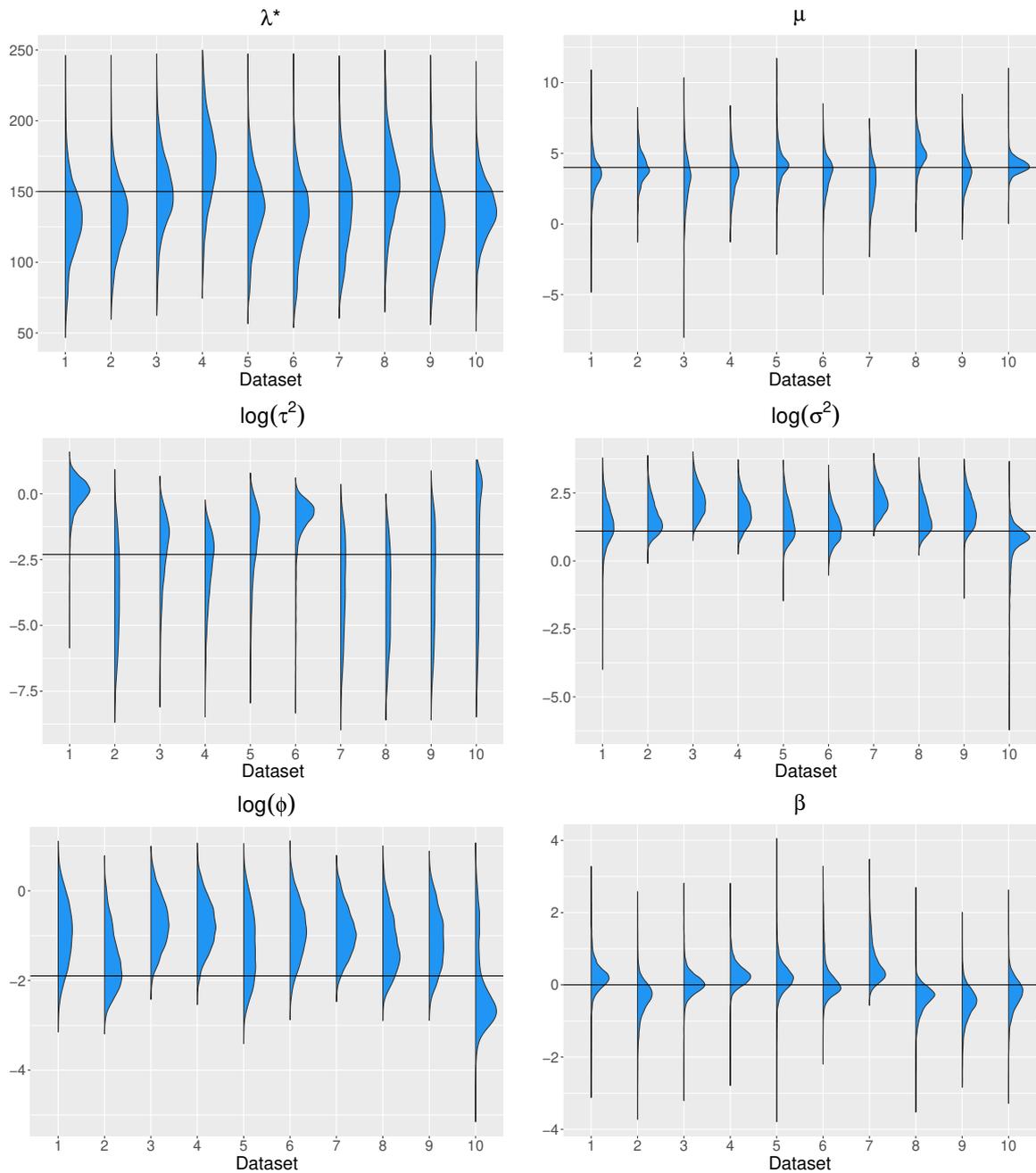


Figura 7.14: Densidades das distribuições a posteriori dos parâmetros do modelo com amostragem preferencial para os dados simulados sem amostragem preferencial. A linha horizontal representa o valor verdadeiro do parâmetro.

de predição REQMPG e PCIPG para ambos modelos. Os resultados são apresentados nas Figuras 7.19 e 7.20. Observamos que os valores são semelhantes para ambos os modelos, tanto para o processo S quanto para o processo Y , indicando que o modelo AP possui poder de capacidade preditiva similar a do modelo SAP na ausência de amostragem preferencial. As Figuras 7.21 e 7.22 apresentam os mapas dos processos S e Y preditos por ambos modelos. Notamos que a predição é equivalente quanto utilizamos o modelo

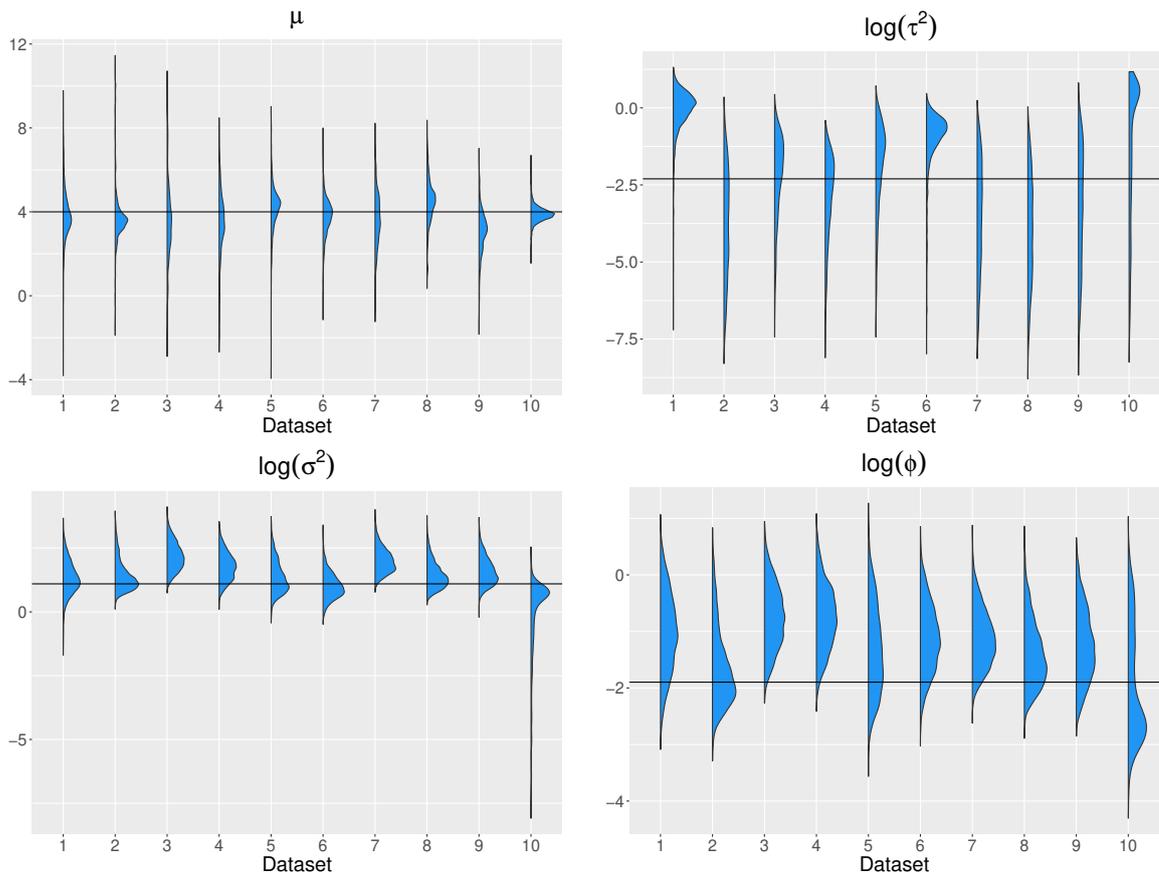


Figura 7.15: Densidades das distribuições a posteriori dos parâmetros do modelo sem amostragem preferencial para os dados simulados com amostragem preferencial. A linha horizontal representa o valor verdadeiro do parâmetro.

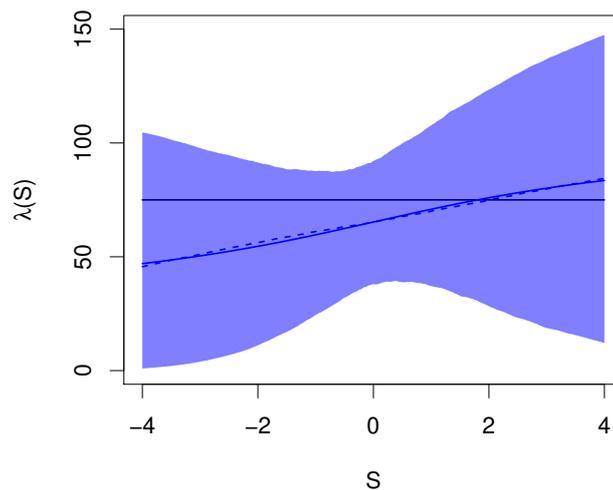


Figura 7.16: Gráfico da função de intensidade estimada pela média (linha contínua) e mediana (linha tracejada) a posteriori e banda de 95% de credibilidade para os dados simulados sem amostragem preferencial do conjunto de dados 1.

com amostragem preferencial ou sem.

Concluimos, nessa análise, que nosso modelo com amostragem preferencial se so-

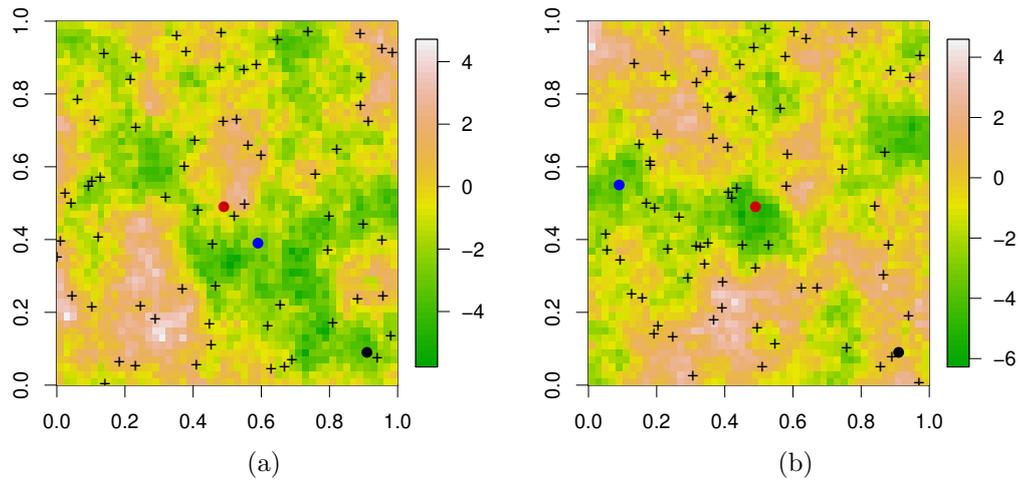


Figura 7.17: Mapa do processo S dos conjuntos de dados (a) 1 e (b) 2 para os dados simulados sem amostragem preferencial. Os pontos em azul, vermelho e preto indicam os locais de x_p , x_{p_1} e x_{p_2} , respectivamente.

bressai, tanto na estimação dos parâmetros quanto na predição, quando os dados provém de amostragem preferencial. No caso da ausência de preferenciabilidade na amostra, nosso modelo tem melhor estimação dos parâmetros e capacidade preditiva tão boa quanto o modelo sem amostragem preferencial.

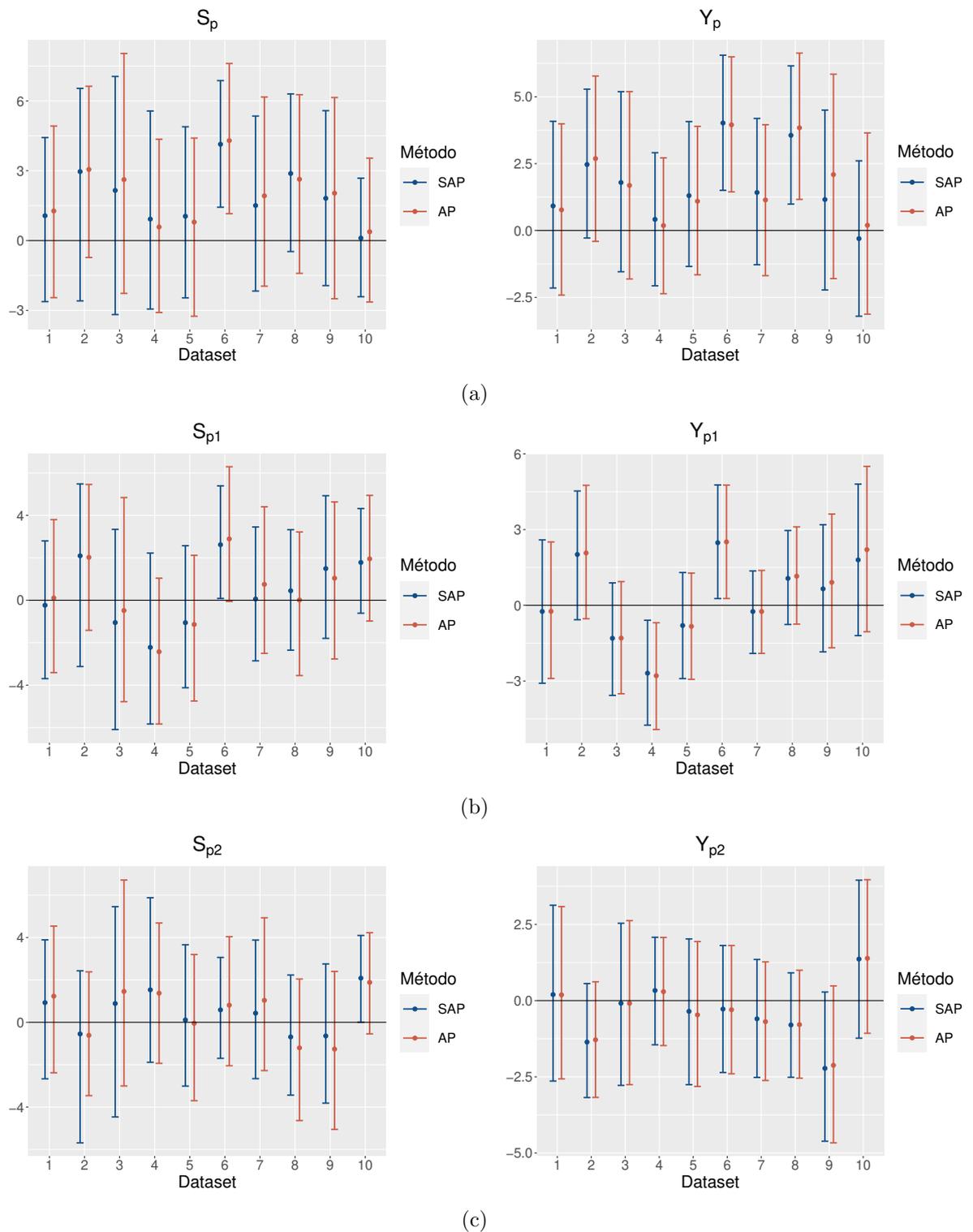


Figura 7.18: Intervalos de credibilidade de 95% do viés de predição dos processos S e Y nos locais (a) x_p , (b) $x_{p1} = (0, 49; 0, 49)$ e (c) $x_{p2} = (0, 91; 0, 09)$, considerando os modelos AP e SAP para os dados simulados sem amostragem preferencial.

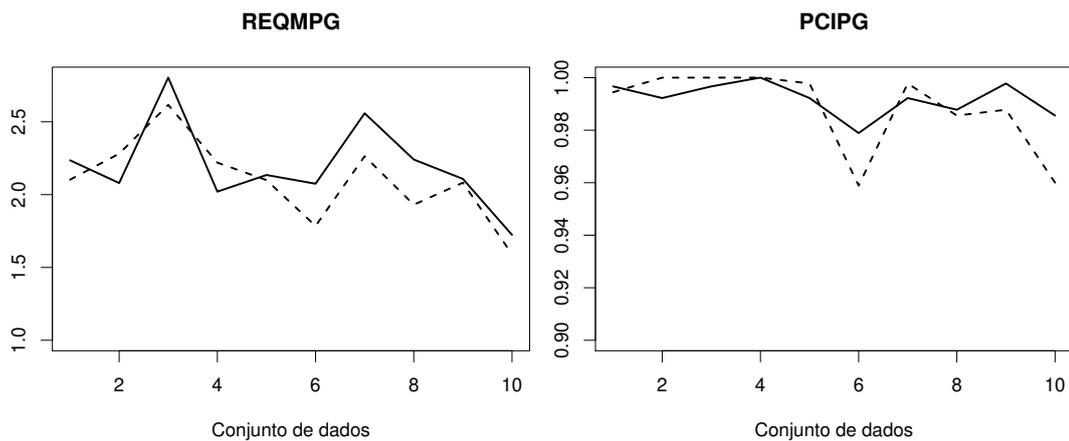


Figura 7.19: Medidas de qualidade de predição para o processo S dos modelos AP (linha contínua) e SAP (linha tracejada) considerando os conjuntos de dados simulados sem amostragem preferencial.

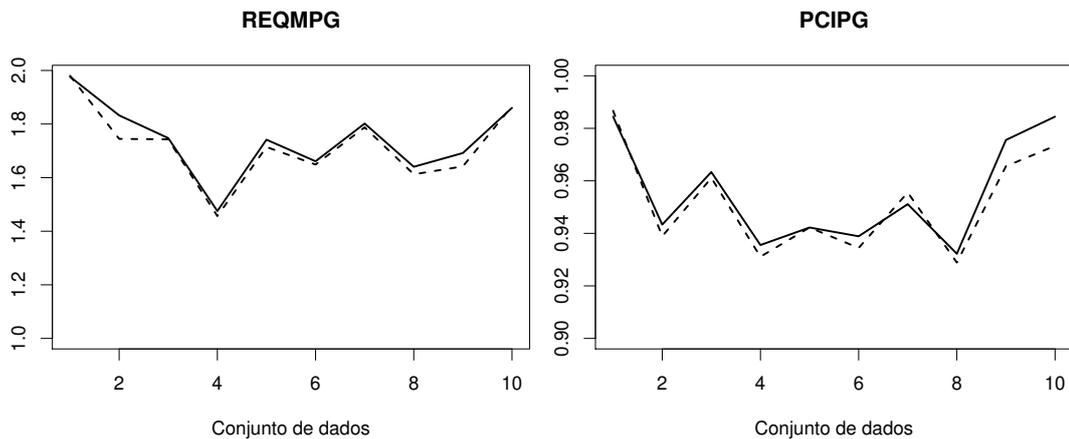


Figura 7.20: Medidas de qualidade de predição para o processo Y dos modelos AP (linha contínua) e SAP (linha tracejada) considerando os conjuntos de dados simulados sem amostragem preferencial.

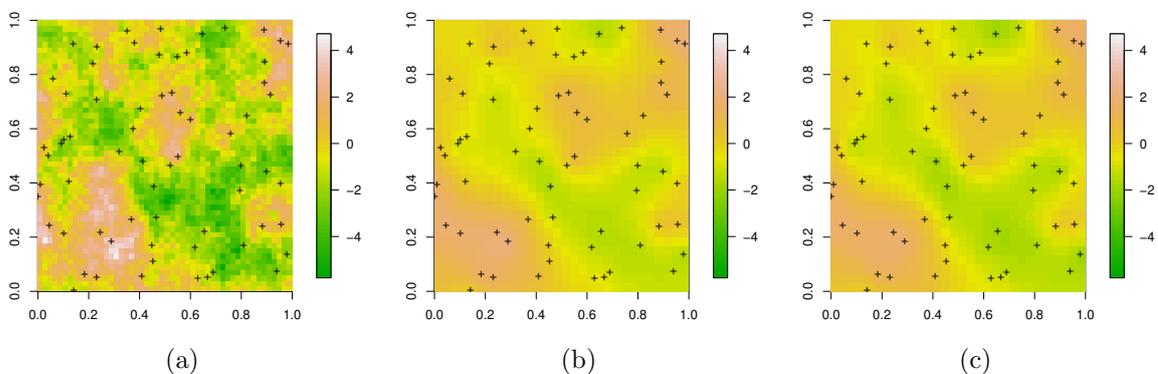


Figura 7.21: Mapa do processo S (a) original, (b) estimado pelo modelo com amostragem preferencial e (c) estimado pelo modelo sem amostragem preferencial, ambos utilizando a mediana da distribuição preditiva a posteriori para o conjunto de dados 1 simulado sem amostragem preferencial.

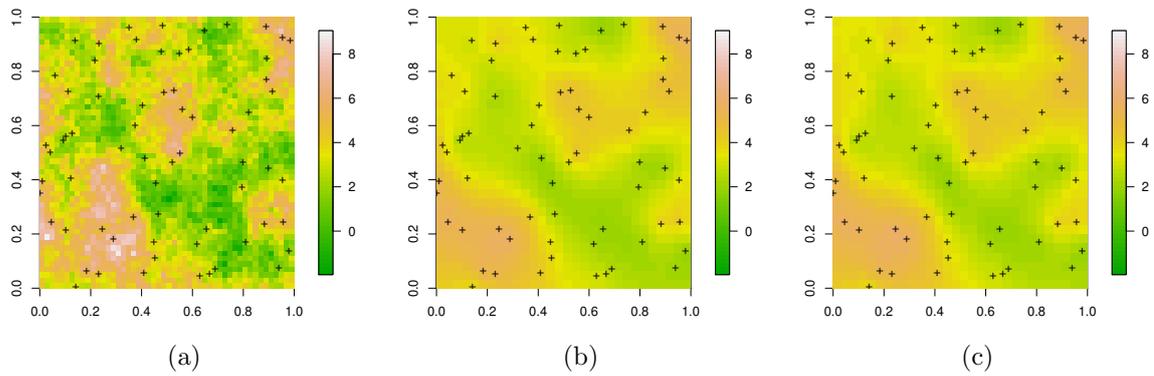


Figura 7.22: Mapa do processo Y (a) original, (b) estimado pelo modelo com amostragem preferencial e (c) estimado pelo modelo sem amostragem preferencial, ambos utilizando a mediana da distribuição preditiva a posteriori para o conjunto de dados 1 simulado sem amostragem preferencial.

7.2 Efeito de fixar o parâmetro de correlação

Para avaliar o efeito de fixar o parâmetro de correlação, foi realizada a estimação dos parâmetros do modelo para vários valores fixos de ϕ , sendo 0,05, 0,15 (verdadeiro) e 0,50. Foram consideradas duas cadeias partindo de valores iniciais distintos, com 200 mil iterações cada. Foram descartadas as primeiras 20 mil iterações de cada cadeia e a amostra a posteriori foi construída a partir de um *lag* igual a 60, totalizando 6 mil observações.

A Figura 7.23 apresenta as distribuições a posteriori dos parâmetros do modelo considerando os valores fixos de ϕ . Percebemos que, quanto maior o valor fixo do parâmetro de correlação, maior são os valores estimados de λ^* . Esse fato ocorre porque, quanto maior o valor de ϕ , maior será a correlação entre os valores de S gerados, o que pode aumentar a quantidade de pontos retidos do processo \tilde{X} e, conseqüentemente, elevar a estimativa de λ^* . Sobre os parâmetros de média e de preferenciabilidade, obtemos estimativas semelhantes, porém há maior variabilidade a posteriori para ϕ igual a 0,50. As estimativas de σ^2 são maiores para valores superiores considerados de ϕ . Zhang (2004) mostra que, para o processo Gaussiano com função de correlação Matérn, há falta de identificabilidade dos parâmetros σ^2 e ϕ , sendo que a razão σ^2/ϕ seria consistentemente estimada quando consideramos a função de correlação exponencial. Dessa forma, a estimativa de σ^2 sofre influência do valor fixado de ϕ , como observamos graficamente.

A superestimação da variância de S para valores de ϕ acima do verdadeiro influencia diretamente a estimação do semivariograma, uma vez que necessita dos valores estimados dos parâmetros da função de covariância. Esse efeito pode ser visualizado na Figura 7.24, em que o semivariograma estimado possui valores superiores à medida que se aumenta o valor fixo de ϕ . Apesar desse efeito, observamos que apenas o valor subestimado de ϕ acarreta em um intervalo de credibilidade a posteriori do semivariograma que não contém o valor real. Os valores de ϕ iguais a 0,15 e 0,50 resultam em IC a posteriori que contêm o semivariograma real, porém o último possui maior variância, chegando à quase o dobro para distâncias maiores que 1 unidade. Notamos esse impacto também na média e na mediana dos semivariogramas, se afastando do semivariograma real. Por isso, deve-se tomar cuidado ao escolher o valor de ϕ que será fixado, assumindo valores razoáveis de acordo com o problema em estudo.

As Figuras 7.25 e 7.26 apresentam as medidas de predição em relação aos processos

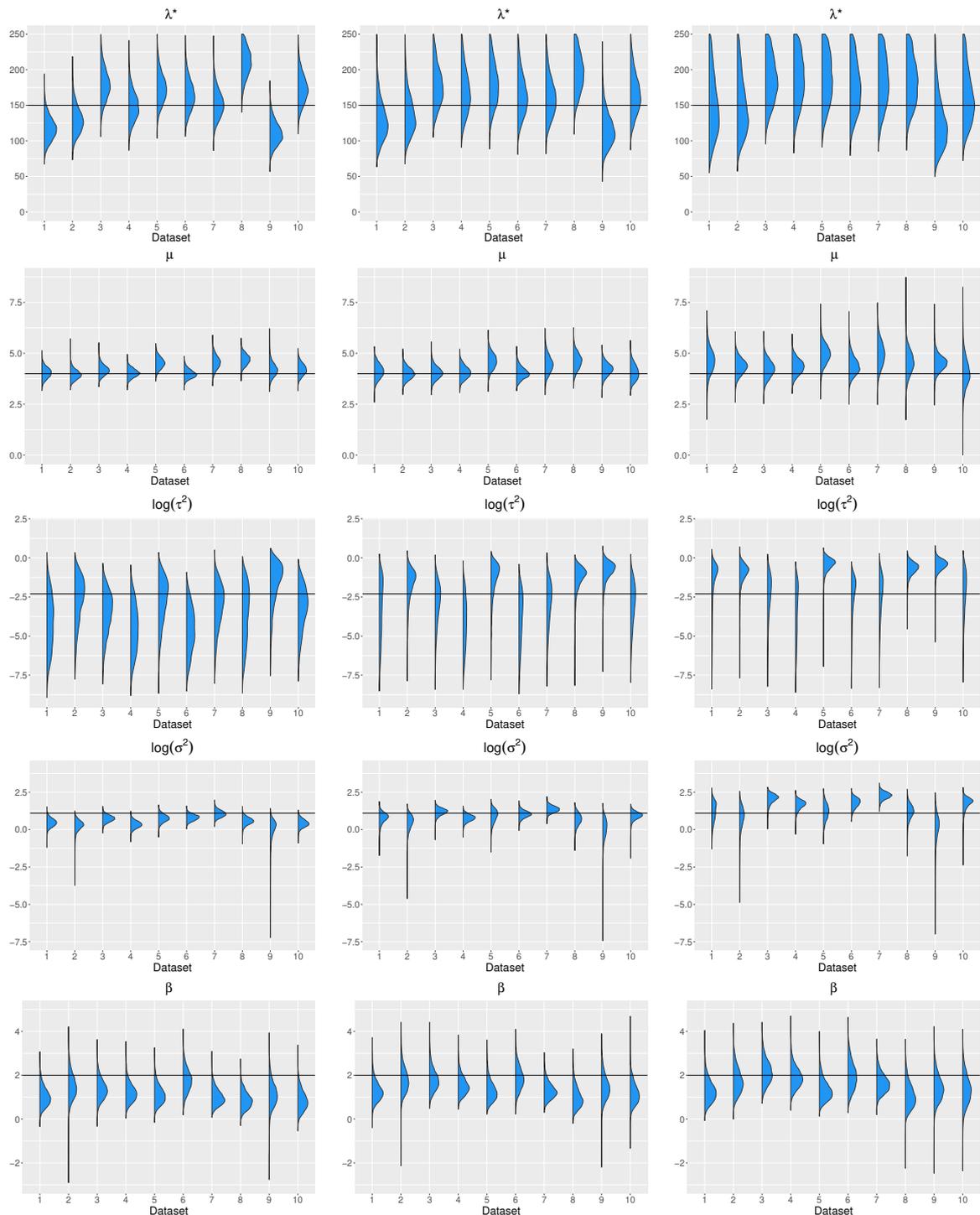


Figura 7.23: Densidades das distribuições a posteriori dos parâmetros do modelo com amostragem preferencial para os dados simulados com amostragem preferencial. O valor de ϕ foi fixado em 0,05 (coluna esquerda) e em 0,50 (coluna direita). A linha horizontal representa o valor verdadeiro do parâmetro.

S e Y para os modelos considerando os vários valores fixos de ϕ . Podemos observar que o modelo que subestima o valor de ϕ apresenta maior valor de EQMPG para todos os conjuntos de dados, para ambos processos. Ainda em relação ao EQMPG, os modelos

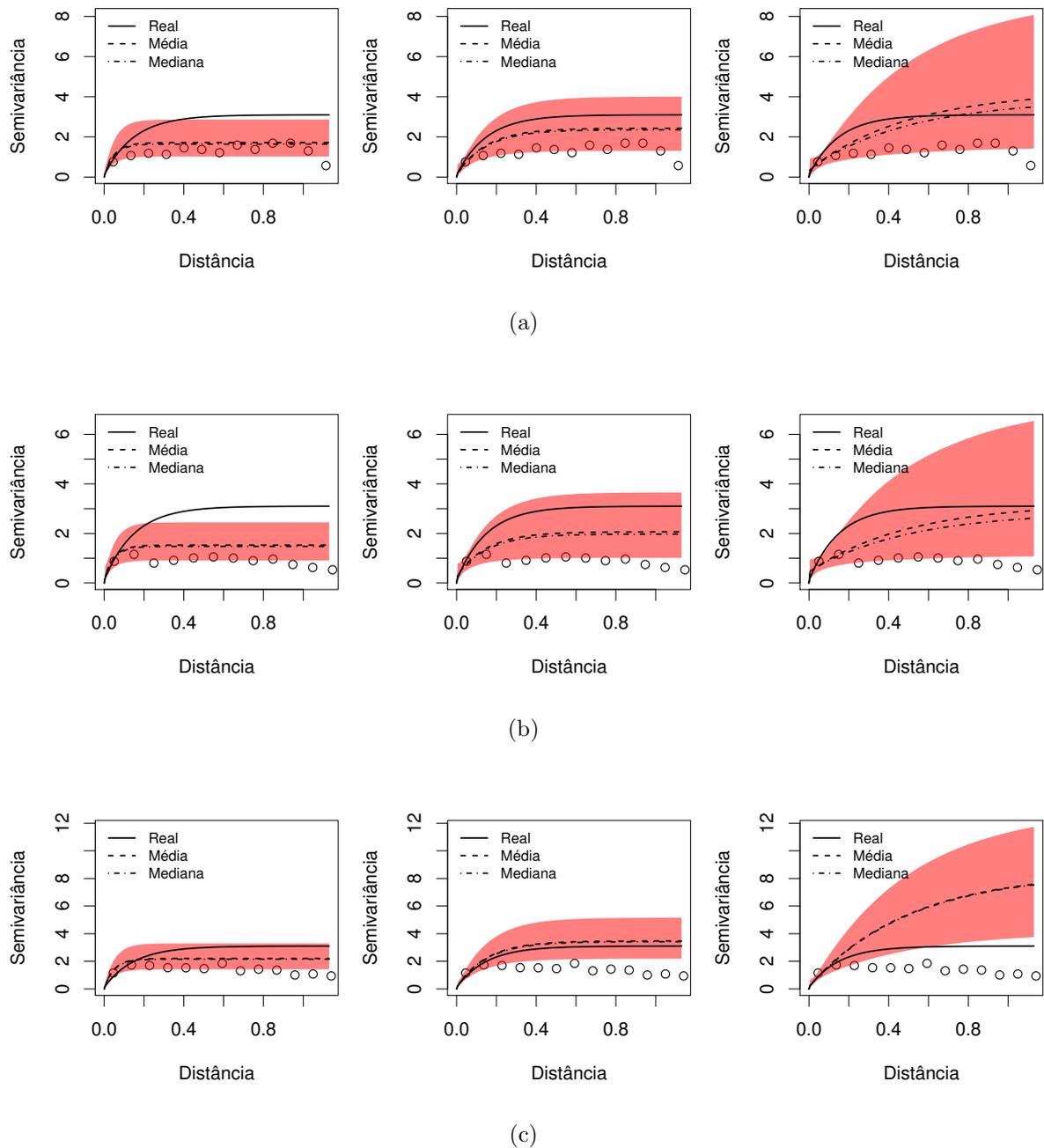


Figura 7.24: Semivariograma estimado e intervalo de credibilidade de 95% (área em vermelho) considerando os valores de ϕ fixos em 0,05 (primeira coluna), 0,15 (segunda coluna) e 0,5 (terceira coluna) para os conjuntos de dados (a) 1, (b) 2 e (c) 3. Os pontos representam o semivariograma amostral e a linha contínua representa o semivariograma real.

com o valor verdadeiro e o valor superestimado de ϕ apresentam medidas semelhantes. Ao analisar a PCIPG, percebemos que o modelo que superestima o parâmetro de correlação possui medidas menores que os demais modelos para alguns conjuntos de dados considerando o processo S . Essa medida melhora quando analisamos a previsão para o

processo Y , que é o nosso interesse principal, e os três modelos possuem performance semelhante para a maior parte dos conjuntos de dados. Esses resultados indicam que fixar o valor do parâmetro de correlação em um valor superior ao verdadeiro não acarreta grandes impactos na predição do processo de interesse.

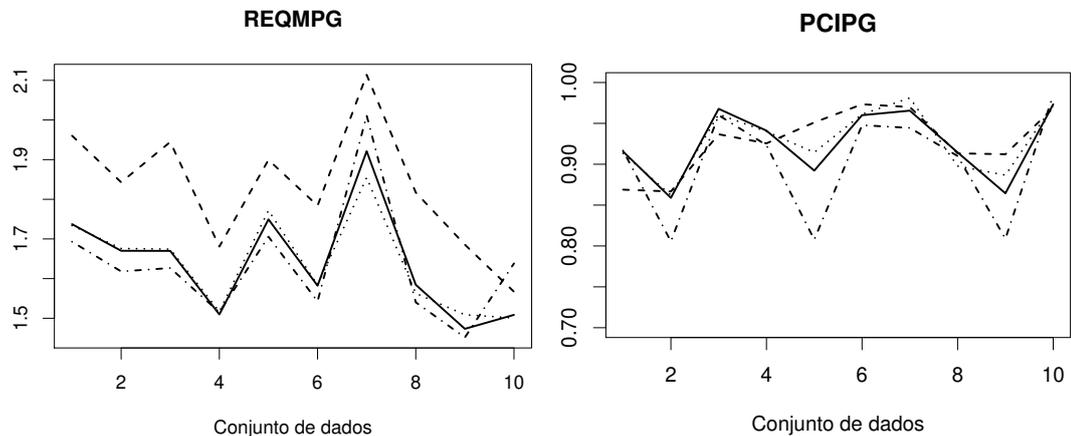


Figura 7.25: Medidas de qualidade de predição para o processo S dos modelos considerando a estimação de ϕ (linha contínua) e com valores de ϕ fixos em 0,05 (linha tracejada), 0,15 (linha pontilhada) e 0,50 (linha ponto-traço).

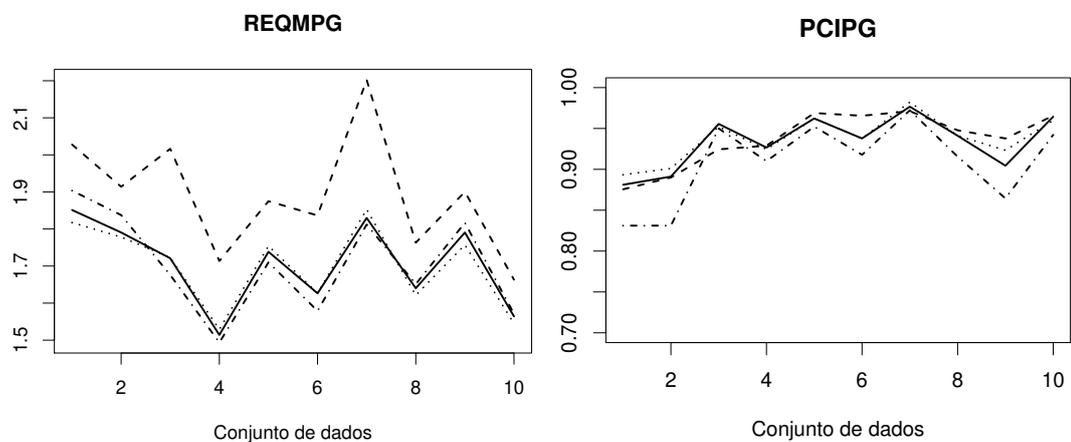


Figura 7.26: Medidas de qualidade de predição para o processo Y dos modelos considerando a estimação de ϕ (linha contínua) e com valores de ϕ fixos em 0,05 (linha tracejada), 0,15 (linha pontilhada) e 0,50 (linha ponto-traço).

As Figuras 7.27 e 7.28 apresentam os mapas dos processos S e Y preditos pelos modelos considerando os vários valores fixos do parâmetro de correlação. Claramente observamos que o modelo que subestima o valor do parâmetro de correlação não obteve boa performance, apresentando valores altos nas regiões que o processo original possui valores baixos, tanto para o processo S quanto para o processo Y . Já os mapas dos modelos que consideram ϕ igual a 0,15 e 0,50 apresentam melhores predições, principalmente nas

regiões onde não há valores amostrados. Percebemos também que a predição é semelhante para os dois últimos, reforçando o resultado de que utilizar um valor superestimado do parâmetro de correlação não resulta grandes impactos sobre os valores preditos do processo de interesse.

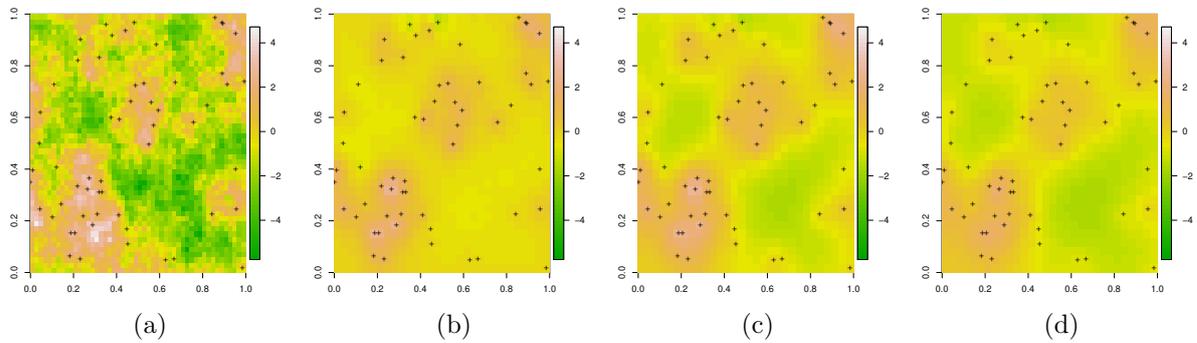


Figura 7.27: Mapa do processo S (a) original e predito pelos modelos fixando ϕ em (b) 0,05, (c) 0,15 (valor verdadeiro) e (d) 0,50 utilizando a mediana da distribuição preditiva a posteriori para o conjunto de dados 1 simulado com amostragem preferencial.

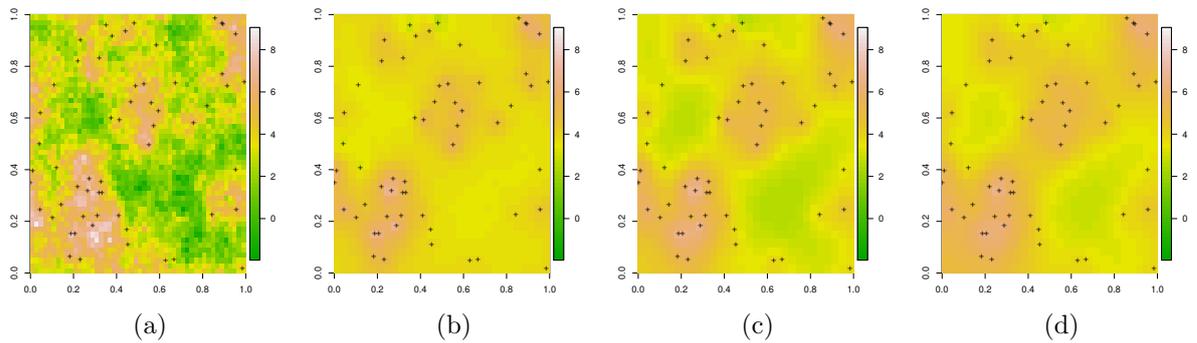


Figura 7.28: Mapa do processo Y (a) original e predito pelos modelos fixando ϕ em (b) 0,05, (c) 0,15 (valor verdadeiro) e (d) 0,50 utilizando a mediana da distribuição preditiva a posteriori para o conjunto de dados 1 simulado com amostragem preferencial.

Para concluir a seção, percebemos que, caso seja de interesse fixar o valor do parâmetro de correlação, devem ser considerados valores razoáveis para o problema em questão. O gráfico de semivariograma pode indicar uma boa estimativa para o valor de ϕ a partir do alcance obtido pelo semivariograma teórico considerado.

7.3 Desempenho do modelo para poucos dados

Para analisar o desempenho do nosso modelo quando há pouca informação sobre o processo em estudo, foi realizado um estudo de simulação considerando os valores de λ^* iguais a 30 e 50. Os valores dos demais parâmetros foram $(\mu, \tau^2, \sigma^2, \phi, \beta) = (4; 0,10; 3; 0,15; 2)$ e a Tabela 7.2 apresenta o número de observações obtidas para cada conjunto de dados. Foram consideradas duas cadeias partindo de valores iniciais distintos, com 200 mil iterações cada. Foram descartadas as primeiras 20 mil iterações de cada cadeia e a amostra a posteriori foi construída a partir de um *lag* igual a 60, totalizando 6 mil observações.

Tabela 7.2: Número de observações obtidas em cada conjunto de dados.

Conjunto de dados	1	2	3	4	5	6	7	8	9	10
$\lambda^* = 30$	12	8	13	10	15	20	16	20	13	13
$\lambda^* = 50$	20	21	20	41	22	32	17	29	23	30

A Figura 7.29 apresenta as densidades das distribuições a posteriori dos parâmetros estimados pelo modelo considerando o valor de λ^* igual a 30. Podemos observar que a estimação do parâmetro λ^* é satisfatória, com grande densidade a posteriori em torno de seu valor verdadeiro. Em relação ao parâmetro de média, notamos que é superestimado na maioria dos conjuntos de dados, evidenciando o efeito de se ter pouca informação nos dados. O efeito pepita e o parâmetro de correlação também são superestimados, mas obtemos estimação satisfatória para σ^2 . O parâmetro de preferenciabilidade β é subestimado em todos os conjuntos de dados e podemos observar que a centralidade da distribuição a posteriori está em torno de zero, indicando erroneamente que os dados não provém de uma amostragem preferencial.

A Figura 7.30 apresenta as densidades das distribuições a posteriori dos parâmetros estimados pelo modelo considerando o valor de λ^* igual a 50. Observamos comportamento semelhante à simulação anterior, sendo os parâmetros μ , τ^2 e ϕ superestimados e σ^2 e β , subestimados. Porém, podemos observar em relação ao parâmetro de preferenciabilidade que sua distribuição a posteriori se deslocou positivamente para quase todos os bancos de dados quando comparado à simulação com λ^* igual a 30. Ou seja, ao agregar mais informação, a estimação do parâmetro β se torna melhor.

As Figuras 7.31 e 7.32 apresentam as medidas de qualidade de predição dos processos S e Y para os modelos. A medida referente ao modelo com λ^* igual a 150 foi

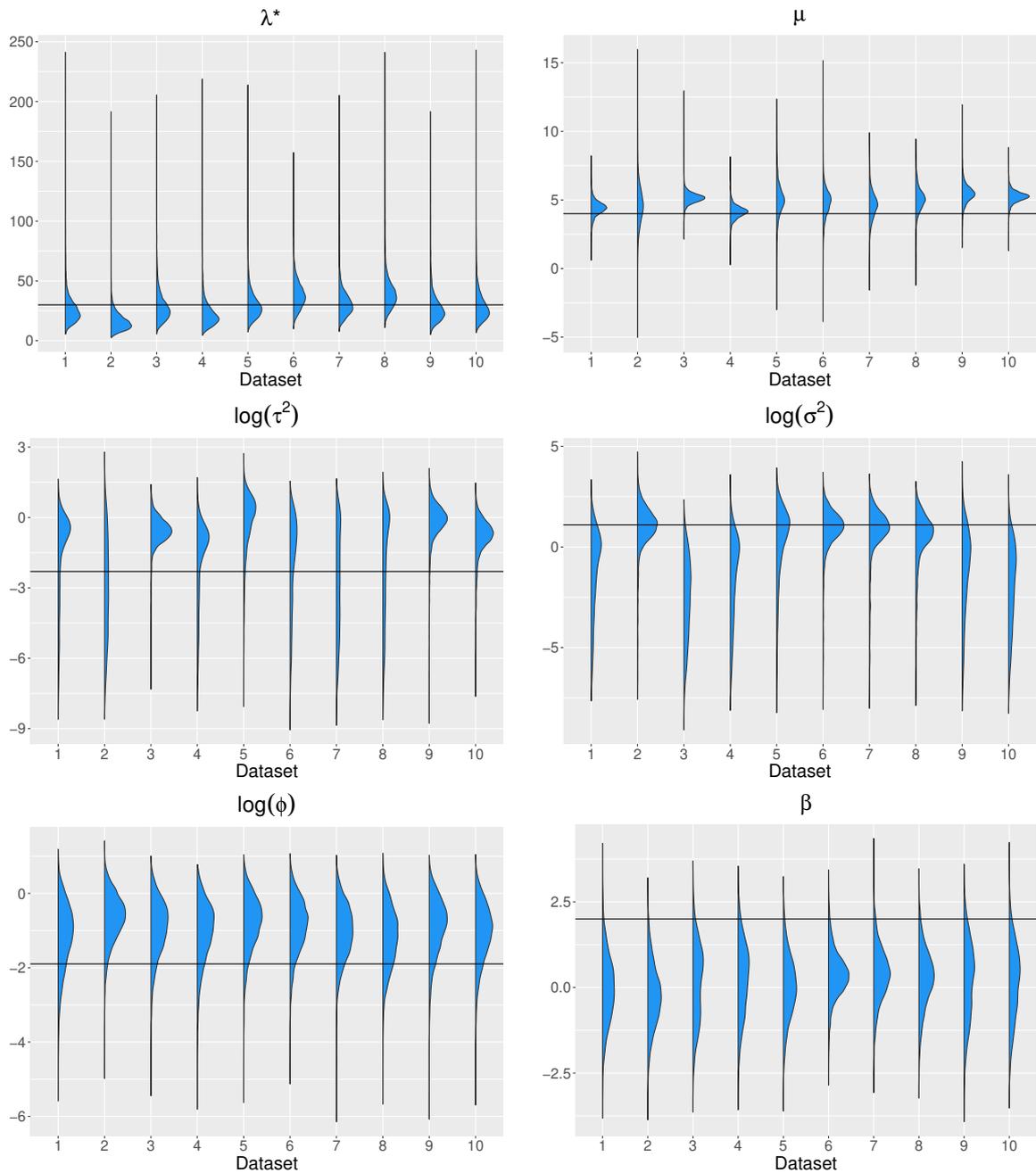


Figura 7.29: Densidades das distribuições a posteriori dos parâmetros do modelo com amostragem preferencial para os dados simulados com amostragem preferencial e λ^* igual a 30. A linha horizontal representa o valor verdadeiro do parâmetro.

apresentada para comparação, uma vez que obtemos boa previsão para esse modelo. Em relação ao EQMPG, observamos que os modelos com poucos dados apresentam valores semelhantes e superiores ao modelo de referência para ambos processos. Analisando os gráficos de PCIPG, notamos que a cobertura de intervalos de previsão para os modelos com poucos dados é semelhante para os processos S e Y e são menores que a cobertura do modelo de referência, principalmente em relação ao processo Y . Esses resultados indicam

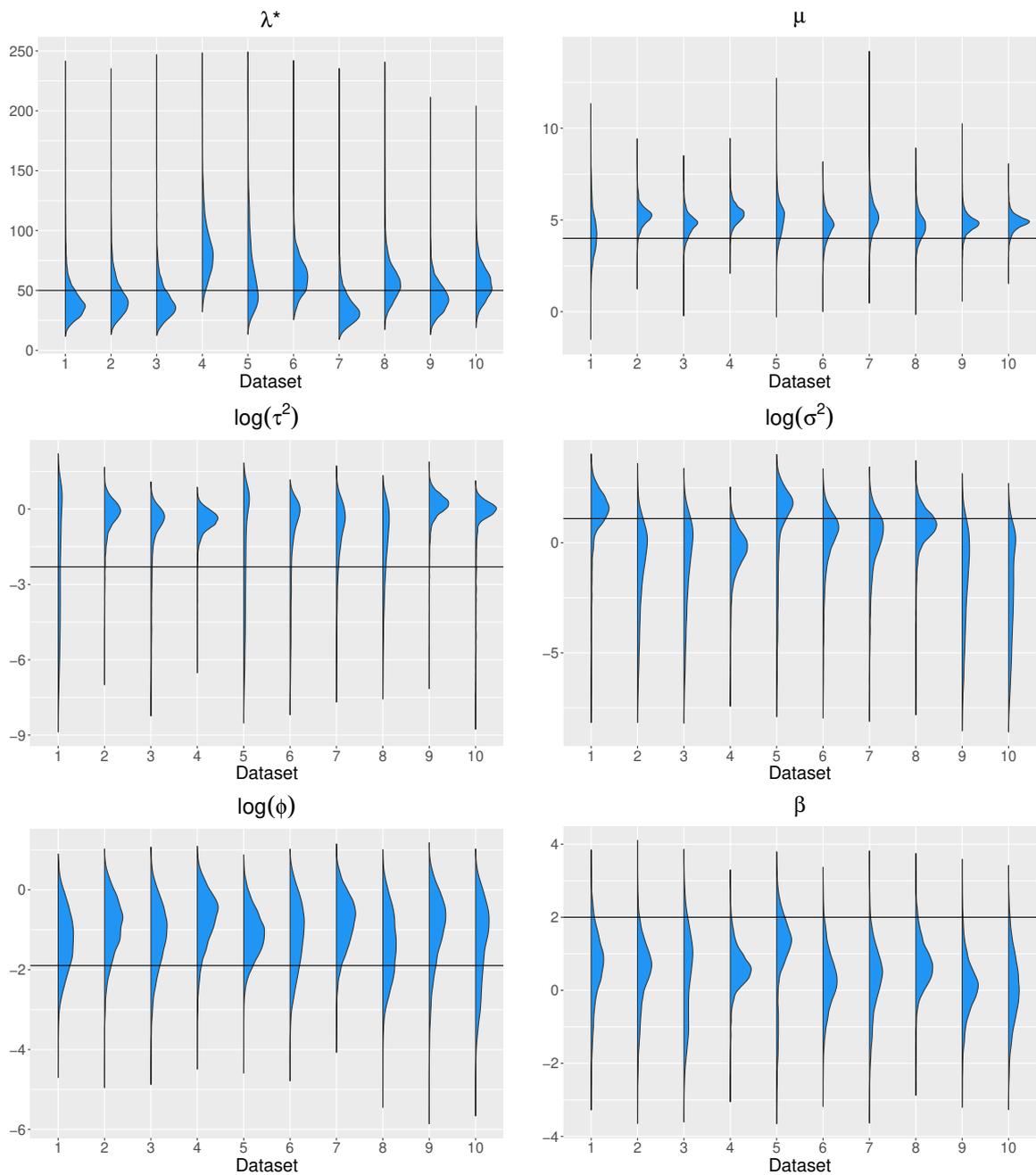


Figura 7.30: Densidades das distribuições a posteriori dos parâmetros do modelo com amostragem preferencial para os dados simulados com amostragem preferencial e λ^* igual a 50. A linha horizontal representa o valor verdadeiro do parâmetro.

que a falta de informação afeta a capacidade preditiva do modelo.

Para concluir a seção, percebemos que a capacidade do modelo de corrigir o viés proporcionado pela amostragem preferencial depende da quantidade de informação dos dados. Quanto mais informação, ou seja, mais locais amostrados, melhor será a estimação. Se a amostra não fornece informação suficiente, o modelo não será capaz de identificar a amostragem preferencial, obtendo estimações viesadas como no modelo tradicional. O

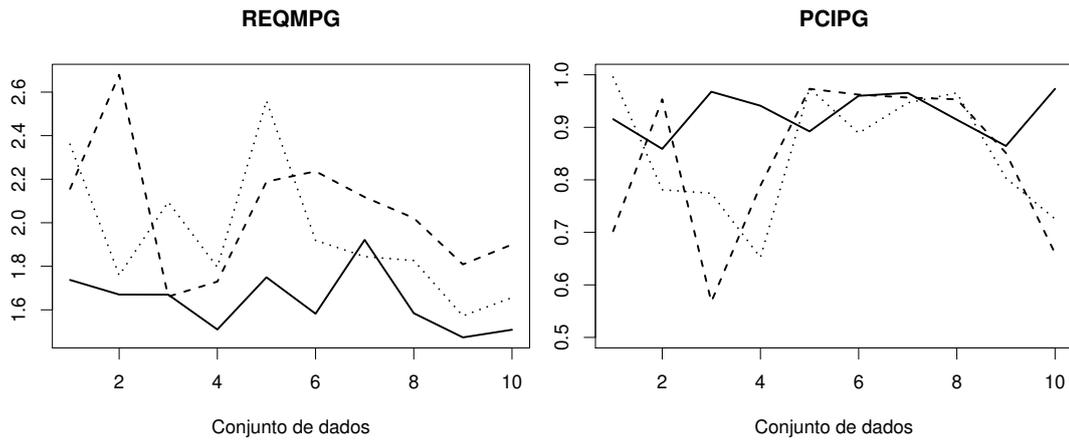


Figura 7.31: Medidas de qualidade de predição para o processo S dos modelos com λ^* igual a 150 (linha contínua), 30 (linha tracejada) e 50 (linha pontilhada), considerando os conjuntos de dados simulados com amostragem preferencial.

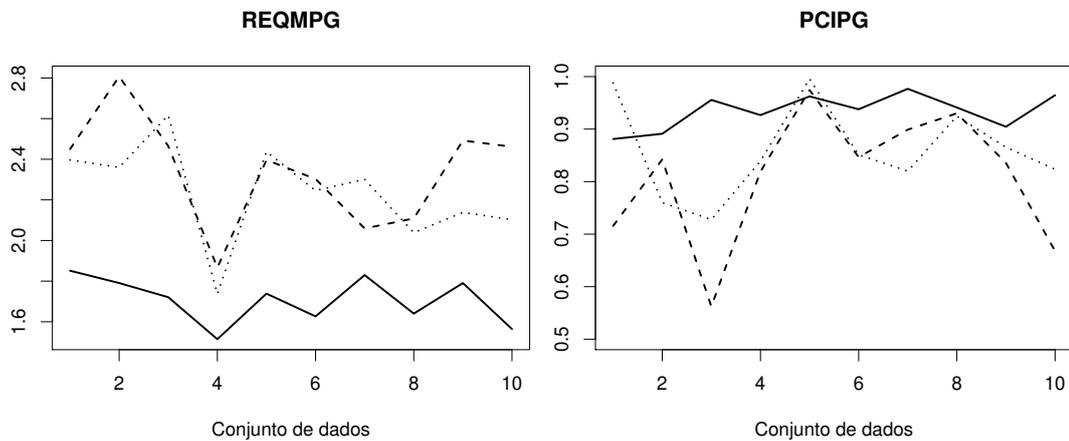


Figura 7.32: Medidas de qualidade de predição para o processo Y dos modelos com λ^* igual a 150 (linha contínua), 30 (linha tracejada) e 50 (linha pontilhada), considerando os conjuntos de dados simulados com amostragem preferencial.

mesmo ocorre com a capacidade preditiva do modelo, fornecendo predições pobres quando há falta de informação nos dados.

7.4 Especificação incorreta da função de intensidade

O modelo escolhido pode não ser o mais correto para representar a geração dos dados do fenômeno em estudo. Dessa forma, vale estudar como nosso modelo se comporta quando os dados são gerados a partir de uma função de intensidade diferente da probito.

A função de intensidade do modelo log-linear é bastante conhecida em geoestatística com amostragem preferencial, visto que é a função considerada no trabalho pioneiro de Diggle et al. (2010) e utilizada por outros, como Ferreira and Gamerman (2015) e Dinsdale and Salibian-Barrera (2019). Assim, nesta seção estudamos o comportamento das estimativas dos parâmetros e valores preditos do nosso modelo quando os dados são gerados a partir da função de intensidade exponencial, isto é,

$$\lambda(x) = \exp\{\gamma + \delta S(x)\}.$$

Consideramos os seguintes valores para os parâmetros: $(\mu, \tau^2, \sigma^2, \phi) = (4; 0,10; 3; 0,15)$. Para gerar o processo pontual, consideramos dois cenários que diferem na intensidade do parâmetro de preferência,

1. $(\gamma, \delta) = (3,5; 1)$ resultando em 89 observações e
2. $(\gamma, \delta) = (1; 2)$ resultando em 61 observações.

A função de intensidade exponencial teórica não é limitada, o que diferencia da função utilizada no nosso modelo que é limitada por λ^* . Dessa forma, espera-se que as estimativas de λ^* sejam altas para compensar a diferença de comportamento das funções. Porém, isso resultará em uma grande demanda computacional, aumentando o tempo de execução do algoritmo exponencialmente. Para evitar esse problema, uma priori gama truncada foi adotada, mas devemos ter em mente que, ao realizar o truncamento, não obteremos taxas estimadas acima do valor escolhido. Logo, analisar o efeito de tal truncamento se torna necessário principalmente sobre a predição em valores não observados, que é o principal objetivo da geoestatística.

Para isso, considerando o cenário 1, a estimação foi realizada para três valores de truncamento do parâmetro λ^* iguais a 150, 250 e 350. Além disso, consideramos realizar a estimação também pelo modelo sem amostragem preferencial (SAP) a fim de comparar com os resultados do nosso modelo quando a função de intensidade é diferente da probito.

As densidades das distribuições a posteriori dos parâmetros do modelo são apresentadas na Figura 7.33.

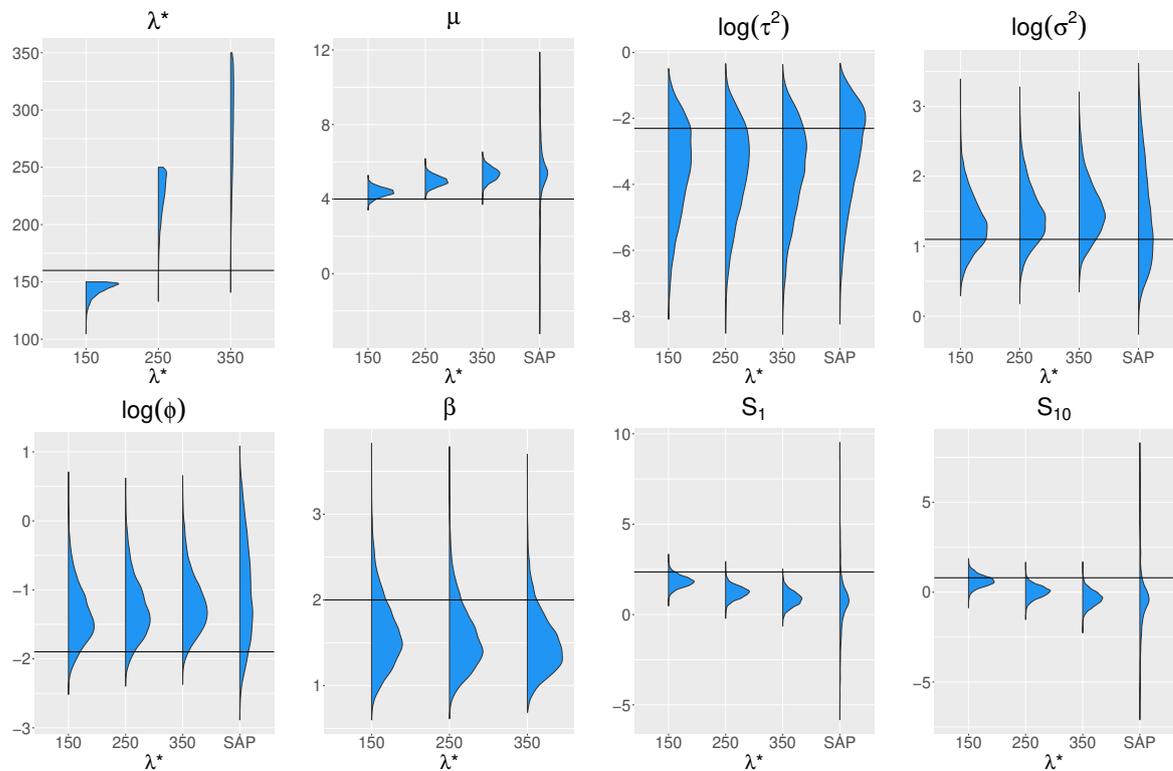


Figura 7.33: Densidades das distribuições a posteriori dos parâmetros do modelo com função de intensidade probito, considerando o cenário 1 de dados simulados com função de intensidade exponencial. A linha horizontal representa o valor verdadeiro do parâmetro.

Podemos observar que a estimação para os parâmetros que definem a estrutura de variância e covariância, τ^2 , σ^2 e ϕ , é semelhante para os três valores de truncamento considerados. O resultado obtido para esses parâmetros pelo modelo SAP apresenta maior variabilidade a posteriori. Em relação ao parâmetro λ^* , aumentar o valor do truncamento resultou em estimativas maiores, como era esperado. Esse fato também resultou em deslocar a distribuição a posteriori de μ para uma região de maiores valores e essa relação pode ser observada na Figura 7.34, em que podemos notar uma tendência crescente nas estimativas desses dois parâmetros, assim como observado na Seção 7.1. Para compensar a estimativa de μ , o modelo também altera a estimativa de S e notamos o mesmo comportamento para S_1 e S_{10} . Em relação ao modelo SAP, a superestimação de μ era esperada, pois não há correção do viés que a amostra apresenta nesse modelo. Sobre o parâmetro de preferência β , notamos estimação semelhante do nosso modelo para os três valores de truncamento considerados. Ou seja, a região de truncamento não afeta muito

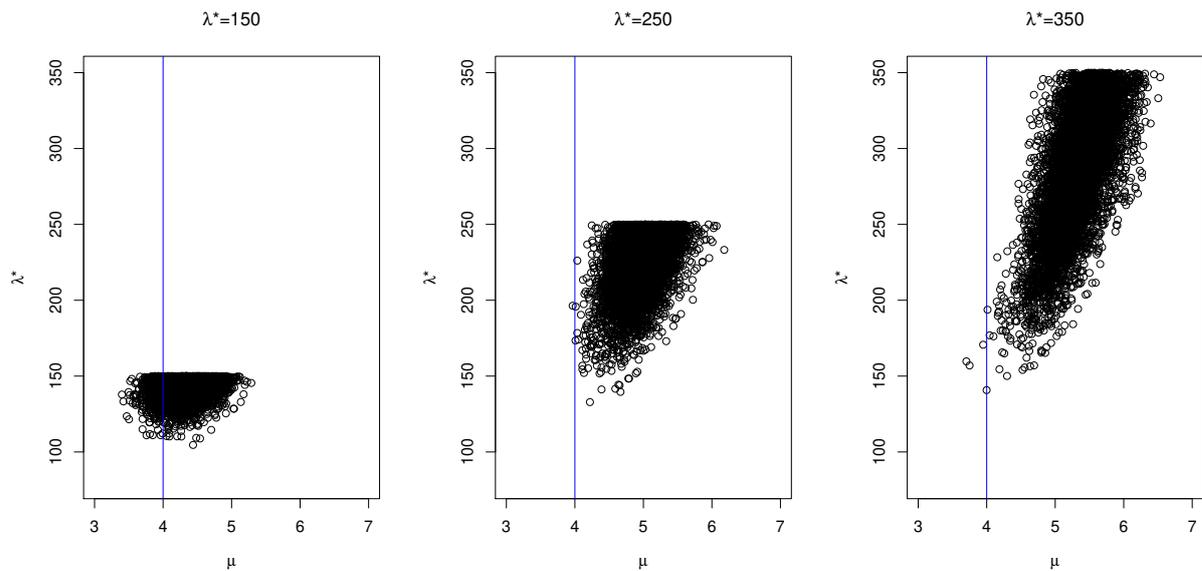


Figura 7.34: Gráfico de dispersão das estimativas de μ e λ^* . A linha vertical representa o valor verdadeiro de μ .

a estimação dos parâmetros, exceto por μ e, obviamente, λ^* . Devemos então analisar esse efeito sobre as predições.

A Figura 7.35 apresenta a função de intensidade estimada para os dados gerados no cenário 1 em função do processo Gaussiano S . Notamos que há diferença significativa entre

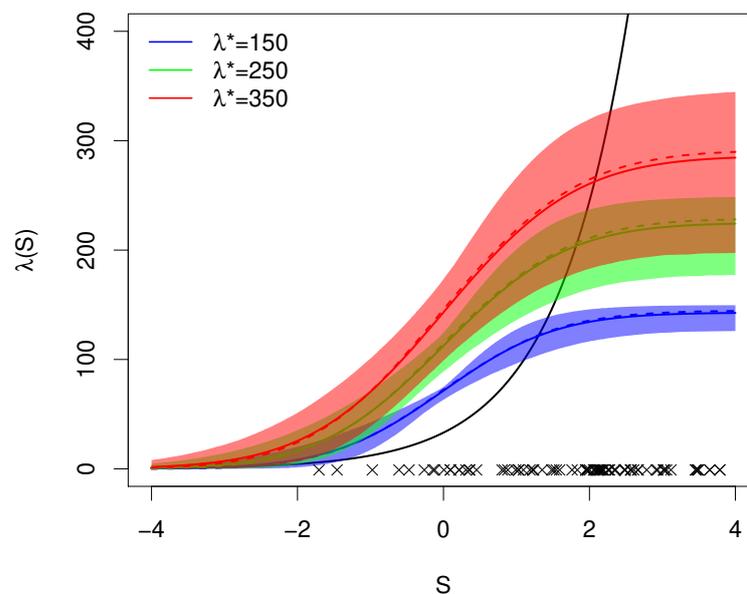


Figura 7.35: Taxa estimada pela média (linha contínua), mediana (linha tracejada) e intervalo de credibilidade de 95% considerando os dados gerados pelo cenário 1. A linha preta representa a taxa exponencial (verdadeira). Para melhor visualização, o eixo vertical foi limitado em 300. Os pontos marcados em “x” correspondem aos valores simulados do processo S na amostra.

a função estimada pelo nosso modelo e a função real, sendo que o intervalo de credibilidade de 95% não contém o valor verdadeiro da função na maior parte apresentada. Isso se dá pelo fato de que a função de intensidade do nosso modelo não possui intercepto dentro da função probito, o que não possibilita a curva de se deslocar para a direita, por exemplo, para “acompanhar” a curva real exponencial. Em outras palavras, para o valor $s = 0$, a função de intensidade obrigatoriamente terá como valor $\lambda^*/2$, o que deixa a curva distante da curva exponencial considerada para a maior parte do intervalo de S considerado.

Para incorporar a incerteza da estimação de S , foram construídos os mapas da função de intensidade estimada a partir da média a posteriori, como pode ser observado na Figura 7.36. Os valores se encontram na escala logarítmica. Percebemos que os valores

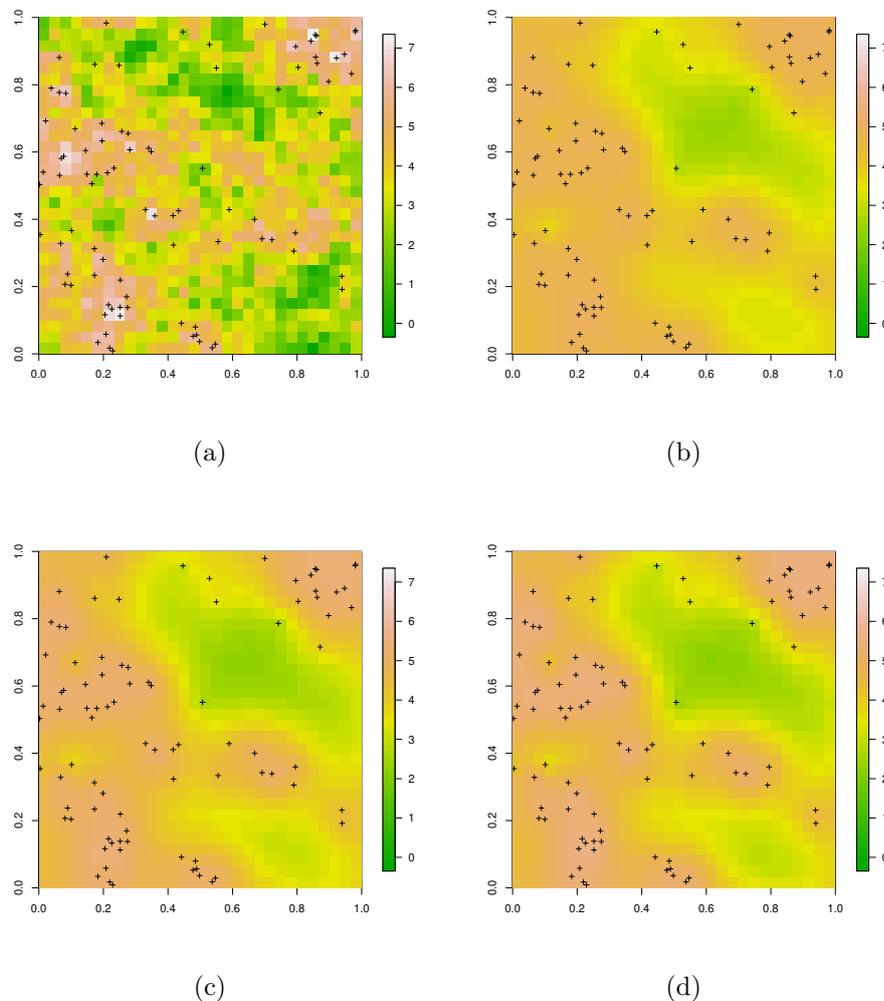


Figura 7.36: Função de intensidade (a) simulada pelo cenário 1 e estimada pela média a posteriori considerando os valores de truncamento para λ^* iguais a (b) 150, (c) 250 e (d) 350. Os valores estão na escala logarítmica.

mais altos da taxa não foram recuperados, como era de se esperar, mas os menores valores

foram superestimados, corroborando com o resultado da Figura 7.35. A estimação da taxa foi semelhante para os três valores de truncamento considerados.

Foram calculadas as proporções de cobertura de intervalos de credibilidade de 95% da taxa estimada, resultando em 73,7%, 82,2% e 85,6% para os valores de truncamento 150, 250 e 350, respectivamente. Esse resultado indica que temos grande proporção de cobertura dos ICs para a função de intensidade, apesar da estimação pobre pela média e mediana a posteriori, como podemos observar na Figura 7.37. Nessa Figura, escolhemos

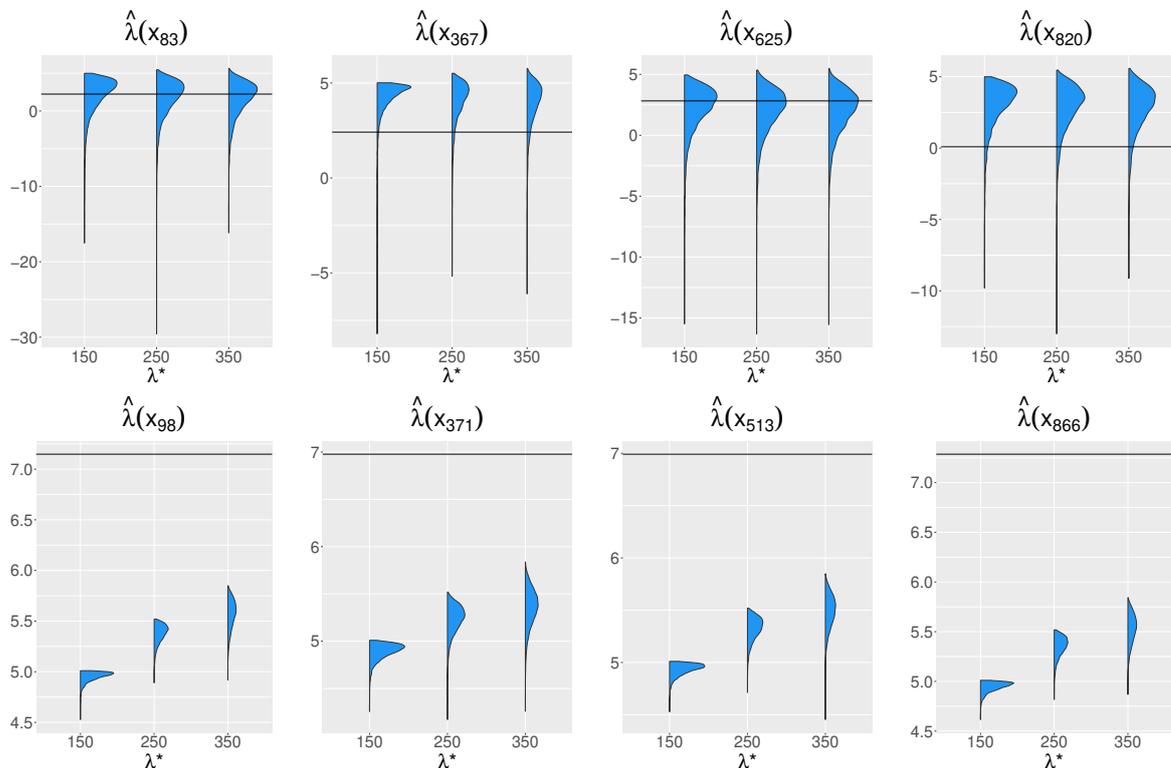


Figura 7.37: Distribuições a posteriori da função de intensidade estimada considerando o cenário 1 para os locais de baixa intensidade real $x_{83} = (0,750; 0,083)$, $x_{367} = (0,217; 0,417)$, $x_{625} = (0,817; 0,683)$, $x_{820} = (0,317; 0,917)$ e de alta intensidade real $x_{98} = (0,250; 0,117)$, $x_{371} = (0,350; 0,417)$, $x_{513} = (0,083; 0,583)$, $x_{866} = (0,850; 0,950)$.

quatro pontos em que a função de intensidade apresenta valores baixos (primeira linha de gráficos) e quatro pontos de valores altos (segunda linha de gráficos). Em relação aos pontos de baixa intensidade, notamos boa estimativa para uns e valores superestimados para outros. Em relação aos pontos de alta intensidade, não conseguimos recuperar o valor verdadeiro, como esperado quando consideramos como taxa real a função exponencial.

Para avaliar a predição para o processo de interesse Y , foram construídos mapas estimados pela média a posteriori dos valores preditos, como podemos observar na Figura 7.38. Notamos boa predição pela média a posteriori e também notamos que os mapas

para ambos valores de truncamento são semelhantes. Notamos também que a predição em regiões sem valor observado é superestimado pelo modelo SAP, ao passo que nosso modelo corrige esse viés mesmo quando a função de intensidade real é exponencial.

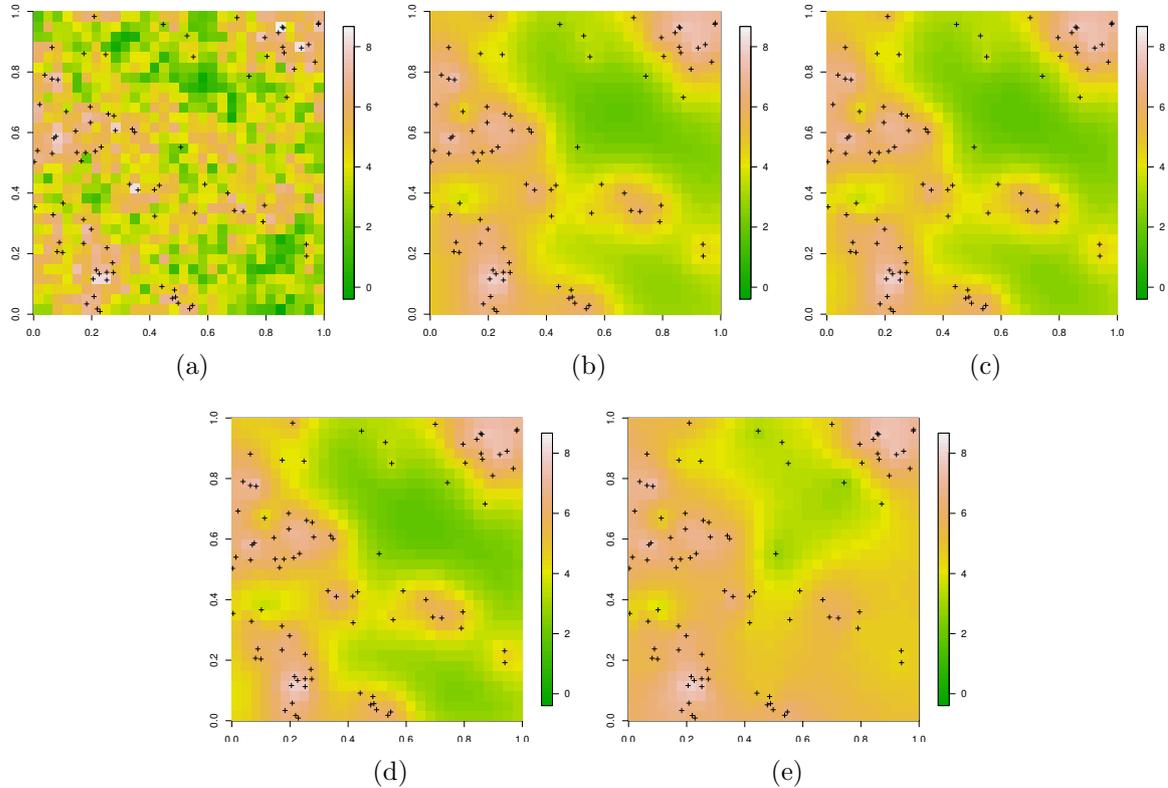


Figura 7.38: Mapa do processo Y (a) simulado pelo cenário 1 e estimado pela média a posteriori da distribuição preditiva considerando os valores de truncamento para λ^* iguais a (b) 150, (c) 250, (d) 350 e considerando (e) o modelo sem amostragem preferencial.

A Figura 7.39 apresenta as distribuições preditivas a posteriori de alguns locais não observados para Y , em regiões sem observações, e percebemos que a predição resultante do modelo considerando todos valores de truncamento de λ^* são muito semelhantes. Isso significa que, mesmo existindo uma diferença na estimação do parâmetro μ para os diferentes truncamentos de λ^* , o processo de predição corrige esse viés pois considera a quantidade $\mu + S(x)$ e a informação do processo pontual. Dessa forma, podemos trabalhar com valores de truncamento não muito altos sem perda de qualidade de predição, tornando o processo de estimação mais eficiente. Note que as distribuições a posteriori dos valores de Y se encontram próximas de seus valores verdadeiros, significando que é possível utilizar o modelo com função de ligação probito para dados originados de uma função de ligação exponencial. Sobre o modelo SAP, as predições são viesadas, como era esperado quando há amostragem preferencial.

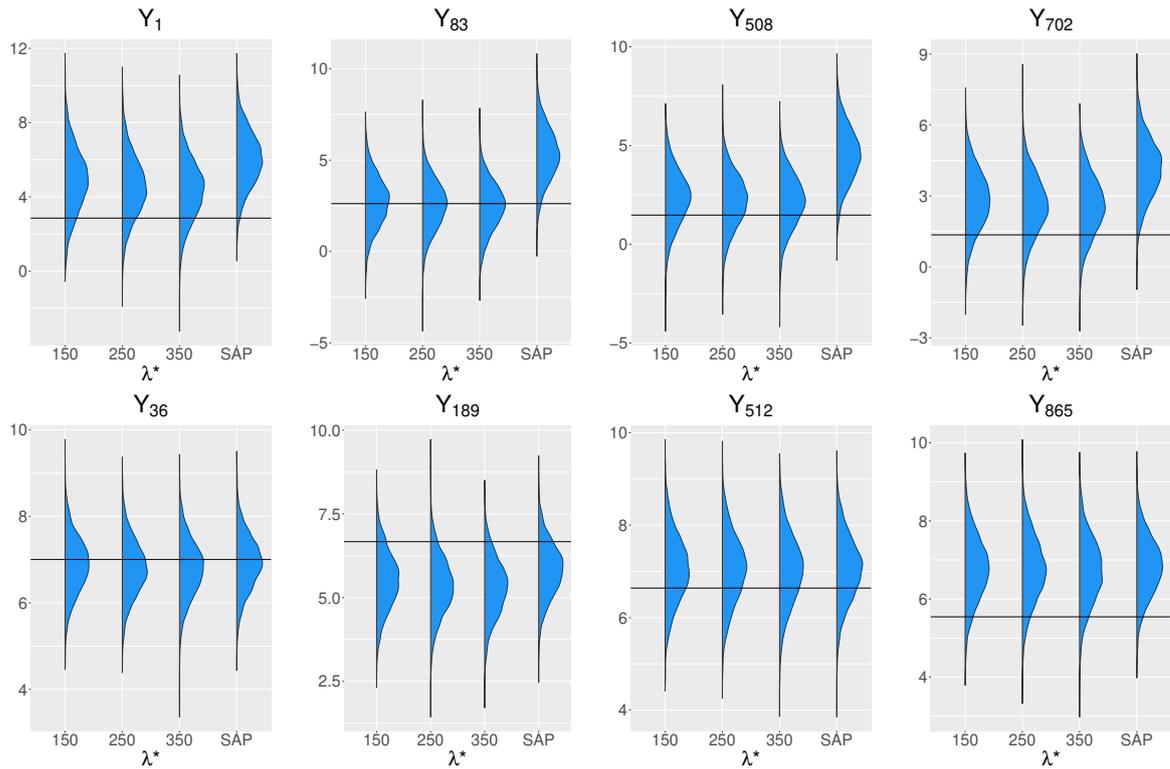


Figura 7.39: Distribuições a posteriori do processo Y para os locais não observados distantes de pontos amostrados (primeira linha) e próximos de pontos amostrados (segunda linha). As coordenadas dos locais escolhidos são $x_{83} = (0,750; 0,083)$, $x_{315} = (0,483; 0,350)$, $x_{620} = (0,650; 0,683)$, $x_{702} = (0,383; 0,783)$, $x_{36} = (0,183; 0,050)$, $x_{189} = (0,283; 0,217)$, $x_{512} = (0,050; 0,583)$ e $x_{865} = (0,817; 0,950)$, considerando o cenário 1.

Observamos, então, que o fato de não conseguirmos boas estimativas para a função de intensidade não impossibilita o modelo de realizar boas previsões para o processo de interesse. Isso se dá pelo fato de que a função de intensidade impacta, principalmente, na geração de locais do processo \tilde{X} . Se observarmos a Figura 7.35, concluímos que, pela função de intensidade exponencial, serão retidos os pontos de \tilde{X} que apresentarem baixíssimos valores de S . Pela função de intensidade probito, pontos que apresentarem baixos valores de S , mas não muito baixo, também tem uma probabilidade considerável de serem retidos e a informação desses pontos é suficiente para realizarmos boa inferência para o processo Y .

Os resultados obtidos para o cenário 2 foram semelhantes. Considerando uma maior força na preferência em relação ao cenário 1, obtemos um viés nos valores preditos para o processo Y pelo nosso modelo pelo fato de haver menos informação de valores baixos (viés pela falta de informação). Os resultados estão no Apêndice A.

Para concluir a seção, podemos dizer que, quando o processo pontual é gerado

pela função de intensidade exponencial, os resultados obtidos através do nosso modelo com função de intensidade probito são satisfatórios para o processo de interesse, que é a predição de Y em locais não observados.

7.5 Modelo com intercepto na função de intensidade

A estimação dos parâmetros do modelo com intercepto na função de intensidade apresentou um problema de identificação, sendo difícil estimar α sob priori pouco informativa. Os dados foram simulados considerando o valor de α igual a 1 e consideramos distribuições a priori para α normais centradas em seu valor verdadeiro com variâncias 1, 0,1.

As distribuições a posteriori dos parâmetros do modelo podem ser observadas na Figura 7.40. Podemos notar que os resultados obtidos para os parâmetros de variabilidade são semelhantes nos dois casos. Porém, para variância a priori igual a 1, a estimação de α se deu em torno de zero e, como consequência, os parâmetros λ^* e μ foram estimados em torno de seus valores verdadeiros. Para a variância a priori igual a 0,1, conseguimos estimação satisfatória para α , mas λ^* é subestimado e μ é superestimado. A relação desses dois parâmetros com α pode ser observada na Figura 7.41. É interessante notar que a tendência crescente entre os valores estimados de λ^* e μ se perde quando, neste modelo, o parâmetro α foi estimado em torno de zero. Nesse caso, a relação de dependência passa para μ e α . Quando a estimação do intercepto se dá longe de zero, não observamos relação entre seus valores estimados e os de μ e voltamos a perceber essa relação entre λ^* e μ . Como resultados preliminares, ainda não está claro o porquê desse comportamento, sendo que um estudo mais aprofundado sobre esse modelo se torna necessário.

Podemos observar na Figura 7.42 a função de intensidade estimada pela média e mediana a posteriori e os intervalos de credibilidade de 95%. Notamos que, ao estimar corretamente o intercepto α , obtemos uma curva estimada mais adequada, pois conseguimos deslocá-la para se ajustar melhor à curva real. A Figura 7.43 apresenta os mapas estimados da função de intensidade. Notamos que os mapas são semelhantes para os dois casos, sendo que para o caso em que α foi estimado como nulo, obtemos valores maiores da taxa nos locais de alta intensidade. Isso é corroborado com o resultado da Figura 7.42, em que o valor máximo estimado pela média a posteriori da função de intensidade é mais próximo do valor real para a variância a priori de α igual a 1.

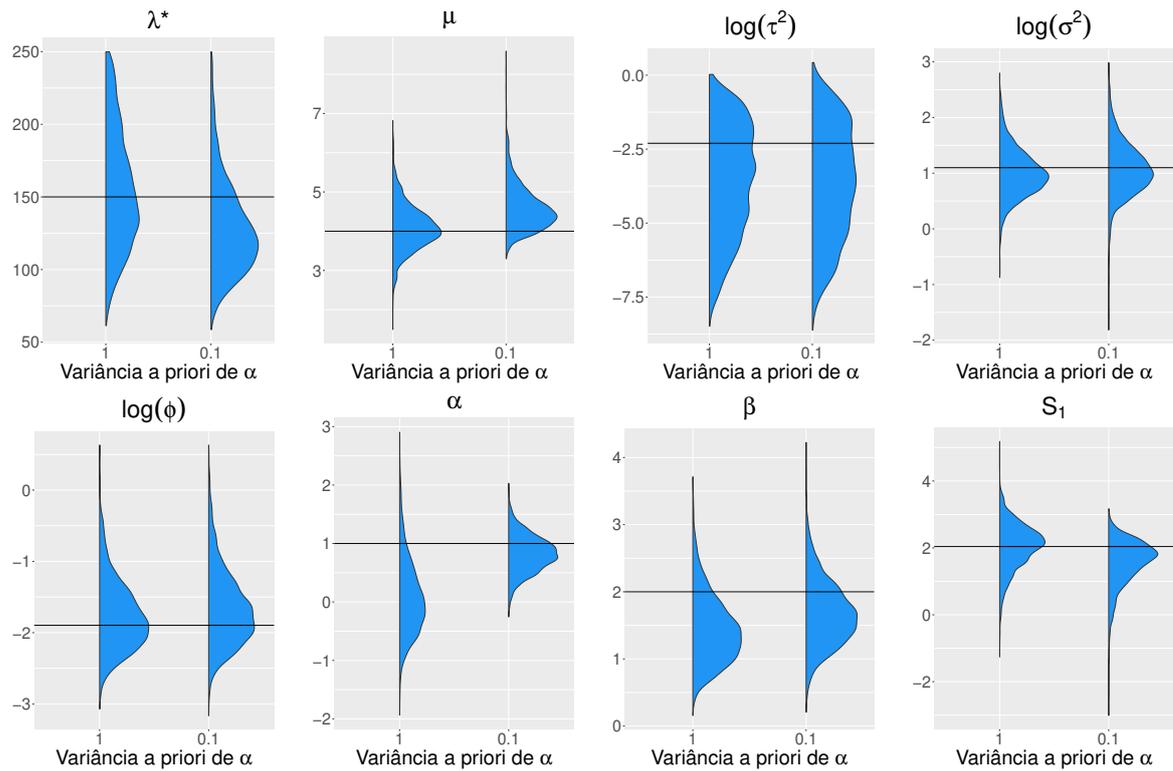


Figura 7.40: Densidades das distribuições a posteriori dos parâmetros do modelo com função de intensidade probito. A linha horizontal representa o valor verdadeiro do parâmetro.

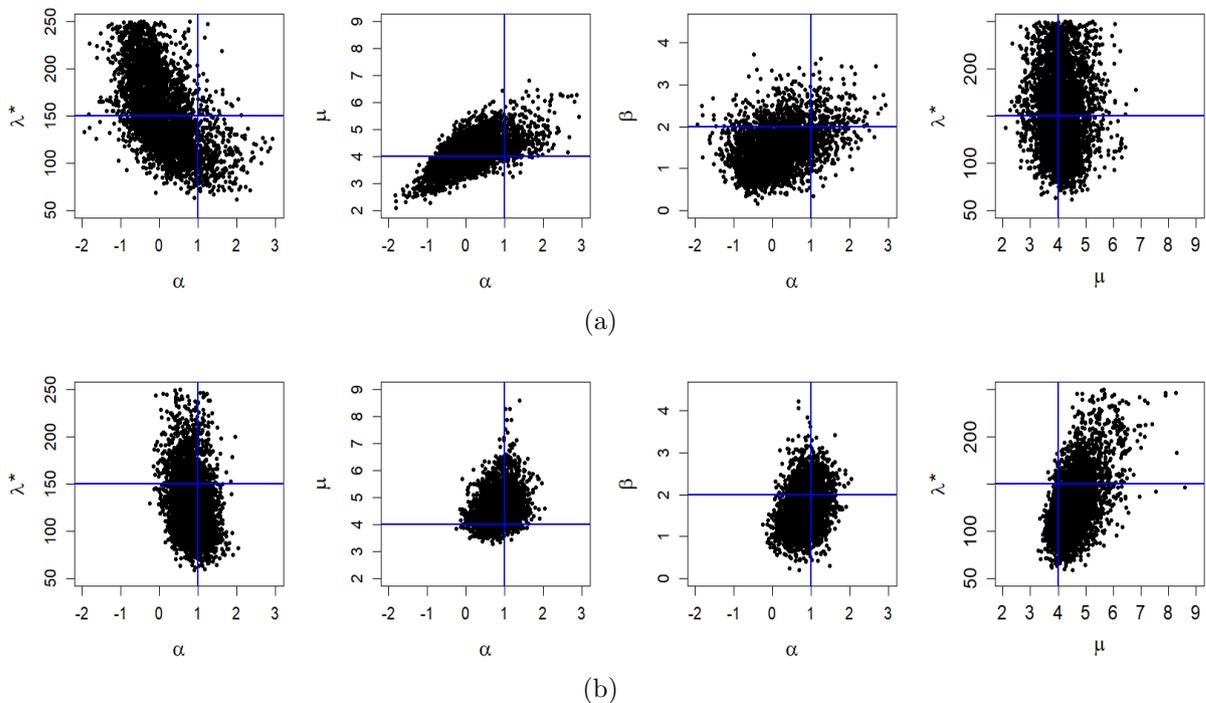


Figura 7.41: Gráfico de dispersão das estimativas dos parâmetros em relação a α para variância a priori de α igual a (a) 1 e (b) 0,1. A linha azul representa o valor verdadeiro do parâmetro.

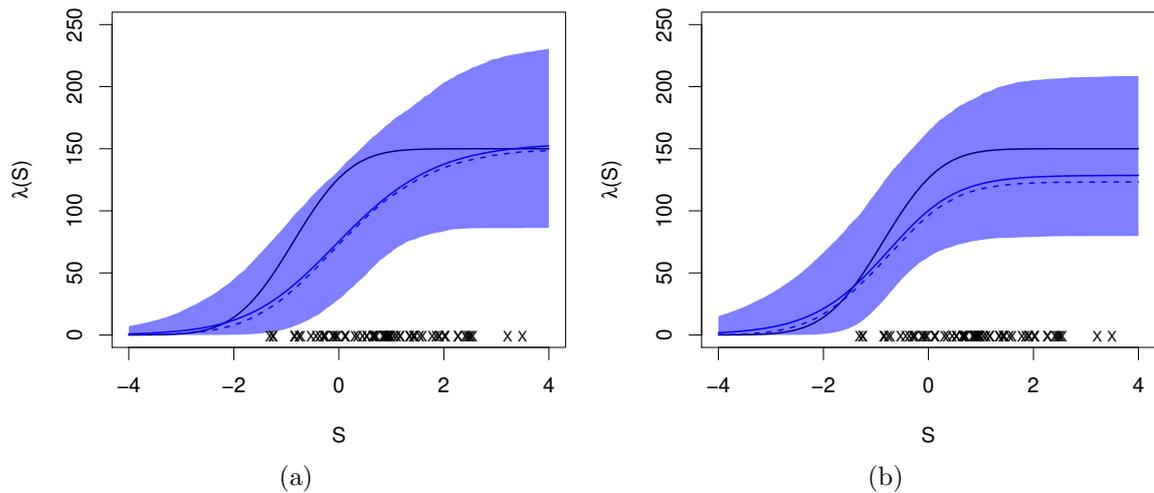


Figura 7.42: Taxa estimada pela média (linha contínua), mediana (linha tracejada) e intervalo de credibilidade de 95% considerando como variância a priori para α os valores (a) 1 e (b) 0,1 para os dados gerados com o valor de α igual a 1. A linha preta representa a taxa real e os pontos marcados em “x” correspondem aos valores simulados do processo S na amostra.

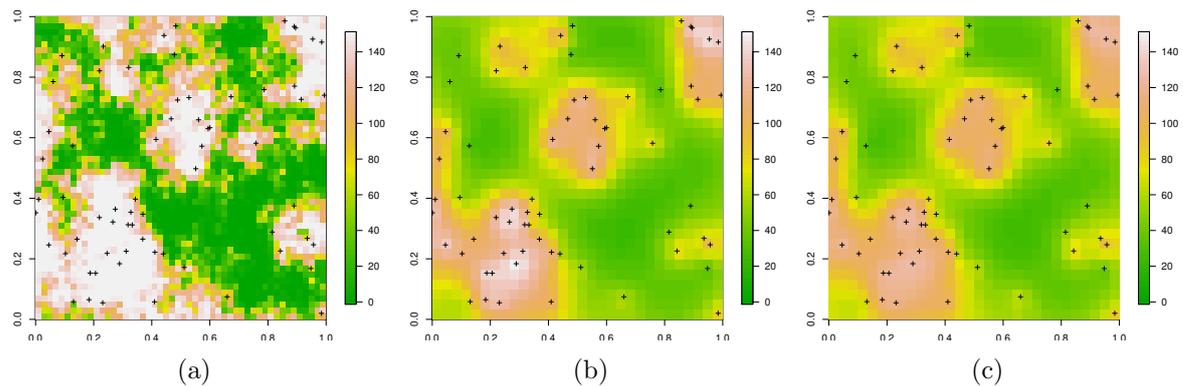


Figura 7.43: Mapas da função de intensidade (a) real e estimada pela média a posteriori considerando as variâncias a priori para α iguais a (b) 1 e (c) 0,1 para os dados gerados com o valor de α igual a 1.

Para analisar a predição da função de intensidade e do processo Y , foram escolhidos locais não amostrados da malha de predição de 30×30 com três características distintas: locais distantes de pontos da amostra, locais próximos de pontos da amostra e locais cuja intensidade real está entre 120 e 130. O objetivo de se escolher pontos com a última característica é analisar a distribuição preditiva da função de intensidade e do processo Y em locais cuja taxa estimada no caso em que α foi estimado em torno de zero está distante da taxa real. As coordenadas dos pontos escolhidos podem ser observadas na Tabela 7.3. A Figura 7.44 apresenta a distribuição a posteriori da função de intensidade estimada nos locais escolhidos. Apesar do comportamento distinto da taxa estimada na Figura 7.42

Tabela 7.3: Locais escolhidos para predição sobre uma malha 30×30 .

Locais distantes	Locais próximos	Taxa real entre 120 e 130
$\mathbf{x}_{83} = (0,750; 0,083)$	$\mathbf{x}_{36} = (0,183; 0,050)$	$\mathbf{x}_9 = (0,283; 0,017)$
$\mathbf{x}_{320} = (0,650; 0,350)$	$\mathbf{x}_{212} = (0,050; 0,250)$	$\mathbf{x}_{210} = (0,983; 0,217)$
$\mathbf{x}_{508} = (0,917; 0,550)$	$\mathbf{x}_{587} = (0,550; 0,650)$	$\mathbf{x}_{435} = (0,483; 0,483)$
$\mathbf{x}_{702} = (0,383; 0,783)$	$\mathbf{x}_{854} = (0,450; 0,950)$	$\mathbf{x}_{664} = (0,117; 0,750)$

para os dois casos considerados, observamos que a estimação da função de intensidade é semelhante quando consideramos a incerteza associada à estimação do processo Gaussiano. Dessa forma, na maioria dos locais escolhidos, obtivemos estimação semelhante da taxa quando α foi estimado em torno de zero e quando foi estimado corretamente.

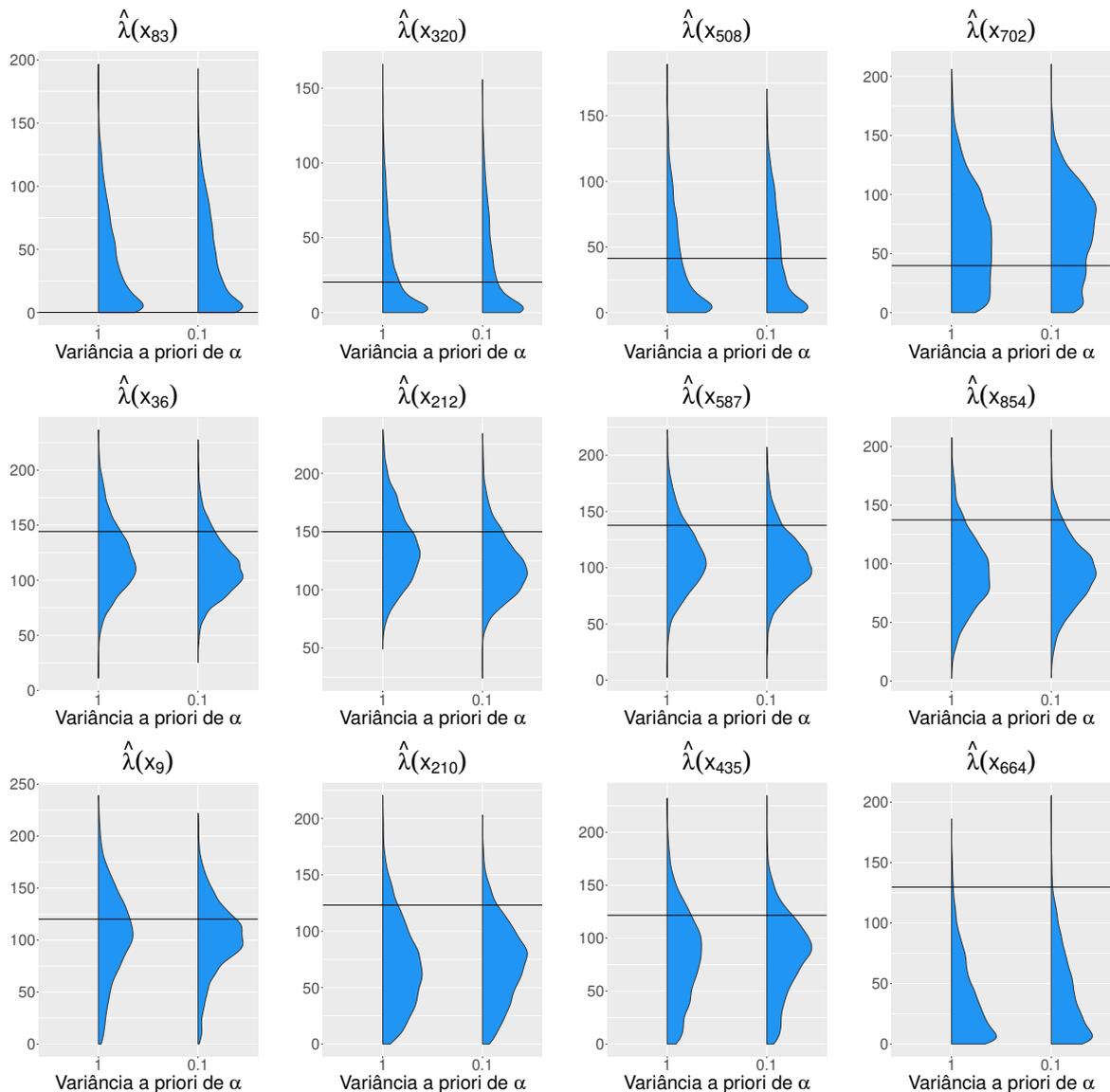


Figura 7.44: Distribuições a posteriori da função de intensidade estimada em locais distantes de pontos amostrados (primeira linha), em locais próximos de pontos amostrados (segunda linha) e em locais cuja intensidade real está entre 120 e 130 (terceira linha).

A Figura 7.45 apresenta o mapa predito do processo Y e podemos notar que a predição para esse processo é similar para os dois casos considerados. Observando a Figura 7.46, concluímos que as distribuições preditivas para Y são muito semelhantes independente se α foi estimado corretamente ou não. Esse resultado é importante pois, apesar de não conseguirmos recuperar bem a função de intensidade no primeiro caso (α estimado em torno de zero), a predição obtida para o processo de interesse Y nesse caso é análoga à predição obtida quando obtemos boa estimativa do intercepto.

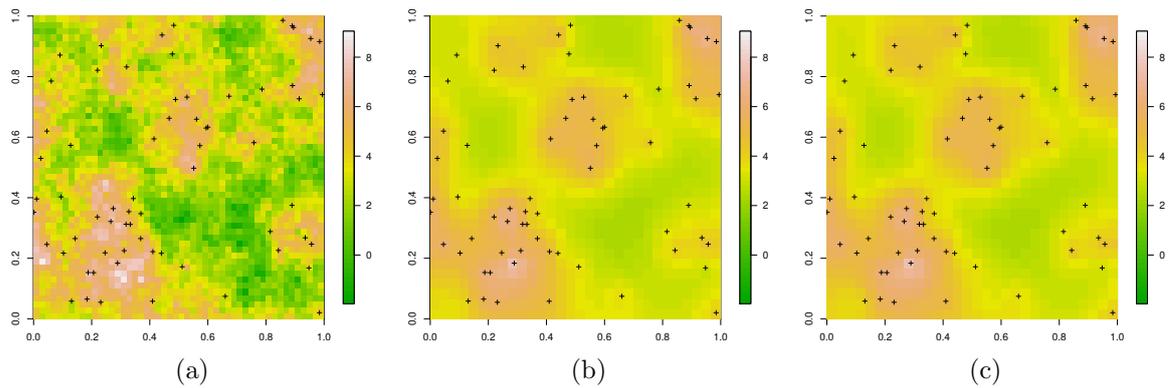


Figura 7.45: Mapas do processo Y (a) simulado e estimado pela média a posteriori considerando as variâncias a priori para α iguais a (b) 1 e (c) 0,1 para os dados gerados com o valor de α igual a 1.

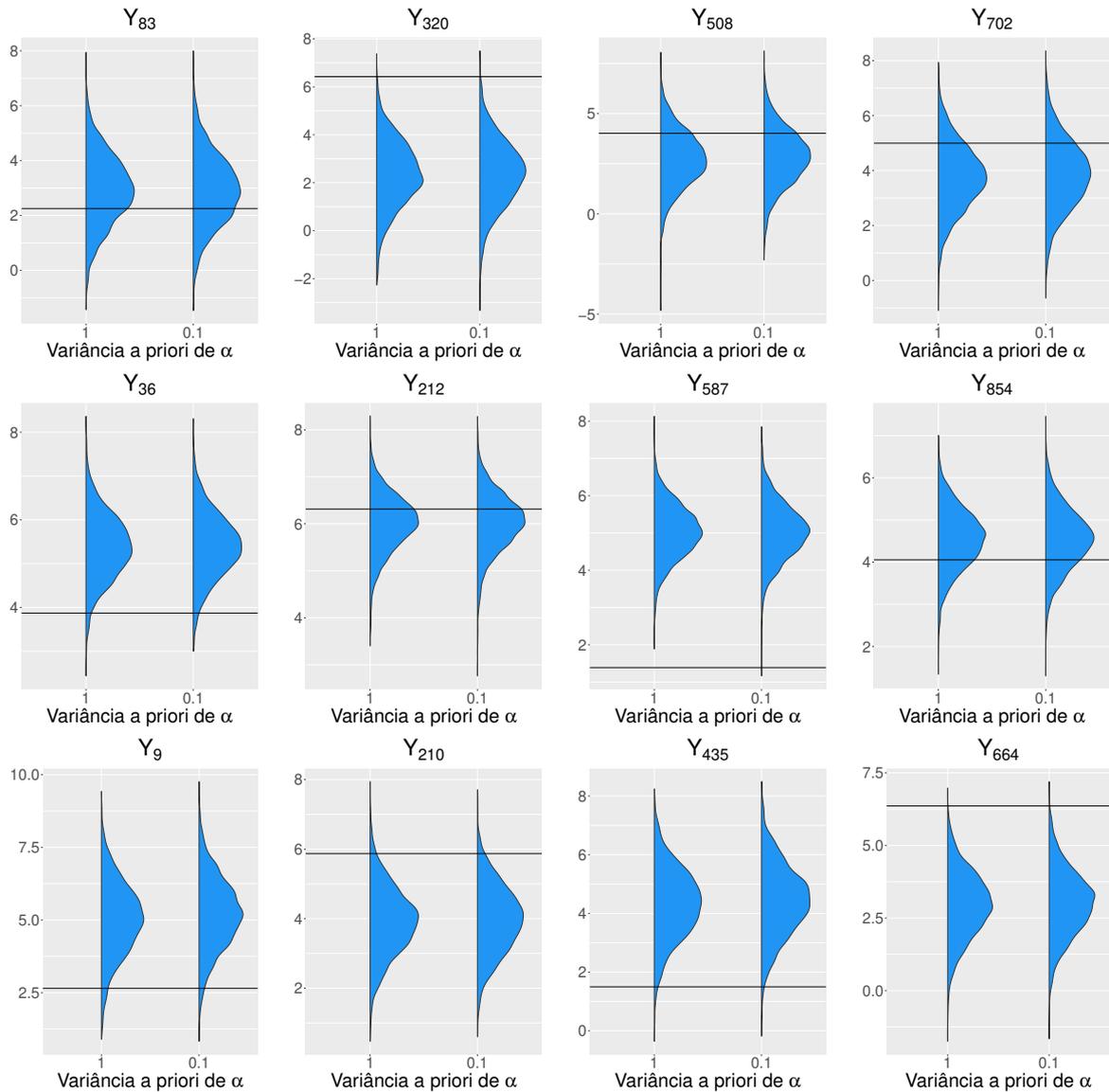


Figura 7.46: Distribuições preditivas a posteriori do processo Y em locais distantes de pontos amostrados (primeira linha), em locais próximos de pontos amostrados (segunda linha) e em locais cuja intensidade real está entre 120 e 130 (terceira linha).

7.6 Modelo com discretização

Para comparar os resultados obtidos pelo nosso modelo com os resultados do modelo que possui aproximação por discretização (modelo aproximado), foi realizada a estimação dos parâmetros do conjunto de dados 1 da Seção 7.1 considerando uma malha regular de tamanho 15×15 sobre o quadrado unitário. A construção do modelo aproximado se deu de forma similar ao modelo apresentado em Ferreira and Gamerman (2015), substituindo a função de intensidade exponencial pela função probito. Dessa forma, podemos identificar o ganho da modelagem que permite realizar inferência exata em comparação com o modelo aproximado.

A Figura 7.47 apresenta as distribuições a posteriori dos parâmetros do modelo. Podemos observar que a estimação é semelhante para os parâmetros λ^* , μ e β . Em relação aos parâmetros de variância e ao de correlação, notamos que o modelo exato fornece uma estimativa a posteriori mais concentrada em torno de seus valores verdadeiros.

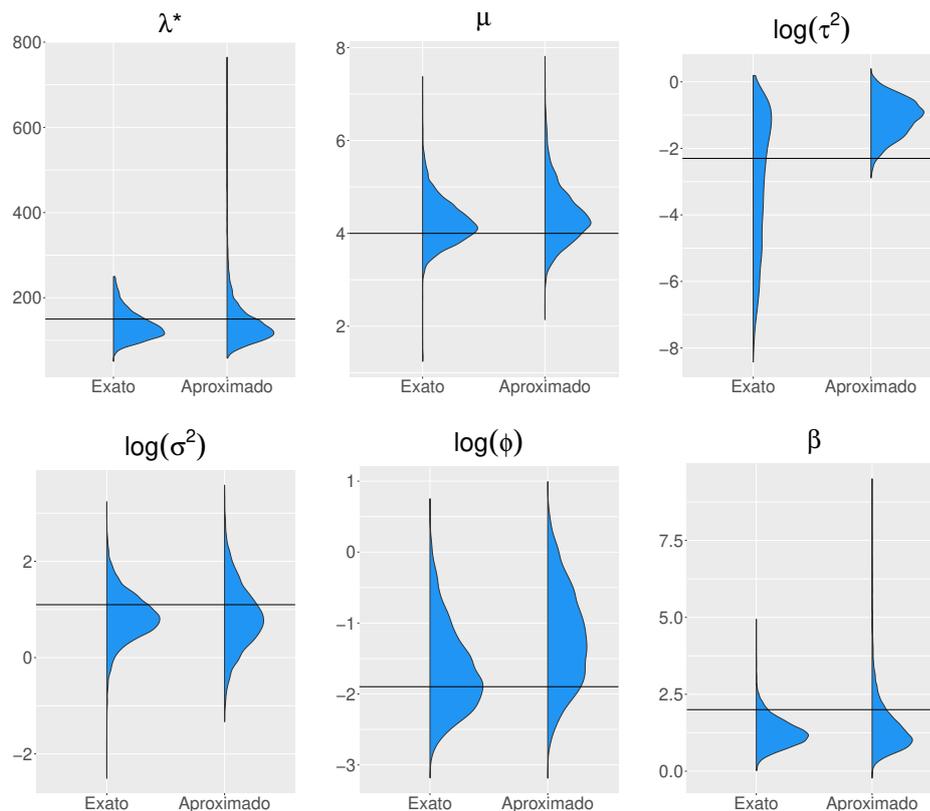


Figura 7.47: Densidades das distribuições a posteriori dos parâmetros dos modelos exato e aproximado com função de intensidade probito, considerando o conjunto de dados 1. A linha horizontal representa o valor verdadeiro do parâmetro.

A Figura 7.48 apresenta a função de intensidade estimada para ambos modelos.

Notamos uma melhor estimação pela média a posteriori do modelo com aproximação, porém este apresenta maior variabilidade a posteriori, como pode ser observado pela banda de credibilidade de 95%. A estimação de maiores valores da taxa acarretou uma estimação maior para a média a posteriori do modelo aproximado, ou seja, mesmo estando mais próxima da taxa real, a incerteza é maior quando comparada com a taxa estimada pelo modelo exato.

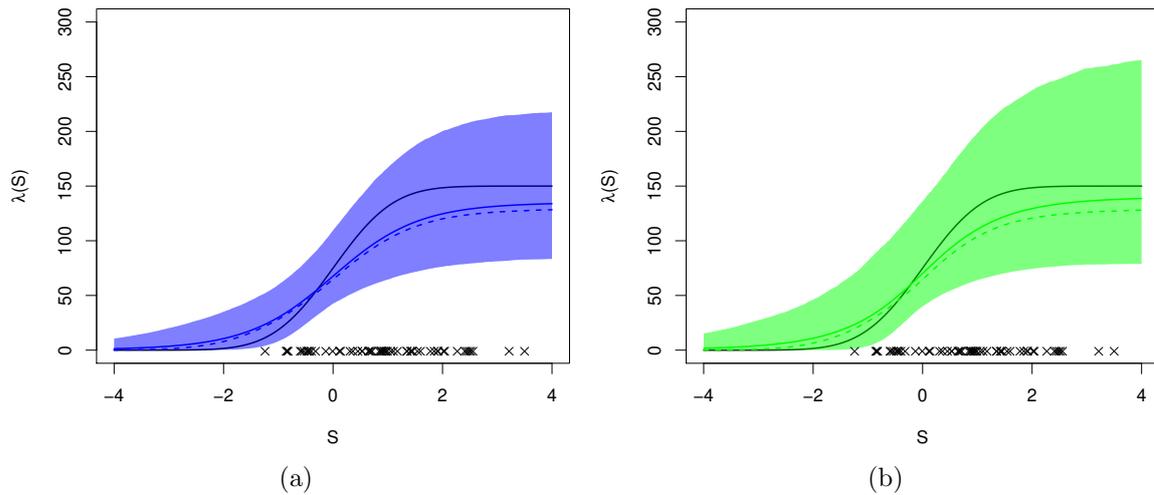


Figura 7.48: Taxa estimada pela média (linha contínua), mediana (linha tracejada) e intervalo de credibilidade de 95% considerando os dados do conjunto de dados 1. (a) modelo exato e (b) modelo com aproximação. A linha preta representa a taxa real. Os pontos marcados em “x” correspondem aos valores simulados do processo S na amostra.

Através de uma malha regular 30×30 , foi realizada a predição para os modelos. A Figura 7.49 apresenta a distribuição a posteriori da função de intensidade estimada em locais não observados na malha de predição que estão distantes ou próximos de pontos amostrados (suas coordenadas estão presentes na Tabela 7.3). Podemos observar estimação semelhante para a função de intensidade nos locais considerados para os dois modelos.

Observamos o mapa predito pelo média a posteriori do processo Y na Figura 7.50 e obtemos melhores predições para o modelo exato, principalmente em regiões que não possuem informação.

As Figuras 7.51 e 7.52 apresentam as medidas de qualidade de predição para ambos modelos a partir de dez conjuntos de dados simulados. Em relação ao processo S , obtemos melhores valores de PCIPG para o modelo exato. Para o processo Y , observamos melhores medidas associadas ao modelo exato para quase todos os conjuntos de dados considerados,

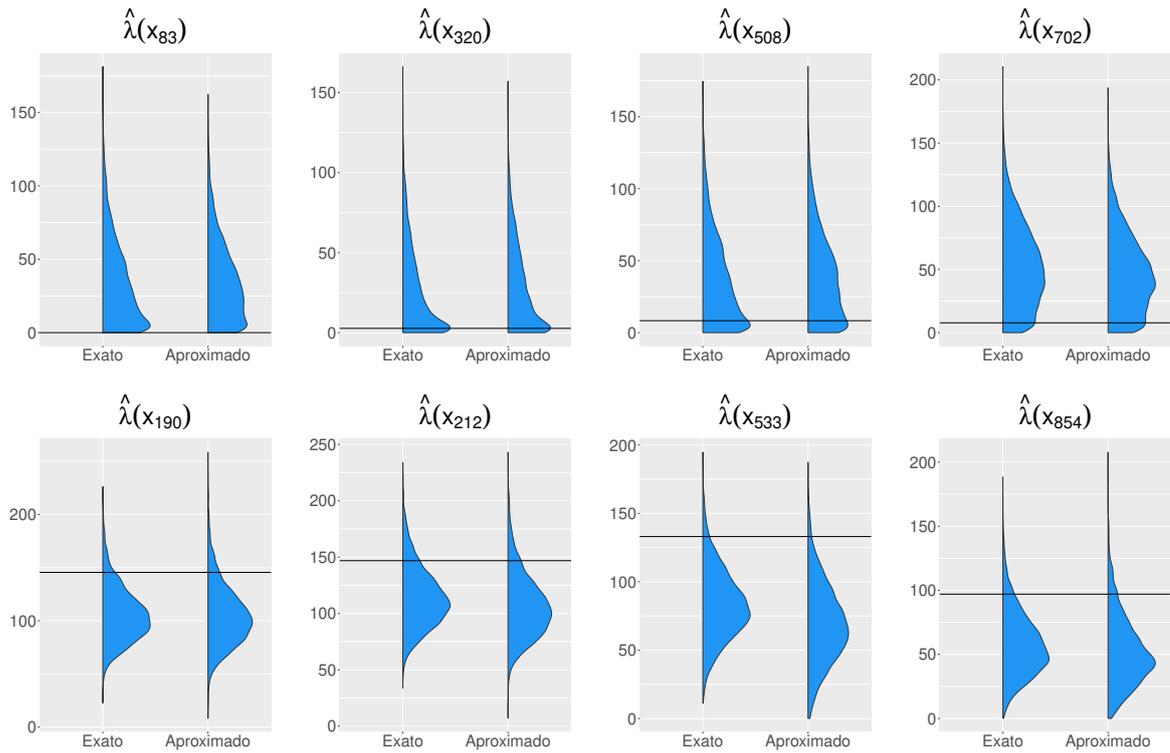


Figura 7.49: Distribuições a posteriori da função de intensidade estimada em locais distantes de pontos amostrados (primeira linha) e em locais próximos de pontos amostrados (segunda linha).

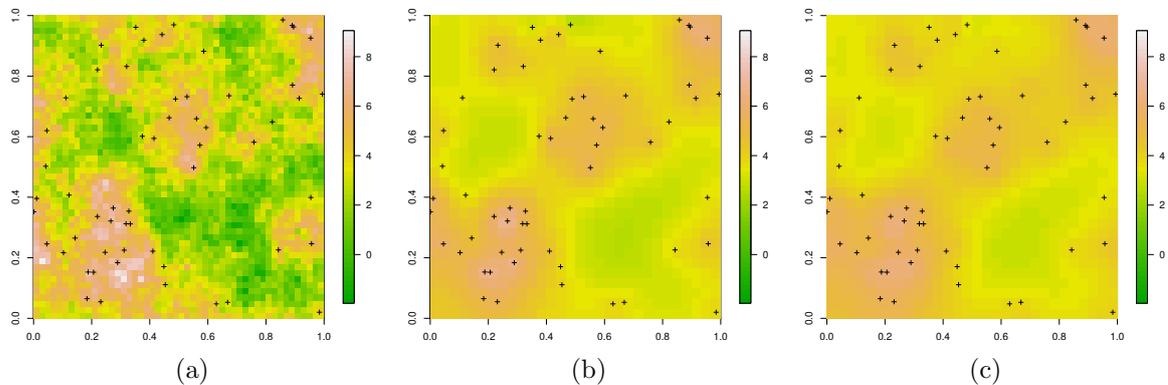


Figura 7.50: Mapa do processo Y (a) simulado pelo conjunto de dados 1 e estimado pela média a posteriori da distribuição preditiva considerando os modelos (b) exato e (c) aproximado.

evidenciando a melhor performance da nossa modelagem.

A Figura 7.53 apresenta a distribuição preditiva a posteriori de Y em alguns locais da malha de predição. Para pontos distantes de locais amostrados, observamos predição semelhante para os dois modelos. Porém, para pontos próximos aos locais da amostra, o modelo exato fornece predição mais precisa, apresentando distribuição preditiva mais concentrada em torno do valor verdadeiro.

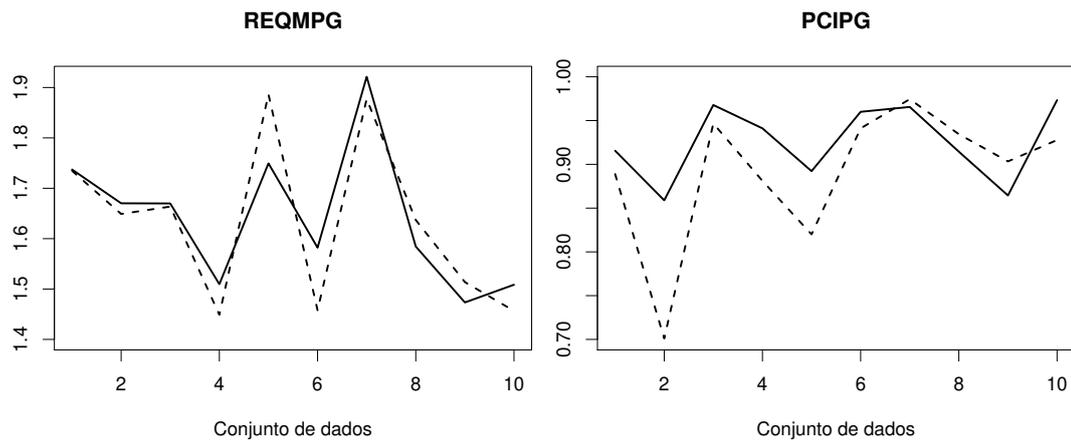


Figura 7.51: Medidas de qualidade de predição para o processo S dos modelos exato (linha contínua) e aproximado (linha tracejada) considerando os conjuntos de dados simulados com amostragem preferencial.

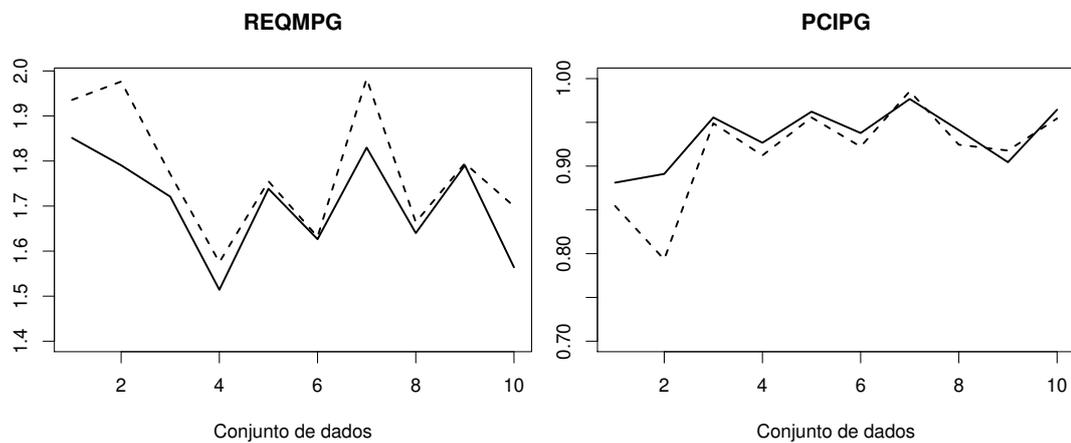


Figura 7.52: Medidas de qualidade de predição para o processo Y dos modelos exato (linha contínua) e aproximado (linha tracejada) considerando os conjuntos de dados simulados com amostragem preferencial.

Logo, podemos concluir que o modelo exato fornece valores preditos do processo de interesse mais precisos, tornando-o mais atraente. Isso ocorre pela capacidade do modelo de realizar inferência sobre locais exatos da região de estudo e não apenas em localizações aproximadas.

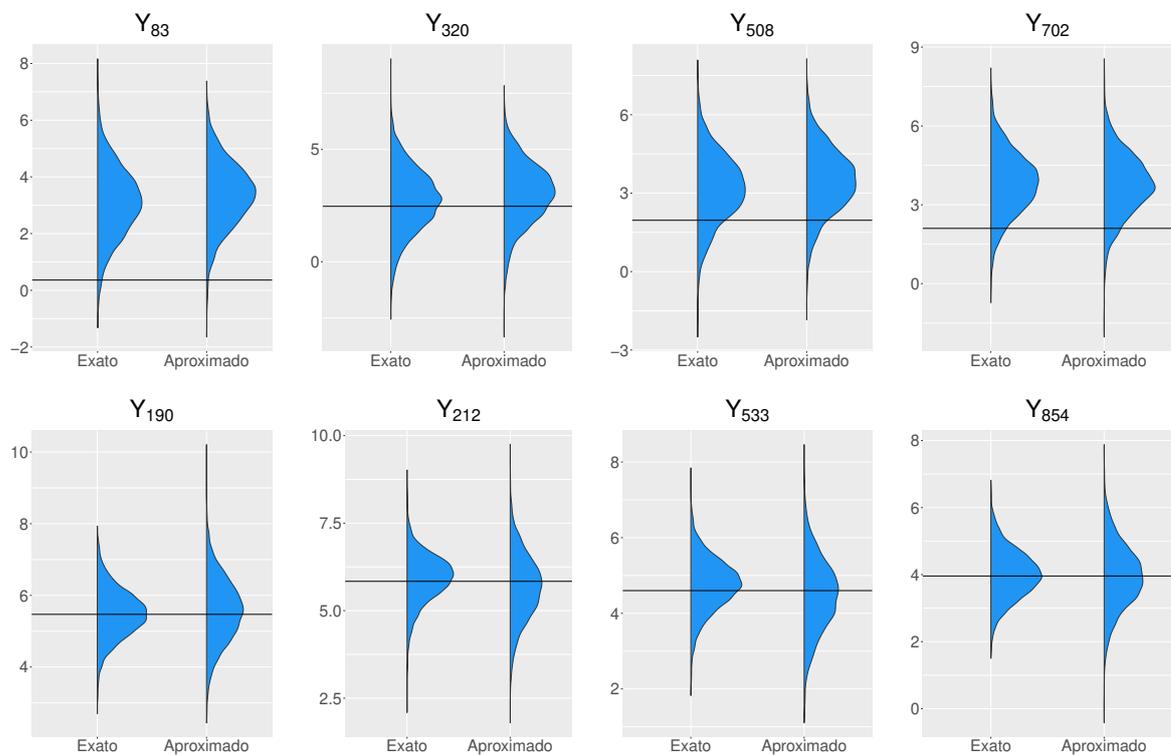


Figura 7.53: Distribuições preditivas a posteriori do processo Y em locais distantes de pontos amostrados (primeira linha) e em locais próximos de pontos amostrados (segunda linha).

7.7 Modelo t-Student

Para avaliar o desempenho do modelo quando incorporamos a distribuição t-Student, foi realizado um estudo de simulação considerando o modelo t no processo S e os parâmetros foram estimados pelos modelos t nos processos S e Y e pelo modelo normal. Os valores assumidos para os parâmetros foram $(\lambda^*, \mu, \tau^2, \sigma^2, \phi, \beta) = (160; 4; 0,10; 3; 0,15; 2)$ e 3 graus de liberdade para a distribuição t. Dessa forma, foram gerados cinco conjunto de dados, resultando em um número de observações igual a 84, 55, 114, 72, 64. Consideramos duas cadeias partindo de valores iniciais distintos, com 200 mil iterações cada. Assumimos um *burn in* de tamanho 20 mil e a amostra a posteriori foi construída a partir de um *lag* igual a 60 de cada cadeia, totalizando 6 mil observações. A análise de convergência das cadeias foi realizada visualmente a partir dos gráficos de valor versus iteração.

As Figuras 7.54 e 7.55 apresentam a distribuição a posteriori dos parâmetros dos modelos. Podemos observar que os resultados obtidos são semelhantes para os três modelos considerados, sendo que o modelo que assume a distribuição t no processo Y fornece maior variância a posteriori. Apesar disso, esse modelo apresenta melhor estimativa do parâmetro de preferência, tendo sua distribuição a posteriori melhor centrada em torno do verdadeiro valor de β .

Em relação a predição, as Figuras 7.56 e 7.57 apresentam os valores de REQMPG e PCIPG para os modelos considerados para os cinco conjuntos de dados. Notamos que a predição realizada pelo modelo correto com distribuição t no processo S é similar à predição fornecida pelo modelo normal, apresentando valores próximos de REQMPG e PCIPG tanto para os valores preditos do processo S quanto para os valores preditos do processo Y . O modelo com distribuição t no processo Y apresenta maior erro de predição, porém apresenta valores de PCIPG muito próximos de 1, indicando que quase todos os intervalos de credibilidade de 95% contém o verdadeiro valor do processo (ou S ou Y). Isso se dá pela maior variância da distribuição a posteriori dos valores preditos por esse modelo, como podemos observar na Figura 7.58 que apresentam as densidades a posteriori dos processos S e Y para o conjunto de dados 5.

Logo, o modelo com processo t em S e o modelo normal possuem performance similares. Veremos posteriormente, na Aplicação 3, que considerar o modelo t-Student (em S) forneceu melhores predições e melhor ajuste comparado ao modelo normal. Porém,

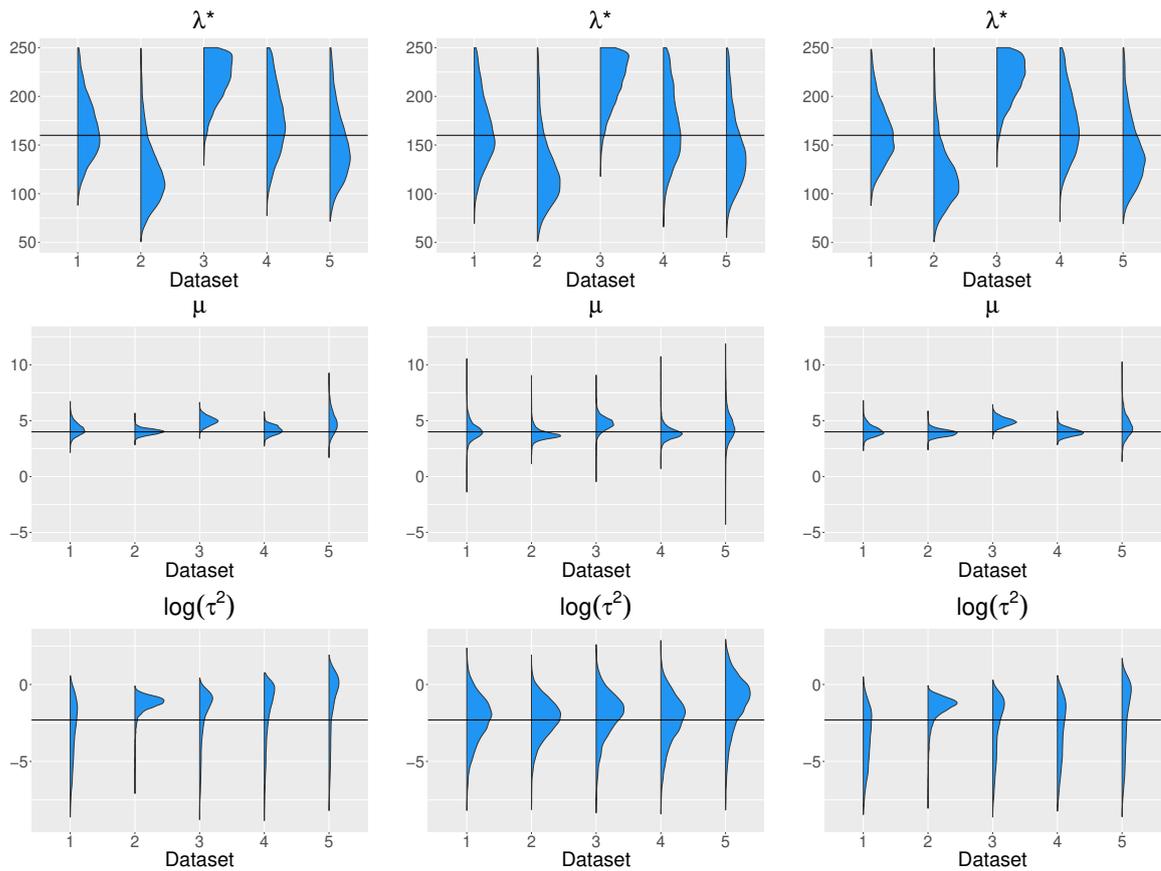


Figura 7.54: Densidades das distribuições a posteriori dos parâmetros λ^* , μ e τ^2 do modelo com distribuição t-Student no processo S (primeira coluna) e no processo Y (segunda coluna) e do modelo Gaussiano (terceira coluna) para os dados simulados com amostragem preferencial e distribuição t-Student no processo S . A linha horizontal representa o valor verdadeiro do parâmetro.

uma análise mais aprofundada se torna necessária para verificar as diferenças na utilização dos dois modelos.

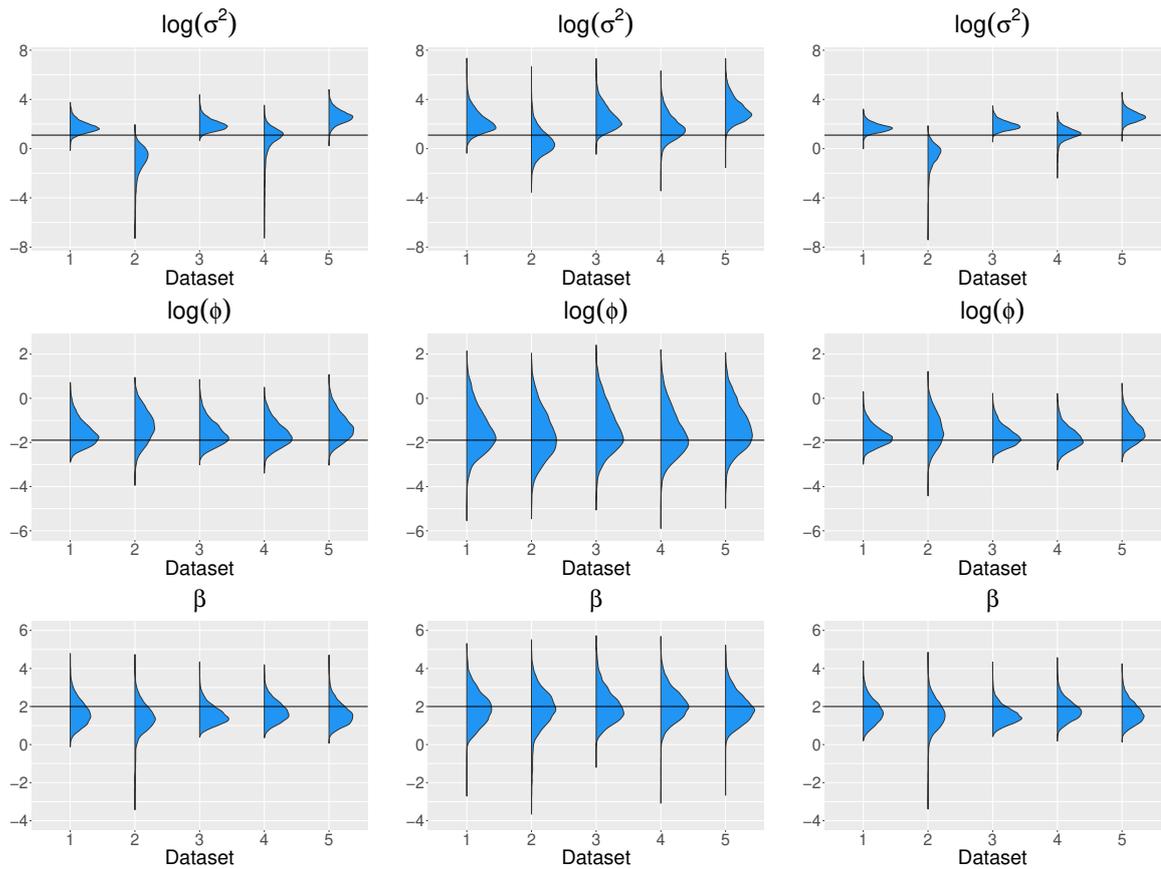


Figura 7.55: Densidades das distribuições a posteriori dos parâmetros σ^2 , ϕ e β do modelo com distribuição t-Student no processo S (primeira coluna) e no processo Y (segunda coluna) e do modelo Gaussiano (terceira coluna) para os dados simulados com amostragem preferencial e distribuição t-Student no processo S . A linha horizontal representa o valor verdadeiro do parâmetro.

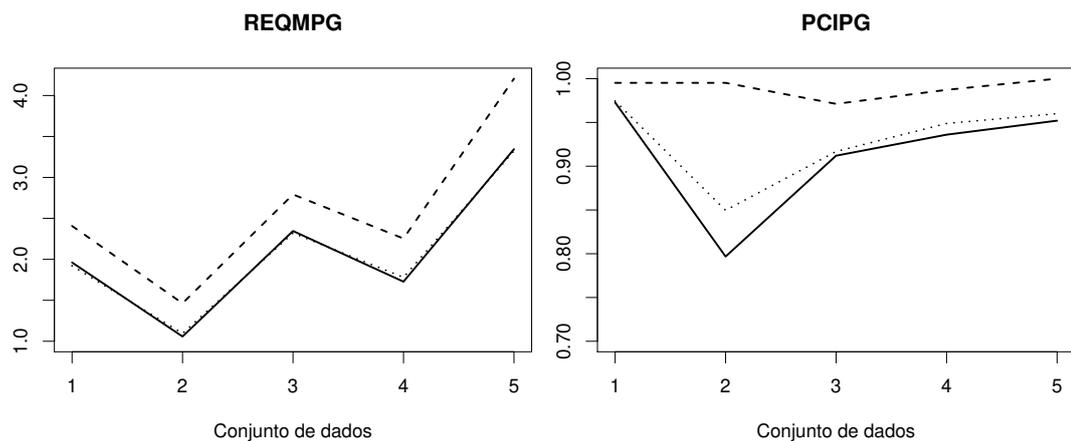


Figura 7.56: Medidas de qualidade de predição para o processo S dos modelos com distribuição t em S (linha contínua), em Y (linha tracejada) e modelo Gaussiano (linha pontilhada), considerando os conjuntos de dados simulados com distribuição t no processo S .

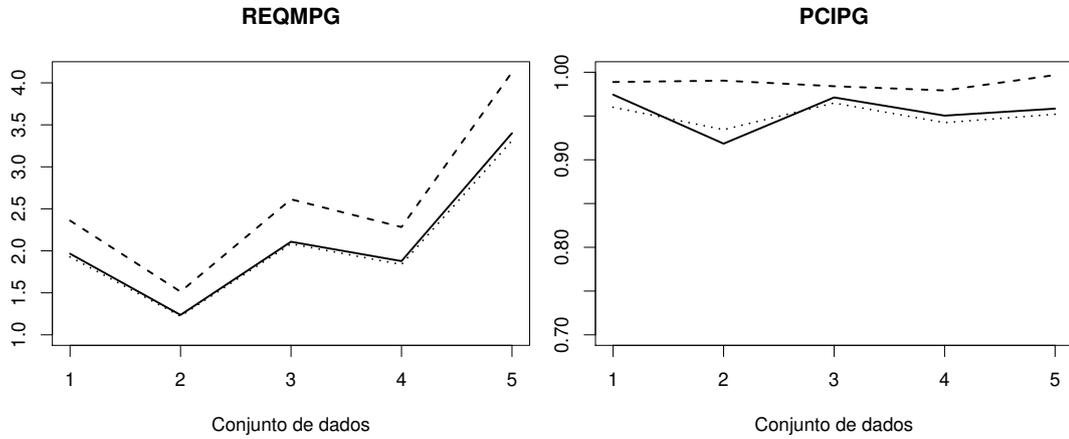


Figura 7.57: Medidas de qualidade de predição para o processo Y dos modelos com distribuição t em S (linha contínua), em Y (linha tracejada) e modelo Gaussiano (linha pontilhada), considerando os conjuntos de dados simulados com distribuição t no processo S .

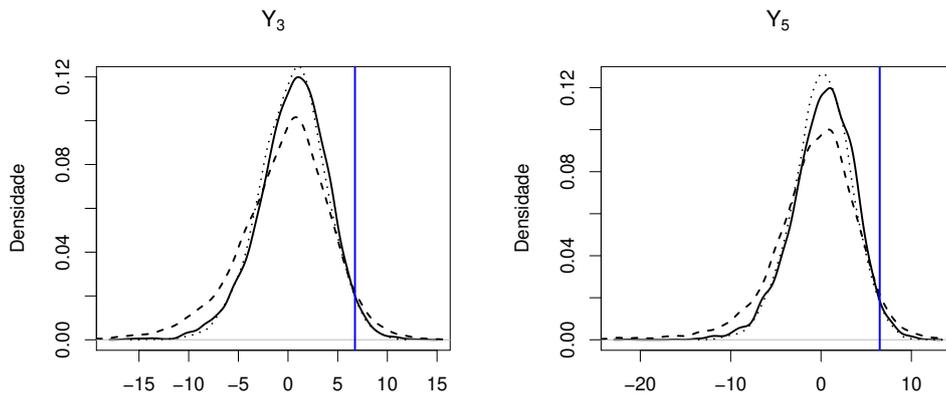


Figura 7.58: Densidades a posteriori de Y_3 e Y_5 preditos pelos modelos com distribuição t em S (linha contínua), em Y (linha tracejada) e modelo Gaussiano (linha pontilhada), considerando os conjuntos de dados simulados com distribuição t no processo S .

Capítulo 8

Aplicações a dados reais

Este capítulo é destinado a aplicação da metodologia discutida em três conjuntos de dados reais. A Aplicação 1 se refere aos dados de poluição em Galícia, que apresenta o estudo em dois anos, sendo que em um deles a amostragem é preferencial e no outro é não preferencial. Apresentamos na Aplicação 2 os dados de radiação da Alemanha, cujo desenho amostral é não preferencial. Por fim, desenvolvemos na Aplicação 3 a análise dos dados de precipitação da cidade do Rio de Janeiro, que apresenta possível amostragem preferencial e *outlier*.

8.1 Aplicação 1: Dados de poluição em Galícia

O conjunto de dados de monitoramento de metais pesados em Galícia é bastante conhecido e utilizado na literatura, veja Fernández et al. (2000), Diggle et al. (2010) e Dinsdale and Salibian-Barrera (2019). Os dados se referem à concentração de chumbo em amostras de musgo, em micrograma por grama de peso seco, realizadas em outubro de 1997 e julho de 2000. O acúmulo de metais pesados em musgos ocorre, em sua maioria, pela deposição pelo ar, o que os torna um indicador de biomonitoramento (Fernández et al., 2000). Os dados estão disponíveis no pacote *PrevMap* (Giorgi and Diggle, 2017) no *software* R.

A Figura 8.1 apresenta o mapa de Galícia e os locais amostrados em ambos os anos da pesquisa. A escolha dos locais para amostragem no estudo de 1997 foi realizada considerando as regiões em que se espera maior concentração de chumbo, tornando o desenho amostral potencialmente preferencial. Em relação ao ano 2000, a amostragem se deu em uma malha aproximadamente regular e esperamos que a preferenciabilidade seja nula (ou quase nula).

A transformação logarítmica foi aplicada aos dados para torná-los mais simétricos. A Figura 8.2 apresenta os gráficos de histograma para os dados transformados e a Tabela 8.1 apresenta estatísticas descritivas. É importante observar que, mesmo após a transformação logarítmica, há duas observações atípicas como evidencia o boxplot dos dados da Figura 8.3.

Tabela 8.1: Estatísticas descritivas para os dados de log-concentração de chumbo em amostras de musgo.

Ano	n	Min	Quartil 1	Mediana	Média	Quartil 3	Max	D.p.	CV
1997	63	0,515	1,091	1,385	1,440	1,851	2,252	0,480	0,333
2000	132	-0,223	0,336	0,587	0,661	0,955	2,163	0,432	0,653

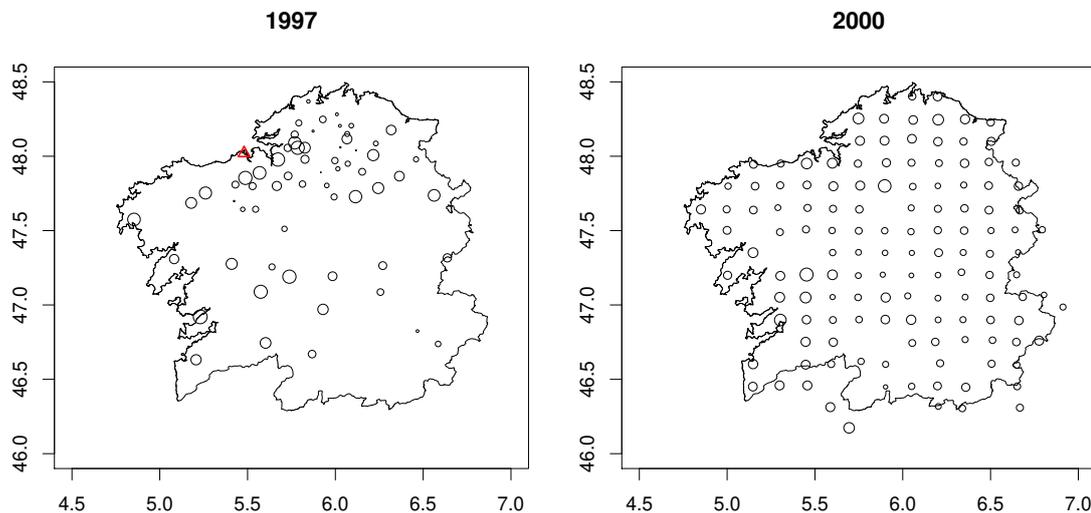


Figura 8.1: Mapa de Galícia e locais amostrados em ambos anos. O diâmetro do círculo é proporcional ao valor amostrado.

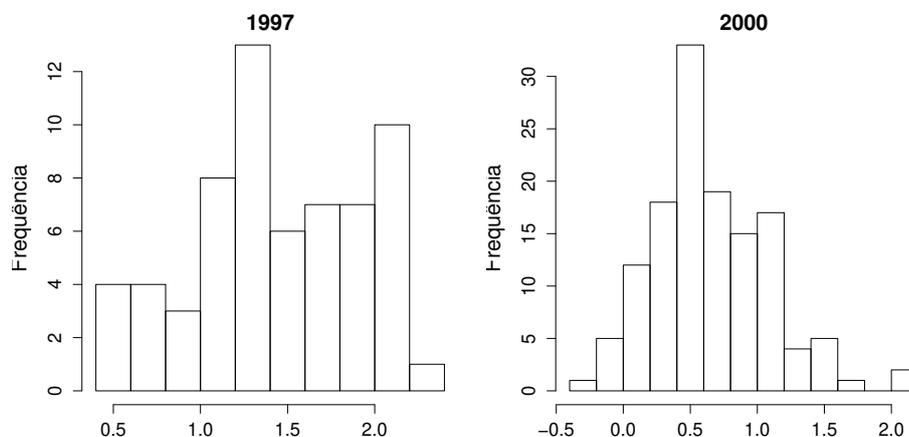


Figura 8.2: Histogramas dos dados de log-concentração de chumbo em musgos.

Para verificar se o processo analisado é isotrópico, construímos os gráficos de semi-variância em função das direções principais para os dois anos que podem ser observados na Figura 8.4. Note que, em todas as quatro direções, o semivariograma apresenta comportamento semelhante, o que indica isotropia. Dessa forma, procedemos com a análise para modelos isotrópicos.

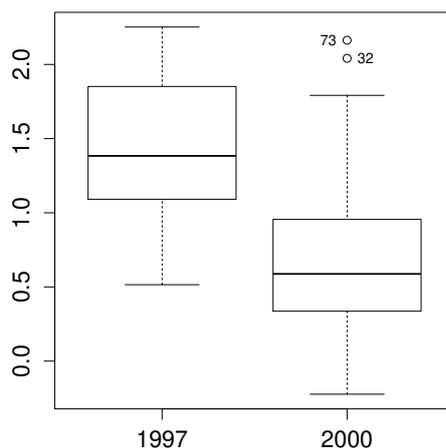


Figura 8.3: Gráfico de boxplot dos dados de log-concentração de chumbo em musgos em Galícia.

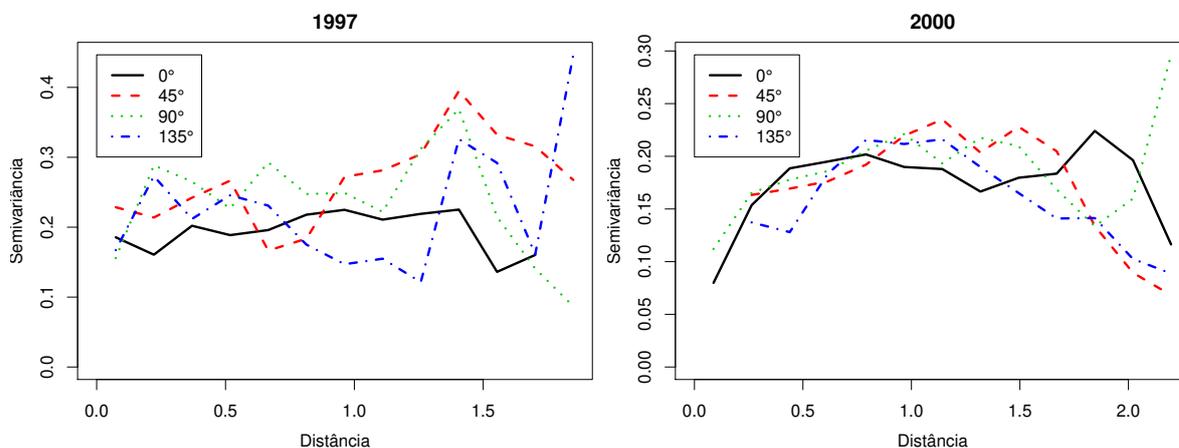


Figura 8.4: Gráfico de semivariograma dos dados de log-concentração de chumbo em musgos em Galícia.

Inferência clássica

A estimação dos parâmetros do modelo foi realizada considerando função de correlação exponencial, tamanho de malha 400 e através dos métodos MCLA, TMB e MCEM. Para comparação, foi considerado o modelo não preferencial Gaussiano (NPG). A Tabela 8.2 apresenta os resultados da estimação. Note que as estimativas dos parâmetros em comum dos modelos NPG e MCLA são iguais, diferindo apenas o erro padrão. Isso ocorre porque a estimação por MCLA utiliza a distribuição não preferencial marginal de Y , que não depende dos valores simulados de S . Analisando os resultados da pesquisa de 1997, podemos observar que o valor de $\hat{\mu}$ é maior para os métodos TMB e MCEM, que consideram a distribuição preditiva correta de $S|\mathbf{x}, \mathbf{y}$. Como temos uma amostragem preferencial

nesse ano, estes métodos fornecem uma estimativa mais correta do parâmetro de média. Por outro lado, os métodos NPG e MCLA resultam em estimativas viesadas, uma vez que não consideram a característica preferencial do processo na estimação desse parâmetro. Sobre o parâmetro de preferência, a estimativa fornecida pelo TMB é maior em comparação as fornecidas por MCEM e possui maior erro padrão. Dessa forma, obtemos estimativas mais precisas utilizando MCEM.

Tabela 8.2: Estimativas dos parâmetros dos modelos ajustados para os dados de concentração de chumbo nos anos de 1997 e 2000 considerando distribuição normal (erro padrão em parênteses).

	1997				2000			
	NPG	MCLA	TMB	MCEM	NPG	MCLA	TMB	MCEM
μ	1,542 (0,113)	1,542 (0,113)	1,708 (0,159)	1,985 (0,050)	0,724 (0,100)	0,724 (0,100)	0,923 (0,191)	0,702 (0,026)
τ^2	0,083 (0,042)	0,083 (0,042)	0,171 (0,039)	0,138 (0,027)	0,000 (0,047)	0,000 (0,001)	0,057 (0,026)	0,056 (0,010)
σ^2	0,147 (0,062)	0,146 (0,062)	0,059 (0,062)	0,194 (0,033)	0,192 (0,058)	0,192 (0,037)	0,174 (0,073)	0,144 (0,015)
ϕ	0,193 (0,120)	0,193 (0,120)	0,521 (0,836)	0,268 (0,046)	0,206 (0,082)	0,206 (0,052)	0,425 (0,260)	0,137 (0,018)
β	-	-1,514 (0,346)	-3,896 (2,785)	-2,754 (0,371)	-	-0,481 (0,254)	-0,847 (0,308)	-0,269 (0,246)

O principal objetivo em geoestatística é prever a variável de interesse (Y) em locais não observados e a Figura 8.5 apresenta os mapas preditos de log-concentração de chumbo em uma malha 20×20 sobre a região de Galícia em 1997. Para a construção dos mapas, utilizamos os valores estimados dos parâmetros como verdadeiros e os plugamos na distribuição preditiva considerada. Para comparação, realizamos a predição para cada combinação de método de estimação (MCLA, TMB e MCEM) e de predição (krigagem, moda da distribuição preditiva de S e algoritmo MH). Como esperado, não conseguimos prever valores longe da média estimada quando utilizamos a krigagem, para cada método de estimação. Por outro lado, quando consideramos a predição pela moda da distribuição preditiva e pelo algoritmo MH desenvolvido, que consideram a distribuição de X , obtemos um intervalo maior de valores preditos e valores maiores nas regiões não amostradas. Em comparação aos métodos de estimação, a predição realizada com os parâmetros estimados pelo MCEM apresentou os maiores valores nos locais distantes de pontos amostrados, também apresentando menor suavidade do mapa. Com os parâmetros estimados pelo TMB, obtemos uma predição mais suave. Vale ressaltar que essa diferença se dá, em grande parte, pelas médias estimadas pelos métodos, que apresentam uma diferença de

quase 0,3 entre a estimada pelo MCEM e pelo TMB. A Figura 8.6 apresenta o erro de predição para as combinações de método de estimação e de predição. Obtemos maior erro de predição através dos parâmetros estimados pelo MCEM, o que é justificado por apresentar maior valor estimado do efeito pepita.

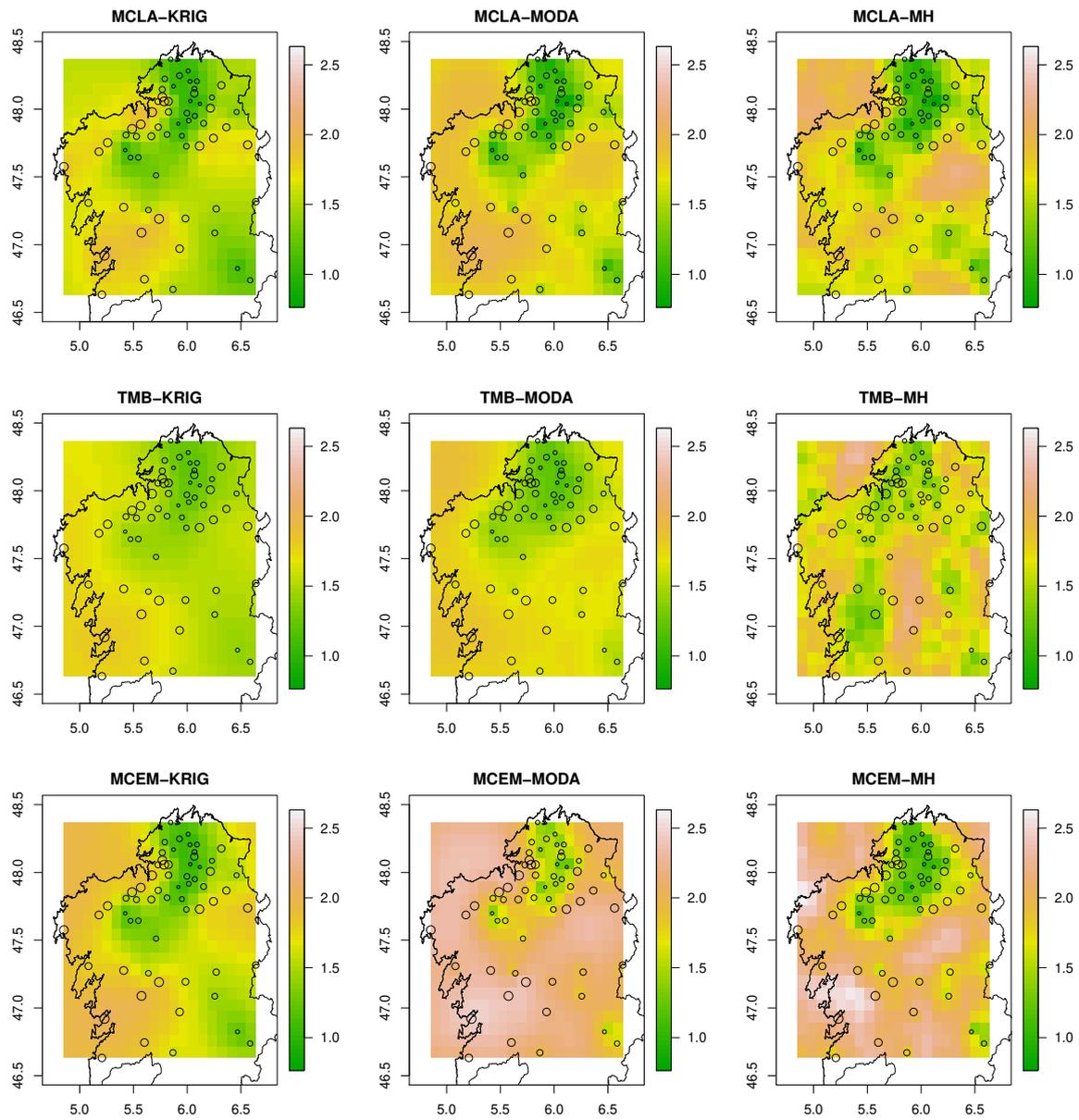


Figura 8.5: Mapas de predição do processo Y para as combinações de métodos de estimação e métodos de predição dos dados de log-concentração de chumbo em Galícia no ano de 1997.

Como sabemos que a amostragem realizada no ano de 2000 é não preferencial, o melhor modelo dentre os discutidos é o modelo NPG, pois sabemos que o parâmetro β é igual a zero. Como os parâmetros estimados por NPG e MCLA são iguais (nos parâmetros comuns), o mapa predito para o modelo NPG é o mesmo que o mapa para o

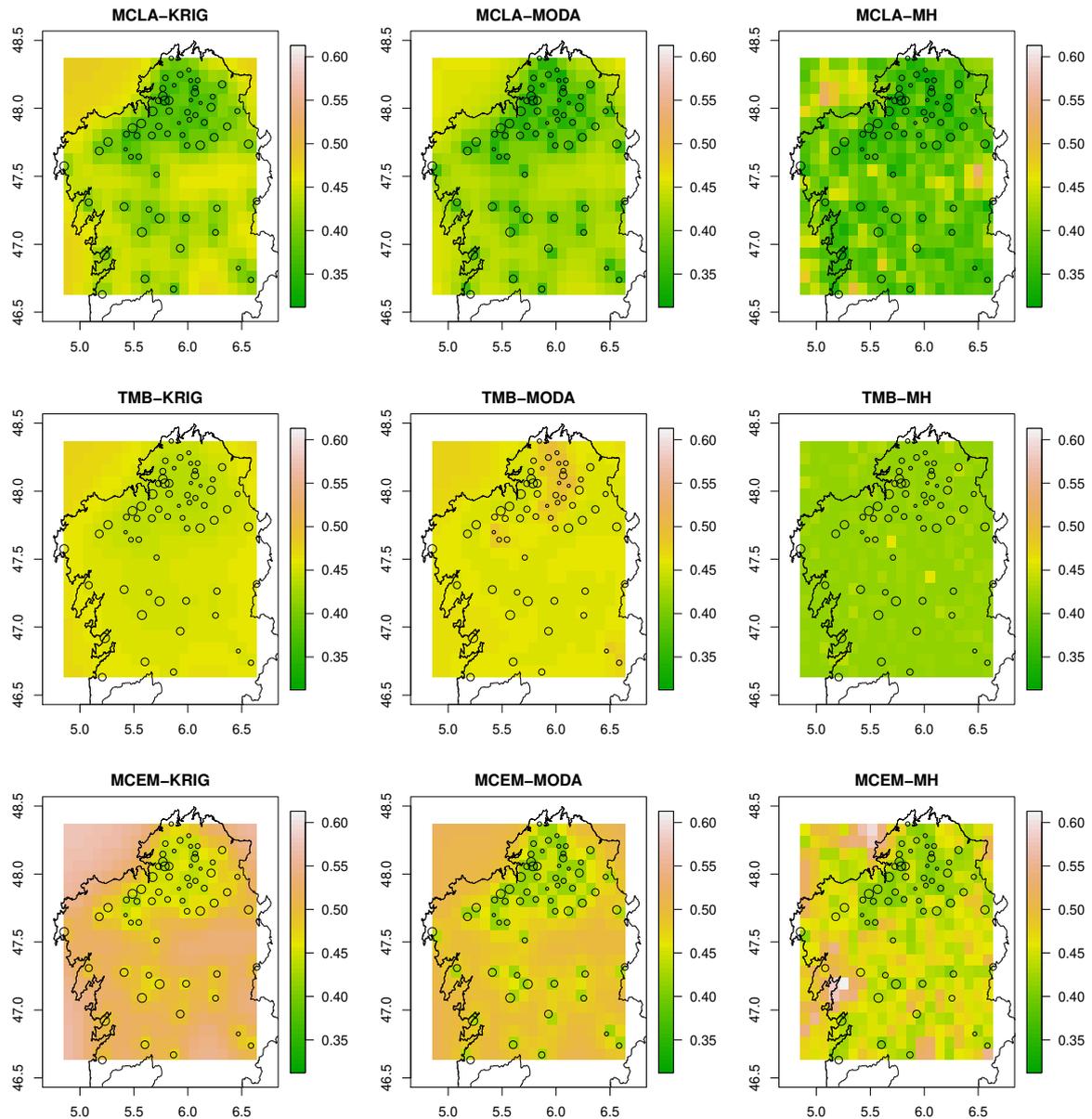


Figura 8.6: Mapas de erro de predição do processo Y para as combinações de métodos de estimação e métodos de predição dos dados de log-concentração de chumbo em Galícia no ano de 1997.

método MCLA combinado com a krigagem. Dessa forma, podemos comparar os mapas obtidos com o mapa do modelo MCLA-KRIG. A Figura 8.7 apresenta os mapas preditos do processo Y para cada combinação de método de estimação e de predição. Notamos que a combinação TMB-MODA, que é fornecida pelo pacote TMB, se aproxima mais da predição do modelo não preferencial. No caso do MCEM, os valores preditos foram superestimados em comparação ao modelo referência. Para esse ano, se considerássemos analisar os dados por um modelo com amostragem preferencial, o mais adequado seria utilizar o pacote TMB para obter os valores preditos. A Figura 8.8 apresenta o erro de predição do processo

Y e observamos que este é maior para os métodos de estimação TMB e MCEM em relação ao modelo de referência. É importante notar que, mesmo utilizando a krigagem, o erro de predição é maior quando consideramos os parâmetros estimados pelos métodos que consideram a amostragem preferencial, o que é esperado pois há mais incerteza no modelo devido ao parâmetro de preferência.

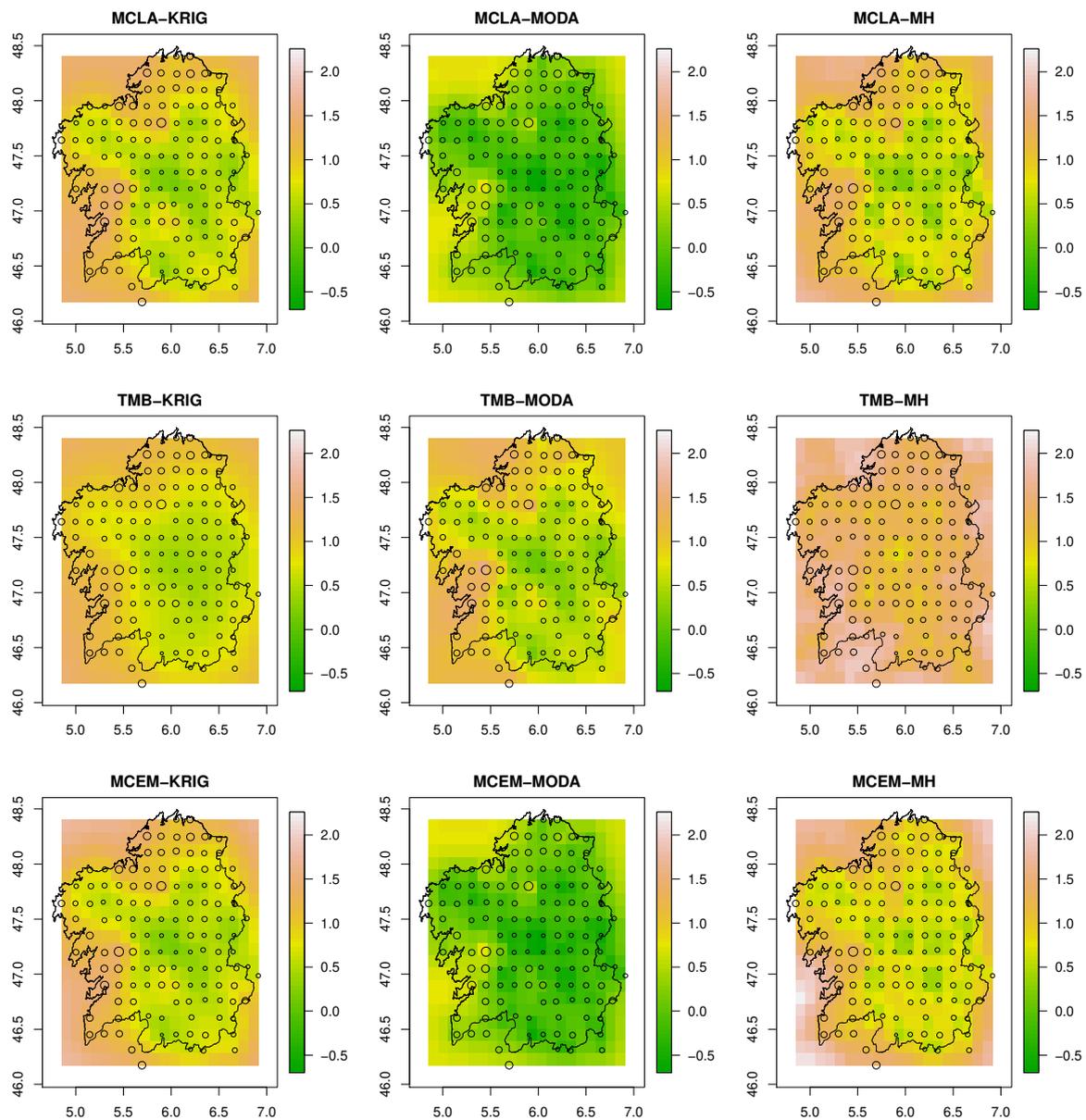


Figura 8.7: Mapas de predição do processo Y para as combinações de métodos de estimação e métodos de predição dos dados de log-concentração de chumbo em Galícia no ano de 2000.

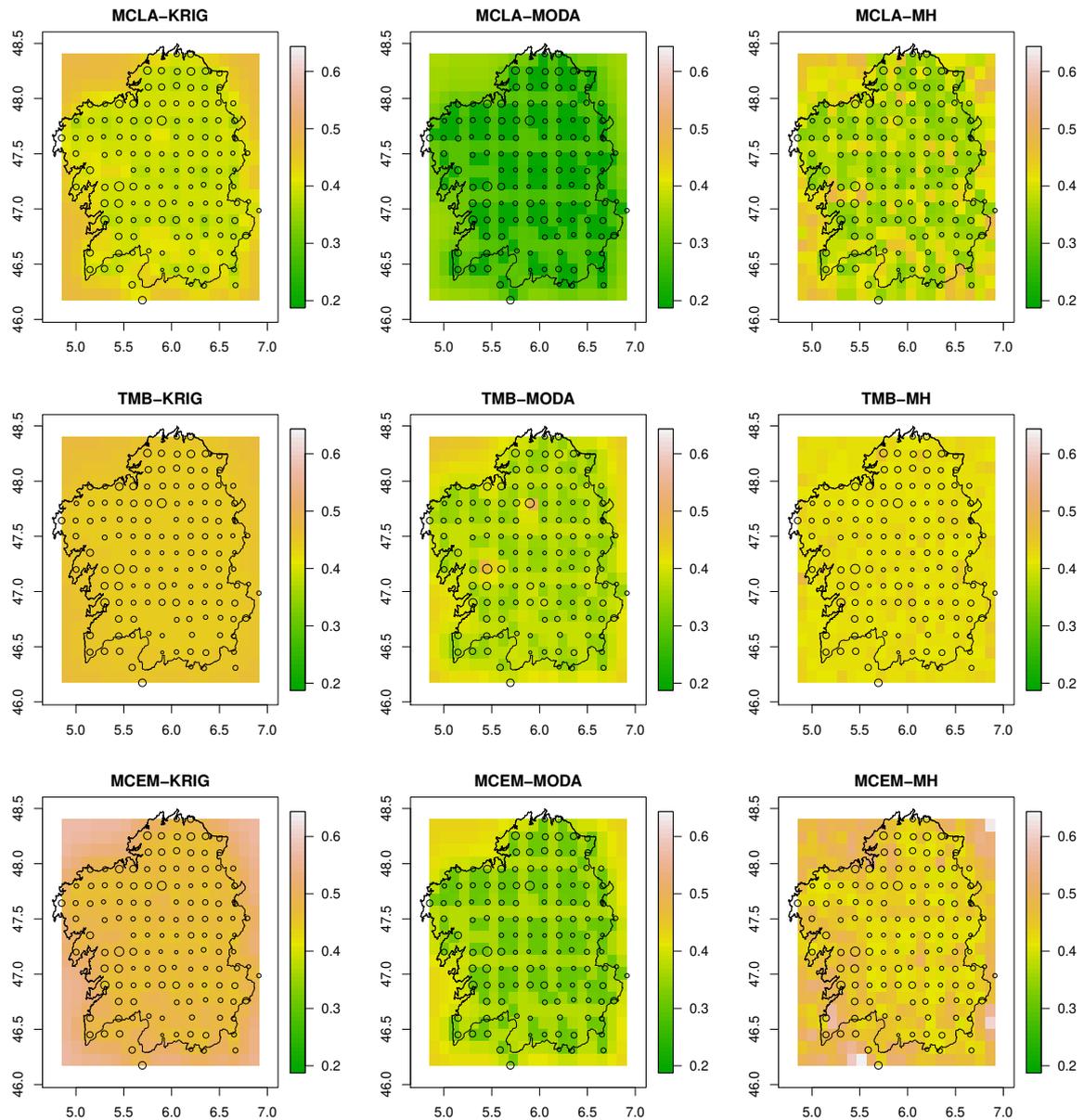


Figura 8.8: Mapas de erro de predição do processo Y para as combinações de métodos de estimação e métodos de predição dos dados de log-concentração de chumbo em Galícia no ano de 2000.

Inferência Bayesiana

Os dados também foram analisados a partir da estatística Bayesiana. Para o modelo com amostragem preferencial, o valor do parâmetro de correlação foi fixado em 0,5. Esse valor foi escolhido através da análise do variograma e da função de correlação exponencial, cujo alcance (distância em que o variograma estabiliza) é dado por $\phi/3$. Para o parâmetro λ^* , foi considerada uma distribuição a posteriori $\text{GammaT}(0,001; 0,001; 500/|B|)$, indicando que o número médio máximo de pontos do processo descartado é igual a 500. A convergência das cadeias de Markov foi verificada por análise gráfica e os gráficos das trajetórias dos parâmetros podem ser observados no Apêndice B.

A Tabela 8.3 apresenta as medidas resumo das distribuições a posteriori dos parâmetros para os modelos com amostragem preferencial (AP) e sem amostragem preferencial (SAP) para os dados do ano de 1997. É interessante notar que ambos modelos apresentaram estimativas semelhantes do parâmetro de média. Como esperado, o parâmetro de preferência foi estimado negativo e o intervalo de credibilidade de 95% não contém o valor zero, indicando que os dados do ano de 1997 possuem amostragem preferencial. As densidades a posteriori dos parâmetros podem ser observadas na Figura 8.9.

Tabela 8.3: Medidas resumo a posteriori para os parâmetros dos modelos com amostragem preferencial (AP) e sem amostragem preferencial (SAP) para os dados de Galícia de 1997.

Modelo	Parâmetro	Média	Mediana	Desvio Padrão	HPD(95%)
AP	λ^*	61,336	58,457	17,910	[34,751 ; 103,215]
	μ	1,590	1,582	0,139	[1,326 ; 1,892]
	τ^2	0,138	0,134	0,046	[0,055 ; 0,235]
	σ^2	0,211	0,173	0,150	[0,030 ; 0,604]
	β	-1,414	-1,355	0,437	[-2,436 ; -0,733]
SAP	μ	1,551	1,547	0,242	[1,018 ; 2,061]
	τ^2	0,105	0,102	0,044	[0,023 ; 0,202]
	σ^2	0,251	0,209	0,168	[0,055 ; 0,690]
	ϕ	0,514	0,430	0,329	[0,116 ; 1,385]

A Figura 8.10 apresenta os mapas de predição pela média a posteriori e o erro de predição a posteriori do processo Y para os dados do ano 1997. Os gráficos estão na mesma escala dos mapas obtidos pela abordagem clássica. Apesar de as estimativas para μ dos modelos AP e SAP serem semelhantes, observamos que o modelo AP fornece valores preditos maiores que o modelo SAP em locais não observados distantes dos pontos da amostra. Logo, notamos que o modelo AP é mais adequado para os dados de

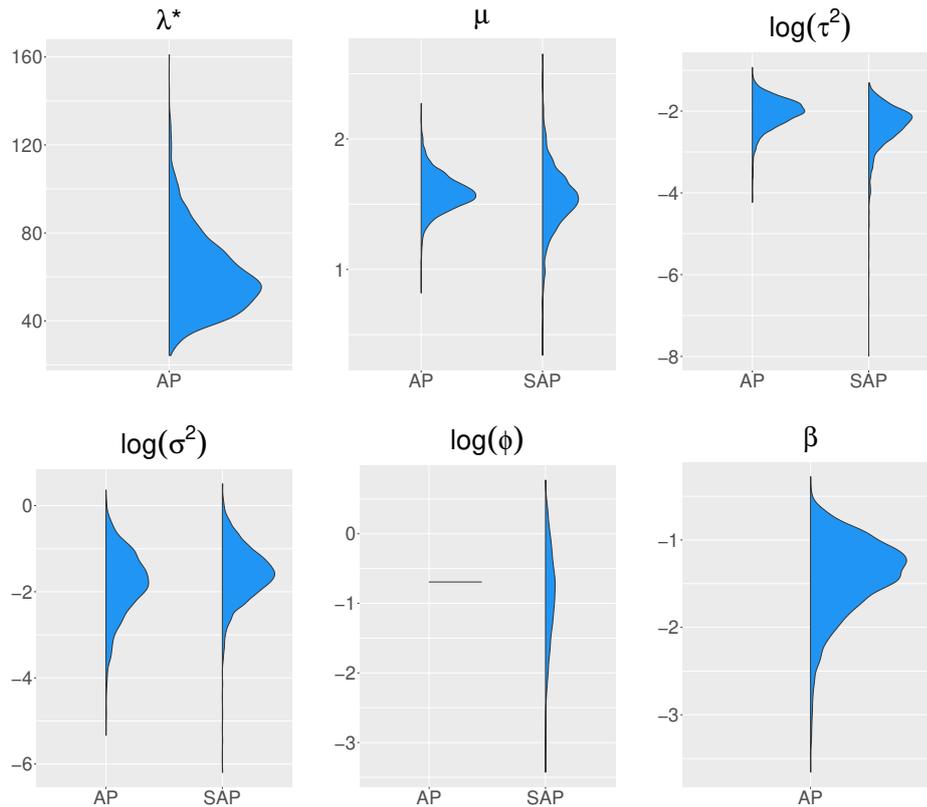


Figura 8.9: Distribuições a posteriori dos parâmetros dos modelos AP e SAP para os dados de Galícia de 1997.

1997. Ao se comparar com os mapas obtidos pela abordagem clássica, notamos que a transição dos menores valores (em locais amostrados) para os maiores valores (em locais não amostrados) se dá de forma mais suave no modelo Bayesiano.

A Tabela 8.4 apresenta as medidas resumo das distribuições a posteriori dos parâmetros para os modelos AP e SAP para os dados do ano 2000. É interessante notar que o modelo AP apresentou um valor superestimado de μ e também indicou presença de amostragem preferencial, apresentando um intervalo de credibilidade de 95% para β que não contém o valor zero. Uma justificativa provável seria que, no ano de 2000, a maior parte dos valores medidos são baixos comparados com a amplitude de valores da amostra. Esse resultado também indica que devemos ter cuidado ao utilizar um modelo com amostragem preferencial sob um plano amostral sem preferência, principalmente quando temos uma amostra em malha regular. As densidades a posteriori dos parâmetros podem ser observadas na Figura 8.11.

A Figura 8.12 apresenta os mapas de predição pela média a posteriori e o erro de predição a posteriori do processo Y para os dados do ano 2000. Observamos o efeito da

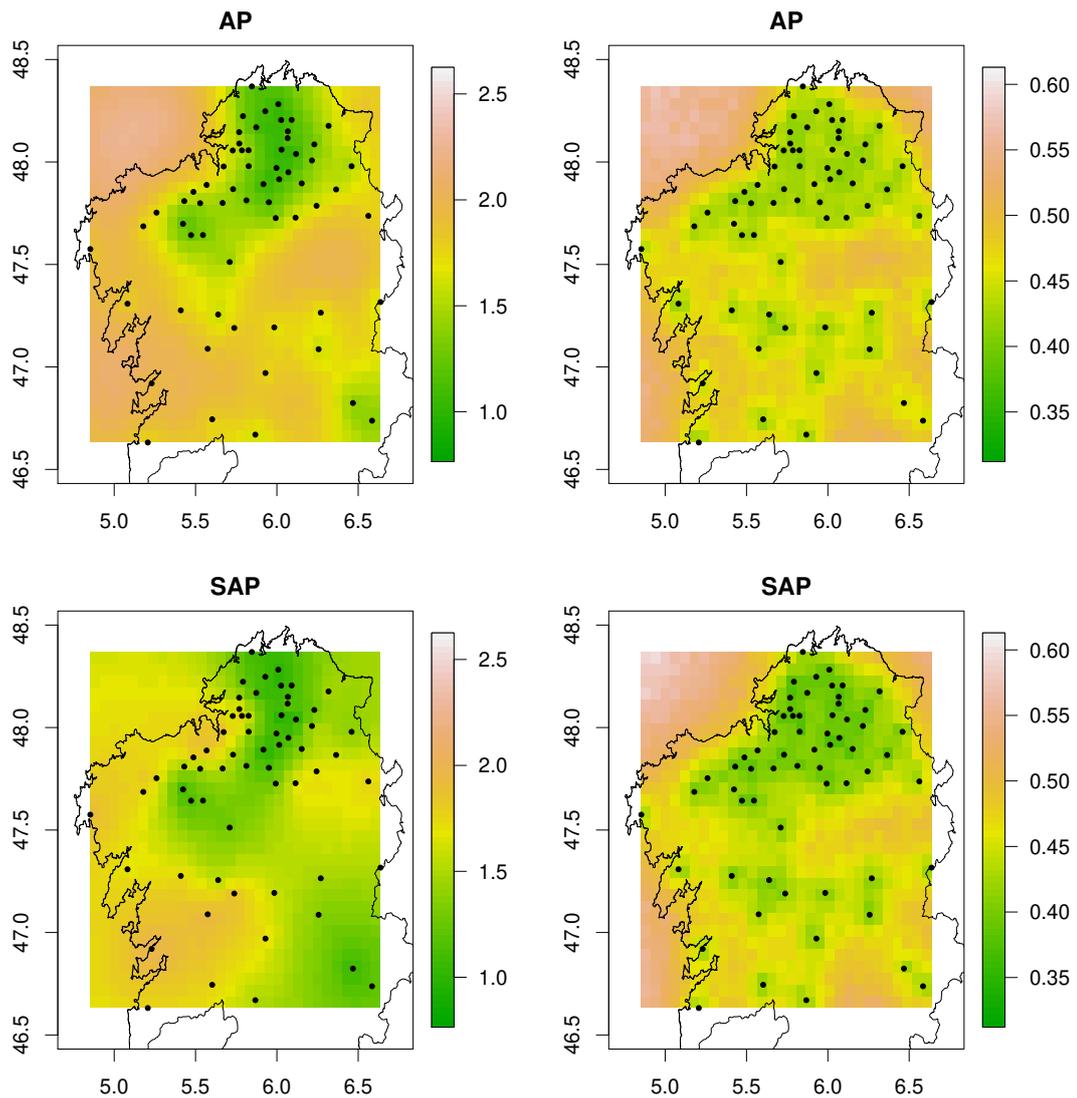


Figura 8.10: Mapa de predição de Y (primeira coluna) e erro de predição (segunda coluna) dos dados de log-concentração de chumbo em Galícia no ano de 1997.

Tabela 8.4: Medidas resumo a posteriori para os parâmetros dos modelos com amostragem preferencial (AP) e sem amostragem preferencial (SAP) para os dados de Galícia de 2000.

Modelo	Parâmetro	Média	Mediana	Desvio Padrão	IC(95%)
AP	λ^*	48,321	47,796	8,129	[34,049 ; 65,615]
	μ	1,042	1,054	0,195	[0,639 ; 1,402]
	τ^2	0,029	0,023	0,024	[0,001 ; 0,084]
	σ^2	0,320	0,324	0,099	[0,135 ; 0,504]
	β	-0,754	-0,697	0,369	[-1,638 ; -0,189]
SAP	μ	0,751	0,741	0,170	[0,442 ; 1,104]
	τ^2	0,018	0,009	0,019	[0,001 ; 0,068]
	σ^2	0,249	0,224	0,104	[0,130 ; 0,520]
	ϕ	0,371	0,314	0,204	[0,165 ; 0,906]

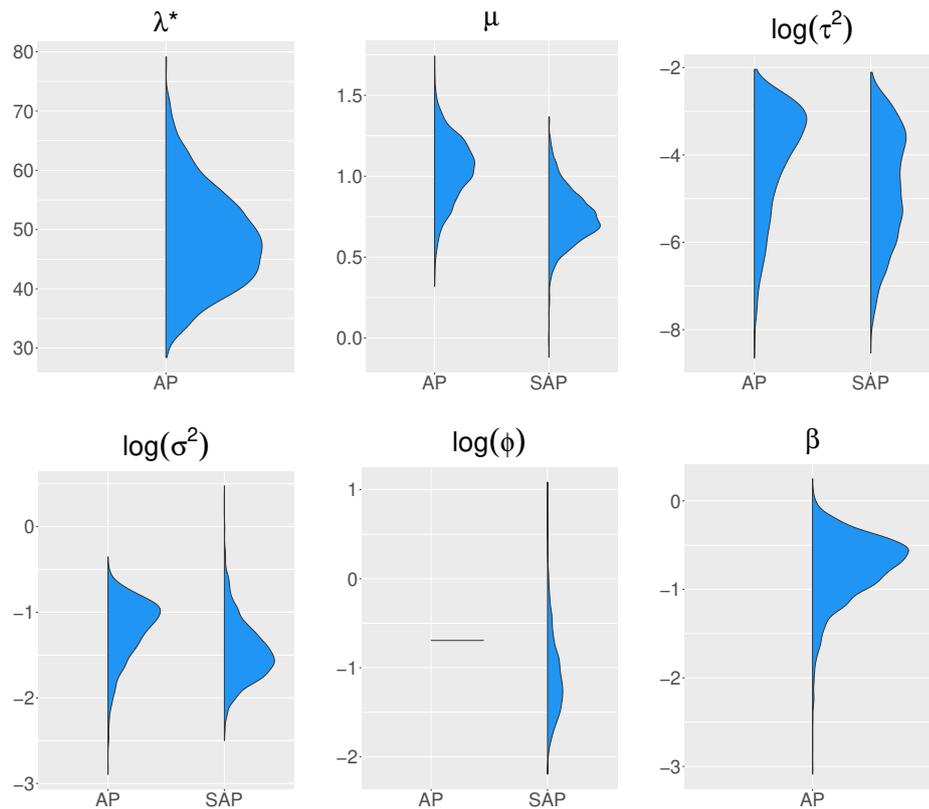


Figura 8.11: Distribuições a posteriori dos parâmetros dos modelos AP e SAP para os dados de Galícia de 2000.

amostragem preferencial pelo modelo AP, fornecendo valores preditos maiores nas regiões distantes da amostra em relação ao modelo SAP. Internamente à malha, as previsões são semelhantes para ambos modelos. O erro de predição para o modelo AP para locais não observados distantes de pontos amostrados é maior que o erro para o modelo SAP. Nesse caso, o melhor modelo seria o SAP, uma vez que sabemos que não há preferência na amostra. Nessa aplicação, o modelo AP não conseguiu captar a característica não preferencial da amostra, assim como nos modelos da abordagem clássica.

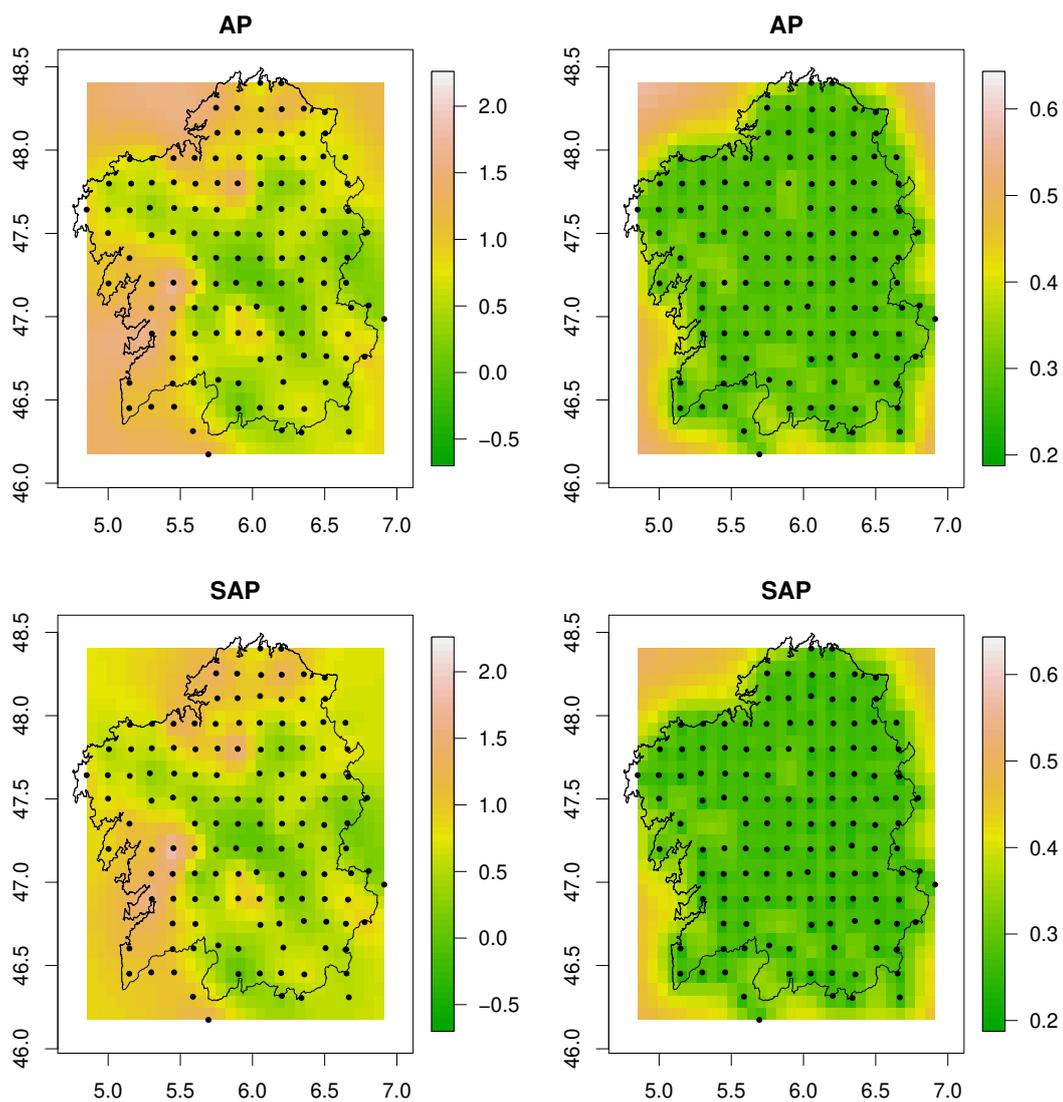


Figura 8.12: Mapa de predição de Y (primeira coluna) e erro de predição (segunda coluna) dos dados de log-concentração de chumbo em Galícia no ano de 2000.

8.2 Aplicação 2: Dados de radiação na Alemanha

O grupo de Monitoramento Radioativo Ambiental (*Radioactivity Environmental Monitoring - REM*) desenvolveu uma plataforma de dados para que os dados de monitoramento radioativo na Europa ficasse disponível para pesquisadores. Esse sistema é chamado de EURDEP (*EUropean Radiological Data Exchange Platform*) e a atualização dos dados se dá quase em tempo real. Dessa forma, se torna interessante o mapeamento automático a partir dos dados atualizados, pois se torna impraticável ter profissionais para construir mapas a cada atualização. Nesse sentido, o grupo REM organizou o exercício de comparação de interpolação espacial (*Spatial Interpolation Comparison - SIC*) de 2004, em que foi fornecido parte de um banco de dados para participantes realizarem a predição para locais não observados (cujos valores haviam sido observados no conjunto de dados completo). Os dados se referem a média diária de taxas de radiação gamma na Alemanha e o conjunto contém 1008 locais de monitoramento, sendo que apenas uma amostra de tamanho 200 foi fornecida. A partir dessa amostra, os participantes deveriam prever a taxa média de radiação gamma nos 808 locais restantes.

A Tabela 8.5 apresenta as estatísticas descritivas para os dados de taxa de radiação gamma. A variável foi medida em nanoSievert/hora - nSv/h e, por simplificação, denotaremos unidades de medida - u.m. Observamos um valor mínimo de 55,8 e máximo de 150,0 u.m e que os valores de média e mediana são semelhantes. O coeficiente de variação de 18,3% indica homogeneidade nos dados.

Tabela 8.5: Estatísticas descritivas para os dados de radiação gama na Alemanha.

n	Min	Quartil 1	Mediana	Média	Quartil 3	Máximo	D.p.	CV
200	55,20	82,20	97,55	96,23	109,50	153,00	17,65	0,183

A Figura 8.13 apresenta uma análise descritiva gráfica. Percebemos que a distribuição dos dados é simétrica e não apresenta um possível dado atípico. Pelo gráfico de postplot, não observamos nenhum ponto com diâmetro “muito” alto. As medidas de localizações foram fornecidas em metros e, para evitarmos possíveis erros computacionais, a unidade de medida foi redimensionada para $\text{km} \cdot 10^{-1}$.

A fim de verificar isotropia do processo, a Figura 8.14(a) apresenta o semivariograma direcional para as direções principais. Observamos comportamento semelhante do semivariograma para todas as quatro direções, indicando isotropia do dados. Podemos,

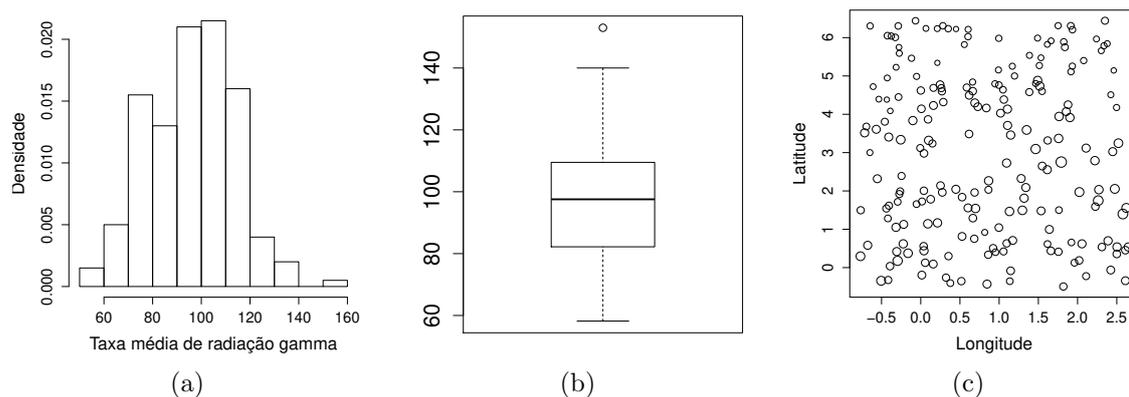


Figura 8.13: Gráficos de (a) histograma, (b) boxplot e (c) postplot dos dados de taxa média de radiação gamma na Alemanha.

então, trabalhar com o semivariograma omnidirecional, que considera apenas o módulo do vetor de distância sem considerar o sentido. A Figura 8.14(b) apresenta o gráfico de envelope simulado construído a partir de permutações aleatórias dos dados (foram consideradas 99). Para cada permutação, o valor do semivariograma empírico é calculado e a banda é construída a partir dos valores máximo e mínimo (considerando também os valores do semivariograma original). Notamos que vários pontos se encontram fora dos limites do envelope, indicando a existência de dependência espacial.

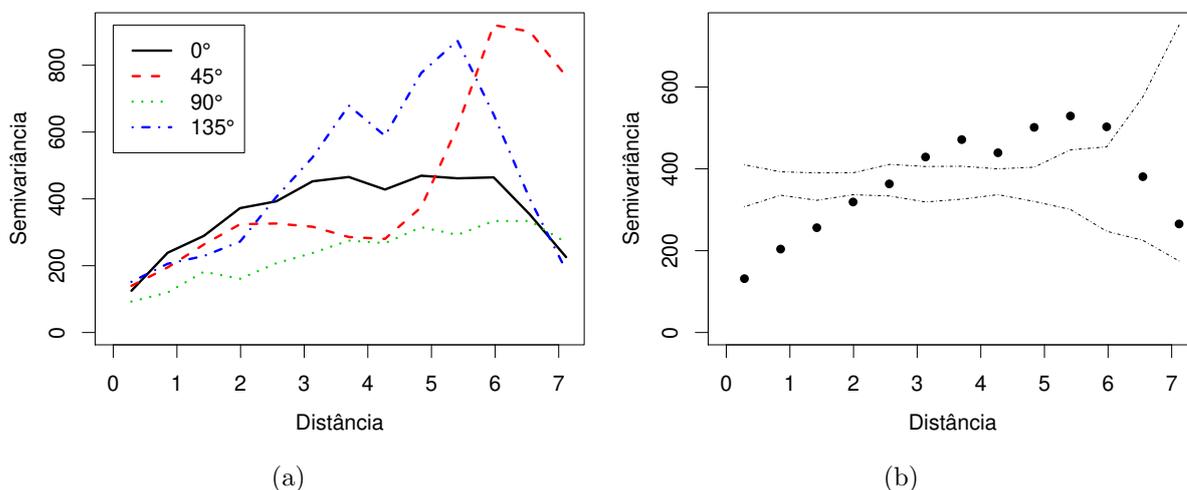


Figura 8.14: Gráficos de (a) semivariograma para as quatro direções principais e (b) envelope simulado do semivariograma omnidirecional dos dados de taxa média de radiação gamma na Alemanha.

A partir dessa análise, prosseguiu-se com a etapa de inferência.

Inferência Bayesiana

Foram ajustados os modelos geoestatístico Bayesiano com amostragem preferencial (AP) e sem amostragem preferencial (SAP) aos dados. O valor de ϕ foi fixado em 2 para ambos modelos e esse valor foi escolhido pela análise do variograma.

Consideramos duas cadeias de Markov com 50 mil iterações cada, descartando as 20 mil primeiras e, das iterações restantes, obtemos uma amostra de tamanho 2 mil (*lag* de 30). O tempo médio para obtenção das duas cadeias foi de 694,0 e 5,7 minutos para os modelos AP e SAP, respectivamente. A convergência das cadeias de Markov foi verificada por análise gráfica e os gráficos das trajetórias dos parâmetros podem ser observados no Apêndice B.

A Tabela 8.6 apresenta as medidas resumo das distribuições a posteriori dos parâmetros dos modelos AP e SAP e a Figura 8.15 apresenta as densidades das distribuições a posteriori dos parâmetros de ambos modelos. Observamos que as estimativas foram semelhantes para ambos modelos, sendo que o modelo AP apresentou estimativas de β em torno de zero, indicando que a amostragem é não preferencial (o intervalo de credibilidade de 95% contém o valor zero), o que era esperado para esses dados.

Tabela 8.6: Medidas resumo a posteriori para os parâmetros dos modelos com amostragem preferencial (AP) e sem amostragem preferencial (SAP) para os dados de radiação na Alemanha.

Modelo	Parâmetro	Média	Mediana	Desvio Padrão	IC(95%)
AP	λ^*	17.012	17.004	1.479	[14.026 ; 20.015]
	μ	94.670	94.571	7.795	[80.165 ; 110.662]
	τ^2	75.979	75.294	12.848	[53.320 ; 104.075]
	σ^2	261.953	256.650	69.170	[148.104 ; 418.897]
	β	-0.084	-0.082	0.111	[-0.315 ; 0.130]
SAP	μ	94.669	94.437	8.496	[78.233 ; 112.714]
	τ^2	76.194	75.96	12.378	[53.181 ; 101.849]
	σ^2	259.59	251.229	67.085	[155.017 ; 403.486]

Realizamos a predição nos 808 locais fornecidos pelo SIC e a Figura 8.16 apresenta o gráfico de valores preditos versus valores reais para ambos modelos. Podemos observar que a maioria dos valores preditos se encontram em torno da bissetriz, sendo que alguns pontos de maior valor real não foram recuperados. As predições realizadas são semelhantes para ambos modelos. Ao calcular o REQMP, obtemos os valores 352,54 e 353,18 para os modelos AP e SAP, respectivamente. Ou seja, apesar de os dados não apresenta-

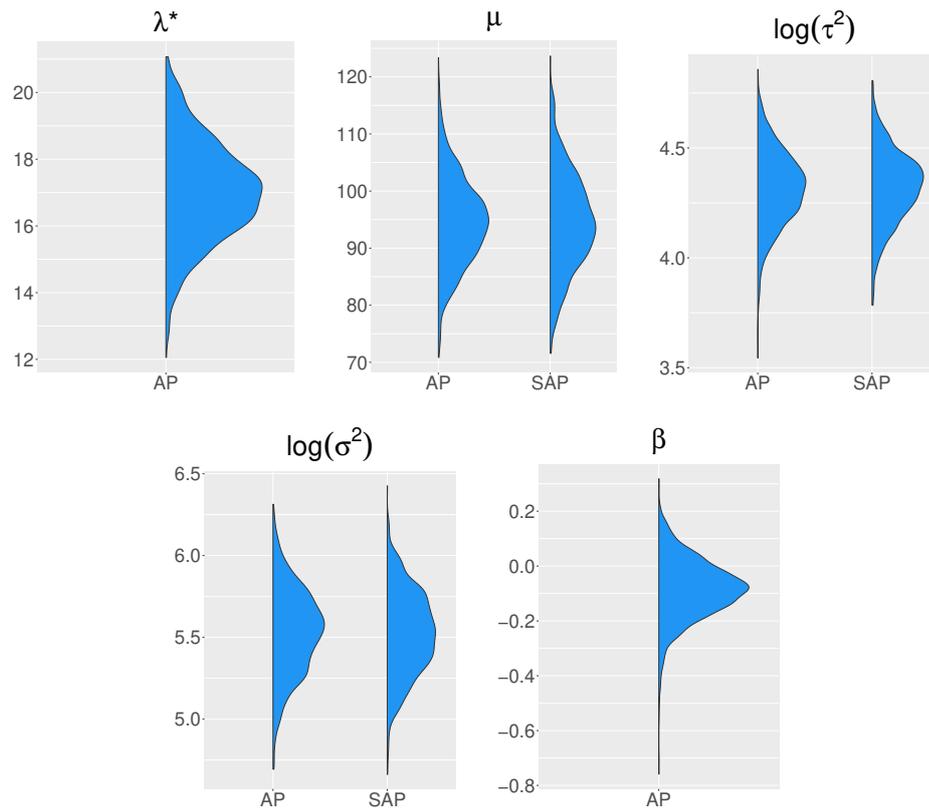


Figura 8.15: Histogramas das distribuições a posteriori dos parâmetros do modelo com amostragem preferencial para os dados de radiação na Alemanha.

rem amostragem preferencial, nosso modelo forneceu previsões levemente melhores que o modelo sem amostragem preferencial.

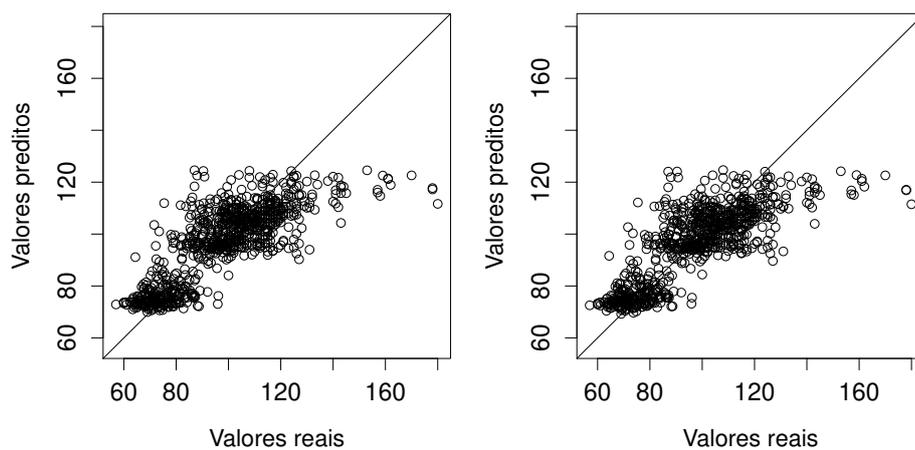


Figura 8.16: Gráfico de valores preditos versus valores reais do processo de interesse Y para os 808 locais fornecidos pelo SIC considerando os modelos (a) AP e (b) SAP.

A Figura 8.17 apresenta os mapas preditos do processo Y por ambos modelos. Podemos observar que o modelo AP fornece valores ligeiramente menores em regiões não

amostradas em comparação com o modelo AP e apresenta menor erro de predição, o que o torna mais atrativo mesmo no contexto não preferencial.

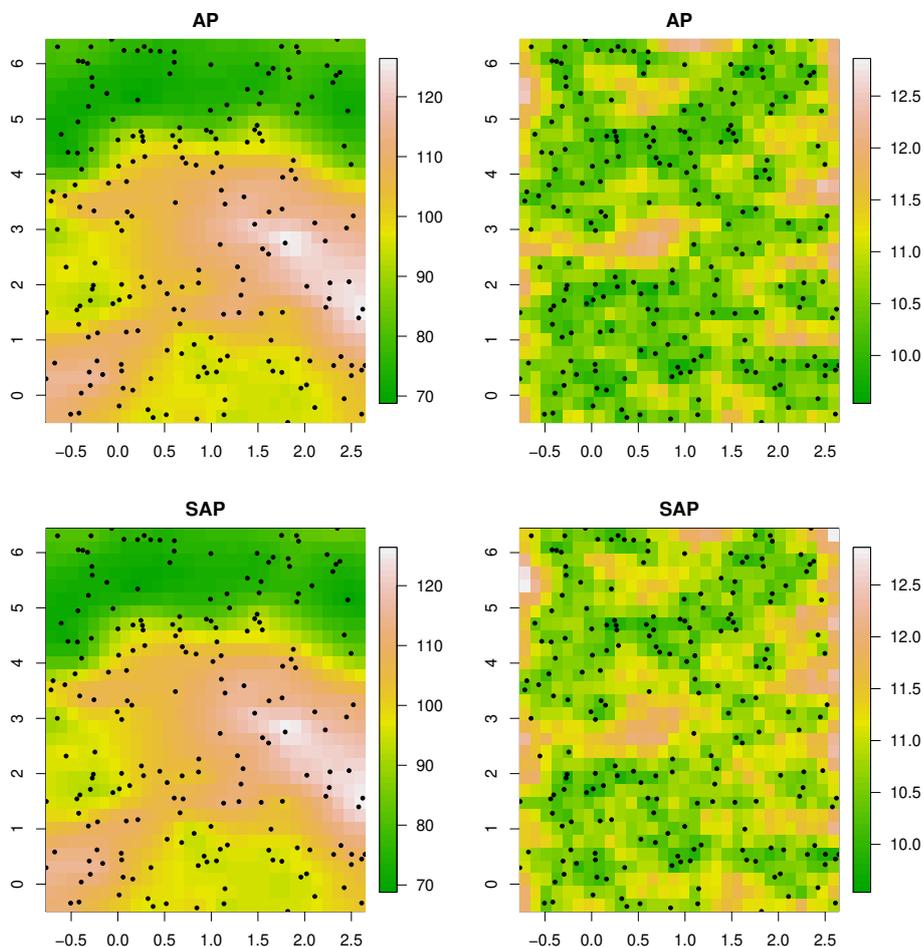


Figura 8.17: Mapas de predição do processo Y (primeira coluna) e erro de predição (segunda coluna) para os dados de radiação na Alemanha.

Em relação à estatística clássica, houve problema de identificação entre os termos de variância, em que o efeito pepita foi estimado em torno da variância total do processo. Foi realizada também a estimação do nosso modelo com aproximação por discretização e notamos o mesmo comportamento. Nessa aplicação, observa-se um cenário em que o efeito pepita não é próximo de zero, podendo ser esse um fator problemático para a estimação de modelos que consideram aproximação por discretização no contexto de amostragem preferencial. Logo, temos evidências da importância e relevância de um modelo exato para os dados com amostragem preferencial.

8.3 Aplicação 3: Dados de precipitação pluviométrica no Rio de Janeiro

Os dados se referem à precipitação total (em mm) observada no mês de Janeiro de 2015 na cidade do Rio de Janeiro. As observações foram obtidas por 33 estações de monitoramento distribuídas pela cidade do Rio de Janeiro e, através do Instituto Pereira Passos, os dados estão disponíveis no site do DATA.RIO (data.rio, acessado em 25 de julho de 2021). A Figura 8.18 apresenta o mapa da cidade do Rio de Janeiro e a localização das estações de monitoramento.

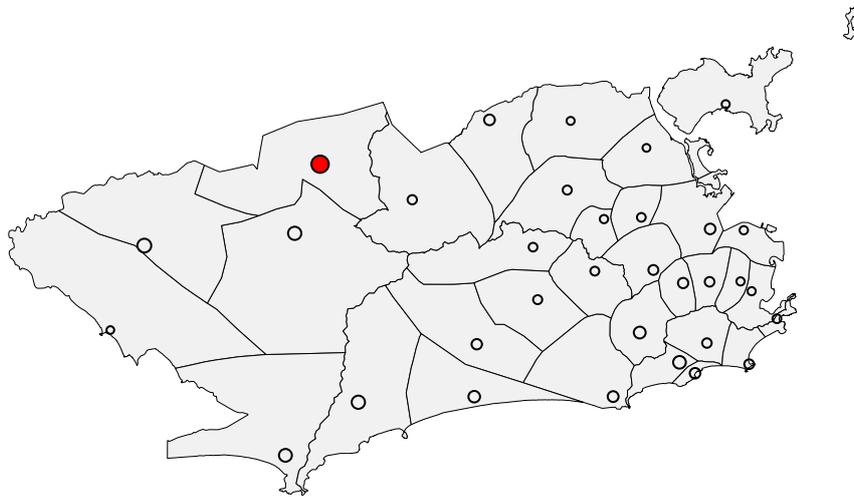


Figura 8.18: Mapa da cidade do Rio de Janeiro e estações de monitoramento alocadas. A estação marcada em vermelho representa uma observação atípica.

A Tabela 8.7 apresenta as medidas descritivas dos dados e a Figura 8.19, os gráficos de histograma e de boxplot. Podemos notar que há uma observação atípica que corresponde ao dado da estação 29 (“Av.Brasil/Mendanha”), com valor de 243,20 mm, destacada na Figura 8.18. Percebemos uma leve assimetria na distribuição dos dados e um coeficiente de variação que indica heterogeneidade.

Tabela 8.7: Estatísticas descritivas para os dados de precipitação na cidade do Rio de Janeiro em Janeiro de 2015.

n	Min	Quartil 1	Mediana	Média	Quartil 3	Máximo	D.p.	CV
33	36,80	59,00	78,80	90,32	104,40	243,20	46,508	0,515

A análise do variograma indicou comportamento semelhante nas quatro direções principais (Figura 8.20 (a)), evidenciando isotropia do processo. Ao verificar o envelope

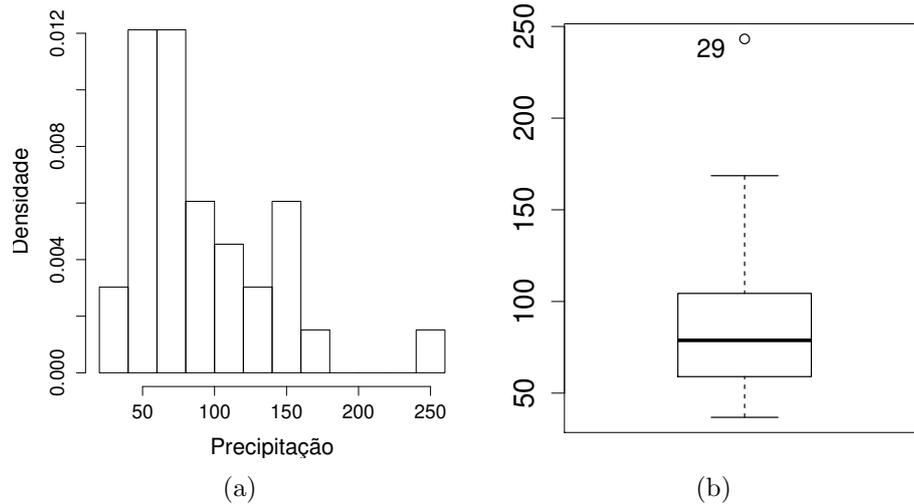


Figura 8.19: Gráficos de (a) histograma e (b) boxplot dos dados de precipitação na cidade do Rio de Janeiro.

simulado para o variograma (Figura 8.20 (b)), obtemos indícios de que há dependência espacial nos dados. Dessa forma, prosseguimos com o processo de inferência.

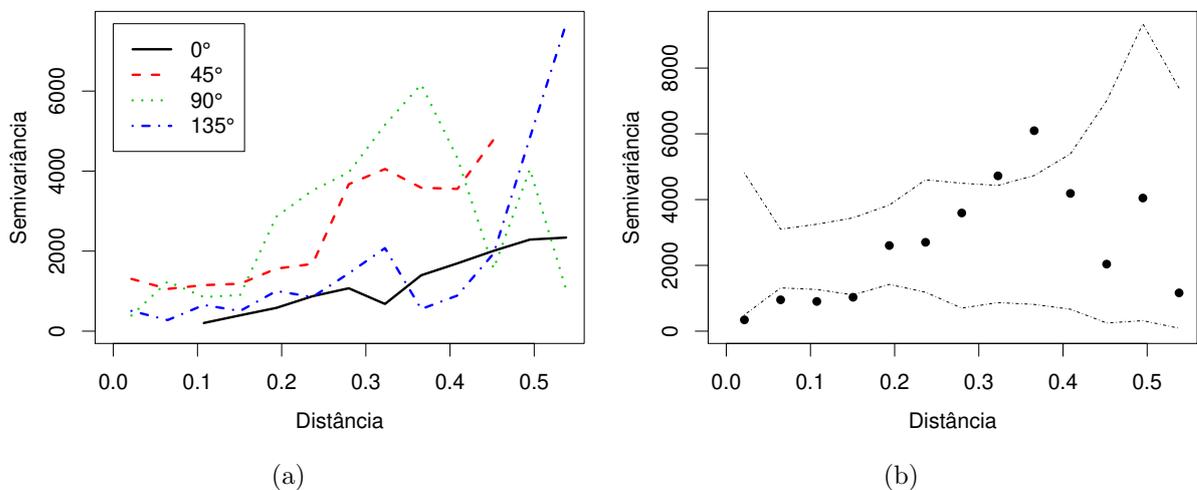


Figura 8.20: Gráficos de (a) semivariograma para as quatro direções principais e (b) envelope simulado do semivariograma omnidirecional dos dados de precipitação na cidade do Rio de Janeiro.

Inferência Bayesiana

Foram ajustados os modelos com amostragem preferencial normal (AP-n) e t-Student com 3 graus de liberdade (AP-t3) e sem amostragem preferencial normal (SAP-n) para os dados. Em relação ao modelo AP, assumimos como distribuição a priori para λ^*

a distribuição $\text{GamaT}(0,001; 0,001; 500/|B|)$ e para ϕ , a distribuição $\text{Gama}(2,10)$.

Consideramos duas cadeias de Markov com 100 mil iterações cada, descartando as 40 mil primeiras e, das iterações restantes, obtemos uma amostra de tamanho 3 mil (*lag* de 40). O tempo médio para obtenção das duas cadeias foi de 26,11, 19,26 e 0,74 minutos para os modelos AP-n, AP-t3 e SAP-n, respectivamente. A convergência das cadeias de Markov foi verificada por análise gráfica e os gráficos das trajetórias dos parâmetros podem ser observados no Apêndice B.

A Tabela 8.8 apresenta as medidas resumo das distribuições a posteriori e a Figura 8.21 apresenta as densidades a posteriori dos parâmetros dos modelos. Podemos observar que as estimações são semelhantes, sendo que o parâmetro de média é estimado ligeiramente maior nos modelos AP. Como obtemos um valor negativo para β , indicando preferência para observações com valores baixos, o modelo tenta corrigir o viés da amostragem preferencial e, conseqüentemente, obtemos estimativa de μ maior comparada ao modelo SAP. A diferença não é grande, pois o valor estimado do parâmetro de preferência é pequeno. Apesar disso, o intervalo de credibilidade de 95% para β não contém o valor zero, tanto para o modelo normal quanto para o modelo t-Student, indicando que há efeito da amostragem preferencial. A estimativa de σ^2 fornecida pelo modelo t-Student

Tabela 8.8: Medidas resumo a posteriori para os parâmetros dos modelos com amostragem preferencial normal (AP) e t-Student com 3 graus de liberdade (AP t3) e sem amostragem preferencial (SAP) para os dados de precipitação na cidade do Rio de Janeiro.

Modelo	Parâmetro	Média	Mediana	Desvio Padrão	IC(95%)
AP-n	λ^*	527,426	485,780	207,956	[262,822 ; 1071,672]
	μ	104,384	108,097	44,971	[2,712 ; 179,106]
	τ^2	247,897	138,637	299,168	[0,874 ; 1079,506]
	σ^2	5436,373	4634,836	3185,463	[1427,003 ; 13882,230]
	ϕ	0,280	0,252	0,141	[0,097 ; 0,627]
	β	-0,823	-0,784	0,424	[-1,738 ; -0,144]
SAP-n	μ	101,622	98,717	48,272	[12,990 ; 218,838]
	τ^2	172,334	85,349	226,493	[0,212 ; 781,080]
	σ^2	5418,782	4594,441	3136,348	[1589,856 ; 13474,450]
	ϕ	0,272	0,241	0,139	[0,089 ; 0,623]
AP-t3	λ^*	520,062	491,884	159,032	[291,047 ; 898,655]
	μ	106,380	106,353	25,497	[55,195 ; 156,755]
	τ^2	485,232	357,891	461,443	[2,662 ; 1624,719]
	σ^2	4389,112	3630,915	3176,343	[718,304 ; 12889,120]
	ϕ	0,275	0,245	0,145	[0,092 ; 0,638]
	β	-0,882	-0,841	0,402	[-1,751 ; -0,209]

apresentou menor valor em relação ao modelo normal, devido a consideração das caudas pesadas, uma vez que a variância é sensível a dados atípicos. Como consequência, o efeito pepita apresentou maior valor no modelo t-Student.

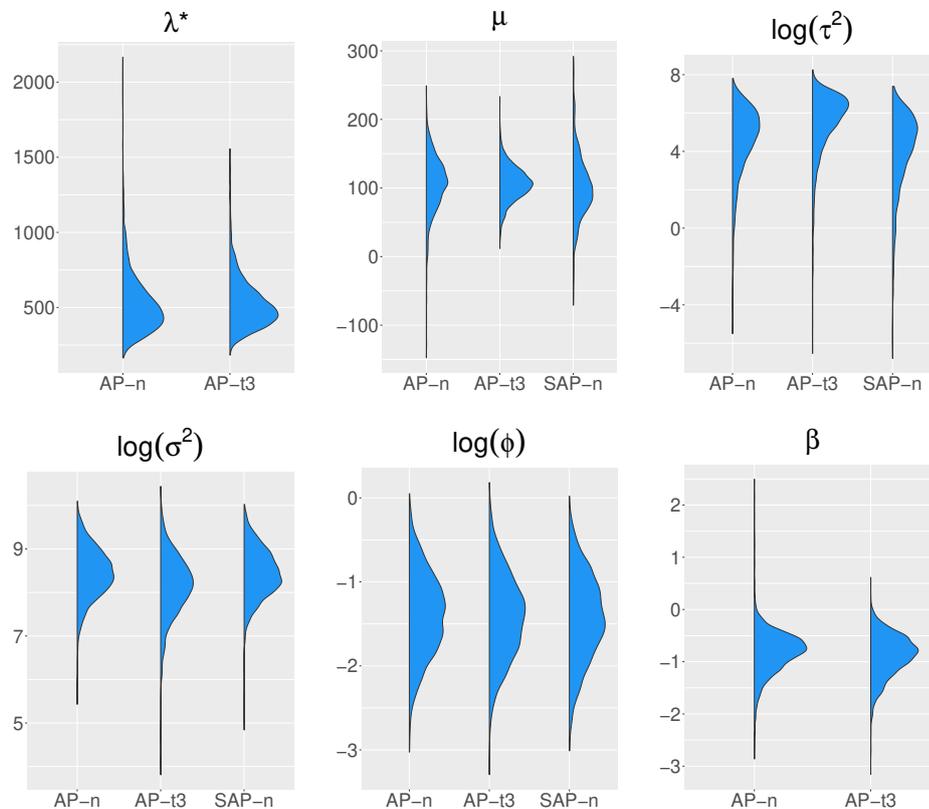


Figura 8.21: Histogramas das distribuições a posteriori dos parâmetros do modelo com amostragem preferencial para os dados de radiação na Alemanha.

A Figura 8.22 apresenta os mapas de predição pela média a posteriori e erro de predição do processo Y . O modelo AP fornece valores preditos ligeiramente maiores nas regiões distantes das estações dentro da cidade do Rio de Janeiro. Como identificamos a presença do efeito da amostragem preferencial, sua predição é mais precisa que a obtida pelo modelo SAP. Analisando o mapa obtido pelo modelo AP t-Student, notamos que os valores preditos em torno da estação 29 (marcada em branco) são ligeiramente menores em relação aos demais modelos, diminuindo o impacto que essa observação causa nos resultados. Em relação ao erro de predição, os modelos apresentam valores similares. Dessa forma, o modelo com caudas mais pesadas que a normal e que considera a amostragem preferencial foi mais adequado aos dados de precipitação.

Os modelos foram ajustados novamente aos dados, porém excluindo o dado da observação 29. Em seguida, foi realizada a predição de Y para essa estação e a distribuição

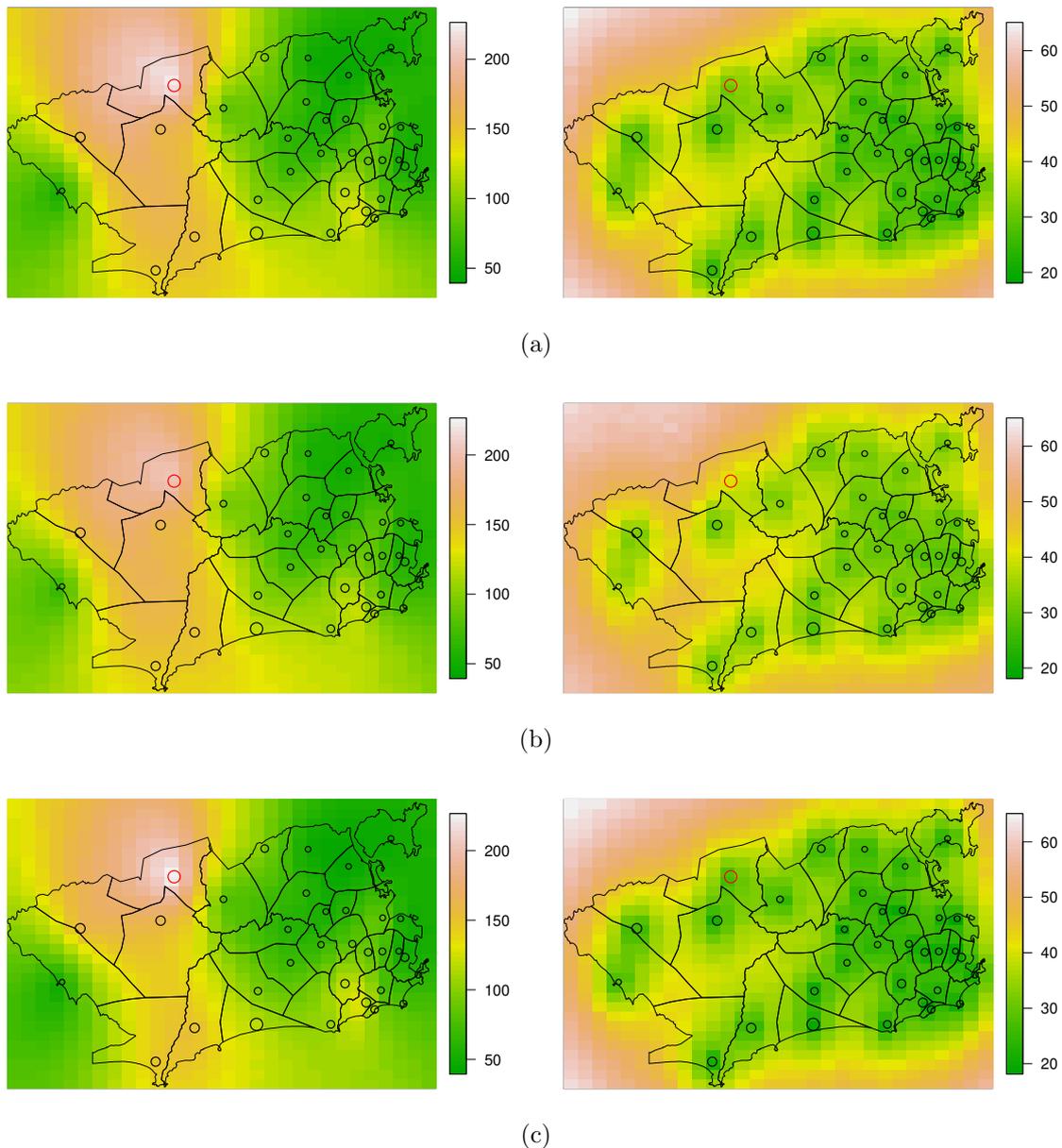


Figura 8.22: Mapa predito do processo Y pelos modelos (a) AP-n, (b) AP-t3 e (c) SAP-n para os dados de precipitação no Rio de Janeiro.

preditiva a posteriori pode ser observada na Figura 8.23. Notamos que, ao retirar a informação da estação 29 do conjunto de dados, os modelos não conseguem capturar bem seu verdadeiro valor. Esse fato era esperado, uma vez esse dado é o único *outlier* e sua exclusão acarreta na falta de informação de valores nessa região. O gráfico à direita da Figura 8.23 apresenta uma ampliação da distribuição em torno do valor verdadeiro de Y_{29} e podemos observar que o modelo que apresentou maior densidade preditiva a posteriori foi o modelo AP-t3 (0,000662), quase o dobro da densidade do modelo AP-n (0,000344) e quase quinze vezes maior que a densidade do modelo AP-n (0,000045). Dessa forma,

evidenciamos novamente a relevância da distribuição t-Student no ajuste aos dados dessa aplicação.

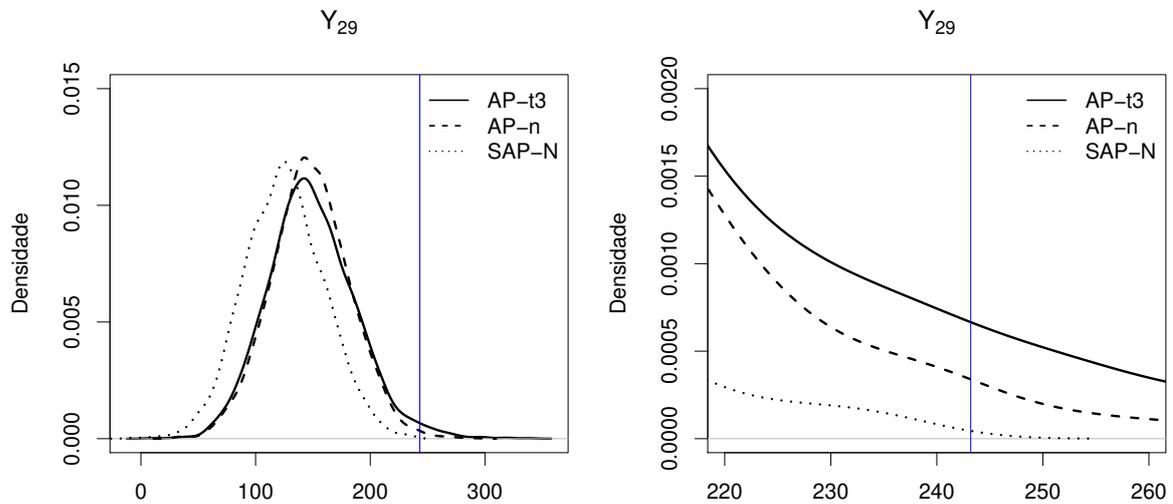


Figura 8.23: Densidades preditivas a posteriori de Y para a estação 29 a partir dos dados em a informação dessa estação. A figura à direita corresponde a uma ampliação em torno do valor verdadeiro de Y_{29} .

Em relação à estatística clássica, assim como na Aplicação 2, houve problema de identificação entre os termos de variância, em que τ^2 foi estimado em torno da variância total do processo. Foi realizada também a estimação do nosso modelo com aproximação por discretização e notamos o mesmo comportamento. Novamente, observamos o cenário em que o efeito pepita não é próximo de zero, podendo ser esse um fator problemático para a estimação de modelos que consideram aproximação por discretização. Logo, temos evidências da importância e relevância de um modelo exato para os dados com amostragem preferencial.

Capítulo 9

Discussão e trabalhos futuros

Quando tratamos dados espacialmente correlacionados, a geoestatística nos auxilia para realizar inferência de forma correta. Se o planejamento amostral não depende do fenômeno em estudo, então as técnicas usuais em geoestatística são suficientes para a análise do problema. Porém, quando os locais da amostra são escolhidos tendo como base características do fenômeno, observamos uma amostragem preferencial e outra abordagem deve ser considerada. Diggle et al. (2010) desenvolvem o modelo geoestatístico com amostragem preferencial e seu trabalho serve como base para esse ramo da geoestatística.

A modelagem se dá considerando um processo latente S responsável pela estrutura espacial dos dados, um processo pontual X que depende de S e representa a realização do plano amostral e o processo de medida Y , que são os valores medidos da variável de interesse nos locais de X . O processo observado é o processo Y e é nele que reside o principal interesse: realizar previsões em locais não observados, normalmente com o objetivo de construção de mapas sobre a região de estudo.

Visto que S é um processo não observado, é natural se pensar no algoritmo EM como uma ferramenta de estimação. Em relação a estatística clássica, na presente tese, desenvolvemos um algoritmo MCEM/SAEM para estimação dos parâmetros do modelo e comparamos com alguns dos já existentes: aproximação da verossimilhança por Monte Carlo (MCLA) e aproximação de Laplace, propostos para esse modelo por Diggle et al. (2010) e Dinsdale and Salibian-Barrera (2019), respectivamente. Para aproximação de Laplace, foi utilizado o pacote TMB (Kristensen et al., 2016), que recorre à diferenciação automática para criação da matriz Hessiana. Nosso algoritmo forneceu resultados semelhantes aos obtidos por aproximação de Laplace e melhor estimação da média e do parâmetro de preferência em relação ao MCLA no estudo de simulações.

A partir dos parâmetros estimados, partimos para a etapa de previsão. Baseados no trabalho de Ferreira and Gamerman (2015), desenvolvemos um algoritmo MH para geração de amostras da distribuição preditiva de S por blocos, tornando o algoritmo mais eficiente com uma redução de cerca de 80% no tempo computacional quando passamos da geração de blocos de tamanho um para blocos de tamanho cinco. Esse resultado

foi primordial para podermos realizar a estimação por MCEM ou SAEM, tornando esse processo factível na prática.

Para realizar predições, plugamos os parâmetros estimados como se fossem os verdadeiros e obtemos os valores preditos do processo S nos locais não observados de interesse. A partir desses valores, obtemos a predição para o processo Y . Dessa forma, podemos combinar métodos de estimação com métodos de predição para obtermos os melhores resultados. A krigagem é o método de predição mais comum utilizado no contexto não preferencial. Mostramos, na Seção 5.5, que a krigagem não forneceu valores preditos razoáveis, sofrendo efeito do viés da amostragem preferencial, isto é, para uma preferência positiva (em que os locais de maior valor tem maior chance de serem escolhidos), os valores preditos foram superestimados. Dessa forma, na presença de amostragem preferencial, a krigagem não é um método de predição adequado, corroborando com os resultados apresentados por Diggle et al. (2010) e Dinsdale and Salibian-Barrera (2019).

Como alternativas, trabalhamos com dois métodos que consideram o processo pontual no processo de estimação: a moda da distribuição preditiva de S e o algoritmo MH desenvolvido. Mostramos que, quando utilizamos os métodos de predição que consideram o processo pontual, obtemos uma correção do viés ocasionado pela amostragem preferencial e valores preditos mais próximos dos valores verdadeiros. Também mostramos que, apesar do método MCLA apresentar viés nas estimativas da média, se utilizamos seus valores estimados dos parâmetros conjuntamente com um processo de predição que considera o processo pontual, obtemos valores preditos corrigidos. Combinando a estimação por aproximação de Laplace ou MCEM com um desses métodos de predição, obtemos uma predição ainda mais precisa.

Tanto para estatística clássica quanto para Bayesiana, os métodos de estimação desenvolvidos recorrem ao uso de discretização da região de estudo para aproximar o processo S que é infinito-dimensional (vide como exemplos Diggle et al. (2010), Pati et al. (2011) e Gelfand et al. (2012)). Essa alternativa é necessária visto que a função de densidade do processo de Cox log-Gaussiano X possui uma integral intratável que depende de S em toda a região de estudo. Apesar da aproximação ser uma solução, esta pode carregar vieses difíceis de mensurar e produzir resultados viesados (Simpson et al., 2016). Dessa forma, motivados pelo trabalho de Gonçalves and Gamerman (2018), desenvolvemos o modelo com função de ligação proibito para a função de intensidade do processo pontual

X , o que tornou possível realizar inferência exata para o modelo geoestatístico com amostragem preferencial sob a abordagem Bayesiana. Dessa forma, evitamos os possíveis erros que a aproximação por discretização possa acarretar.

A inferência para o modelo Bayesiano se deu via amostragem por MCMC. Uma limitação encontrada foi o custo computacional na geração de processos Gaussianos, que são da ordem de n^3 . Dessa forma, controlar o parâmetro λ^* se torna necessário em aplicações cuja geração de pontos do processo descartado esteja alta, tornando o algoritmo ineficiente. Como alternativa, Gonçalves and Gamerman (2018) propõe utilizar uma distribuição truncada em um valor razoável como distribuição a priori que possa tornar a execução do algoritmo factível. Essa solução deve ser analisada para cada aplicação pois, como consequência, podemos não gerar dados do processo descartado suficientes para realizar uma boa inferência. Além disso, foi necessário controlar o parâmetro de preferência a priori para podermos identificar λ^* e β .

Assim como no trabalho de Gonçalves and Gamerman (2018), a particular escolha da função probito na função de intensidade possibilitou a construção de um algoritmo mais eficiente, gerando amostras do processo Gaussiano a partir de uma distribuição condicional completa Skew Normal e não por uma geração aproximada por MH, como é realizada em Ferreira and Gamerman (2015).

O modelo Bayesiano exato desenvolvido não possui intercepto na função de ligação e, como consequência, a função de intensidade quando o processo Gaussiano é nulo será, obrigatoriamente, metade de seu supremo. Dessa forma, se a função de intensidade original não passa pelo ponto $(0, \lambda^*/2)$, nosso modelo não será capaz de realizar boa estimação da taxa em torno de $S = 0$. Dessa forma, um modelo com intercepto na função de intensidade poderia auxiliar no ajuste da taxa estimada. Realizamos uma análise com este modelo na Seção 7.5 e apresentamos alguns resultados preliminares. Observamos dificuldade de identificar o parâmetro α (intercepto) sob priori pouco informativa, sendo necessária uma priori informativa para obtermos uma estimação satisfatória desse parâmetro. A partir da estimação correta do intercepto, fomos capazes de ajustar melhor a função de intensidade estimada em relação à real. Porém, quando comparamos a predição para o processo Y do modelo com boa e ruim estimação do intercepto, obtemos resultados semelhantes para ambos os casos. Esse resultado traz evidência de que, mesmo com uma aproximação não tão boa da função de intensidade, podemos obter boas predições do mo-

delo. Verificamos esse “efeito” também quando consideramos como função de intensidade a função exponencial, em que não obtemos um bom ajuste da função de intensidade mas obtemos boa predição para o processo Y . Como são resultados preliminares, um estudo mais aprofundado sobre essa modelagem deve ser realizada.

A partir de estudos de simulação, mostramos que o modelo Bayesiano exato para o problema de amostragem preferencial se mostrou eficaz tanto na estimação dos parâmetros do modelo quanto na predição do processo de interesse Y . A eficiência do modelo foi verificada no estudo de simulação em situações em que a amostragem preferencial está presente e também no caso não preferencial, sendo um modelo que produziu ótimos resultados em ambas situações. Também mostramos que o modelo se sobressai em relação ao modelo tradicional (não preferencial) quando há amostragem preferencial. Sobre a má especificação da função de intensidade, isto é, quando os dados provém de uma função de intensidade diferente da probito, analisamos o comportamento do modelo quando o processo de Cox log-Gaussiano é o verdadeiro e verificamos que, apesar de não conseguirmos estimar bem a função de intensidade na maior parte dos valores do processo S , obtemos ótimos resultados sobre a predição do processo Y , que é o principal interesse. Quando comparamos nosso modelo exato com o modelo com aproximação, ambos com função de intensidade probito, observamos que o modelo exato fornece valores preditos mais precisos que o modelo com aproximação, o que o torna mais atrativo.

Quando consideramos dados com observações atípicas, necessitamos de outra distribuição além da normal, seja para o processo latente ou para o processo de medida (ou ambos). Nesta tese, desenvolvemos o modelo geoestatístico t-Student com amostragem preferencial, sob as abordagens clássica e Bayesiana. Em ambas, consideramos que o processo latente S segue uma distribuição t-Student. No caso clássico, através de estudos de simulação, obtivemos boa estimação pelos métodos MCEM e SAEM. Em relação à aproximação de Laplace, a alocação de memória pelo pacote TMB foi alta, chegando a utilizar 26 Gb de memória RAM quando consideramos uma malha de estimação 20×20 , sendo a principal limitação desse método. Além disso, obtivemos estimativas ruins por aproximação de Laplace, sendo que nosso método MCEM/SAEM se sobressaiu para esse modelo. Em relação à predição, os valores preditos foram semelhantes quando considerados os parâmetros estimados de ambos métodos de estimação. Porém, a predição realizada à partir dos valores dos parâmetros fornecidos pela aproximação de Laplace

apresentou alta variância em comparação com a predição realizada à partir dos valores dos parâmetros fornecidos pelo MCEM, sendo a última mais precisa. Dessa forma, nossa metodologia de estimação se mostrou melhor tanto na estimação quanto para fornecer predição à partir de suas estimativas. Sob a abordagem Bayesiana, obtivemos apenas resultados preliminares, sendo que uma melhor análise do modelo se faz necessária.

Sobre a análise com dados reais, em relação aos dados de biomonitoramento de metais pesados em Galícia, observamos que os modelos clássico e Bayesiano foram capazes de identificar a amostragem preferencial da pesquisa de 1997, fornecendo resultados semelhantes. Para o ano 2000, ambos modelos acusaram uma preferência negativa na ausência de amostragem preferencial. Esse efeito pode ter sido causado pelo fato de que a amostra do ano 2000 possui poucos valores “altos” em relação à amostra como um todo. Em relação aos dados de radiação da Alemanha, o modelo clássico e o modelo Bayesiano, ambos com aproximação por discretização, não foram capazes de identificar corretamente a variância do processo latente, aumentando a estimativa no efeito pepita. Já o modelo Bayesiano exato foi capaz de identificar corretamente os valores das duas variâncias separadamente. O uso do termo “corretamente” aqui se faz no sentido de as estimativas serem coerentes com a análise do variograma. Problemas computacionais foram encontrados na estimação dos parâmetros por MCLA e por aproximação de Laplace, como valores não finitos para a função de verossimilhança, não sendo possível obter as estimativas por esses métodos. Para a aplicação dos dados de precipitação na cidade do Rio de Janeiro, novamente a estimação do modelo clássico por MCEM não foi capaz de identificar corretamente os termos de variância. O modelo Bayesiano exato estimou coerentemente esses parâmetros, de acordo com a análise do variograma. Em relação à essa aplicação, observamos um dado atípico, que pode viesar as estimativas do modelo normal. Dessa forma, ajustando um modelo t-Student, mostramos que o impacto dessa observação é reduzido em seus arredores, melhorando a predição pelo modelo. Em relação aos modelos com aproximação por discretização, pretendemos realizar uma análise mais aprofundada para esses casos, a fim de que esses modelos sejam capazes de identificar corretamente o efeito pepita e a contribuição.

Como trabalhos futuros, planejamos implementar formas de melhorar a simulação de processos Gaussianos nos algoritmos propostos. Sabemos que esse é um dos principais problemas computacionais e sofremos dele sob ambas abordagens, clássica e Bayesiana.

Dinsdale and Salibian-Barrera (2019) utilizam campo Markoviano Gaussiano para aproximar o processo Gaussiano S através da abordagem SPDE (*Stochastic Partial Differential Equation*). Uma limitação dessa alternativa é a restrição do valor de κ da função de covariância Matérn para valores inteiros, excluindo, por exemplo, a função de covariância exponencial.

Uma extensão do nosso modelo seria considerar a distribuição da família de contornos elípticos no modelo, em que a distribuição t-Student é um caso particular. De Bastiani et al. (2015) desenvolvem o modelo geoestatístico com essa família de distribuições para o caso não preferencial, cuja distribuição marginal de Y possui forma fechada. No contexto de amostragem preferencial, a distribuição marginal de Y não possui forma analítica e estender o modelo para considerar a distribuição de família de contornos elípticos se torna um desafio, principalmente sob a abordagem clássica.

Almejamos futuramente a possibilidade de estimar os graus de liberdade da distribuição t-Student, que foram considerados fixos nesta tese. Uma forma de realizar essa estimação é considerar modelos com diferentes graus de liberdade (uma sequência de valores, por exemplo) e escolher o melhor modelo segundo algum critério. Para essa alternativa precisamos estimar os parâmetros de vários modelos, o que pode ser muito custoso, principalmente sob a abordagem Bayesiana. Logo, estimar o parâmetro ν seria mais eficiente. Fonseca et al. (2008) consideram como distribuições a priori para os graus de liberdade as prioris obtidas pela regra de Jeffreys e Jeffreys independente, mostrando para o modelo de regressão t-Student que as distribuições a posteriori obtidas são próprias. Essa pode ser uma alternativa para aplicarmos no nosso modelo para poder estimarmos os graus de liberdade.

Mostramos que o ajuste do modelo t-Student forneceu melhores resultados na presença de *outlier*, mas ainda não temos uma ferramenta para medir o impacto dessa observação nos resultados do modelo. Assim, pretendemos desenvolver a análise de diagnóstico para os modelos propostos.

Planejamos realizar uma análise mais aprofundada sobre o modelo com intercepto na função de intensidade. Os resultados presente na tese foram obtidos à partir de uma análise preliminar e alguns fatores devem ser estudados, como a identificabilidade dos parâmetros α e λ^* . Também podemos considerar outras modelagens, como acrescentar um parâmetro aditivo na função de intensidade, denotemos por α^* . Dessa forma, a taxa

do processo de Poisson está no intervalo α^* e $(\alpha^* + \lambda^*)$. Nesse contexto, o parâmetro λ^* perde a interpretação de ser o supremo da função de intensidade, porém tornamos o modelo mais flexível. Pretendemos também comparar a eficiência do nosso modelo com função de ligação probito e modelos com outras funções de limitadas, como a logística e a complemento log-log.

Uma extensão natural do nosso modelo seria considerar o tempo como variável. As Aplicações 2 e 3 possuem, em seu banco de dados, medidas repetidas, mensuradas em diferentes meses. Essa informação poderia ser incorporada para melhorar os resultados de predição, considerando as medidas de tempos anteriores. Dessa forma, planejamos desenvolver o modelo espaço-temporal com amostragem preferencial.

Seria interessante realizar a aplicação do modelo em outros conjuntos de dados, a fim de maior validação da importância desse trabalho. Por fim, pretendemos desenvolver um pacote na linguagem R integrada com a linguagem Cpp a fim de possibilitar qualquer pesquisador a utilizar a metodologia proposta nesta tese, tanto sob enfoque clássico quanto Bayesiano.

Referências Bibliográficas

- Beskos, A., Papaspiliopoulos, O., Roberts, G. O., and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based inference for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society*, 68(3):333–382.
- Biondi, F., Myers, D., and Avery, C. (1994). Geostatistically modelling stem size and increment in an old-growth forest. *Canadian Journal of Forest Research*, 24:1354–1368.
- Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York.
- De Bastiani, F., Cysneiros, A. M. A., Uribe-Opazo, M., and Galea, M. (2015). Influence diagnostics in elliptical spatial linear models. *Test*, 24(2):322–340.
- Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics*, 27(1):94 – 128.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B.*, 39(1):1–38.
- Diggle, P. and Ribeiro, P. (2007). *Model-based Geostatistics*. Springer Series in Statistics. Springer.
- Diggle, P. J., Menezes, R., and Su, T.-I. (2010). Geostatistical inference under preferential sampling (with discussion). *Journal of the Royal Statistical Society: Series C*, 59(2):191–232.
- Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C.*, 47(3):299–350.
- Dinsdale, D. and Salibian-Barrera, M. (2019). Methods for preferential sampling in geostatistics. *Journal of the Royal Statistical Society: Series C*, 68(1):181–198.
- Eddelbuettel, D. and François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18.
- Fernández, J., Rey, A., and Carballeira, A. (2000). An extended study of heavy metal deposition in Galicia (NW Spain) based on moss analysis. *The Science of the Total Environment*, 254(1):31–44.

- Ferreira, G. and Gamerman, D. (2015). Optimal design in geostatistics under preferential sampling (with discussion). *Bayesian Analysis*, 10(3):711 – 735.
- Fonseca, T. C. O., Ferreira, M. A. R., and Migon, H. S. (2008). Objective Bayesian analysis for the Student-t regression model. *Biometrika*, 95(2):325–333.
- Galarza, C. E., Lachos, V. H., and Bandyopadhyay, D. (2017). Quantile regression in linear mixed models: A stochastic approximation EM approach. *Statistics and its Interface*, 10(3):471–482.
- Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall. Taylor & Francis.
- Gelfand, A. E., Sahu, S. K., and Holland, D. M. (2012). On the effect of preferential sampling in spatial prediction. *Environmetrics*, 23(7):565–578.
- Giorgi, E. and Diggle, P. (2017). PrevMap: An R package for prevalence mapping. *Journal of Statistical Software*, 78.
- Gonçalves, F. B. and Gamerman, D. (2018). Exact Bayesian inference in spatiotemporal Cox processes driven by multivariate Gaussian processes. *Journal of the Royal Statistical Society: Series B.*, 80(1):157–175.
- Griewank, A. and Walther, A. (2008). *Evaluating derivatives: principles and techniques of Algorithmic Differentiation*. Society for Industrial and Applied Mathematics, Philadelphia, second edition.
- Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software*, 70(5):1–21.
- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. Series B. *Journal of the Royal Statistical Society*, 44(2):226–233.
- López-Granados, F., Jurado-Expósito, M., Peña-Barragán, J., and García-Torres, L. (2005). Using geostatistical and remote sensing approaches for mapping soil properties. *European Journal of Agronomy*, 23(3):279–289.
- Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58(8):1246–1266.

- Møller, J. and Waagepetersen, R. P. (2003). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall. Taylor & Francis.
- Pati, D., Reich, B. J., and Dunson, D. B. (2011). Bayesian geostatistical modelling with informative sampling locations. *Biometrika*, 98(1):35–48.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall. Taylor & Francis.
- Shaddick, G. and Zidek, J. V. (2014). A case study in preferential sampling: Long term monitoring of air pollution in the UK. *Spatial Statistics*, 9:51–65. Revealing Intricacies in Spatial and Spatio-Temporal Data: Papers from the Spatial Statistics 2013 Conference.
- Simpson, D., Illian, J. B., Lindgren, F., Sørbye, S. H., and Rue, H. (2016). Going off grid: computationally efficient inference for log-Gaussian Cox processes. *Biometrika*, 103(1):49–70.
- Tierney, L., Kass, R. E., and Kadane, J. B. (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association*, 84(407):710–716.
- Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46:234–240.
- Wei, G. C. G. and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man’s data augmentation algorithms. *Journal of the American Statistical Association*, 85(411):699–704.
- Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261.
- Zhang, H. and Wang, Y. (2010). Kriging and cross-validation for massive spatial data. *Environmetrics*, 21(3-4):290–304.

APÊNDICE A — Especificação incorreta da função de intensidade: cenário 2

Nesta seção estudamos o comportamento das estimativas dos parâmetros e valores preditos do nosso modelo quando os dados são gerados a partir da função de intensidade exponencial, isto é,

$$\lambda(x) = \exp\{\gamma + \delta S(x)\}.$$

para o caso em que $(\gamma, \delta) = (1; 2)$, o que chamamos de cenário 2. Consideramos os seguintes valores para os parâmetros: $(\mu, \tau^2, \sigma^2, \phi) = (4; 0,10; 3; 0,15)$. A amostra simulada resultou em 61 observações.

O cenário 2 apresenta maior força de preferência na amostragem que o cenário 1 ($\delta = 1$). A estimação dos parâmetros foi realizada utilizando o modelo probito com distribuição a priori truncada em 150 para λ^* e também o modelo sem amostragem preferencial (SAP). As distribuições a posteriori dos parâmetros do modelo podem ser observadas na Figura A.1 e, assim como no caso anterior, observamos estimação satisfatória para os parâmetros de covariância. Nesse cenário, o modelo SAP subestima σ^2 e ϕ . O parâmetro μ é superestimado devido a amostra ser composta, em sua maioria, por valores altos, não possuindo informação dos menos valores. Note que o viés é maior para o modelo SAP, como esperado. Novamente, devemos analisar se esse viés do parâmetro (e dos valores de S) impactam negativamente na predição do processo Y .

A Figura A.2 apresenta a função de intensidade estimada (em função de S) pelo modelo e observamos grande diferença em comparação com a curva real. Novamente, esse viés é ocasionado pela falta de intercepto na função de intensidade do nosso modelo. Observe que a amostra é constituída, em sua maioria, de valores de S entre 2 e 4, apresentando um viés forte da amostragem preferencial nesse cenário.

Para incorporar a incerteza da estimação de S , foram construídos os mapas da função de intensidade estimada a partir da média e da mediana a posteriori, como pode ser observado na Figura A.3. Os valores se encontram na escala logarítmica. Percebemos que os valores mais altos da taxa não foram recuperados, como era de se esperar, mas os menores valores foram superestimados pela média, corroborando com o resultado da

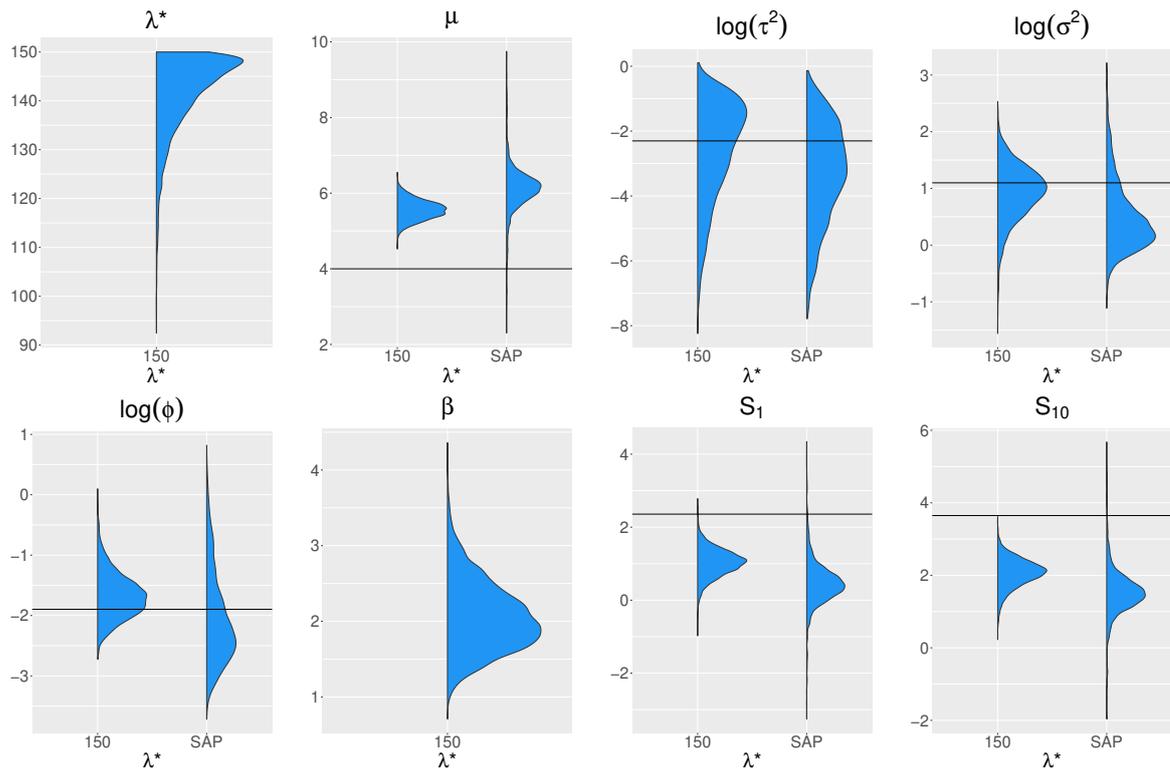


Figura A.1: Densidades das distribuições a posteriori dos parâmetros do modelo com função de intensidade probito, considerando o cenário 2 de dados simulados com função de intensidade exponencial. A linha horizontal representa o valor verdadeiro do parâmetro.

Figura A.2. Nesse caso, um melhor estimador seria a mediana a posteriori.

Foi calculada a proporção de cobertura de intervalo de credibilidade de 95% da taxa estimada, resultando em 87,7%. Esse resultado indica que temos grande proporção de cobertura dos ICs para a função de intensidade, apesar da estimação pobre pela média e mediana a posteriori, como podemos observar na Figura A.4. Nessa Figura, escolhemos quatro pontos em que a função de intensidade apresenta valores baixos (primeira linha de gráficos) e quatro pontos de valores altos (segunda linha de gráficos). Em relação aos pontos de baixa intensidade, notamos boa estimativa para uns e valores superestimados para outros. Em relação aos pontos de alta intensidade, não conseguimos recuperar o valor verdadeiro, como esperado quando consideramos como taxa real a função exponencial.

Para avaliar a predição para o processo de interesse Y , foram construídos mapas estimados pela média a posteriori dos valores preditos, como podemos observar na Figura A.5. Notamos que os menores valores do processo não são recuperados pelo nosso modelo, porém são mais próximos que a predição obtida pelo modelo SAP. Note que, dos locais que compõe a amostra, não há valores baixos do processo, o que induz uma falta de

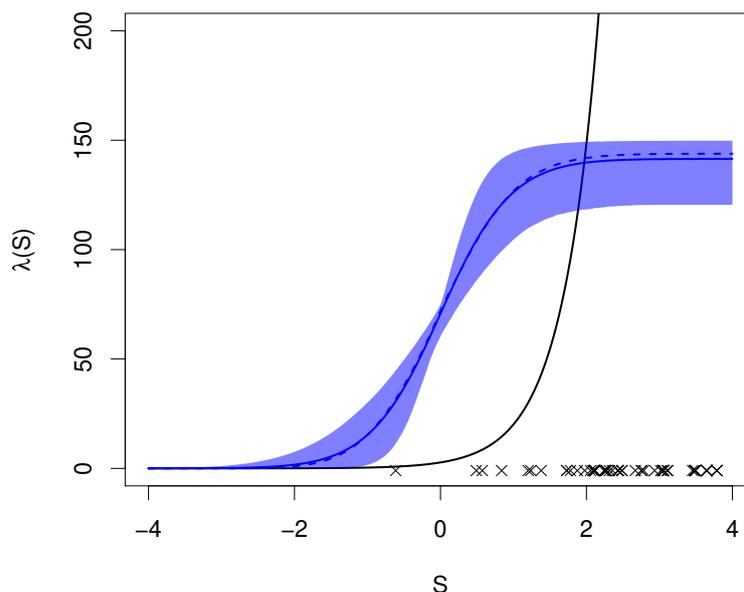


Figura A.2: Taxa estimada pela média (linha contínua), mediana (linha tracejada) e intervalo de credibilidade de 95% considerando os dados gerados pelo cenário 2. A linha preta representa a taxa exponencial (verdadeira). Para melhor visualização, o eixo vertical foi limitado em 200. Os pontos marcados em “x” correspondem aos valores simulados do processo S na amostra.

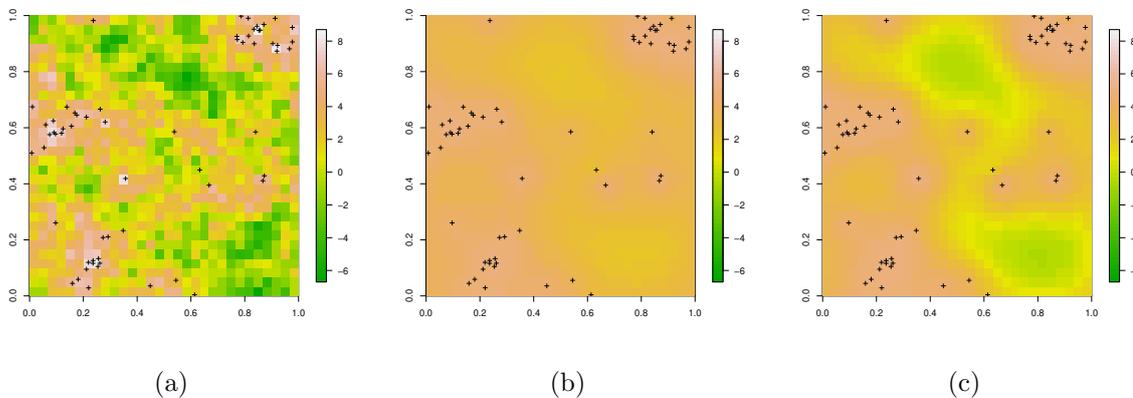


Figura A.3: Função de intensidade (a) real e estimada pela (b) média e (c) mediana a posteriori para os dados gerados pelo cenário 2. Os valores estão na escala logarítmica.

informação e resulta em uma predição viesada. Dessa forma, podemos dizer que nosso modelo forneceu boas predições visto a amostra fornecida.

A Figura A.6 apresenta as distribuições preditivas a posteriori de alguns locais não observados para Y , em regiões sem observações. Notamos que as distribuições a posteriori dos valores preditos de Y se encontram mais próximas de seus valores verdadeiros no modelo AP do que no modelo SAP. Apesar de apresentar viés nas predições pelo nosso modelo, esse erro é resultante da forte preferência na amostragem que produz falta de

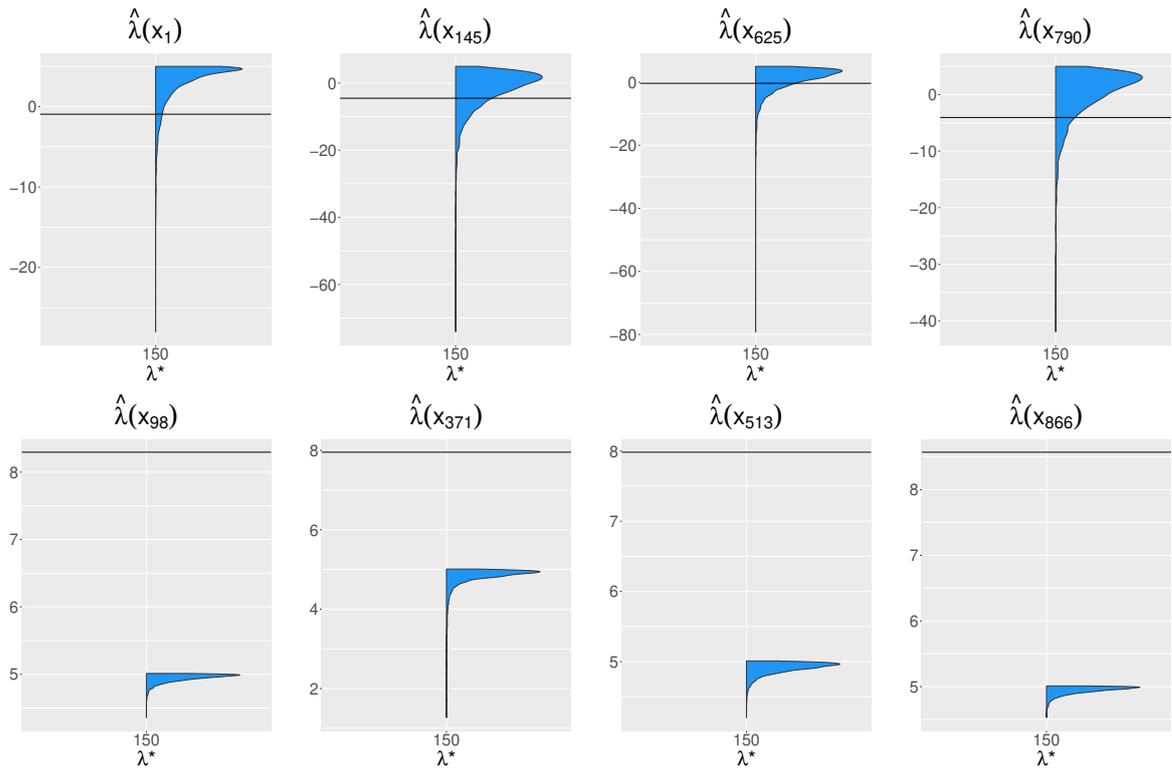


Figura A.4: Distribuições a posteriori da função de intensidade estimada considerando o cenário 2 para os locais de baixa intensidade real $x_1 = (0,017; 0,017)$, $x_{145} = (0,817; 0,150)$, $x_{625} = (0,817; 0,683)$, $x_{790} = (0,317; 0,883)$ e de alta intensidade real $x_{98} = (0,250; 0,117)$, $x_{371} = (0,350; 0,417)$, $x_{513} = (0,083; 0,583)$, $x_{866} = (0,850; 0,950)$.

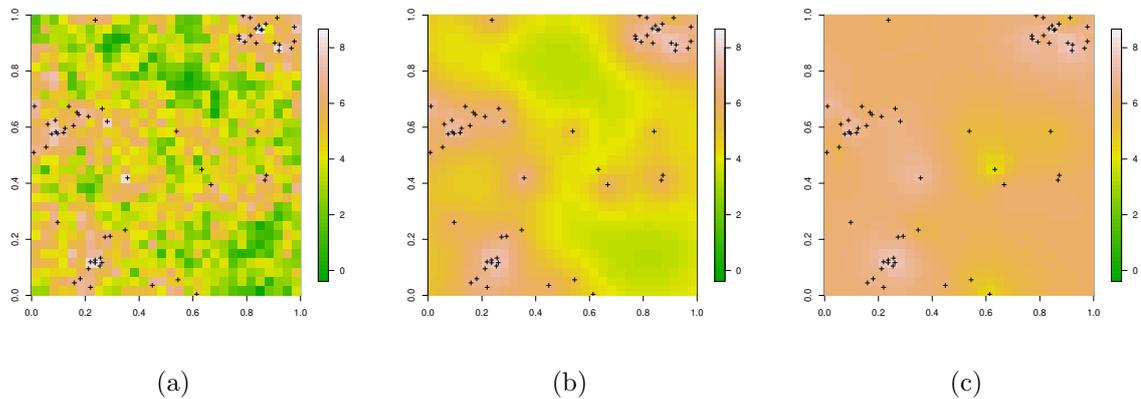


Figura A.5: Mapa do processo Y (a) real e estimado pela média a posteriori considerando (b) o modelo probito e (c) o modelo sem amostragem preferencial, para os dados gerados pelo cenário 2.

informação dos menores valores dos processos.

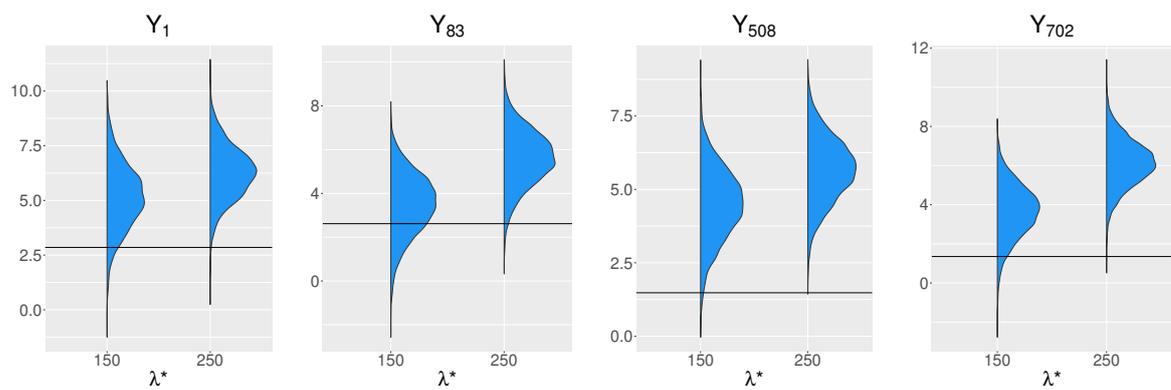


Figura A.6: Distribuições a posteriori do processo Y para os locais não observados $x_1 = (0,017; 0,017)$, $x_{83} = (0,750; 0,083)$, $x_{508} = (0,917; 0,550)$ e $x_{702} = (0,383; 0,783)$, considerando o cenário 2.

APÊNDICE B — Trajetórias das iterações das cadeias de Markov por estimação Bayesiana

Apresentamos a seguir alguns gráficos de trajetórias das cadeias de Markov no processo de estimação dos parâmetros do modelo Bayesiano. As Figuras B.1 e B.2 apresentam as trajetórias das cadeias para o conjunto de dados 1 simulado na Seção 7.1 para os modelos com e sem amostragem preferencial, respectivamente.

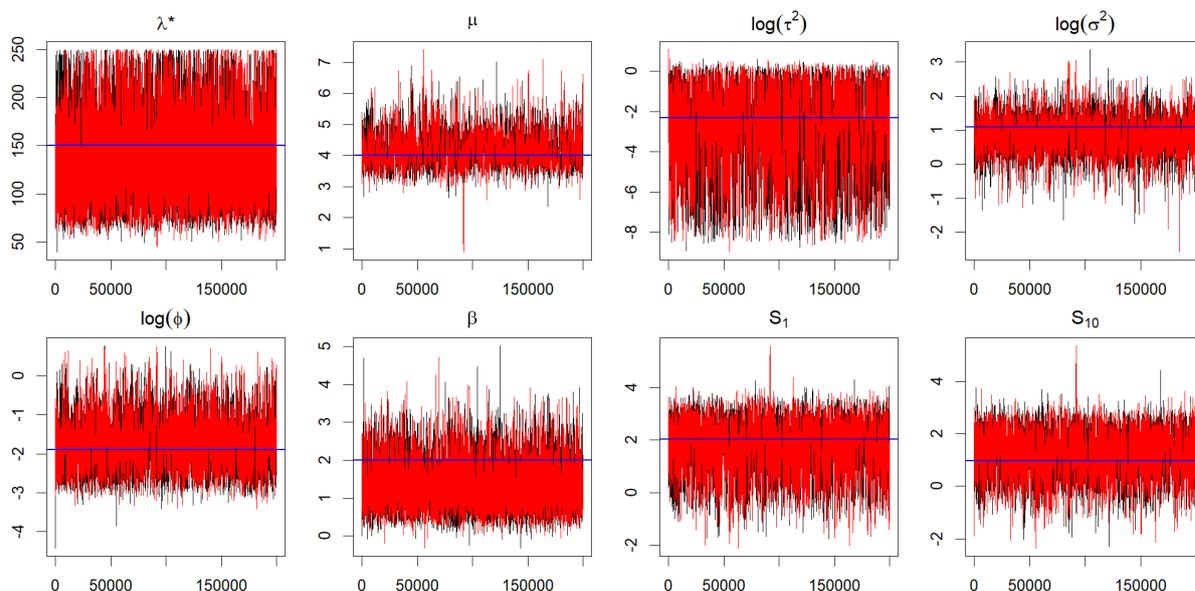


Figura B.1: Gráficos da trajetória dos parâmetros considerando o modelo AP e os dados simulados com amostragem preferencial, para o conjunto de dados 1. A linha horizontal representa o valor verdadeiro do parâmetro.

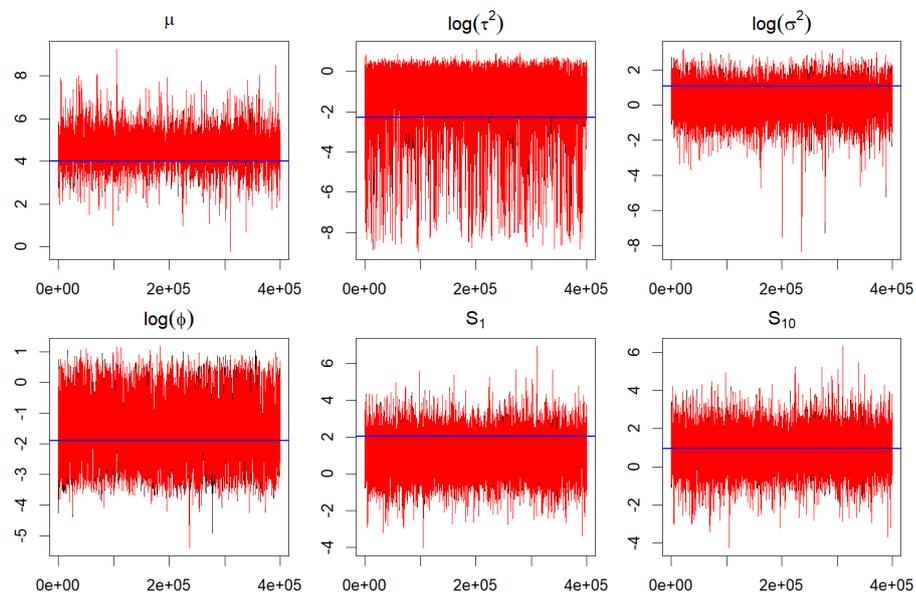


Figura B.2: Gráficos da trajetória dos parâmetros considerando o modelo SAP e os dados simulados com amostragem preferencial, para o conjunto de dados 1. A linha horizontal representa o valor verdadeiro do parâmetro.

As Figuras B.3 e B.4 apresentam as trajetórias das cadeias para o conjunto de dados simulado na Seção 7.5 para os modelos com intercepto considerando como variância a priori para α os valores 1 e 0,1, respectivamente.

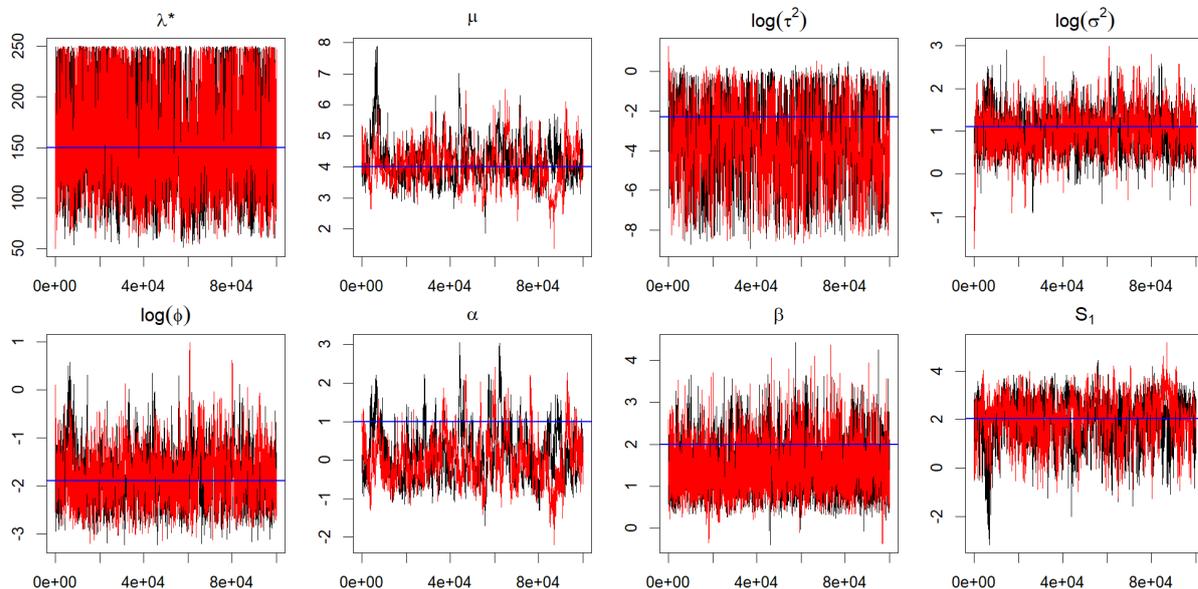


Figura B.3: Gráficos da trajetória dos parâmetros considerando o modelo AP e os dados simulados com intercepto igual a 1, para variância a priori de α igual a 1. A linha horizontal representa o valor verdadeiro do parâmetro.

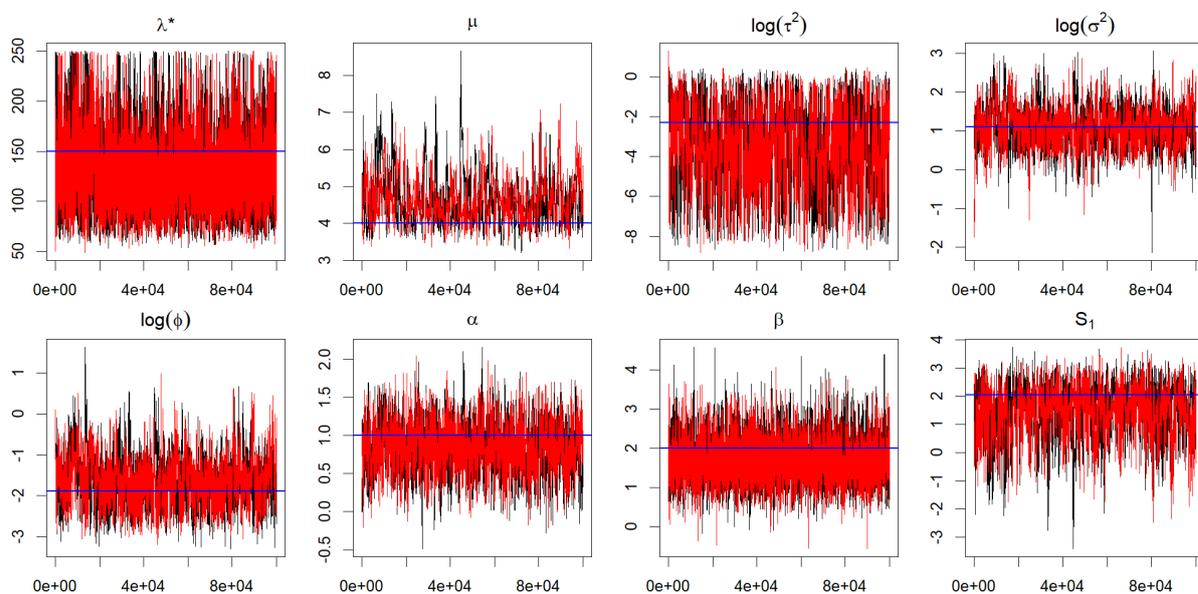


Figura B.4: Gráficos da trajetória dos parâmetros considerando o modelo AP e os dados simulados com intercepto igual a 1, para variância a priori de α igual a 0,1. A linha horizontal representa o valor verdadeiro do parâmetro.

As Figuras B.5 e B.6 apresentam as trajetórias das cadeias para a Aplicação 1 sobre os dados de Galícia no ano de 1997, considerando os modelos com e sem amostragem preferencial, respectivamente.

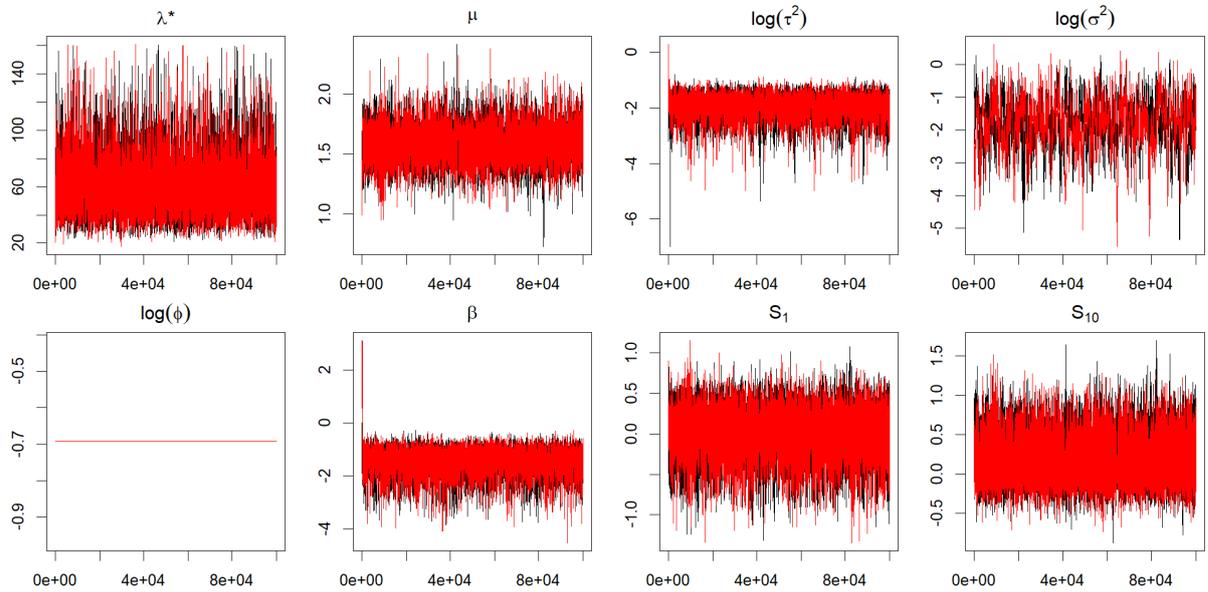


Figura B.5: Gráficos da trajetória dos parâmetros considerando o modelo AP e os dados de Galícia no ano de 1997.

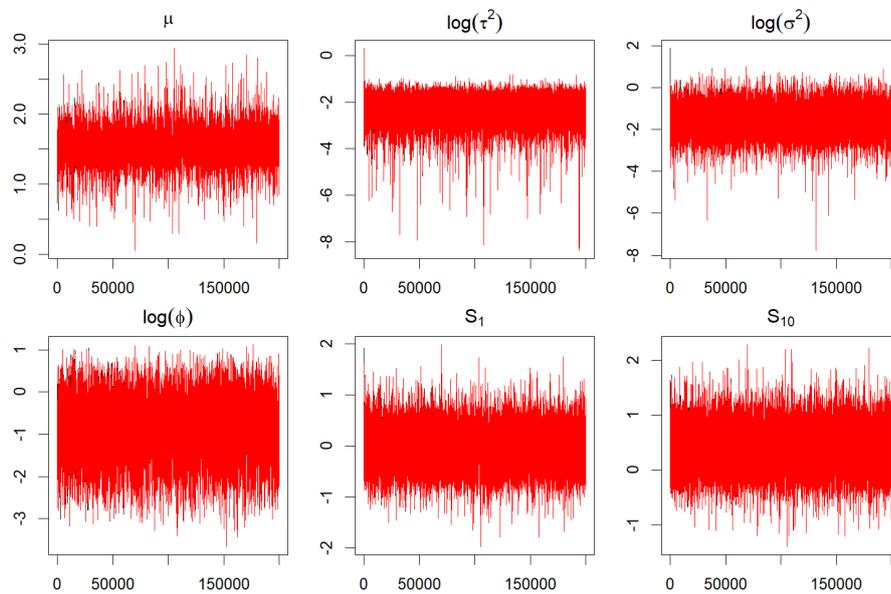


Figura B.6: Gráficos da trajetória dos parâmetros considerando o modelo SAP e os dados de Galícia no ano de 1997.

As Figuras B.7 e B.8 apresentam as trajetórias das cadeias para a Aplicação 1 sobre os dados de Galícia no ano de 2000, considerando os modelos com e sem amostragem preferencial, respectivamente.

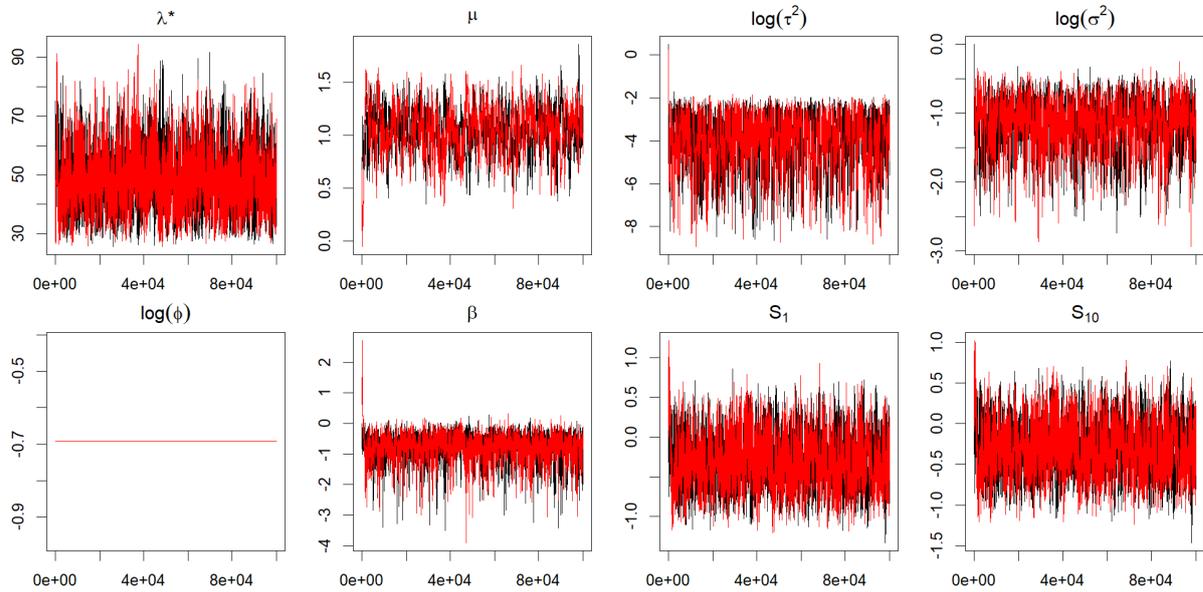


Figura B.7: Gráficos da trajetória dos parâmetros considerando o modelo AP e os dados de Galícia no ano de 2000.

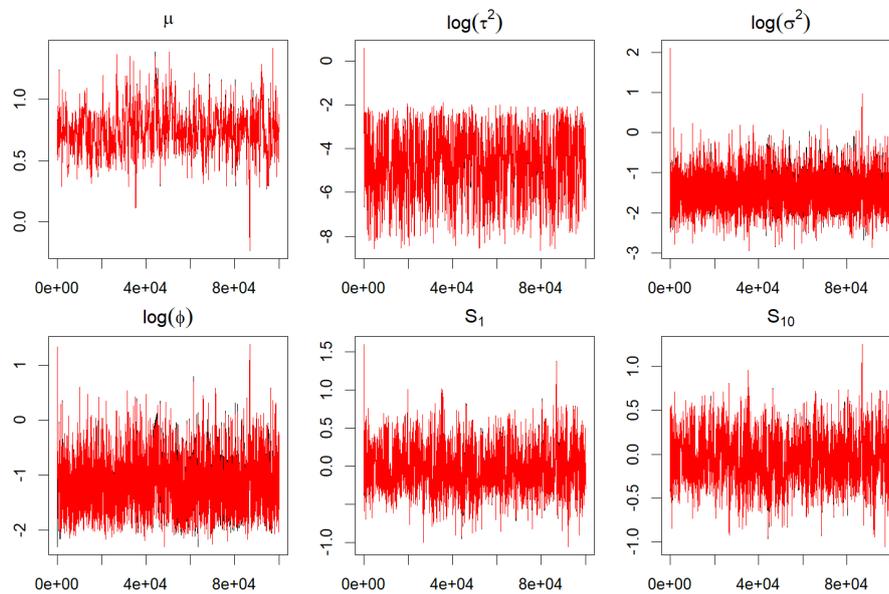


Figura B.8: Gráficos da trajetória dos parâmetros considerando o modelo SAP e os dados de Galícia no ano de 2000.

As Figuras B.9 e B.10 apresentam as trajetórias das cadeias para a Aplicação 2 sobre os dados de radiação na Alemanha, considerando os modelos com e sem amostragem preferencial, respectivamente.

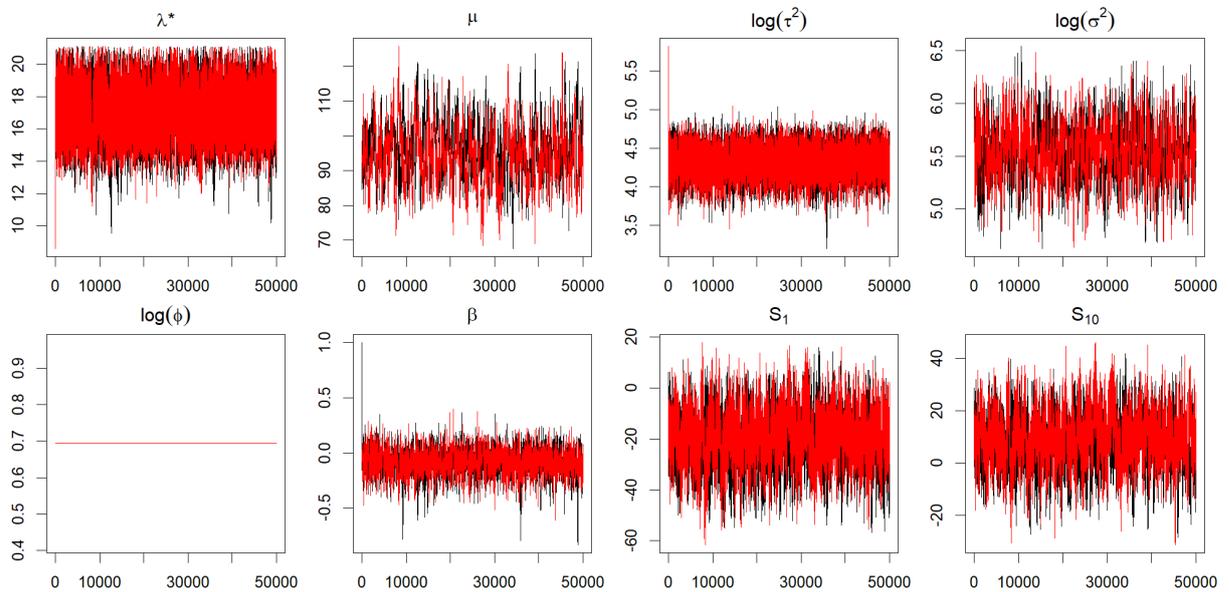


Figura B.9: Gráficos da trajetória dos parâmetros considerando o modelo AP e os dados de radiação na Alemanha.

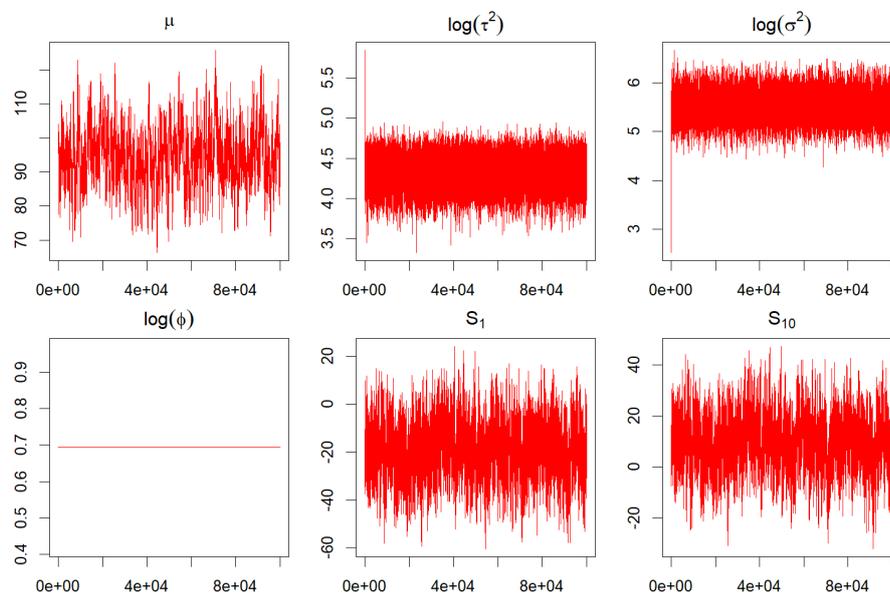


Figura B.10: Gráficos da trajetória dos parâmetros considerando o modelo SAP e os dados de radiação na Alemanha.

As Figuras B.11 a B.13 apresentam as trajetórias das cadeias para a Aplicação 3 sobre os dados de precipitação na cidade do Rio de Janeiro, considerando os modelos AP normal, SAP normal e AP t-Student com 3 graus de liberdade, respectivamente.

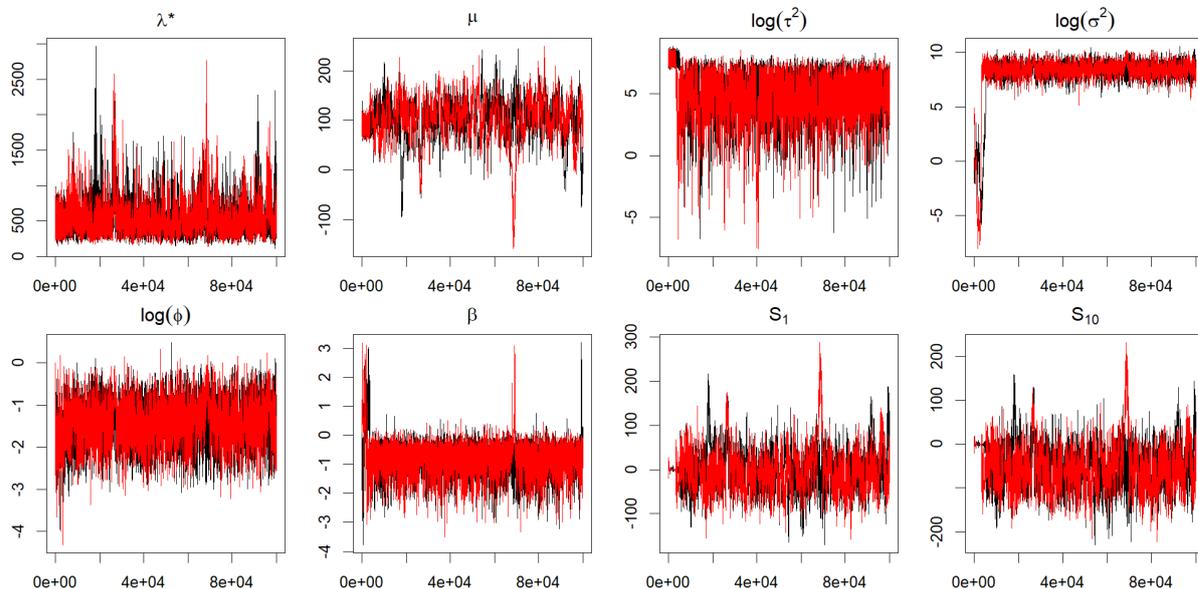


Figura B.11: Gráficos da trajetória dos parâmetros considerando o modelo AP-n e os dados de radiação na Alemanha.

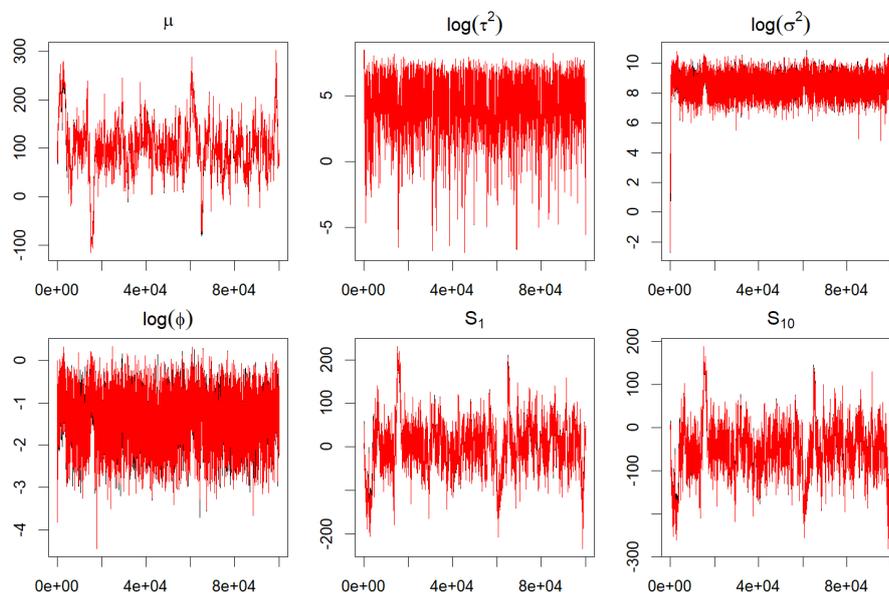


Figura B.12: Gráficos da trajetória dos parâmetros considerando o modelo SAP-n e os dados de radiação na Alemanha.

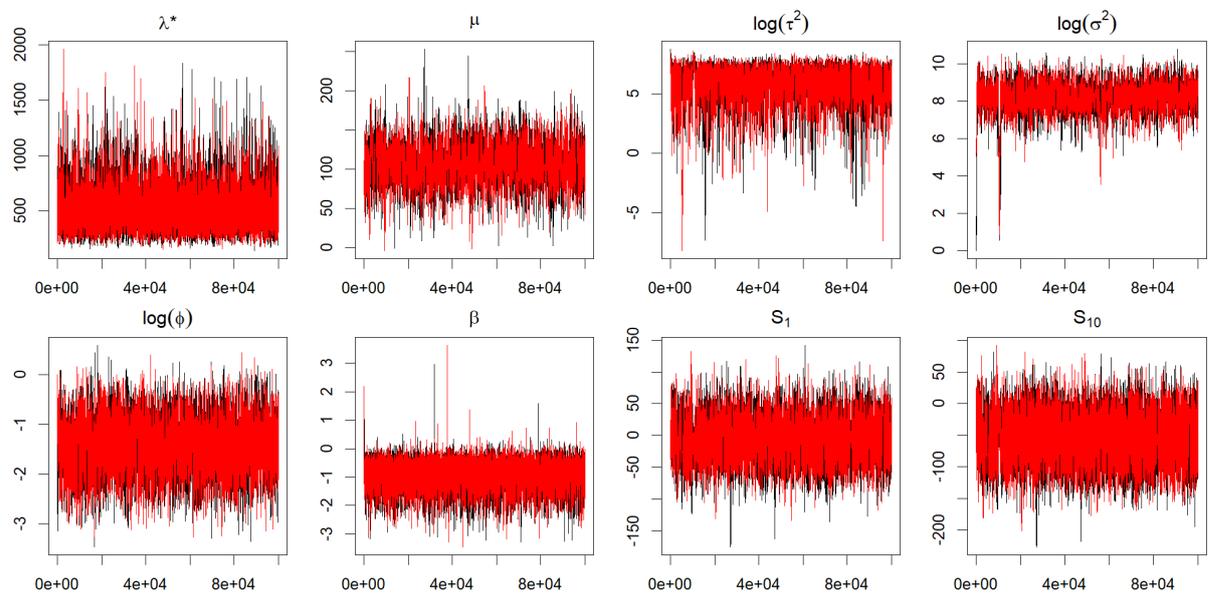


Figura B.13: Gráficos da trajetória dos parâmetros considerando o modelo AP-t3 e os dados de radiação na Alemanha.