



UNIVERSIDADE EDUARDO MONDLANE

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE MATEMÁTICA E INFORMÁTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA APLICADA

DISSERTAÇÃO DE MESTRADO

**MODELAÇÃO BAYESIANA MISTA: UMA APLICAÇÃO A
CASOS E ÓBITOS POR MALÁRIA NO ESPAÇO-TEMPO
EM MOÇAMBIQUE NO PERÍODO DE 2017 A 2023**

FRANCISCO SALMINA

Maputo

2026

UNIVERSIDADE EDUARDO MONDLANE
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE MATEMÁTICA E INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA APLICADA

DISSERTAÇÃO DE MESTRADO

**MODELAÇÃO BAYESIANA MISTA: UMA APLICAÇÃO A
CASOS E ÓBITOS POR MALÁRIA NO ESPAÇO-TEMPO
EM MOÇAMBIQUE NO PERÍODO DE 2017 A 2023**

FRANCISCO SALMINA

Dissertação submetida ao Programa de Pós-Graduação em Estatística Aplicada da Universidade Eduardo Mondlane, como parte dos requisitos para a obtenção do grau de Mestre em Estatística Aplicada.

Supervisor: Doutor Adelino Martins (Hasselt University)

Maputo
2026

UNIVERSIDADE EDUARDO MONDLANE
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE MATEMÁTICA E INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA APLICADA

DISSERTAÇÃO DE MESTRADO

**MODELAÇÃO BAYESIANA MISTA: UMA APLICAÇÃO A CASOS E
ÓBITOS POR MALÁRIA NO ESPAÇO-TEMPO EM MOÇAMBIQUE NO
PERÍODO DE 2017 A 2023**

FRANCISCO SALMINA

Dissertação submetida ao Programa de Pós-graduação em Estatística Aplicada da Universidade Eduardo Mondlane, como parte dos requisitos para a obtenção do grau de Mestre em Estatística Aplicada.

Aprovado em _____ de _____ de 2026, por:

Doutor Adelino José Chingore Juga
(Arguente - DMI - Universidade Eduardo Mondlane)

Prof. Doutor Alberto Chicafo Mulenga
(Presidente - DMI - Universidade Eduardo Mondlane)

Doutor Adelino Martins
(Supervisor - DMI - Universidade Eduardo Mondlane)

Maputo
Março 2026

Dedicatória

A Deus, pelo dom da vida.

AGRADECIMENTOS

Expresso a minha gratidão primordial a Deus por me conceder vida e saúde, além de possibilitar a superação desta etapa significativa em minha jornada.

Agradeço à Universidade Eduardo Mondlane - UEM, em especial à Faculdade de Ciências, Departamento de Matemática e Informática - DMI, Programa de Pós-Graduação em Estatística Aplicada, pela valiosa oportunidade de participar deste programa.

Ao Doutor Adelino Martins, expresso profundo agradecimento por seu apoio, amizade e orientação durante o processo de pesquisa. Faltam-me palavras para expressar adequadamente minha gratidão a si.

Os agradecimentos são extensivos aos demais professores do curso de pós-graduação em Estatística Aplicada na UEM, pelos valiosos ensinamentos compartilhados ao longo do curso, proporcionando uma experiência educacional enriquecedora e inspiradora, meu muito obrigado.

À minha mãe, Salmina Litebe Maculuve, uma mulher verdadeiramente extraordinária e de fibra, que nos enche de orgulho e satisfação com sua presença. Agradeço profundamente por todo o apoio incondicional e incentivo em todos os momentos.

À minha esposa, Márcia Rodrigues Pondja, e ao meu filho, Francisco Salmina Júnior, expresso minha profunda gratidão pela compreensão e paciência nos momentos difíceis e extenuantes desta etapa.

À toda minha amada família, composta por meus irmãos, tios, tias, cunhados e sobrinhos, expresso meu profundo apreço.

Aos companheiros e colegas nomeadamente José Júnior, Kelvin Sozinho e Lucas Jango e os amigos, cuja lista é longa, que ao longo do curso, proporcionaram debates enriquecedores ao conhecimento profissional, expresso minha profunda gratidão. A troca de ideias e experiências entre nós contribuiu significativamente para o crescimento mútuo e para a construção de um ambiente de aprendizado valioso.

Epígrafe

*A natureza está repleta de padrões; é nosso papel decifrá-los para o benefício da humanidade
(parafraseado) Alfred Russel Wallace (1823–1913).*

Resumo

A vigilância espaço-temporal de dados de saúde constitui uma componente essencial na identificação, monitorização e controlo de problemas de saúde pública. Em Moçambique, a malária é uma das principais causas de morbidade e mortalidade, e, apesar dos esforços de redução da doença, com a implementação, por exemplo, do Programa Nacional de Combate a Malária, ainda continua a afectar maior número de pessoas, em especial para crianças e mulheres grávidas. O modelo de Poisson é recorrentemente usado para modelar casos de malária com o enfoque espacial, porém a literatura reporta limitações do modelo que, na sua aplicação, podem apresentar resultados com viés, como é o caso de, na sua forma clássica, não incorporar explicitamente dependências espaciais ou temporais e não permitir modelar heterogeneidade espacial não observada entre distritos, o que pode comprometer a precisão das estimativas. Diante dessas limitações, esta dissertação propõe uma modelação de dados provenientes de uma base secundária de casos e óbitos por malária, obtida junto ao Ministério da Saúde, com o objectivo de analisar os padrões espaço-temporais de casos e óbitos por malária em Moçambique no período de 2017 a 2023, utilizando a classe de modelos Bayesiano misto. A aplicação deste modelo visa por um lado estimar o risco relativo da doença global e específica (no caso por distrito), captando assim a heterogeneidade espacial e temporal da incidência e óbito da malária. Por outro lado, verificar a influência dos factores climáticos na disseminação da malária. Os resultados mostraram que a classe do modelo Bayesianos apresentam melhoria na redução do viés, relativamente aos modelos de Poisson simples usados recorrentemente, com melhor capacidade preditiva, superando alguma das limitações tidas nos modelos recorrentemente usados, indicando distritos com maior número de casos de malária (Nhamatanda, Mafambisse, Sussundenga, Alto Molocue e Lugela). Os resultados mostraram também uma influência de escoamento superficial de água de chuva estagnada, da temperatura e da precipitação total nos casos e óbitos por malária. De ponto de vista prático, os resultados do modelo orientam políticas públicas mais eficazes, apoiando os programas nacionais de controlo da malária na alocação de recursos, na planificação de campanhas de prevenção e na implementação de medidas focadas em zonas e períodos críticos.

Palavras-chave: Cadeias de Markov Monte Carlo; Dados epidemiológicos; Malária; Modelação espaço-temporal; Modelos mistos.

Abstract

Spatio-temporal surveillance of health data constitutes an essential component in the identification, monitoring, and control of public health problems. In Mozambique, malaria is one of the leading causes of morbidity and mortality, and, despite efforts to reduce the disease, including the implementation of, for example, the National Malaria Control Program, it continues to affect a large number of people, particularly children and pregnant women. The Poisson model is frequently used to model malaria cases with a spatial focus; however, the literature reports limitations of this model which, in its application, may produce biased results, such as, in its classical form, not explicitly incorporating spatial or temporal dependencies and not allowing the modelling of unobserved spatial heterogeneity across districts, which may compromise the accuracy of the estimates. Given these limitations, this dissertation proposes the modelling of data from a secondary database of malaria cases and deaths, obtained from the Ministry of Health, with the objective of analysing the spatio-temporal patterns of malaria cases and deaths in Mozambique over the period from 2017 to 2023, using the class of Bayesian mixture models. The application of this model aims, on the one hand, to estimate the relative risk of the disease both globally and specifically (in this case, by district), thereby capturing the spatial and temporal heterogeneity of malaria incidence and mortality. On the other hand, it seeks to assess the influence of climatic factors on the spread of malaria. The results showed that the class of Bayesian models presents improvements in bias reduction compared to the simple Poisson models commonly used, with better predictive capacity, overcoming some of the limitations of the frequently used models, and identifying districts with a higher number of malaria cases (Nhamatanda, Mafambisse, Sussundenga, Alto Molócuè, and Lugela). The results also showed an influence of surface runoff from stagnant rainwater, temperature, and total precipitation on malaria cases and deaths. From a practical perspective, the model results guide more effective public policies, supporting national malaria control programmes in resource allocation, the planning of prevention campaigns, and the implementation of targeted measures in critical areas and periods.

Keywords: Markov Chain Monte Carlo; Epidemiological data; Malaria; Spatio-temporal modeling; Mixture models.

Conteúdo

1	INTRODUÇÃO	1
1.1	Contextualização	1
1.2	Problema de pesquisa	2
1.3	Objectivos	4
1.3.1	Objectivo geral	4
1.3.2	Objectivos específicos	4
1.4	Relevância	4
1.5	Organização da dissertação	5
2	REVISÃO DA LITERATURA	6
2.1	A malária no mundo e a sua evolução	6
2.2	Epidemiologia de malária em Moçambique	9
2.2.1	Conceito da malária	9
2.2.2	Factores que condicionam a malária	10
2.3	Épocas chuvosas e zonas de inundações	11
2.4	Vacinas e erradicação de malária	11
2.5	Estudos similares	12
2.6	Inferência Bayesiana	14
2.6.1	Estimação linear Bayesiana	15
2.6.2	Cadeias de Markov e Monte Carlo	16
2.6.3	Amostragem de Gibbs	17
2.7	Modelos hierárquicos Bayesianos	18
2.7.1	Modelo Poisson-Gama	18
2.7.2	Modelo Poisson-Lognormal	19
2.7.3	Modelos autoregressivos condicionais	20
2.8	Fundamentos estatísticos para a modelação epidemiológica	22
2.8.1	Taxa bruta	22
2.8.2	Modelos lineares generalizados	23
2.9	Análise espacial	24
2.9.1	Estrutura de vizinhança espacial	26
2.9.2	Autocorrelação espacial global univariada	27

2.9.3	Autocorrelação espacial global bi-variada	28
3	METODOLOGIA	29
3.1	Material	29
3.1.1	Descrição do local de estudo	29
3.1.2	População e fonte de dados	29
3.1.3	Aspectos éticos	30
3.2	Métodos	30
3.2.1	Distribuição Poisson-Lognormal não misto	32
3.2.2	Distribuição Poisson-Lognormal – modelo misto com tendência global e específica	32
3.2.3	Distribuição Poisson-Lognormal – modelo misto com tendência especial e temporal	34
3.2.4	CrITÉRIOS de selecção de modelos	36
4	RESULTADOS E DISCUSSÃO	37
4.1	Análise descritiva	37
4.1.1	Distribuição geográfica dos casos e óbitos por malária	38
4.1.2	Análise da autocorrelação espacial	42
4.2	Estimação de modelos dos casos e óbitos por malária	44
4.2.1	Estimação do modelo para casos de malária (global e específica)	44
4.2.2	Estimação do modelo para óbitos por malária	48
4.2.3	Evolução temporal de casos e óbitos por malária.	48
4.2.4	Estimação de risco relativo	51
4.2.5	Teste de convergência - Monte Carlo via Cadeias de Markov	53
4.3	Discussão	54
5	CONCLUSÕES RECOMENDAÇÕES	57
5.1	Conclusões	57
5.2	Recomendações	59
	Referências	60

Lista de Figuras

1	Número estimado de casos de malária por país e área em 2023;	8
2	Países e áreas com casos indígenas em 2000 e seu status em 2023a	9
3	Divisão administrativa e localização de Moçambique;	31
4	Distribuição dos casos de malária no período de 2017 a 2023, apresentados da esquerda para a direita e de cima para baixo. Cada painel representa um ano distinto, permitindo observar a evolução temporal e a variação geográfica da incidência da doença ao longo destes sete anos.	39
5	Distribuição óbitos por malária no período de 2017 a 2023, apresentados da esquerda para a direita e de cima para baixo. Cada painel representa um ano distinto, permitindo observar a evolução temporal e a variação geográfica da taxa de mortalidade ao longo destes sete anos.	41
6	Efeito espacial	47
7	Efeito temporal global	47
8	Histograma de parâmetro de mistura kappa	48
9	Linhas que mostram distritos com alta incidência de malária de 2017 a 2023 em função do tempo.	50
10	Linhas que mostram distritos com alta incidência de óbitos por malária de 2017 a 2023 em função do tempo.	50
11	Risco relativo de casos de malária no lado direito e de óbitos, a esquerda da figura em Moçambique de 2017 à 2023.	51
12	Tendência comum de casos estimados por malária em Moçambique no período de 2017 a 2023.	52
13	Tendência comum de taxa de óbito por malária em Moçambique no período de 2017 a 2023.	53
A1	Tendência dos casos de malária nos distritos com baixa incidência.	65
A2	Tendência dos obitos nos distritos com baixa incidência.	65
A3	Tendência dos casos de malária nos distritos incidência média.	66
A4	Tendência dos obitos por malária nos distritos com incidência média.	66
A5	A esquerda, os gráficos de convergência de casos e a direita, óbitos por malária das cadeias de Markov, onde t_2m e $ruoff$ representam respectivamente a temperatura média e o escoamento superficial.	67

A6	Convergência de parâmetros.	67
A7	Funções de autocorrelação	68
A8	Os mapas ilustram a distribuição espacial da precipitação total no período de 2017 a 2022, apresentados da esquerda para a direita e de cima para baixo. Cada painel representa um ano distinto, permitindo observar a evolução temporal e a variação geográfica da influência das variáveis climáticas na propagação de casos e morbidade por malária ao longo destes seis anos.	69
A9	Os mapas ilustram a distribuição espacial da temperatura média total no período de 2017 a 2022, apresentados da esquerda para a direita e de cima para baixo. Cada painel representa um ano distinto, permitindo observar a evolução temporal e a variação geográfica da influência das variáveis climáticas na propagação de casos e morbidade por malária ao longo destes seis anos.	70
A10	Tendência comum de contaminação por malária em Moçambique no período de 2017 a 2023.	71

Lista de Tabelas

1	Resumo de artigos e modelos estatísticos usados em estudos espaço-temporais de casos de malária.	13
2	Descrição das variáveis do estudo.	30
3	Descrição das variáveis de clima.	32
4	Estatísticas descritivas anuais de casos e óbitos por malária.	37
5	Valores do teste de Moran's I para autocorrelação espacial de casos e óbitos por malária em Moçambique.	43
6	Resultados do Moran Local para os distritos com significância estatística para casos e óbitos por malária simultaneamente. (p -valor < 0.05)	43
7	Critério de Informação WAIC para os Modelos de Casos e Óbitos de Malária.	44
8	Parâmetros estimados para casos e óbitos por malária do modelo 2.	49

Lista de Abreviaturas

AIC	<i>Akaike Information Criterion</i>
BIC	<i>Bayesian Information Criterion</i>
BYM	Besag-York-Mollié
CAR	<i>Conditional Autoregressive</i>
CISMI	Centro de Investigação em Saúde de Manhiça
DIC	<i>Deviance Information Criterion</i>
DNA	Ácido Desoxirribonucleico
ECMWF	<i>European Centre for Medium-Range Weather Forecasts</i>
GIS	<i>Geographic Information System</i>
GLM	<i>Generalized Linear Models</i>
GLMM	<i>Generalized Linear Mixed Models</i>
GMP	<i>Global Malaria Program</i>
ICAR	<i>Intrinsic Conditional Autoregressive</i>
INLA	<i>Integrated Nested Laplace Approximation</i>
IWLS	<i>Iterative Weighted Least Squares</i>
JAGS	<i>Just Another Gibbs Sampler</i>
MCMC	<i>Markov Chain Monte Carlo</i>
MISAU	Ministério da Saúde
MV	Máxima Verossimilhança
OMS	Organização Mundial da Saúde
OPAS	Organização Pan-Americana da Saúde
PNCM	Plano Nacional de Controlo da Malária
USAID	<i>United States Agency for International Development</i>
WAIC	<i>Watanabe–Akaike Information Criterion</i>

1 INTRODUÇÃO

1.1 Contextualização

A malária continua a constituir um dos maiores desafios de saúde pública a nível mundial, afectando as populações de países tropicais e subtropicais. Apesar dos avanços registados nas últimas décadas em termos de prevenção, diagnóstico e tratamento, a doença permanece responsável por milhões de casos e centenas de milhares de mortes anualmente. De acordo com a Organização Mundial da Saúde (OMS), estima-se que, em 2022, tenham ocorrido cerca de 249 milhões de casos de malária em todo o mundo, resultando em aproximadamente 608 mil mortes, sendo que cerca de 95% desses casos ocorreram no continente africano (OMS, 2023). A elevada carga da doença nesta região está associada a factores ambientais favoráveis à proliferação do vector *Anopheles*, condições socioeconómicas desfavoráveis e limitações estruturais nos sistemas de saúde.

Moçambique insere-se entre os países com maior incidência de malária a nível do continente africano, sendo a doença uma das principais causas de morbilidade e mortalidade no país. A malária afecta particularmente crianças menores de cinco anos e mulheres grávidas, grupos populacionais mais vulneráveis às complicações associadas à infecção (OMS, 2020).

A dinâmica de transmissão da malária é caracterizada por elevada complexidade, resultante da interacção entre múltiplos factores ambientais, biológicos, demográficos e socioeconómicos. Elementos como temperatura, precipitação, humidade e presença de corpos de água influenciam directamente o ciclo de vida do mosquito vector e a sobrevivência do parasita *Plasmodium*. Paralelamente, factores humanos, como a densidade populacional, as condições de habitação, o acesso a serviços de saúde e a utilização de medidas preventivas, desempenham igualmente um papel determinante na propagação da doença. Estas interacções produzem padrões heterogéneos de incidência, tanto no espaço como no tempo, tornando essencial a adopção de abordagens analíticas capazes de capturar essa variabilidade (Gething et al., 2011).

Nos últimos anos, o desenvolvimento de métodos estatísticos avançados tem permitido melhorar a análise e interpretação de dados epidemiológicos. Em particular, os modelos estatísticos espaço-temporais têm sido amplamente utilizados para investigar padrões de distribuição de doenças infecciosas, permitindo identificar áreas de maior risco, detectar aglomerados espaciais e compreender a evolução temporal da transmissão.

Entender e modelar a propagação da malária de forma eficaz requer uma abordagem capaz de capturar simultaneamente as dimensões espaciais e temporais da doença. A mobilidade humana emerge como um factor central nesse processo, uma vez que deslocamentos entre regiões podem intensificar a transmissão e dificultar as estratégias de controlo (Gething et al., 2011; Almeida, 2009 e Barbieir, 2005). Neste contexto, conceitos como a variabilidade espaço-temporal e a transmissão secundária ganham importância para a compreensão e o combate à doença.

A modelação estatística espaço-temporal, em particular através de abordagens bayesianas mistas, permite incorporar a variabilidade nos dados de saúde pública, considerando efeitos fixos e aleatórios, bem como dependências espaciais e temporais. Esta capacidade torna os modelos Bayesianos mistos uma ferramenta apropriada para a análise de padrões complexos de propagação da malária.

No contexto de Moçambique, onde a carga da doença permanece elevada e a sua distribuição apresenta grande heterogeneidade geográfica, a aplicação de modelos espaço-temporais pode fornecer evidências relevantes para o planeamento e a avaliação das políticas de controlo da malária. Assim, o presente estudo tem como objectivo geral modelar o risco relativo espaço-temporal de casos e óbitos por malária em Moçambique no período de 2017 a 2023, utilizando abordagens bayesianas que permitam captar a estrutura espacial e temporal subjacente à transmissão da doença.

1.2 Problema de pesquisa

A malária continua a ser um dos principais problemas de saúde pública em Moçambique, representando uma das principais causas de morbilidade e mortalidade no país (Centro de Investigação em Saúde de Manhiça [CISM]), 2023).

O Centro de Investigação em Saúde de Manhiça, através de um dos seus objectivos principais, “conceber e implementar estratégias inovadoras para o controlo e/ou eliminação da malária em diferentes níveis de transmissão em Moçambique” e alinhado com o objectivo do Plano Nacional de Controlo de Malária (PNCM), “reduzir a morbi-mortalidade por malária na população em geral e particularmente nas mulheres grávidas e crianças menores de 5 anos de idade”, aplicam esforços do controlo da doença porém, ainda existe persistência de forma heterogênea em diferentes regiões, influenciada, de um lado, por factores ambientais, climáticos e socioeconómicos e por

outro, pela mobilidade humana.

A distribuição espaço-temporal da malária não é uniforme, variando significativamente entre distritos e ao longo das estações do ano, o que dificulta a implementação das estratégias eficazes de combate. Embora existam estudos sobre a epidemiologia da malária tal como por exemplo: Boulieri et al. (2020); Arambepola et al. (2020), Armando et al. (2025), Plucinski et al. (2018), Jana et al. (2022), Zacarias e Andersson, (2011) e Escaramís et al. (2011), pouco exploram de forma aprofundada a interação entre factores ambientais e padrões de deslocamento da população, simultaneamente, na propagação da doença. Os autores acima referidos, apesar de usar modelos Poisson com estrutura hierárquica com uma estimação global que segue, especialmente, um priori autorregressivo condicional (*Condiciona autoregressive* [CAR]), eles usam um efeito aleatório para a distribuição espacial e outro efeito aleatório para o tempo, assumindo que a dinâmica da doença apresenta dependência estruturada tanto no espaço quanto no tempo. Assim, não sendo suficiente a captura simultânea do efeito espacial e temporal.

O estudo recente feito pelos autores Armando et al. (2025), empregou um modelo Bayesiano espaço-temporal para prever a incidência mensal de malária em Moçambique, utilizando o método INLA (*Integrated Nested Laplace Approximation*) com o objectivo central de capturar simultaneamente as variações espaciais entre distritos e as flutuações temporais mensais e anuais da transmissão empregando regressão com distribuição binomial negativa, porém apresenta uma limitação muito crucial, que é assumir a dependência espacial (Besag–York–Mollié 2-BYM2) e dependência espacial (RW1) uniformes para todos os distritos o que torna transições temporais suaves mês a mês. Um outro ponto, é o modelo usado pelos autores, diminuir a precisão quando usado lags mais longos. Um outro estudo que se destaca é o feito pelos autores Zacarias e Andersson (2011) que implementam um modelo Bayesiano hierárquico espaço-temporal com distribuição Poisson para os casos observados de malária e consideram a população (casos esperados) como um desvio no entanto, o modelo assume relação linear por intervalos, mas existe não-linearidades entre clima e malária.

Na presente pesquisa, pretende-se usar o modelo Poisson-Lognormal porém com estimação global e específica do risco relativo de acordo com a modelação proposta de Boulieri et al, (2020). A estimação global e específica permitirá modelar a distribuição espaço-temporal da malária em Moçambique entre 2017 e 2023, para compreender os padrões de disseminação da doença, buscar-

se responder à seguinte questão de pesquisa: Como a modelagem espaço-temporal pode contribuir para a compreensão da disseminação da malária em Moçambique, considerando a mobilidade humana e factores ambientais?

1.3 Objectivos

1.3.1 Objectivo geral

Modelar a distribuição do risco relativo espaço-tempo de casos e óbitos por malária em Moçambique no período de 2017 a 2023.

1.3.2 Objectivos específicos

- Descrever padrões de distribuição e correlação espacial e temporal de casos e óbitos por malária;
- Verificar a influência dos factores climáticos (temperatura média, precipitação e escoamento superficial das águas) na taxa de incidência de casos e óbitos por malária;
- Decompor a tendência global e específica de casos e óbitos por malária;
- Estimar o risco relativo e a taxa de mortalidade por malária.

1.4 Relevância

A análise espacial da propagação da malária tem sido pouco realizada. Durante a pesquisa bibliográfica poucos estudos sobre o tema foram encontrados, tratando a maioria da propagação da doença quanto à migração populacional e não à mobilidade humana e nem à óbitos por malária.

O estudo pretende mostrar, no âmbito social, que a análise dos padrões espaço-temporais de disseminação e óbitos por malária pode em muito contribuir para com as acções de controle da doença. Assim como, trazer uma nova óptica sobre estas acções, incentivando a adopção de novas práticas de vigilância epidemiológica de modo a prever e impedir a disseminação da malária e suas consequências, morte. Quanto ao âmbito académico, a aplicação de modelos bayesianos mistos em dados espaço-temporais pode aprimorar metodologias para análise de doenças transmissíveis, contribuindo para a literatura sobre epidemiologia computacional; e por outro lado, o estudo pode gerar modelos preditivos mais robustos, auxiliando na identificação de áreas de alto

risco e contribuindo para estratégias de controle e prevenção.

1.5 Organização da dissertação

A dissertação está organizada por capítulo. O primeiro capítulo busca apresentar a introdução e contextualizar sobre o tema em análise. Apresenta também os objectivos que preconizam este estudo. No capítulo 2 é apresentada a revisão da literatura. Nela está contida uma revisão da literatura com foco nas questões de interesse ao estudo; no capítulo 3, referente a metodologia, são descritos os recursos utilizados (dados e programas computacionais) e a metodologia usada; no capítulo 4 são apresentados os resultados e discussão sobre os resultados, ou seja, os padrões espaciais de como a malária se propaga em Moçambique: os grupos detectados, os fluxos gerados e a análise epidemiológica destes; finalmente, no capítulo 5, são apresentadas as conclusões e recomendações.

2 REVISÃO DA LITERATURA

A malária continua a ser uma das principais causas de mortalidade em Moçambique, revelando-se persistente apesar dos esforços de controlo. Em 2022, representou cerca de um quarto de todas as consultas externas, ultrapassando os 12 milhões de casos confirmados. Entre 2018 e 2022, os casos aumentaram anualmente, com excepção de 2021, ano em que se registou uma ligeira redução possivelmente associada ao impacto da pandemia de COVID-19. Nas crianças dos 6 aos 59 meses, a prevalência da doença manteve-se estável, embora a anemia malárica tenha aumentado de forma expressiva no mesmo período. Além disso, desde 2015, segundo a Agência dos Estados Unidos para o Desenvolvimento Internacional (*United States Agency for International Development* [USAID]), verifica-se um crescimento acentuado de número de casos nas províncias do norte, particularmente Manica, Nampula, Zambézia e Cabo Delgado (USAID, 2023).

2.1 A malária no mundo e a sua evolução

A origem da malária tem sido amplamente discutida entre investigadores das ciências biomédicas, sendo considerada uma das doenças infecciosas mais antigas. A génese e a transmissão dos parasitas da malária (*Plasmodium*) têm sido debatidas pelos vários autores como Rich et al. (2009), com base em argumentos, designadamente a sua origem não humana – através dos chimpanzés, outros primatas e galinhas –, a sua origem humana ou coespeciação, o grau de imunidade das populações, a adaptação dos vectores às condições climáticas e o tipo de fixação das populações, entre outros (Carter & Mendis, 2002).

A existência de cinco parasitas que, actualmente, infectam os humanos – *Plasmodium malariae*, *Plasmodium ovale*, *Plasmodium vivax*, *Plasmodium falciparum* e *Plasmodium knowlesi* – e as análises genéticas apontam para que os parasitas tenham evoluído a partir de África, segundo Escalante et al. (2005), com a população dos primeiros homínídeos que divergiram dos grandes primatas, saindo da savana africana, há 5 milhões de anos (Packard, 2007). Entre 100 000 e 50 000 a.C., os primeiros homens modernos iniciam a migração que viria a dar origem aos primeiros euroasiáticos modernos, deixando para trás vários insectos e animais tropicais, que serviam de repositório e transmissor, assim como várias doenças tropicais, nomeadamente a febre-amarela, tripanossomíase, cegueira dos rios e a elefantíase (Webb, 2009). Os parasitas da malária, transportados no seu sangue, viriam a encontrar, no novo ambiente, uma vasta variedade de insectos da

espécie *Anopheles*, como vector de transmissão. Julga-se que a malária terá afectado as grandes civilizações milenares, havendo documentos chineses com referências à doença, em 2700 a.C., tábuas de argila da Mesopotâmia, a partir de 2 000 a.C., papiros egípcios de 1 570 a.C. e textos hindus de 600 a.C. (Cox, 2010).

O crescente contacto entre os povos, por motivos comerciais e militares, assim como a existência progressiva de aglomerados populacionais que requereram a prática da agricultura de um modo extensivo e, ainda, a domesticação de animais viriam a alterar a ecologia das regiões e a imunidade dos indivíduos aos quais os parasitas e os vectores se adaptaram (Webb, 2009). A domesticação de animais e a agricultura criaram uma nova ordem epidemiológica. De acordo com as evidências apresentadas por Webb (2009), no seu livro “Humanity’s Burden: A Global History of Malaria”, ela teve uma vasta dispersão, a nível mundial, à excepção de alguns países cujas condições ambientais não permitiram a reprodução do parasita e onde a fixação humana era muito reduzida. Esta doença que, actualmente, possui uma elevada morbimortalidade nos países tropicais africanos, há menos de um século afectava quase toda a população mundial (Sequeira, 2017).

Conhecido o agente da doença e o seu vector de transmissão, as medidas de controlo da malária viriam a polarizar-se entre dois campos: por um lado, aqueles que defendiam que a sua eliminação passaria por uma reforma social e pelo desenvolvimento agrícola eficaz (Ângelo Celli e o Coronel S.P. James destacaram-se nesta posição), num momento em que as actividades humanas configuravam a ecologia e a epidemiologia da malária, através das guerras, da exploração mineira, da construção de estradas e caminhos-de-ferro, da progressiva industrialização e dos movimentos populacionais forçados e voluntários (Packard, 2007). Por outro lado, aqueles que defendiam que o controlo da doença passaria por soluções na esfera biomédica, como a administração de anti-maláricos e a utilização de insecticidas e larvicidas, para eliminar os mosquitos e seus criadouros.

Com o um dos lemas, “Abordar a desigualdade na resposta global à malária”, a cada ano, o *World malaria report* fornece ferramentas vitais para avaliar o progresso global e as lacunas na luta contra a malária. O relatório do ano 2024 fornece um instantâneo crítico e actualizado dos esforços para controlar e eliminar a doença em 83 países no mundo todo.

Analisado o relatório foi notório que introduz, pela primeira vez, um capítulo dedicado a necessidade de uma resposta mais inclusiva e eficaz, com foco em atingir as populações mais vulneráveis

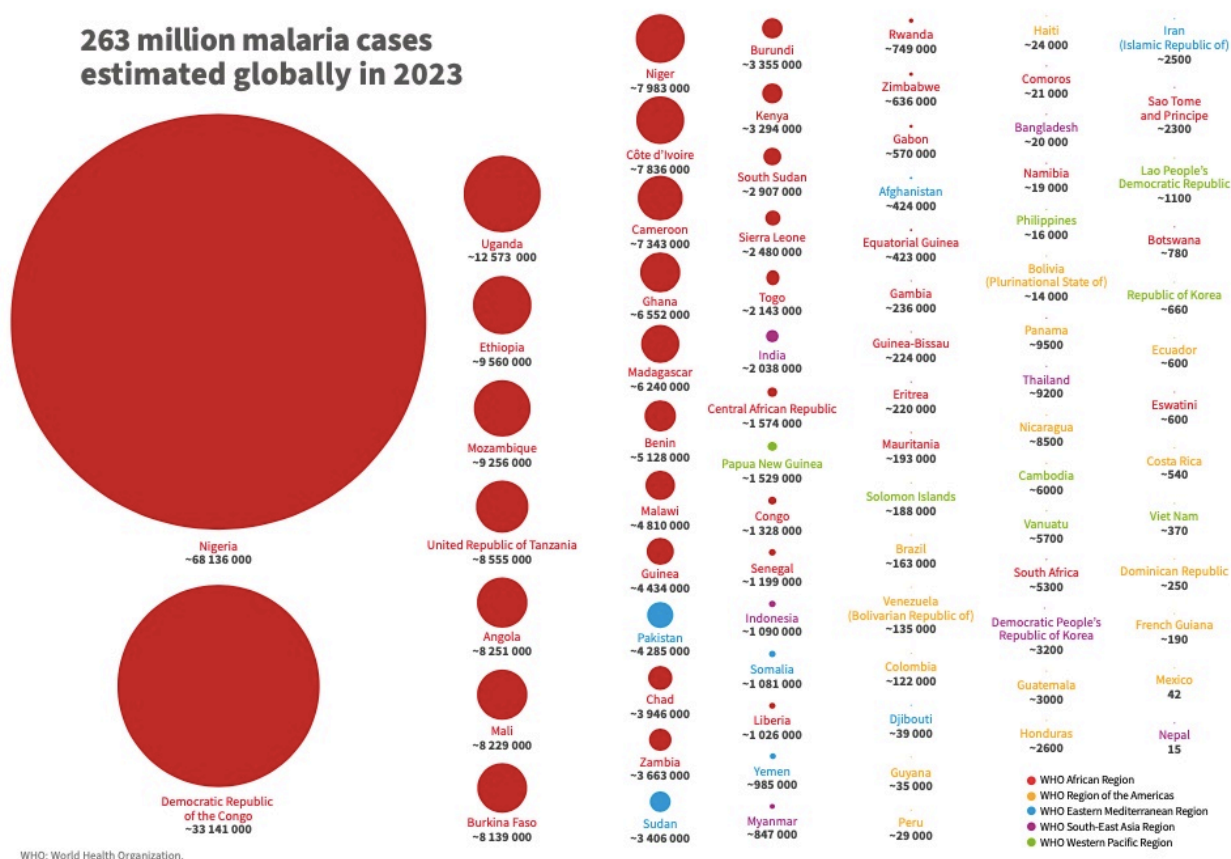


Figura 1: Número estimado de casos de malária por país e área em 2023; Fonte: OMS (2023).

à malária. Grupos com alto risco de infecção por malária incluem crianças menores de 5 anos, mulheres e meninas, povos indígenas, migrantes, pessoas com deficiência e pessoas em áreas remotas com acesso limitado a cuidados de saúde. A Figura 1 apresenta o número de casos clínicos de malária e mortes estimadas ocorrido no período de 2000 a 2023, e a incidência de casos de malária e taxa de mortalidade.

Globalmente, em 2023, houve quase 263 milhões de casos estimados de malária em 83 países endêmicos de malária (Figura 2), um aumento de 11 milhões de casos em comparação com 2022. Entre 2000 e 2019, o número de casos anuais estimados de malária permaneceu estável, variando entre 227 milhões e 248 milhões nos 108 países que eram endêmicos de malária em 2000. Desde 2020, o número estimado de casos de malária tem aumentado constantemente, e a maior parte desse aumento ocorreu em países da Região Africana da OMS (89,7%) e da Região do Mediterrâneo Oriental da OMS (15,5%). Os principais países que contribuíram para o aumento de casos entre 2022 e 2023 foram Etiópia (+4,5 milhões), Madagascar (+2,7 milhões), Paquistão (+1,6 milhões), Nigéria (+1,4 milhões) e República Democrática do Congo (+600 000).

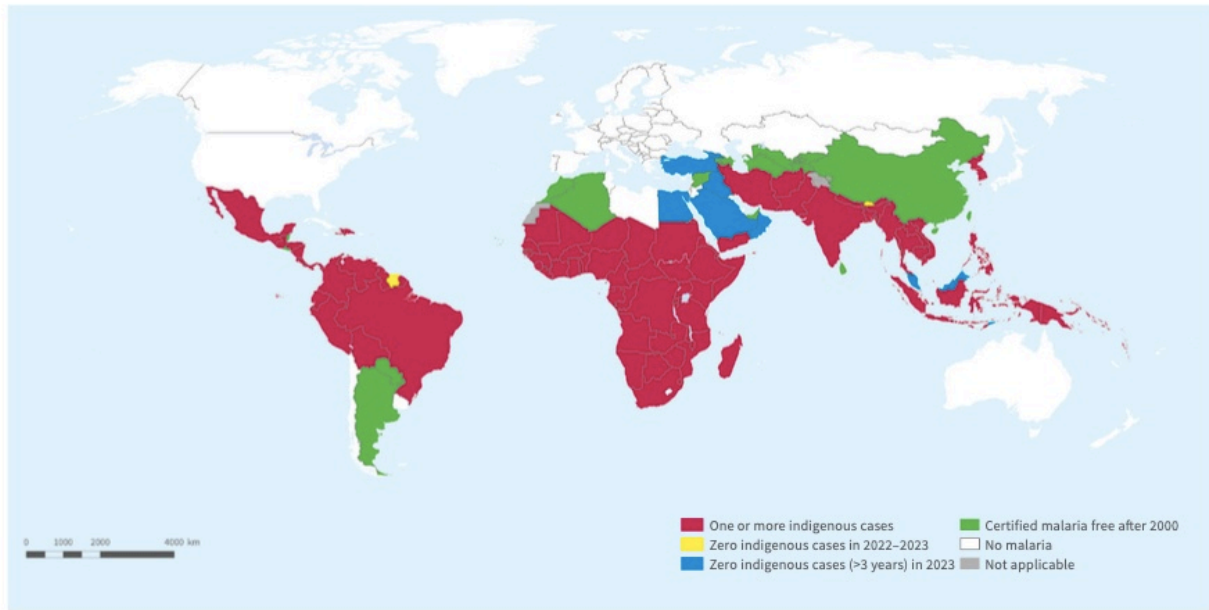


Figura 2: Países e áreas com casos indígenas em 2000 e seu status em 2023a
 Fonte: OMS (2024)

2.2 Epidemiologia de malária em Moçambique

Segundo o PNCM, em Moçambique, a malária é uma doença endémica, em todo o país, variando entre zonas hiperendémicas, ao longo do litoral, zonas mesoendémicas, nas terras planas do interior, e de algumas zonas hipoendémicas, nas terras altas do interior. Esta variação está, principalmente, relacionada com o aumento da precipitação e temperatura e, conseqüentemente, dos criadouros dos mosquitos anofelinos (Zacarias & Anderson, 2010). Segundo o Centro de Investigação em Saúde de Manhica, Moçambique é um dos dez países com maior peso da malária no mundo, com a doença sendo responsável em cerca de 29% (42% em crianças menores de 5 anos). Dados do inquérito de indicadores de malária realizado em 2018, pela instituição referida acima, indicam que a prevalência de malária em crianças menores de 5 anos foi de 39% em Moçambique, variando de 1% na Cidade e Província de Maputo até 57% na província de Cabo Delgado. A transmissão é perene, com picos durante e após estações chuvosas (Dezembro a Abril). No entanto, as intensidades de transmissão variam dependendo da quantidade de chuva e da temperatura do ar observada em cada ano e também dependendo das condições locais específicas.

2.2.1 Conceito da malária

A malária é uma doença infecciosa febril aguda, causada por protozoários do gênero *Plasmodium* e transmitida ao homem, através da picada de mosquitos fêmeas do gênero *Anopheles* (Avila, 1996

e Souza et al., 1997). Portanto, sua ocorrência está condicionada à interação de três fatores: o parasito (plasmódio), o hospedeiro (homem) e o vector (mosquito).

As espécies de plasmódio associadas à infecção humana são: *Plasmodium falciparum*, *P. vivax*, *P. malariae* e *P. ovale*. A malária por *P. ovale* ocorre apenas no continente africano.

O mosquito torna-se transmissor da doença quando infectado ao sugar o sangue de uma pessoa com gametócitos circulantes. Gametócito é a forma sexuada dos parasitos (plasmódio da malária) encontrados nos humanos infectados.

Os parasitos surgem na corrente sanguínea humana em períodos que variam, após a picada do mosquito infectado, de poucas horas para o *P. vivax* e de 7 a 12 dias para o *P. falciparum*. Enquanto que, o espaço de tempo correspondente entre a picada do mosquito e o aparecimento dos primeiros sintomas (período de incubação) é de 08 a 12 dias para o *P. falciparum*, 13 a 17 para o *P. vivax* e de 28 a 30 para o *P. malarie* (Ferreira, 2004).

Nos casos com complicações podem ainda ocorrer forte dor abdominal, sonolência e redução da consciência, podendo levar ao coma nos casos de malária cerebral. Os adultos não imunes, bem como crianças e gestantes, podem apresentar manifestações mais graves da infecção, podendo ser letal. De um modo geral, as formas brandas são causadas pelo *P. malarie* e *P. vivax* e as formas clínicas mais graves pelo *P. falciparum* (Ferreira, 2004).

2.2.2 Factores que condicionam a malária

De uma forma geral, as causas da malária são sempre múltiplas, um processo altamente complexo que envolve o agente, o hospedeiro e o ambiente. No relatório da Organização Pan-Americana de Saúde (OPAS) de 2008 citado por Shimada e Imakawa (2010), incorpora os factores que condicionam a transmissão da malária tais como: ambientais (vegetação, clima, hidrologia, topografia); sócio-demográficos (migrações, mobilidade humana e densidade populacional); biológicos (ciclo vital dos insectos vectores de agentes infecciosos); médico sociais (estado imunológico da população, efectividade dos sistemas locais de saúde e dos programas específicos de controle de doenças) e as origens históricas da doença na região.

Para Shimada & Imakawa (2010), os factores ambientais desempenham um papel importante na transmissão da malária e podem ser divididos em dois grupos: meio ambiente natural e meio ambiente modificado pela acção humana (antrópico). O meio natural inclui a topografia, hidrologia, tipo de solo e clima. Enquanto que o meio antrópico pode ter influência das migrações populacionais, da mobilidade humana, da ocupação desordenada nas áreas rurais, do uso inadequado de corpos de água, etc. O impacto de cada um desses grupos varia por área geográfica e depende do contexto socioeconómico em que são observados. Além disso, a distribuição de vectores é regulada por factores climáticos, como temperatura, umidade relativa do ar e regime das chuvas que criam condições favoráveis para a sua reprodução, desenvolvimento e longevidade.

Para Martens et al. (1995), a precipitação exerce influência sobre a malária porque parte do ciclo do mosquito ocorre na água. As chuvas em quantidade moderada podem facilitar a proliferação dos mosquitos vectores, mas se ocorrerem, em excesso, destroem os criadouros dos mosquitos, eliminando as larvas dos mesmos.

2.3 Épocas chuvosas e zonas de inundações

Um estudo realizado por Américo (2023) com o objectivo de mapear os riscos de infecção por doenças sensíveis ao clima (malária e diarreia) na época chuvosa 2023 em Moçambique indica que a precipitação é um factor que contribuiu, significativamente, para o aumento de casos de malária e cólera no país. O estudo indica que temperaturas a partir de 22 a 35 graus são favoráveis para o aumento de casos de malária e diarreia, porém, quando está acima de 38, há tendência de redução.

A esporogonia no mosquito tem o tempo de sobrevivência de aproximadamente 10 a 20 dias e o mosquito permanece infeccioso por 1 a 2 meses. A fase esquizogónica inicia com a inoculação do parasita no sangue do hospedeiro vertebrado por meio da picada de uma fêmea do mosquito do género *Anopheles*.

2.4 Vacinas e erradicação de malária

A nível Mundial, o Programa Global de Malária da OMS (traduzido de *Global Malaria Program* [GMP]) é responsável por coordenar os esforços do controlo e eliminação da malária.

A escala exacta de perdas económicas atribuídas à malária em Moçambique não é bem conhe-

cida. Porém é evidente que a malária contribui para elevadas perdas económicas, altas taxas de absentismo escolar e uma fraca produtividade agrícola, principal meio de subsistência da maioria da população rural.

A actividade de controlo da malária em Moçambique remonta da década de 50 aquando do início do programa global de erradicação da malária. Contudo, só foi em 1982 que foi criado o Programa Nacional de Controlo da Malária (PNCM) com a designação actual, com objectivo principal de reduzir a morbi-mortalidade por malária na população em geral e particularmente nas mulheres grávidas e crianças menores de 5 anos de idade e grupos socialmente desfavorecidos.

Os principais fármacos antimaláricos são classificados, de acordo com o modo de acção, em dois grandes grupos: O primeiro grupo inclui os alcalóides derivados da cinchona, as aminoquinolinas e as acridinas. Acredita-se que esses fármacos podem interferir no metabolismo da glicose em diferentes pontos (Rey, 2001) e também na habilidade do parasita de digerir a hemoglobina, impedindo, assim, que ele se alimente ou intoxicando-o com altos níveis de ferriprotoporfirina-IX, que é um subproduto tóxico da digestão da hemoglobina (Robert et al., 2001; Silva et al., 2005). O segundo grupo inclui as pirimidinas e biguanidas e envolve a interferência na síntese do ácido tetra-hidrofólico, importante cofactor no processo de síntese de DNA e de aminoácidos (Delfino, 2002). A OMS recomenda vacinas contra malária para a prevenção da malária por *P. falciparum* em crianças que vivem em áreas onde a malária é endêmica, priorizando áreas de transmissão moderada e alta. Duas vacinas contra malária são recomendadas para uso: RTS,S e R21.

2.5 Estudos similares

Para enriquecer e sustentar a pesquisa, a análise dos artigos seleccionados foi fundamentada em alguns tópicos necessários para qualquer pesquisa, tendo em vista a busca pela coerência entre um determinado problema de pesquisa e os objectivos traçados, pressupostos epistemológicos e metodológicos que contribuem para sua explicação, descrição e/ou compreensão. Nessa direcção, para a discussão aqui apresentada utilizou-se como parâmetros o campo científico no qual o estudo está circunscrito, suas bases epistemológicas, o tipo de desenho de pesquisa, a natureza dos dados colectados/produzidos, os modelos para o seu desenvolvimento.

De forma geral, na Tabela 1, embora os estudos apresentem problemas e resultados relevantes

Tabela 1: Resumo de artigos e modelos estatísticos usados em estudos espaço-temporais de casos de malária.

Título do Artigo	Autores (Ano)	Palavras-Chave	Modelo Usado
A Bayesian mixture modeling approach for public health surveillance	Boulieri, A.; Bennett, J.E.; Blangiaro, M. (2018)	Bayesian hierarchical analysis; Mixture modeling; Public health surveillance; Spatio-temporal modeling.	Modelo de mistura Bayesiano hierárquico com componente espacial (BYM) e temporal (ICAR/RW1)
Malaria Temporal Variation and Modelling Using Time-Series in Sussundenga District, Mozambique	Ferrão, J.L. et al. (2021)	Malaria; modelling; temporal; Sussundenga.	ARIMA (2,2,1)(1,1,1) com sazonalidade semanal (s2)
Spatio-temporal modeling of weekly malaria incidence in children under 5 for early epidemic detection in Mozambique	Colborn, K.L. et al. (2018)	Malaria early warning system; spatio-temporal modelling; exceedance probability; Mozambique.	Modelo Bayesiano com covariáveis climáticas e efeitos aleatórios Gaussianos
Childhood malaria case incidence in Malawi between 2004 and 2017	Chirombo, J. et al. (2020)	Malaria, Climate, Statistical model, Spatio-temporal, Vectors	GLMM com efeitos espaciais (CAR) e temporais
Spatio-temporal analysis of mortality among children under 5 in Manhiça (Moçambique)	Escaramis, G. et al. (2011)	Child mortality, spatio-temporal, Poisson regression, CAR, Mozambique	Regressão de Poisson com CAR e tendência temporal polinomial
Spatial and temporal patterns of malaria incidence in Mozambique	Zacarias, O. P., & Andersson, M. (2011)	Malaria; spatio-temporal modelling; climatic covariates; Bayesian hierarchical model; Mozambique	Modelo Bayesiano hierárquico espaço-temporal com efeitos CAR, RW1 no tempo e interação espaço-tempo AR(1)

social e cientificamente, em função dos modelos usados e propostos, apresentam lacunas ou falhas em relação aos aspectos de efeitos espaciais. Apenas nos artigos de Bennett et al. (2020) e Zacarias e Andersson (2011) é considerado o efeito espacial nas suas propostas, respectivamente do modelo mistura e efeito CAR e RW1, que por sinal, serão bases para a pesquisa.

2.6 Inferência Bayesiana

A estimação pelo método de Máxima Verossimilhança (MV) é clássico e amplamente utilizado para a estimação de parâmetros estatísticos. A sua lógica consiste em seleccionar os valores dos parâmetros que maximizam a probabilidade de se observar o conjunto de dados disponíveis. Formalmente, para observações independentes $y = (y_1, \dots, y_n)$ com densidade $f(y_i | \theta)$, a função de verossimilhança é definida por

$$L(\theta | y) = \prod_{i=1}^n f(y_i | \theta) \quad (1)$$

e o estimador de máxima verossimilhança é dado por

$$\hat{\theta} = \arg \max_{\theta} L(\theta | y) \quad (2)$$

Apesar da sua simplicidade, a MV apresenta limitações práticas. Em particular, os seus resultados dependem fortemente de amostras grandes para garantir propriedades assintóticas desejáveis, como consistência e normalidade aproximada (Casella & Berger, 2011). Em situações com dados escassos, incompletos ou dependentes, especialmente em contextos espaciais, temporais ou hierárquicos, os estimadores podem tornar-se instáveis ou enviesados. Outro ponto crítico é que a MV não incorpora conhecimento prévio, dependendo exclusivamente dos dados observados, o que pode ser insuficiente em contextos epidemiológicos ou ambientais onde a incerteza é elevada. (Gelman et al., 2013).

Nesse cenário, a inferência Bayesiana apresenta vantagens substanciais. Em vez de fornecer apenas um estimador pontual, como faz a MV, a abordagem Bayesiana combina informação prévia $\pi(\theta)$ com a verossimilhança para obter a distribuição a posteriori.

A distribuição posterior contém toda a informação sobre o parâmetro de interesse. No entanto, é desejável resumir essa informação posterior de forma concisa, e uma forma simples de caracterizar a informação a posteriori é calcular probabilidades $p(a < \theta < b | y)$ para diversos valores de a e b . No entanto, resumir a informação a posteriori em probabilidades $p(a < \theta < b | y)$ é útil, mas não é suficiente. Na abordagem frequentista, os resultados da maioria das análises estatísticas consistem numa estimativa do parâmetro juntamente com o seu erro padrão ou intervalo de confiança (IC) a 95%.

A informação de que dispomos sobre θ (quantidade de interesse, não observável), resumida probabilisticamente através de $p(\theta)$, pode ser aumentada observando-se uma quantidade aleatória X relacionada com θ . A distribuição amostral $p(x|\theta)$ define esta relação. A ideia de que após observar $X = x$ a quantidade de informação sobre θ aumenta é bastante intuitiva e o teorema de Bayes é a regra de actualização utilizada para quantificar este aumento de informação,

$$p(\theta|x) = \frac{p(\theta, x)}{p(x)} = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int p(\theta, x)d\theta}. \quad (3)$$

Se observamos, $\frac{1}{p(x)}$ não depende de θ , então funciona como função normalizadora de $p(\theta|x)$, pelo que, para um valor fixo de x , a função $L(\theta, x) = p(x|\theta)$ fornece a verossimilhança de cada um dos valores possíveis de θ quando $p(\theta)$ é distribuição a priori de θ .

Segundo Lesaffre e Lawson (2012) a teoria da inferência Bayesiana está fundamentada no teorema de Bayes (equação 3). Os autores destacam o problema geral da inferência Bayesiana, como por exemplo, o facto de a distribuição a posteriori poder ser convenientemente resumida em termos de esperanças de funções particulares do parâmetro θ (que usam integrais). O cálculo dos tais valores esperados segundo a distribuição de θ , exige uma habilidade de integrar as funções, muitas vezes complexas e multidimensionais, que é extremamente importante em inferência Bayesiana. No entanto, a inferência exata somente será possível se estas integrais puderem ser calculadas analiticamente, caso contrário devemos usar aproximações. Essas aproximações, importantes para o estudo, é o caso de Cadeias de Markov e Monte Carlo (MCMC).

2.6.1 Estimação linear Bayesiana

É claramente difícil formular alguma informação a priori de forma muito acurada, segundo Albert (2009), e pode ocorrer que seja possível especificar com alguma convicção apenas médias, variâncias e covariâncias. Entretanto, a posteriori vai depender da completa especificação da distribuição priori. O estimador linear Bayesiano de um parâmetro é um estimador cujo valor depende apenas das médias e covariâncias, sem requerer uma completa especificação da priori.

Em estatística frequentista é necessário assegurar que as inferências feitas não são indevidamente sensíveis ao modelo pressuposto, que é imposto de forma possivelmente arbitrária (Carlin et al., 1992). Isto é, as inferências devem ser robustas a aspectos da escolha de modelos, sobre os quais temos pouca convicção. Isto também ocorre no contexto Bayesiano, mas há um tópico adicional sobre robustez também à escolha da priori, a análise de sensibilidade. A distribuição posteriori

será sempre o que de facto é relevante para inferências, para qualquer especificação da distribuição priori. Entretanto, espera-se que prioris que não sejam muito dissimilares levam a posterioris que são também razoavelmente similares. Portanto, é indesejável que a análise dependa demais de uma específica escolha de priori, uma vez que a priori é, na melhor das hipóteses, apenas um resumo grosseiro das reais convicções anteriores que se possa ter.

Desenvolvimentos nas últimas décadas tem reconhecido que técnicas de simulação podem ser usadas para gerar amostras cuja distribuição é a posteriori (Albert, 2009). Isto não é simples, uma vez que as distribuições posterioris podem ser de alta dimensão, com forma analítica indisponível e complexas estruturas de dependências. Devido a tais factos, uma diversidade de técnicas têm sido desenvolvidas utilizando técnicas de MCMC. A técnica MCMC talvez mais simples e mais utilizada é o amostrador de Gibbs cujo algoritmo em sua forma básica, é descrito em 2.6.2.

2.6.2 Cadeias de Markov e Monte Carlo

Segundo Gelman et al. (2013) apud Lesaffre e Lawson (2012), MCMC são um conjunto de métodos matemáticos e computacionais que usam cadeias de Markov para gerar amostras aleatórias de distribuições de probabilidade complexas ou desconhecidas.

- Uma cadeia de Markov é um modelo que descreve um sistema que muda de estado de forma aleatória, mas seguindo certas probabilidades que dependem apenas do estado actual, e não dos estados anteriores;
- Um método de Monte Carlo é uma técnica que usa números aleatórios para aproximar quantidades de interesse, como médias, integrais, máximos, etc.
- Ao combinar cadeias de Markov e Monte Carlo, é possível criar algoritmos que exploram o espaço de estados do sistema de forma eficiente e convergem para a distribuição desejada.
- Alguns exemplos de algoritmos de MCMC são: Metropolis Hastings e amostragem de Gibbs.

Para compreender o MCMC, é necessário conhecer os princípios básicos do método de Monte Carlo. O método de Monte Carlo é utilizado para aproximar uma característica da distribuição de probabilidade de uma variável aleatória X (por exemplo, o seu valor esperado), quando não é possível calculá-la analiticamente. Com um computador, geramos uma amostra x_1, \dots, x_n de

valores independentes da distribuição de X . Em seguida, utilizamos a distribuição empírica da amostra para efectuar os nossos cálculos. A distribuição empírica é a distribuição de probabilidade discreta que atribui a probabilidade $1/n$ a cada um dos valores x_1, \dots, x_n .

Em geral, qualquer cálculo que queiramos efectuar sobre a distribuição real de X é aproximado pelo cálculo correspondente efectuado sobre a distribuição empírica. Este tipo de procedimento funciona geralmente bem, no sentido em que a aproximação converge para a quantidade verdadeira à medida que a dimensão da amostra n aumenta. Isto é conhecido como o **princípio do plug-in**.

2.6.3 Amostragem de Gibbs

Amostragem de Gibbs é uma técnica que permite gerar amostras de uma distribuição alvo, difícil de simular ou calcular, usando as distribuições condicionais da distribuição alvo. No amostrador de Gibbs a cadeia irá sempre se mover para um novo valor, isto é, não existe um mecanismo de aceitação-rejeição.

- Pode ser usada para gerar amostras da distribuição conjunta de probabilidades de duas ou mais variáveis aleatórias, usando as distribuições condicionais de cada variável.
- As amostras geradas pela amostragem de Gibbs podem ser usadas para estimar quantidades de interesse, como parâmetros, distribuições posteriores, etc.

Suponha que o vector de parâmetros de um modelo seja $\theta = (\theta_1, \theta_2, \dots, \theta_d)^T$ e deseja-se simular da distribuição conjunta $p(\theta|y)$.

- Definir a distribuição alvo, $p(\theta|y)$, onde θ é o vector de parâmetros e y é o vector de dados.
- Definir as distribuições condicionais da distribuição alvo, $p(\theta_i|\theta_{-i}, y)$, onde θ_i é o i -ésimo componente do vector de parâmetros e θ_{-i} é o vector de parâmetros sem o i -ésimo componente.
- Gerar uma amostra inicial do vector de parâmetros, $\theta^{(0)}$.
- Repetir até atingir o número desejado de amostras, ou seja, atingir a sua distribuição estacionária, que corresponde à distribuição posterior conjunta dos parâmetros.

Sob condições bastante gerais a sequência de θ 's obtidos desta maneira forma uma cadeia de Markov (dado o valor actual, o valor subsequente de θ é independente de todos os valores anteriores),

e a distribuição estacionária da cadeia é a verdadeira distribuição a posteriori (Gill, 2008).

Desta forma, em uma aplicação típica, a cadeia é inicializada, rodada por um longo período até que a amostra resultante pareça ter atingido a distribuição estacionária, e, a partir deste ponto, conhecido como aquecimento (*burn-in*¹) da cadeia, a sequência subsequente de amostras é analisada como sendo uma amostra da distribuição a posteriori. O algoritmo de Gibbs requer que seja possível simular as distribuições condicionais de cada parâmetro dadas as observações e os valores actuais dos demais parâmetros na cadeia, o que dá origem e uma diversidade de técnicas de simulação.

2.7 Modelos hierárquicos Bayesianos

Os modelos hierárquicos possuem dois ou mais níveis de hierarquia. Mesmo após o agrupamento dos níveis em um único modelo, ele ainda permite a incorporação de efeitos aleatórios associados a cada nível hierárquico. Estes efeitos nada mais são que erros aleatórios das diferenças entre as unidades de cada nível, relacionados com o distrito de interesse a ser estudada (Goldstein, 1999). Teoricamente, os níveis de um modelo hierárquico são formados por uma amostra aleatória das unidades a serem consideradas. Assim, estas estruturas de dados podem ser estudadas pelos modelos lineares hierárquicos, tornando possível especificações separadas de cada nível para posteriormente serem reunidas em um único modelo (Raudenbush & Bryk, 2002).

Para melhor entendermos o tipo de modelagem hierárquica que será realizada nesta dissertação, basear-se-á no modelo hierárquico descrito por Casella e Berger (2011).

2.7.1 Modelo Poisson-Gama

O modelo Poisson–Gama surge da combinação hierárquica de um modelo Poisson (casos, eventos), e uma distribuição Gama para modelar a heterogeneidade ou sobre-dispersão. A combinação resulta em uma distribuição binomial negativa, que é mais flexível do que a Poisson quando os dados apresentam maior variabilidade do que o previsto. O modelo assume que o número de casos segue uma distribuição Poisson, mas que o seu parâmetro de intensidade varia segundo uma distribuição Gama (Hilbe, 2011). A formulação do modelo assume dois níveis, o nível 1, que é o

¹descarta os primeiros elementos de uma amostra MCMC, devido a discrepância entre a distribuição alvo

Modelo de Poisson,

$$Y_{it} | \theta_{it} \sim \text{Poisson}(E_{it}\theta_{it}), \quad (4)$$

onde Y_i representa o número de casos observados no distrito i , e λ_i é a taxa média de incidência. e o nível 2, que assume heterogeneidade via distribuição Gama:

$$\lambda_i \sim \text{Gamma}(\alpha, \beta), \quad (5)$$

onde α e β são os parâmetros da distribuição Gamma que permitem:

$$\mathbb{E}[Y_i] = \mu, \quad \text{Var}(Y_i) = \mu + \frac{\mu^2}{r}, \quad (6)$$

permitindo variância superior à média.

A inclusão de covariáveis ocorre através de:

$$\log(\lambda_i) = \mathbf{x}_i^\top \boldsymbol{\beta}, \quad (7)$$

onde \mathbf{x}_i é o vector de covariáveis (caso de: precipitação, temperatura, humidade relativa) e $\boldsymbol{\beta}$ são os efeitos fixos.

2.7.2 Modelo Poisson-Lognormal

O modelo Poisson–Lognormal constitui uma extensão ainda mais flexível, assumindo que o parâmetro Poisson segue uma distribuição Lognormal, o que permite acomodar níveis elevados e irregulares de heterogeneidade (Aitchison & Ho, 1989). A distribuição Lognormal produz caudas mais pesadas do que a Gama, o que torna este modelo especialmente apropriado para cenários em que ocorrem valores extremos ou variações abruptas entre áreas ou períodos temporais (Ghosh et al., 2006).

A formulação do modelo segue, também, dois níveis. O primeiro nível é de Poisson

$$Y_{it} | \theta_{it} \sim \text{Poisson}(E_{it}\theta_{it}) \quad (8)$$

e o segundo, Lognormal

$$\log(\theta_{it}) = \mu + \epsilon_{it}, \quad \epsilon_{it} \sim N(0, \sigma^2), \quad (9)$$

ou equivalentemente:

$$\theta_{it} \sim \text{Lognormal}(\mu, \sigma^2). \quad (10)$$

A inclusão das covariáveis é feita através de

$$\log(\theta_{it}) = \mathbf{x}_{it}^\top \boldsymbol{\beta} + u_i, \quad (11)$$

onde $u_i \sim N(0, \sigma^2)$ representa a heterogeneidade não observada.

Covariáveis comuns incluem variáveis climáticas defasadas, variáveis demográficas e factores ambientais.

2.7.3 Modelos autoregressivos condicionais

Os modelos (Condicional Autoregressive) CAR foram introduzidos por Besag (1974) e são fundamentais para capturar dependência espacial entre distritos adjacentes. O efeito espacial de cada distrito é condicionado aos efeitos dos distritos vizinhas, reflectindo a tendência de regiões próximas exibirem padrões epidemiológicos semelhantes.

A formulação do modelo CAR segue a distribuição condicional do efeito espacial u_i , que é dado por:

$$u_i | u_{-i} \sim N\left(\frac{1}{n_i} \sum_{j \sim i} u_j, \frac{\sigma^2}{n_i}\right), \quad (12)$$

onde n_i é o número de distritos vizinhas e $j \sim i$ indica vizinhança geográfica. O termo u_i representa o efeito aleatório espacial associado ao distrito i , captando heterogeneidade espacial não observada. O termo u_i denota o conjunto de efeitos espaciais de todos os outros distritos, exceto i .

A componente espacial pode ser incorporada em modelos de risco através de:

$$\log(\lambda_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{u}_i + \mathbf{v}_i \quad (13)$$

onde:

- u_i é o efeito espacial estruturado (CAR),
- v_i é o efeito não estruturado (ruído independente).

Esta decomposição constitui a base dos modelos BYM e BYM2, amplamente usados em epidemiologia espacial.

2.7.3.1 Tipos de modelos autoregressivos condicionais

Os modelos CAR são, segundo NIMBLE (de Valpine et al., 2017) e NIMBLE (NIMBLE Development Team, 2023), amplamente utilizados para capturar dependência espacial em modelos hierárquicos Bayesianos. A principal ideia é especificar distribuições condicionais para cada área geográfica, permitindo que o valor do efeito espacial em cada região dependa dos valores observados nas regiões vizinhas. Contudo, existem duas formulações principais: o Intrinsic CAR (ICAR) e o Proper CAR, que diferem em termos de estrutura matemática e propriedades da distribuição conjunta.

Modelo ICAR (Intrinsic CAR): O modelo ICAR é definido pelas distribuições condicionais:

$$\theta_i \mid \theta_{-i}, \tau \sim \text{Normal} \left(\frac{1}{n_i} \sum_{j \in \partial i} \theta_j, \frac{1}{n_i \tau} \right), \quad (14)$$

onde ∂i representa o conjunto de vizinhos da área i , n_i o número de vizinhos e τ o parâmetro de precisão global. A matriz de precisão resultante é dada por $Q = \tau(D - W)$, onde W é a matriz de adjacência e D é uma matriz diagonal com $D_{ii} = n_i$. Neste caso, $\det(Q) = 0$, o que implica que a distribuição conjunta é **imprópria**. Isto significa que a soma das componentes espaciais é indefinida, exigindo a imposição de restrições (por exemplo, somatório igual a zero) para garantir identificabilidade. Apesar disso, o ICAR é amplamente utilizado devido à sua capacidade de produzir suavização espacial eficaz.

Modelo Proper CAR: No modelo Proper CAR, a estrutura condicional é ajustada pela introdução de um parâmetro adicional de correlação espacial ρ , garantindo que a distribuição conjunta é própria. As distribuições condicionais são dadas por:

$$\theta_i \mid \theta_{-i}, \tau, \rho \sim \text{Normal} \left(\rho \frac{1}{n_i} \sum_{j \in \partial i} \theta_j, \frac{1}{n_i \tau} \right). \quad (15)$$

A matriz de precisão conjunta torna-se $Q = \tau(D - \rho W)$, e a distribuição conjunta é própria se $|\rho| < \frac{1}{\lambda_{\max}(D^{-1/2} W D^{-1/2})}$, onde λ_{\max} é o maior valor próprio da matriz normalizada. Ao contrário do ICAR, o proper CAR possui determinante não nulo e, portanto, define uma densidade multivariada válida: $\det(Q) \neq 0$.

A escolha entre ICAR e proper CAR depende dos objetivos da análise: enquanto o ICAR é mais simples e tradicionalmente usado em mapeamento de doenças, o proper CAR oferece maior robustez estatística e controlo explícito da intensidade de dependência espacial.

2.8 Fundamentos estatísticos para a modelação epidemiológica

2.8.1 Taxa bruta

É o estimador mais simples para o risco de ocorrência de um evento, definindo-se como a razão entre o número de eventos ocorridos na área e o número de pessoas expostas à ocorrência desse evento. Um grande problema associado ao uso de taxas, porém, é a alta instabilidade que elas possuem para expressar o risco de um determinado evento quando ele é raro e a população da região de ocorrência é pequena. As variações bruscas que ocorrem com estas taxas podem nada ter a ver com o fenómeno e sim com uma variabilidade associada às observações. Flutuações aleatórias casuais, como a ocorrência de um ou dois casos do evento a mais ou a menos num distrito, causam variações substanciais nas taxas brutas se a sua população for pequena, efeito este não verificado em distritos de população grande. Além disso, para situações em que não ocorrem casos do evento em algumas regiões, a taxa bruta estima o risco de ocorrência do evento como zero, algo não real tratando-se de dados como doenças. Mapas de eventos baseados directamente nessas estimativas brutas são de difícil interpretação e frequentemente geram falsas conclusões.

Para evitar essas falsas conclusões, o estimador Bayes é aplicado e esse, inclui efeitos espaciais, calculando a estimativa localmente, utilizando somente os vizinhos geográficos da área na qual se deseja estimar a taxa, convergindo em direcção a uma média local em vez de uma média global. As taxas corrigidas são menos instáveis, pois levam em conta no seu cálculo não só a informação da área, mas também a informação de sua vizinhança. Mapas baseados nessas estimativas são mais

interpretativos e informativos.

2.8.2 Modelos lineares generalizados

Modelos lineares generalizados (GLM) estendem modelos de regressão ordinária para abranger distribuições de resposta não normais e funções de modelação da média. Três componentes especificam um modelo linear generalizado: Um componente aleatório identifica a variável de resposta Y e sua distribuição de probabilidade; um componente sistemático especifica variáveis explicativas usadas em uma função preditora linear; e uma função de ligação especifica a função de $E(Y)$ que o modelo equivale ao componente sistemático. Nelder e Wedderburn (1972), introduziram a classe de GLMs, embora muitos modelos na classe estivessem bem estabelecidos naquela época.

2.8.2.1 Regressão logística

O uso da regressão logística cresceu muito nos últimos tempos, desde a sua utilização inicial na área da epidemiologia, o método atravessou as fronteiras e é agora aplicado em diversas áreas, não se limitando apenas a uma área específica da investigação, mas evoluindo simultaneamente, no que diz respeito aos aspectos estatísticos do próprio modelo de regressão logística.

Para a aplicação da regressão logística é importante entender que se procura o melhor e o mais parcimonioso modelo estatístico para um determinado grupo de variáveis, denominadas de co-variáveis. A regressão logística é utilizada principalmente para prever variáveis dependentes binárias, mas com a variável de resposta discreta, ela não pode ser modelada diretamente por regressão linear. Portanto, ao invés de prever uma estimativa pontual do evento em si, ele constrói o modelo para prever as chances da sua ocorrência (Delen, 2011).

A selecção de modelos para regressão logística é idêntica aos modelos de regressão comum, sendo que a sua dificuldade dilata com o aumento das variáveis explicativas e em razão de produzir efeitos e interações (Agresti, 2002). A partir da análise de uma variável ganhamos a capacidade de avaliar o tipo de modelo a ser usado na regressão, podendo distinguir entre regressão linear e categórica, ou seja, na formulação do modelo há que ter em consideração a tipologia e natureza da variável resposta. Neste sentido, e perante a eventualidade de uma análise de dados quando temos uma variável dependente do tipo nominal dicotómico, a escolha correta é o modelo logístico (Hosmer & Lemeshow, 2000).

2.8.2.2 Regressão de Poisson

Uma das técnicas bastante utilizada em estudos de previsão epidemiológicos, são os Modelos Lineares Generalizados, concretamente a Regressão de Poisson. Na Regressão de Poisson, o componente aleatório é especificado pela distribuição de Poisson $y|\mu, x \sim Poisson(\mu)$, $y = 0, 1, 2, \dots$ com $\mu = exp(x^T \beta)$ para x e β dois vectores de predictores e coeficientes lineares, respectivamente.

Quando todos os predictores são discretos (categóricos), o modelo de regressão de Poisson é equivalente ao modelo *log-linear*.

Os estimadores de máxima verossimilhança para β são obtidos encontrando os valores que maximizam a log-verossimilhança

$$L(\beta) = \sum_{i=1}^n [y_i(x_i^T \beta) - exp(x_i^T \beta)] \quad (16)$$

pelo que, as estimativas são obtidas usando algoritmos iterativos como Newton-Raphson, Mínimos Quadrados ponderados iterativos (*Iterative Weighted least Squares* [IWLS]).

- Se $\beta_j > 0$, a média (μ) aumenta com x_j ;
- Se $\beta_j < 0$, a média (μ) diminui com x_j .

O aumento unitário em x_j faz variar a média (μ) por um factor $exp(\beta_j)$ (como as chances na regressão logística).

Como um dos grandes problemas de Regressão de Poisson é a super-dispersão, podemos usar um offset para ajustar as observações colectadas em diferentes janelas de tempo, espaço, etc.

2.9 Análise espacial

A análise espacial, segundo Alexandre e Vinícios (2020), consiste em métodos e técnicas que buscam identificar padrões de associação espacial (dependência e heterogeneidade espaciais) nos fenómenos socioeconómicos com o objectivo de compreender suas estruturas e dinâmicas no espaço, contribuindo para a formulação e avaliação de políticas públicas e privadas. Os seus efeitos espaciais estão directamente relacionados com a interação dos agentes através do espaço (Dependência espacial) e com a instabilidade estrutural através do espaço (Heterogeneidade espacial).

Para Almeida (2012), dados espaciais denotam a variação de algum fenômeno tendo a preocupação em determinar onde ocorre tal variação (localização). Os dados espaciais mostram a magnitude da variação do atributo do fenômeno em estudo e fornece a referência explícita em termos de localização geográfica deste atributo. Ainda para o mesmo autor, existem três tipos de dados espaciais:

- **Dados geoestatísticos**

São mais adequados para a distribuição contínua no espaço. Exemplo: pressão atmosférica, temperatura. São geralmente são poucos usados para fenômenos socioeconômicos.

- **Dados na forma de pontos**

São geralmente usados para amostras de eventos no espaço com a distribuição discreta. São amplamente usados em áreas como epidemiologia, medicina e criminologia.

- **Dados na forma de áreas (polígonos)**

Os autores Haddad et al. (2017), defendem que a análise de dados espaciais deve ser realizada de forma cuidadosa, pois pode ser sensível a problemas muito específicos:

- **Falácia ecológica**

Que é a inferência ecológica sobre o comportamento do indivíduo a partir da análise de dados agregados;

- **Problema da unidade de área modificável**

Ou seja, os resultados de dependência espacial são influenciados pelo número de unidades espaciais, e mesmo quando as escalas espaciais são constantes, os resultados são afetados pelas várias formas de fazer combinações das unidades espaciais.

- **Efeito de proximidade espacial**

O que significa que as observações próximas da fronteira podem estar correlacionadas espacialmente com regiões que não fazem parte da área de estudo.

- **Influência de valores atípicos espaciais**

Observações espaciais podem apresentar uma dependência espacial distinta do restante das outras observações vizinhas.

Uma solução proposta para os problemas de sensibilidades, é a definição de matrizes de ponderação espacial.

Uma matriz de ponderação espacial é uma matriz de dimensão $n \times n$. Reflete o arranjo das interações espaciais por meio de pesos espaciais $w_{(ij)}$ que representam o **grau de conexão** entre as regiões segundo algum critério de proximidade.

2.9.1 Estrutura de vizinhança espacial

Para os autores Haddad et al. (2017), na análise espacial, a definição da matriz de vizinhança é um elemento fundamental, pois descreve as relações de proximidade entre as unidades geográficas. Define-se pela matriz baseada na distância geográfica que tem-se conectividade equilibrada (mesmo número de vizinhos para todas as unidades espaciais).

$$w_{ij}(k) = \begin{cases} 1 & d_{ij} \geq d_i(k) \\ 0 & d_{ij} < d_i(k) \end{cases} \quad (17)$$

em que $d_i(k)$ é uma distância de corte tal que todas as regiões tenham o mesmo número de vizinhos (k). A métrica da distância d_{ij} é também chamada de distância euclidiana na recta, tal que;

$$d_{ij} = \sqrt{(u_i - v_i)^2 + (u_j - v_j)^2} \quad (18)$$

em que u_i é a ordenada e v_i é a abcissa, projectadas no mapa. Por outro lado, a distância euclidiana no arco será dada por:

$$d_{ij} = R \cos^{-1}[\sin\theta_1 \sin\theta_2 + \cos\theta_1 \cos\theta_2 \cos(k_1 - k_2)] \quad (19)$$

em que R é o raio da Terra em torno do Equador (6378km), θ é a latitude e k é a longitude. A normalização da matriz w gera matriz de pesos espaciais utilizada na análise espacial. Essa normalização é formulada como:

$$w_{ij}^* = \frac{w_{ij}}{\sum_j w_{ij}} \quad (20)$$

Observa que a normalização na linha da matriz de pesos espaciais pode produzir assimetria se o número de vizinhos é diferente para cada localização, por isso que a região i exerce uma influência distinta sobre a região j do que esta sobre a primeira região ($w_{ij}^* \neq w_{ji}^*$).

Existem diferentes critérios para a definição da vizinhança, entre os quais se destacam:

- **Contiguidade (Queen ou Rook):** duas áreas são consideradas vizinhas se partilham fronteira (Rook) ou fronteira/ponto (Queen); A matriz de vizinhança espacial $W = (w_{ij})$ é definida segundo o critério de contiguidade do tipo Queen, tal que:

$$w_{ij} = \begin{cases} 1, & \text{se } \partial A_i \cap \partial A_j \neq \emptyset \\ 0, & \text{caso contrário} \end{cases}$$

- **K-vizinhos mais próximos (KNN):** cada área é considerada vizinha das K áreas geograficamente mais próximas, com base na distância;
- **Distância fixa:** áreas são consideradas vizinhas se estiverem dentro de um determinado raio.

Neste estudo, optou-se pelo critério de contiguidade do tipo Queen, por ser mais adequado para dados administrativos, uma vez que preserva a estrutura geográfica real das interações espaciais entre distritos.

2.9.2 Autocorrelação espacial global univariada

O primeiro passo num estudo de análise exploratória de dados espaciais é testar a hipótese de que os dados espaciais sejam distribuídos aleatoriamente, ou seja, que os valores de um atributo numa região não dependem dos valores deste atributo nas regiões vizinhas.

Moran (1950) propôs a elaboração de um coeficiente de autocorrelação espacial, usando a medida de autocovariância na forma de produto cruzado. Coeficiente I , segundo Moran (1950), para uma variável aleatória é dado por:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{(\sum_{i=1}^n \sum_{j=1}^n w_{ij}) \sum_{j=1}^n (y_i - \bar{y})^2} \quad (21)$$

Uma forma de visualizar a autocorrelação espacial é baseada no diagrama de dispersão de Moran. Esse diagrama representa a relação linear entre Wz e z tal como em um modelo de regressão simples:

$$Wz = \alpha + \beta z + \epsilon \quad (22)$$

Onde o coeficiente angular deste modelo é equivalente ao I-Morano, ou seja, $\hat{\beta} = I = \frac{z'Wz}{z'z}$.

Se o coeficiente angular da recta for positivo, há evidências de que a autocorrelação espacial é positiva, caso contrário é negativa.

2.9.3 Autocorrelação espacial global bi-variada

O objectivo é verificar se os valores de uma variável observada numa dada região guardam uma associação com os valores de outra variável observada em regiões vizinhas. Aqui, o coeficiente I de Moran para autocorrelação de duas variáveis aleatórias padronizadas:

$$I_{kl} = \frac{z'_k W z_l}{z'_k z} \quad (23)$$

em que $z_i = \frac{y_i - \bar{y}}{\sigma_i}$ para todo $i = k, l$.

O diagrama de dispersão de Moran no caso bi-variado é baseado no modelo de regressão simples abaixo:

$$W z_2 = \alpha + \beta z_1 + \epsilon \quad (24)$$

Em que o coeficiente angular é equivalente ao I-Moran para o caso bi-variado, ou seja, $\hat{\beta} = I^{z^1 z^2} = \frac{z'_1 W z_2}{z'_1 z_2}$.

3 METODOLOGIA

3.1 Material

3.1.1 Descrição do local de estudo

No estudo utiliza-se dados de 154 distritos distribuídos por 11 províncias de Moçambique, em três regiões, Norte, Centro e Sul do país (Figura 3). A unidade espacial de análise corresponde aos distritos administrativos do país, cuja delimitação geográfica foi obtida a partir de ficheiros vectoriais (shapefile). O shapefile utilizado contém os limites administrativos dos distritos de Moçambique e foi obtido a partir do portal de geotecnologia focado em recursos naturais, *Geographic Information System* (GIS)². Oficialmente designado República de Moçambique, Moçambique fica localizado na costa oriental de África Austral, limitado a norte pela Tanzânia, a noroeste pelo Malawi e Zâmbia, a oeste pelo Zimbábwe, a leste pelo canal de Moçambique e Oceano Índico, e a sul e sudeste pela África do Sul e Eswatini. O país tem uma área total de 81590 km^2 e uma população estimada em mais de vinte e sete milhões de habitantes. Caracteriza-se por ser um país com rios geralmente sazonais, com cheias durante a época das cheias e baixos níveis de água durante a época seca.

3.1.2 População e fonte de dados

A população do presente estudo é composta por dados documentados e registados na base de dados do MISAU sobre casos e óbitos por malária no país durante o período de 2017 a 2023. A fonte de colecta de dados é o MISAU, através da disponibilização e autorização da base de dados que contém as variáveis de estudo pelo Programa Nacional de Controle de Malária (PNCM). Os dados climáticos (Tabela 3) utilizados neste estudo foram adquiridos através do Centro Europeu de previsões meteorológicas a médio prazo (*European Centre for Medium-Range Weather Forecasts* [ECMWF]). Ao seu todo, a base de dados é composta por 1078 observações e 10 variáveis que são descritas nas Tabelas 2 e 3.

Encontradas as taxas de detecção, após a utilização do modelo, as suas distribuições são apresentadas em mapa (Figura 3), que em muitas das aplicações, porém, em vez de observações associadas a pontos com exata posição de ocorrência do evento, os dados estão dispostos a partir de contagens por áreas, agregados por regiões geograficamente definidas, como regiões administrativas ou sanitárias. Essa agregação dos dados pode ocorrer por conveniência ou pode simplesmente reflectir a

²<https://forest-gis.com>

forma como os dados foram disponibilizados. Em caso de dados agregados, a análise de dispersão espacial do risco de ocorrência de um acontecimento normalmente é feita por via de mapas de índices ou taxas de incidência, onde as áreas são sombreadas de acordo com os valores desta taxa.

3.1.3 Aspectos éticos

Este estudo observa os princípios éticos fundamentais aplicáveis à investigação científica, particularmente no âmbito da saúde pública. A modelação dos dados foi feita a partir de bases de dados secundárias já existentes, provenientes de fontes oficiais, como o MISAU. Por esta razão, não houve contacto directo com seres humanos nem recolha de informações identificáveis, garantindo-se o anonimato e a confidencialidade dos dados utilizados. No que respeita à autoria, o trabalho foi desenvolvido unicamente pelo autor no âmbito da conclusão do curso de mestrado, respeitando rigorosamente os critérios de integridade académica. Todas as as informações dos outros investigadores ou trabalhos, foram devidamente citados, assegurando a transparência e o respeito pelas normas de coautoria científica. Quanto ao financiamento, esta dissertação foi realizada sem patrocínio externo directo. Todos os custos associados à pesquisa foram assumidos pelo próprio autor, não havendo, portanto, conflito de interesses a declarar. Por fim, os resultados obtidos foram utilizados exclusivamente para fins académicos e científicos. Nas publicações em revistas ou em eventos, foram assegurados o cumprimento dos critérios éticos de divulgação e o respeito pela origem e pela confidencialidade dos dados utilizados.

3.2 Métodos

Todo o processamento de dados deste estudo, fora de alguns gráficos, foi realizado utilizando o pacote estatístico R, Ihaka e Gentleman (1996), que é um sistema para computação estatística e gráfica, distribuído sobre licença geral pública e aberta. Os modelos são aplicados utilizando o método de simulação de Gibbs através da função JAGS (*Just Another Gibbs Sampler*) e o pacote

Tabela 2: Descrição das variáveis do estudo.

Variáveis	Descrição	Classificação
Ano	Período anual em que se observou o caso de malária	Discreta
Província	Província da ocorrência do caso de malária	Nominal
Distrito	Distrito da ocorrência do caso de malária	Nominal
Casos de Malária	Reportadas pelo serviço de saúde local	Discreta
Óbitos por Malária	Casos de óbitos por malária por distrito	Discreto



Figura 3: Divisão administrativa e localização de Moçambique;
 Fonte: África Turismo, 2014

Tabela 3: Descrição das variáveis de clima.

Variáveis	Descrição	Classificação
t2m	Temperatura média mensal	Contínua
tp	Quant. total de água na forma de chuva ou outro tipo	Contínua
Latitude	Posição que relaciona a Terra e o equador	Contínua
Longitude	Posição que relaciona a Terra e meridiano de Greenwich	Contínua
Runoff	Escoamento superficial da água da chuva estagnada	Contínua

rjags. O pacote *rjags* implementados usando o software acima referido na sua versão 4.5.0.

3.2.1 Distribuição Poisson-Lognormal não misto

Este modelo é usado para estimar a tendência temporal global. Mais especificamente considere Y_{it} e E_{it} sejam o número de casos ou óbitos e contagens esperadas respectivamente, no i -ésimo distrito e no t -ésimo ano, e denotado por μ_{it} o valor esperado de Y_{it} , então o modelo de regressão de Poisson (que denominamos de “Modelo 1”) é definido como:

$$Y_{it} \sim \text{Poisson}(\mu_{it}) \quad (25)$$

$$\log(\mu_{it}) = \log(E_{it}) + \alpha_0 + \alpha_1 t + \alpha_2 t^2 + x_i \beta \quad (26)$$

onde α_0 representa risco de casos ou mortalidade relativo logarítmico de base em todos os distritos, e α_1 e α_2 , representam a evolução global de casos ou da taxa de mortalidade relativa logarítmica, β é o vector de parâmetros das covariáveis, e x_i corresponde à i -ésima linha da matriz de covariáveis. No entanto, o problema do Modelo 1 aqui definido é estimar a tendência temporal (t) global e assumir que é mesma para todos os distritos.

3.2.2 Distribuição Poisson-Lognormal – modelo misto com tendência global e específica

No modelo de Poisson-Lognormal misto, o parâmetro μ_{it} é modelado como uma mistura de dois modelos alternativos, um modelo comum (global) e um modelo específico de distrito. O modelo comum assume uma tendência temporal global para todos os distritos, enquanto o modelo específico de distrito estima uma tendência temporal para cada uma delas. Este modelo permitiu distinguir entre distrito que seguem a tendência de tempo esperada e aquelas que apresentam comportamento não comum. Seja Y_{it} o número de casos ou óbitos por malária no distrito i no tempo t , assume-se,

no primeiro nível da verossimilhança

$$Y_{it} \sim \text{Poisson}(\mu_{it}E_{it})$$

e no segundo nível, assume-se o modelo linear

$$\log(\mu_{it}) = \log(E_{it}) + \eta_{it}$$

em que E_{it} representa a população (offset).

A decomposição do efeito linear, do segundo nível, da hierarquia modela μ_{it} como (Boulieri et al., 2020):

$$\log(\mu_{it}) = \kappa_i \mu_{it}^{((C))} + (1 - \kappa_i) \mu_{it}^{((AS))} \quad (27)$$

$$\mu_{it}^{((C))} = \alpha_0 + \eta_i + \gamma_t \quad (\text{componente global})$$

$$\mu_{it}^{((AS))} = \nu_i + \Psi_{(i_t)} \quad (\text{componente específica do distrito})$$

e $\kappa_i \sim \text{Bernoulli}(\phi_{it})$ é o parâmetro de mistura que selecciona entre a estimativa $\mu_{it}^{((C))}$ da componente comum e a estimativa $\mu_{it}^{((AS))}$ da componente específica do distrito. O modelo comum segue uma abordagem padrão de mapeamento de doenças, onde o risco relativo consiste em um componente espacial global η_i e um componente temporal global γ_t . O componente espacial η_i recebe uma priori de convolução, amplamente conhecida como BYM, proposta por Besag (1974). Trata-se de uma priori gaussiana, $\eta_i \sim N(\nu_i, \sigma_\eta^2)$, onde ν_i é o termo da estrutura espacial que segue um Prior Autorregressivo Condicional Impróprio (Intrinsic Conditional Autoregressive prior – ICAR)

$$\nu_i | \nu_{j_{ji}} \sim \text{Normal}\left(\frac{1}{n_i} \sum_{j=1}^N w_{ij} \nu_j, \frac{\sigma_\nu^2}{n_i}\right) \quad (28)$$

onde, w_{ij} representa os elementos da matriz de adjacência W (com $w_{ij} = 1$ se os distrito i e j partilharem fronteiras, e $w_{ij} = 0$ caso contrário) e n_i é o número de vizinhos do distrito i , analogamente o componente temporal γ_t segue um prior de passeio aleatório Gaussiano de ordem 1 (RW1), que é o análogo temporal do prior ICAR, e α_0 é i intercepto global. Para terminar a especificação dos modelos 27 a 30, o parâmetro de mistura κ_i é estipulado a priori como: $\kappa_i \sim \text{Bernoulli}(0.95)$, supondo que, inicialmente, cerca de 5% dos distritos são invulgares. A média

posterior de κ_i , reflecte a probabilidade posterior de que distrito i seja comum ou incomum durante todo o período temporal. Em suma, os prioris do modelo são definidos no terceiro nível como $\alpha_0 \sim \mathcal{N}(0, 1000)$, $\eta_i \sim \text{CAR}(\tau_\eta)$, $\nu_i \sim \mathcal{N}(0, \tau_\nu^{-1})$, $\gamma_t \sim \mathcal{N}(0, \tau_\gamma^{-1})$, $\delta_t \sim \mathcal{N}(0, \tau_\delta^{-1})$, $\psi_{it} \sim \mathcal{N}(0, \tau_\psi^{-1})$, $\pi_i \sim \mathcal{N}(0, \tau_\pi^{-1})$

e os hiperparâmetros: $\tau_\eta, \tau_\nu, \tau_\gamma, \tau_\delta, \tau_\psi, \tau_\pi \sim \text{Gamma}(0.001, 0.001)$ sob o Posterior: $p(\Theta|Y) \propto p(Y|\mu) \cdot p(\Theta)$.

3.2.3 Distribuição Poisson-Lognormal – modelo misto com tendência especial e temporal

Os modelos mais usuais são os denominados CAR (*Conditional Auto Regressive*), onde se inclui na regressão linear um efeito aleatório condicionado pela vizinhança. Dada a flutuação aleatória contagens de casos e óbitos por malária, o modelo mais adequado considera simultaneamente a estimativa dos parâmetros da regressão e a estabilização dos indicadores, modelos estes usualmente ajustados utilizando ferramentas desenvolvidas no campo da inferência Bayesiana.

A outra classe de modelo de Poisson-Lognormal é o misto, que permite que o parâmetro de mistura κ_i , lide com series temporais longas passa a ser κ_{it} permitindo que a probabilidade do distrito i ser comum ou incomum mude ao longo do tempo. Além disso, a média posterior κ_{it} fornece informações sobre onde e quando é observado comportamento incomum. Estes dois pressupostos do parâmetro κ_{it} são o diferencial para o estudo em causa, a presença da modelagem de tempo e espaço em simultâneo. O parâmetro κ_{it} segue um modelo hierárquico (Boulieri, 2020):

$$\kappa_{it} \sim \text{Bern}(\phi_{it}), \quad (29)$$

onde

$$\text{logit}(\phi_{it}) = \pi_i + \delta_t + \text{logit}\left(\frac{\tau}{1-\tau}\right), \quad (30)$$

com

$$\tau \sim U(0.9, 1), \quad (31)$$

o componente espacial π_i é atribuído a um ICAR, $\pi_i \sim \text{ICAR}(W, \sigma_\pi^2)$ e o componente temporal δ_t é atribuído ao temporal do ICAR, $\delta_t \sim \text{ICAR}(Q, \sigma_\delta^2)$, em que W e Q são as matrizes de adjacência espacial e temporal, respectivamente, e $\text{logit}\left(\frac{\tau}{1-\tau}\right)$. A especificação acima, contraposta limitação e, assim, temos em conta as dependências das observações incomuns no espaço e no tempo.

Esta versão do modelo constitui uma extensão hierárquica do modelo Poisson–AR(1) apresentado anteriormente. O objectivo é capturar possíveis comportamentos anómalas (surtos) na incidência de malária, introduzindo um mecanismo de mistura que distingue dois regimes latentes no processo gerador dos dados.

O primeiro regime, denominado regime normal, representa os casos esperados segundo os efeitos espaciais, temporais e climáticos observados. O segundo, denominado regime de surto, modela situações excepcionais em que a incidência de casos é anormalmente elevada, sendo controlado por um componente estocástico adicional.

Dessa forma, o modelo incorpora uma estrutura de mistura Bernoulli, permitindo identificar padrões espaço-temporais irregulares e avaliar a probabilidade de ocorrência de surtos locais de malária. A formulação geral foi, anteriormente dada por:

$$y_{it} \sim \text{Poisson}(\mu_{it}), \text{ fazendo } \log(\mu_{it}) = \log(E_{it}) + \eta_{it} \quad (32)$$

$$\eta_{it} = \beta_0 + \beta_{\text{runoff}} x_{\text{runoff},it} + \beta_{\text{t2m}} x_{\text{t2m},it} + \beta_{\text{sp}} x_{\text{tp},it} + u_{it} \quad (33)$$

$$u_{it} = \kappa_{it} v_{it} + (1 - \kappa_{it}) w_{it} \quad (34)$$

$$\kappa_{it} \sim \text{Bernoulli}(\phi_{it}), \quad \text{logit}(\phi_{it}) = \pi_i + \delta_t \quad (35)$$

$$w_{it} \sim N(0, \tau_w^{-1}), \quad v_{it} \sim N(0, \tau_v^{-1}) \quad (36)$$

onde y_{it} representa o número de casos de malária no distrito i e tempo t ; μ_{it} é o valor esperado ajustado pelo offset populacional E_{it} ; κ_{it} é a variável indicadora que define o regime (*normal* ou *surto*); ϕ_{it} é a probabilidade de estar no regime de surto, modelada pelos efeitos espaciais π_i e temporais δ_t e τ_v e τ_w são os parâmetros de precisão associados aos efeitos aleatórios dos dois regimes. Nos modelos de contagem utilizados, considerou-se a população em risco como termo de exposição. Assim, foi incorporado no modelo um termo de offset, definido como o logaritmo da população de cada distrito e período temporal, isto é, $\log(E_{it}) = \log(\text{população}_{it})$. Esta abordagem permite modelar taxas de casos e óbito por malária em vez de contagens absolutas, tornando comparáveis os riscos entre distritos com diferentes tamanhos populacionais.

3.2.4 Critérios de selecção de modelos

A avaliação e comparação de modelos estatísticos pode ser feita através de diversos critérios amplamente utilizados na literatura. Entre os mais conhecidos destacam-se o Akaike Information Criterion (AIC), o Bayesian Information Criterion (BIC), o Deviance Information Criterion (DIC) e, em contextos Bayesianos modernos, medidas totalmente baseadas na distribuição preditiva, como o Watanabe–Akaike Information Criterion (WAIC). Cada uma destas abordagens possui vantagens e limitações, sobretudo quando aplicada a modelos hierárquicos ou espaço-temporais, nos quais a estrutura de dependência pode tornar alguns critérios menos adequados. Neste trabalho, opta-se por utilizar o WAIC, proposto por Watanabe (2010), por ser um critério totalmente Bayesiano, válido mesmo para modelos complexos, hierárquicos e com distribuições não regulares. O WAIC baseia-se na previsão fora da amostra, utilizando a expectativa sobre a distribuição posterior dos parâmetros, o que o torna mais robusto do que critérios como o DIC em cenários onde a posterior pode ser multimodal ou não aproximadamente normal.

O WAIC, avalia o poder preditivo de um modelo Bayesiano, ajustando uma penalização pela complexidade do modelo, com base na variabilidade da log-verossimilhança a posteriori. É uma forma de estimar o erro de previsão fora da amostra (*out-of-sample predictive accuracy*), similar ao que o cross-validation faz. Para o trabalho usa-se usaremos somente o $WAIC_2$, definido como:

$$WAIC_2 = -2\log\hat{p}(\kappa) + 2\rho(WAIC)_2 \quad (37)$$

onde:

$$\hat{p}(\kappa) = \prod_{i=1}^n \frac{1}{m} \sum_{l=1}^m L_i(\theta^{(l)}). \quad (38)$$

Aquele que apresentar o menor valor do WAIC é escolhido como o melhor modelo. Uma das principais vantagens práticas do WAIC é que a sua aplicação depende apenas dos valores da verossimilhança avaliada ao longo das amostras da cadeia MCMC. Assim, depois da convergência da cadeia estar estabelecida, o cálculo do WAIC torna-se direto, bastando armazenar a verossimilhança ponto a ponto. Esta característica torna o WAIC particularmente adequado para modelos Bayesianos ajustados por simulação, dispensando aproximações assintóticas como as utilizadas no AIC ou BIC. Para detalhes teóricos adicionais sobre a derivação e propriedades do WAIC, consulte Watanabe (2010) e Gelman et al. (2013), que discutem a sua relação com validação cruzada e descrevem o seu bom desempenho em modelos hierárquicos e espaço-temporais.

4 RESULTADOS E DISCUSSÃO

4.1 Análise descritiva

A Tabela 4 apresenta as principais medidas descritivas dos casos e óbitos anuais de malária no período de 2017 a 2023. Observa-se um crescimento progressivo na média dos casos de malária, que passou de aproximadamente 64.672 casos em 2017 para 80.841 em 2023. De modo semelhante, a mediana evoluiu de 56.514 para 64.568 no mesmo intervalo temporal, indicando que a distribuição dos casos se deslocou para valores mais elevados ao longo dos anos. A moda (valor mais frequente) também apresentou aumento expressivo, subindo de 124.239 em 2017 para 190.709 em 2023. Esses resultados sugerem uma tendência geral de aumento da incidência média de malária no período analisado.

No que se refere à dispersão dos dados, tanto a variância quanto o desvio-padrão exibem valores elevados e, em geral, crescentes até 2022. Esse comportamento indica que a distribuição dos casos se tornou mais heterogênea, reflectindo uma maior variabilidade espacial e temporal da ocorrência da doença. Destaca-se que em 2020 e 2022 as variâncias ultrapassam (5×10^9), com desvios-padrão superiores a 70.000 casos, o que demonstra forte amplitude entre as observações. Os valores máximos acompanham a tendência de crescimento observada nas medidas centrais, passando de 355.018 em 2017 para 634.408 em 2022, com ligeira redução em 2023 (519.176). Essa diminuição pontual pode estar relacionada a factores ambientais menos favoráveis à proliferação do vector, a acções de controle mais eficazes ou a oscilações naturais no ciclo de transmissão da doença.

Tabela 4: Estatísticas descritivas anuais de casos e óbitos por malária.

Ano	Casos				Óbitos			
	Mínimo	Moda	Média	Desvio padrão	Mínimo	Moda	Média	Desvio padrão
2017	1058	56514	64672.1	46743.4	0	3	7.25	13.9
2018	1070	61778	67553.3	54898.5	0	2	6.34	15.2
2019	946	61550	71346.4	59369.5	0	1	4.77	11.2
2020	530	61145	74117.5	71564.5	0	1	3.68	9.29
2021	0	47170	66076.6	69209.6	0	1	2.66	7.05
2022	853	64272	81173.6	78056.6	0	0	2.75	8.48
2023	724	64568	80841.6	72378.5	0	0	2.17	7.76

Por outro lado, na mesma Tabela 4 observa-se uma tendência geral de redução tanto na média quanto nos valores máximos de óbitos ao longo do período analisado, contrastando com o aumento dos casos de malária. Embora os casos de malária tenham aumentado no mesmo período, o número de óbitos mostra tendência inversa. Essa divergência entre incidência e mortalidade indica que, apesar da expansão da transmissão, o controle clínico e a resposta do sistema de saúde melhoraram substancialmente. Assim, a elevação dos casos pode estar mais associada a maior capacidade de detecção e vigilância epidemiológica do que necessariamente a um agravamento do quadro sanitário.

4.1.1 Distribuição geográfica dos casos e óbitos por malária

Os mapas da Figura 4 e 4 apresentam a distribuição espacial do número de casos de malária em Moçambique, desagregado por distrito e representado através de uma escala cromática. A intensidade da cor revela uma considerável heterogeneidade na distribuição de doenças ao longo dos distritos. Os mapas 4 e 4 são apresentados utilizando uma escala categorizada em quatro intervalos de intensidade, correspondentes ao número de casos por distrito, expressos em unidades de 10^5 (centenas de milhares). As classes utilizadas foram: $0-1 \times 10^5$, $1-3 \times 10^5$, $3-5 \times 10^5$ e valores superiores a 5×10^5 casos. Esta classificação permite evidenciar diferenças espaciais relevantes na carga de malária, facilitando a comparação entre distritos e entre os diferentes anos analisados.

Observando-se as figuras, evidencia-se uma forte heterogeneidade geográfica na incidência da malária. No entanto, evidencia-se as províncias do centro e norte como as mais afectadas, com especial ênfase para áreas em torno da Sofala, Manica, Zambézia, Nampula e Cabo Delgado, que apresentam contaminações mais elevadas ao longo dos sete anos de análise. Em termos concretos, a concentração alarmante da doença está nos distritos de Nhamatanda, Mafambisse, Sussundenga, Distrito de Manica, Cidade de Chimoio, Alto Molocue, Morrumbala, Mecubuli e Lugela, que apresentam os maiores números absolutos de casos, ultrapassando 450.000 registos no período de estudo.

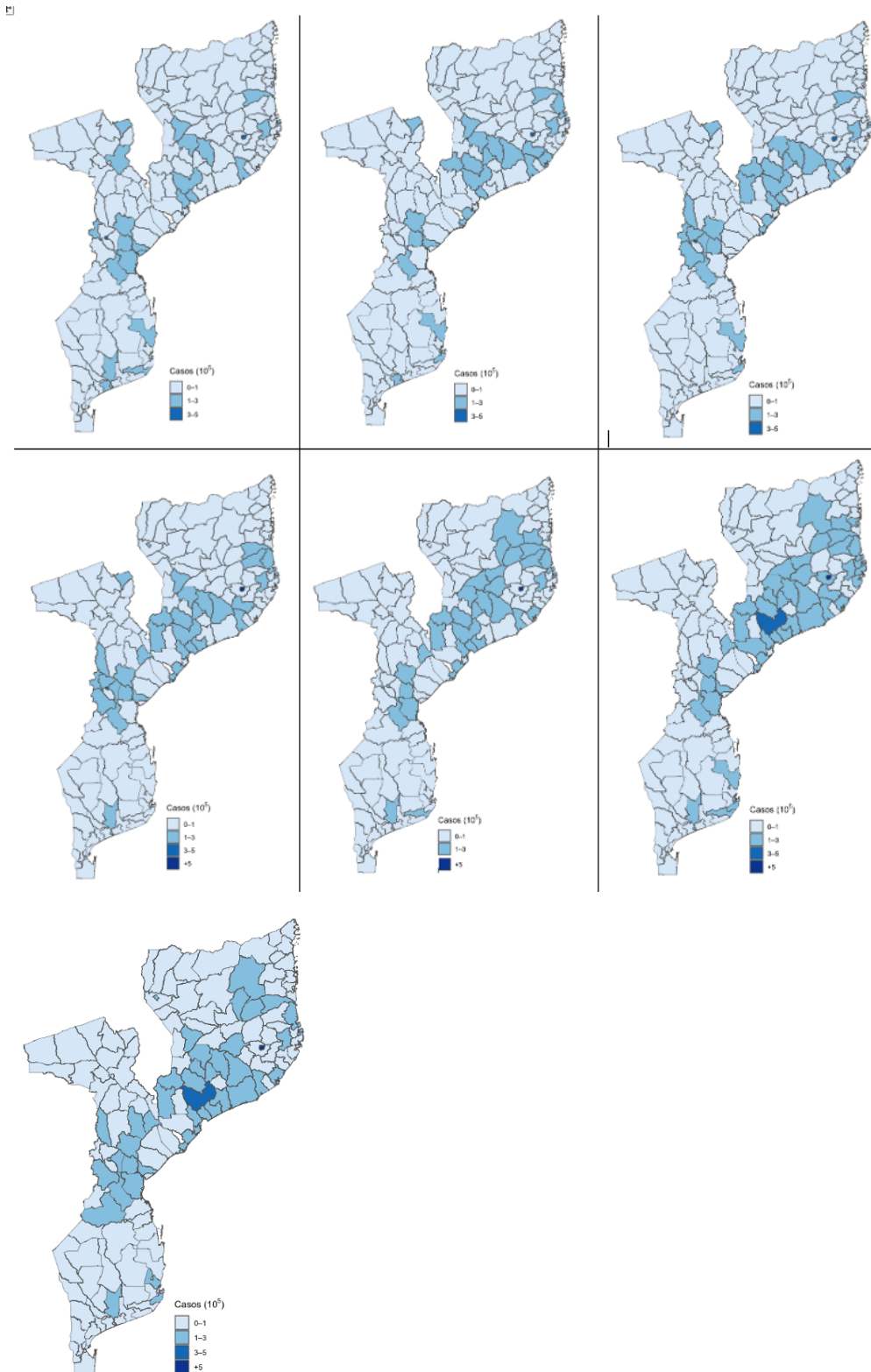


Figura 4: Distribuição dos casos de malária no período de 2017 a 2023, apresentados da esquerda para a direita e de cima para baixo. Cada painel representa um ano distinto, permitindo observar a evolução temporal e a variação geográfica da incidência da doença ao longo destes sete anos.

Simultaneamente, os distritos de Maganja da Costa, Pebane e Inhassunge demonstram a maior propensão (incidência per capita), indicando uma carga desproporcional da doença considerando

o tamanho populacional destas regiões. Estes territórios, pelas suas condições climáticas quentes e húmidas (vide o mapa da Figura 7 e 8 nos apêndices) em que os distritos tiveram picos de temperaturas médias altas nos anos de 2017, 2019, 2020 e 2022), favorecem a proliferação do mosquito vector *Anopheles*, o que explica parte dessa elevada carga de doença. Por outro lado, as províncias de Maputo, Gaza e Inhambane, apresentam uma carga de casos inferior. No entanto é notável os distritos de Massingir, Malehice e Chidenguele com uma tendencia de contaminação por malária. O menor número de casos de Malária na zona sul, evidencia um padrão que pode estar relacionado com a maior urbanização, melhores infraestruturas de saúde e condições climáticas menos favoráveis à transmissão da malária. Esses factores serão analisados nos próximos subcapítulos. O maior número de distritos nas províncias de Nampula, Sofala e Tete apresentam valores intermédios, reflectindo uma carga epidemiológica significativa, ainda que não tão elevada quanto a observada em outras províncias. Em síntese, o mapa sugere que a malária permanece um problema de saúde pública com maior impacto no centro e norte de Moçambique, enquanto o sul apresenta uma situação epidemiológica comparativamente mais favorável.

Por outro lado, os mapas apresentados no gráfico 5 e 5 mostra os distritos do centro e norte do país com a maior concentração da carga de mortalidade, destacando-se, em especial, a província da Zambézia e Niassa.

Embora o número absoluto de mortes seja, em geral, muito mais reduzido do que o de infecções, algumas áreas do centro e norte do país concentram valores relativamente mais elevados, particularmente na província da Zambézia (como Mocuba e Gurue) e em partes de Manica e Niassa, com valores de 50 ou mais óbitos. Em contraste, os distritos do sul, ainda que registem casos de malária, mostra mortalidade residual, reflectindo desigualdades regionais não apenas na transmissão, mas também na resposta do sistema de saúde. Um facto, é o distrito de Liupo, em Nampula, que, ao longo dos sete anos, apresentou casos de óbitos quase acima de 70, por ano.

De forma geral, isso indica que existem regiões que constituem áreas críticas, onde a elevada transmissão se associa a uma maior letalidade, o que pode estar relacionado com factores como a dificuldade de acesso aos serviços de saúde, limitações no diagnóstico precoce e escassez de infraestruturas hospitalares. No entanto, a comparação mostra também que nem sempre os distritos com muitos casos apresentam proporcionalmente mais óbitos. Em partes de Nampula e Cabo Delgado, por exemplo, a incidência é bastante alta, mas a mortalidade registada é relativamente baixa. Este contraste sugere que, apesar da elevada transmissão, existe maior capacidade de res-

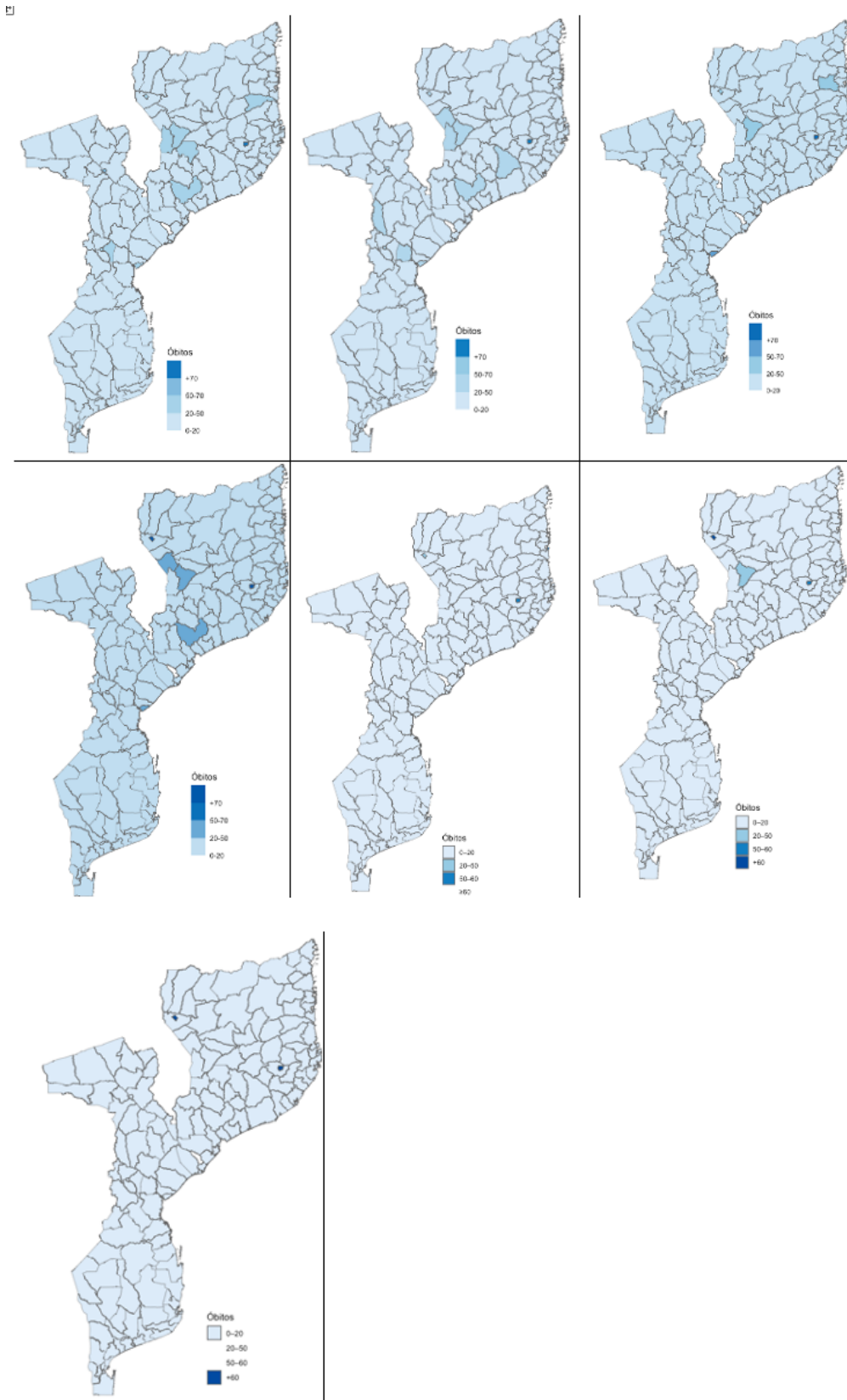


Figura 5: Distribuição óbitos por malária no período de 2017 a 2023, apresentados da esquerda para a direita e de cima para baixo. Cada painel representa um ano distinto, permitindo observar a evolução temporal e a variação geográfica da taxa de mortalidade ao longo destes sete anos.

posta dos serviços de saúde locais, conseguindo evitar a progressão da doença para formas graves e, conseqüentemente, reduzir o número de mortes.

Por outro lado, as figuras dos apêndices A8 e A9 das variáveis climáticas padronizadas, revela padrões sazonais críticos para a compreensão da dinâmica de transmissão da malária. Evidencia-se alguma sincronia temporal entre picos de incidência de malária e variáveis climáticas, sugerindo, deste modo, forte influência ambiental na transmissão da doença e por conseguinte, mortes. Ao mesmo tempo, existem picos coincidentes que indicam que condições climáticas específicas (no caso, temperatura e precipitação) precedem o aumento de casos. Como pode se observar, durante o período de análise, com a exceção do ano 2020, houve uma elevada onda de acúmulo de água em poças, valas, lagoas temporárias e coleções de água parada, que funcionam como criadouros ideais para o *Anopheles*. É evidente também que esse acúmulo foi mais favorecido na zona centro e norte. Tendo, na zona sul, província de Gaza e Inhamabene ser afectado no ano 2023, em particular alguns distritos de Maputo cidade e Província em todos os anos. A chuva por si só não basta para assumir-se a criação plena do mosquito causador da Malária; é necessário que a temperatura esteja dentro de um intervalo óptimo (cerca de 20–30 °C) para permitir o desenvolvimento do parasita *Plasmodium* dentro do mosquito. Nas figuras do apêndice 8 observa-se uma significativa heterogeneidade na expressão sazonal entre os distritos. No entanto, em média, a temperatura está no intervalo acima indicado. Neste caso, a precipitação aumenta a umidade relativa do ar, favorecendo a sobrevivência do mosquito adulto.

4.1.2 Análise da autocorrelação espacial

Os resultados que constam na Tabela 5 referem à Estatística de Moran I que indicam uma autocorrelação espacial positiva moderada, ou seja, distritos com muitos casos estão próximos de outros distritos com muitos casos, e o mesmo para distritos com poucos casos. O valor esperado de Moran's I sob hipótese nula da distribuição aleatória dos casos, confirma que a comparação faz sentido, pois o valor está próximo de zero (-0.007). Por outro lado, o desvio padrão (z-score) = 5.719, que está bem alto (> 1.96), indica forte evidência contra a hipótese nula.

Para verificar a tendência local, analisou-se as estatísticas de Moran, em que mediu-se a autocorrelação espacial local, ou seja, identificou-se distritos onde os valores (casos de malária) estão agrupados de forma significativa, seja em clusters de alto ou clusters de baixo. A Tabela 6 mostra distritos que estão em grupos de alto risco de malária (distritos com casos e óbitos altos cercados

Tabela 5: Valores do teste de Moran's I para autocorrelação espacial de casos e óbitos por malária em Moçambique.

Estatística / Informação	casos de malária	óbitos por malária
Moran's I (observado)	0.287	0.0177
Expectativa sob aleatorização	-0.007	-0.007
Variância (sob H ₀)	0.03	0.0020
Desvio padrão (z)	5.719	0.047
p-valor (teste: <i>greater</i>)	5.373×10^{-9}	0.297
Hipótese alternativa	(positiva)	(negativa)

Notas: Resultado do teste Moran I calculado com método de aleatorização.

de vizinhos também altos). Para o Moran I negativo, sugere-se grupos espaciais, ou seja, distritos com valores muito diferentes dos seus vizinhos mas, que podem estar cercados em um distrito com casos baixos cercado por distritos com casos altos.

Por outro lado, na mesma Tabela 5, é evidente que os óbitos por malária não apresentam um padrão espacial agregado significativo, isto é, não se observam grupos espaciais fortes quando analisamos o conjunto total de distritos. A distribuição espacial dos óbitos tende a ser aproximadamente aleatória sob o ponto de vista global.

Tabela 6: Resultados do Moran Local para os distritos com significância estatística para casos e óbitos por malária simultaneamente. (p-valor < 0.05)

Distrito	I_i	$Z(I_i)$	p-valor	Distrito	I_i	$Z(I_i)$	p-valor
Boane	1.097	2.342	0.019	Mapai	1.041	2.133	0.033
Chicualacuala	1.113	2.126	0.033	Massangena	0.858	2.005	0.045
Chigubo	0.767	2.090	0.037	Milange	1.372	2.070	0.038
Cidade Da Matola	0.887	2.079	0.038	Mocuba	2.451	2.357	0.018
Derre	-0.635	-3.273	0.001	Mocubela	0.243	2.297	0.022
Gondola	0.477	3.018	0.003	Molumbo	-0.346	-1.982	0.047
Ile	1.305	3.201	0.001	Mulevala	-0.190	-3.157	0.002
Lugela	1.047	2.684	0.007	Namaacha	1.288	1.975	0.048
Mabalane	0.963	2.235	0.025	Namacurra	2.269	3.264	0.001
Maganja Da Costa	1.648	2.838	0.005	Nicoadala	1.182	2.610	0.009
Manhica	0.831	1.959	0.050	Rapale	0.131	3.006	0.003

De forma geral, as Tabelas 5 e 6 mostram que a malária não está distribuída aleatoriamente no espaço, ou seja, a transmissão segue padrões espaciais claros, o que não é evidente para óbitos por malária.

4.2 Estimação de modelos dos casos e óbitos por malária

Tabela 7: Critério de Informação WAIC para os Modelos de Casos e Óbitos de Malária.

Modelo (Equação)	WAIC	
	Casos de malária	Óbitos por malária
Modelo 1 (28)	54321.0	4591.2
Modelo 2 (29)	3813.06	1053.3
Modelo 3 (35)	6281.38	—

No presente estudo procedeu-se a simulação e estimação de modelos estatísticos descritos no 3.2, recorrendo a diferentes funções de ligação e estruturas hierárquicas, com o objectivo de identificar a especificação mais apropriada para caracterizar simultaneamente a dinâmica espaço e temporal dos casos de malária e dos óbitos associados. Entre os modelos avaliados incluíram-se aqueles especificamente propostos pelos autores principais que fundamentam a nossa pesquisa, permitindo assim uma comparação sistemática entre alternativas metodológicas consistentes. A selecção do modelo final baseou-se, como já mencionado na metodologia, no WAIC como métrica principal. Conforme evidenciado na Tabela 7, a comparação dos valores de WAIC, aliada à análise da necessidade de captura explícita das dependências temporais e espaciais, permitiu discriminar, entre todas as opções simuladas, aquelas que melhor representavam os padrões epidemiológicos subjacentes. Os resultados indicam que o modelo (2) Poisson-Lognormal com tendência temporal e efeito específico apresentou o melhor desempenho global, sendo, portanto, o mais adequado para descrever a variabilidade espaço-temporal na incidência da doença, cuja estrutura permite representar de forma mais parcimoniosa a evolução temporal conservando, contudo, a heterogeneidade distrital observada. Deste modo, e em conformidade com os critérios metodológicos previamente estabelecidos, os resultados apresentados nos pontos subsequentes referem-se exclusivamente às estimativas derivadas deste modelo óptimo. Os resultados abaixo descritos permitem compreender como os diferentes factores climáticos influenciam a dinâmica da malária em Moçambique e contribuem para a interpretação robusta dos padrões observados ao longo do período em estudo.

4.2.1 Estimação do modelo para casos de malária (global e específica)

Os resultados do modelo indicado na Tabela 8 apresentam as estimativas simuladas com os dados para a incidência de malária nos distritos e anos em estudo. A estimativa de intercepto (β_0) apresenta média posterior de -0.031 , com desvio padrão de 0.059 e intervalo de confiança de aproximadamente $[-0.144, 0.085]$. Este resultado indica que, dado o conjunto de covariáveis, o nível médio do logaritmo do número esperado de casos de malária encontra-se próximo de zero,

sugerindo que as variáveis incluídas (escoamento, temperatura e pressão) ajustam adequadamente a média do processo.

No que diz respeito às covariáveis climáticas, o coeficiente para o *runoff* (escoamento) apresenta média posterior de 1,325 com elevada (desvio padrão de aproximadamente 2.51 e intervalo de confiança [2.09, 6.53]). Este intervalo indica efeito positivo do escoamento, fornecendo evidências suficiente para confirmar a associação entre escoamento e incidência de malária. O coeficiente associado à temperatura média (*t2m*) é estimado em -0.014 (desvio padrão ≈ 0.001 ; intervalo de credibilidade $[-0.016, -0.012]$), revelando um efeito negativo e estatisticamente consistente: à medida que a temperatura média aumenta, a incidência de malária tende a diminuir ligeiramente.

O parâmetro de autocorrelação temporal ρ apresenta média posterior de 0.774 (desvio padrão ≈ 0.228 ; intervalo de confiança [0.222, 0.992]), evidenciando uma forte correlação temporal entre anos consecutivos.

No que concerne às precisões associadas aos efeitos hierárquicos, o parâmetro τ_δ (efeitos temporais) apresenta média de aproximadamente 0.623 (desvio padrão ≈ 2.173), enquanto τ_π (efeitos espaciais) apresenta média de 0.007 (desvio padrão ≈ 0.006). A baixa precisão espacial implica elevada variabilidade entre distritos ($\sigma_\pi^2 \approx 1/\tau_\pi$), o que corrobora a existência de heterogeneidade geográfica significativa na incidência da doença.

Os parâmetros de precisão associados à mistura (k_1) indicam comportamentos distintos entre regimes. O parâmetro τ_v , relativo ao *regime de surto*, apresenta média de 0.746 (desvio padrão ≈ 0.786), enquanto τ_w , relativo ao *regime normal*, apresenta média de 1.516 (desvio padrão ≈ 0.967). Como esperado, o regime de surto exibe menor precisão (maior variabilidade), reflectindo a natureza mais dispersa dos eventos de surto em comparação ao regime normal.

Em síntese, o modelo de mistura identifica dois regimes latentes de comportamento da malária: um regime normal, caracterizado por variações regulares associadas a factores climáticos, e um regime de surto, associado a flutuações excepcionais na incidência. Os parâmetros ρ , τ_π e τ_δ indicam, respectivamente, uma forte dependência temporal, heterogeneidade espacial acentuada e variabilidade temporal moderada.

Na vertente global, o intercepto global (α_0) apresentou média posterior de 0.012 (IC95%: -0.087

a 0.082), com desvio-padrão de 0.060. O intervalo de confiança inclui o zero, indicando ausência de evidência para um efeito global significativo. Esse resultado sugere que, após o controle pelos efeitos espaciais e temporais, o risco médio de malária no país manteve-se estável ao longo do período analisado. Assim, as variações na incidência parecem estar mais associadas a padrões específicos de distritos e períodos do que a uma tendência global comum.

A Figura 6 apresenta a distribuição dos efeitos espaciais (η_i), que representam as variações residuais entre distritos após o ajuste pela tendência global e temporal. Observa-se que a maioria dos distritos apresenta valores próximos de zero, indicando semelhança em relação ao padrão médio nacional. Contudo, há distritos com valores negativos acentuados ($\eta_i < -4$), sugerindo risco inferior ao esperado, e poucos distritos com valores positivos ($\eta_i > 2$), indicando risco elevado. Esses resultados evidenciam heterogeneidade espacial na incidência de malária, com alguns distritos apresentando padrões distintos do comportamento médio nacional.

A Figura 7 apresenta a tendência temporal global (γ_t) estimada pelo modelo misto de Poisson-Lognormal, representando a variação média do risco de malária ao longo do tempo em todos os distritos. Observa-se uma redução gradual do efeito temporal entre os períodos 2017 e 2020, atingindo o valor mínimo no período 5 ($\gamma_t \approx -0.35$), o que indica um menor risco relativo nesse intervalo. A partir do ano 2022, verifica-se um aumento acentuado do efeito temporal, sugerindo uma redução da incidência de casos e óbitos, seguido de estabilização no ano 2023. De modo geral, o padrão temporal global indica flutuações moderadas no risco de contaminação e óbitos por malária, possivelmente associadas a variações climáticas e à eficácia diferenciada das medidas de controle implementadas ao longo dos anos.

A distribuição dos parâmetros de mistura (κ_i), desempenham um papel crucial no modelo misto Poisson-Lognormal, pois indicam o grau de adesão de cada distrito à tendência temporal global ($\mu_{it}^{(C)}$) em comparação com um comportamento específico de distrito ($\mu_{it}^{(AS)}$). A Figura 8 apresenta a distribuição dos parâmetros de mistura (κ_i), que indicam o grau de adesão de cada distrito à tendência temporal global ($\mu_{it}^{(C)}$) em relação ao modelo específico de distrito ($\mu_{it}^{(AS)}$). Observa-se que a maioria dos distritos apresenta valores de κ_i concentrados entre 0.5 e 0.8, sugerindo uma forte influência da tendência temporal comum. Poucos distritos exibem valores inferiores a 0.3, indicando padrões temporais próprios e possivelmente associados a factores locais, intervenções de saúde pública ou diferenças socioeconômicas. Esses resultados demonstram que, embora o comportamento temporal da malária em Moçambique seja amplamente coerente entre os distritos, per-

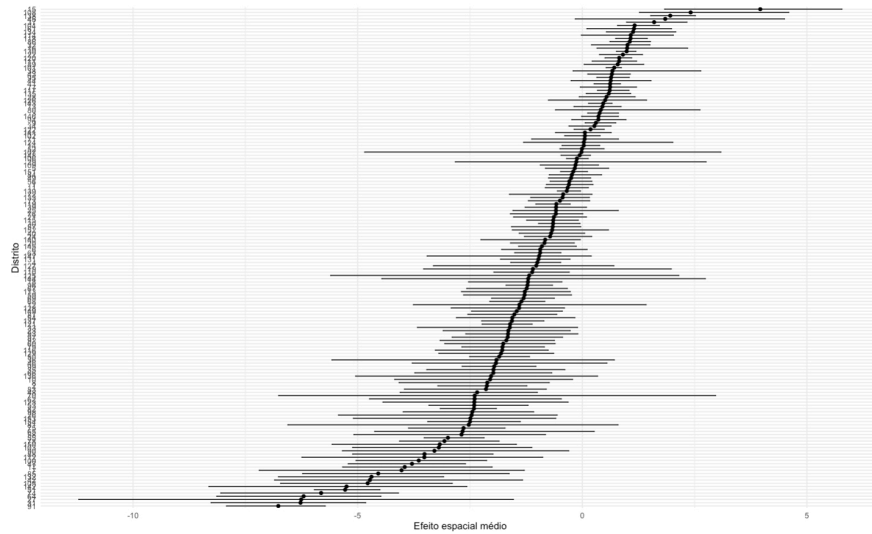


Figura 6: Efeito espacial

sistem heterogeneidades que justificam a inclusão de componentes espaço-temporais específicos no modelo.

No presente estudo, distritos como Alto Molócuè, Buzi, Cidade de Nampula, Inhassoro, Mocuba, Ribáuè e Vilankulos apresentaram valores de κ_i superiores a 0.5, evidenciando uma forte aderência à tendência temporal global. Isso significa que, nesses locais, as oscilações anuais na incidência de malária foram semelhantes às tendências observadas no país como um todo, sugerindo que as políticas de controlo e os factores climáticos (temperatura, escoamento e chuvas) tiveram impacto uniforme sobre esses distritos.

Em contraste, distritos como Bilene, Chicualacuala, Chigubo, Mabalane, Macomia, Muidumbe e Quissanga apresentaram valores de κ_i inferiores a 0.5, caracterizando comportamentos temporais específicos. Nesses casos, a variação temporal da malária mostra-se influenciada por condições locais particulares, diferenças climáticas, padrões de precipitação e escoamento de águas após

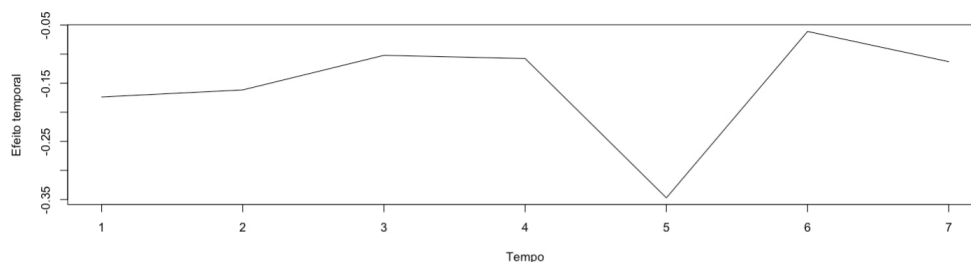


Figura 7: Efeito temporal global

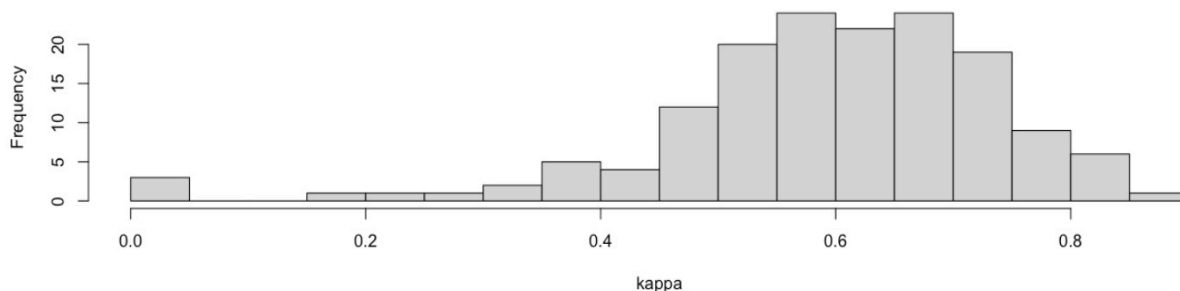


Figura 8: Histograma de parâmetro de mistura kappa

chuvas.

4.2.2 Estimação do modelo para óbitos por malária

Estimando a taxa de mortalidade por malária em função de variáveis climáticas, incorporando a dependência espacial entre distritos vizinhos, resultados (Tabela 8) indicam que a temperatura média apresentou um efeito negativo e estatisticamente significativo sobre a mortalidade por malária ($\beta = -0.651$; $IC_{95\%} = [-1.085, -0.139]$), implicando que distritos mais quentes tendem a registrar menores taxas de óbitos. O risco relativo estimado foi $\exp(-0.651) \approx 0.52$, o que representa uma redução aproximada de 48% na taxa de mortalidade associada a um aumento de um desvio-padrão na temperatura média. No entanto, o escoamento superficial (`runoff`) apresentou um efeito positivo e significativo ($\beta = 0.826$; $IC_{95\%} = [0.324, 1.291]$), indicando que distritos com maior escoamento apresentam maior mortalidade por malária. O risco relativo correspondente é $\exp(0.826) \approx 2.28$, sugerindo que um aumento de um desvio-padrão em `runoff` está associado a um aumento de aproximadamente 128% no risco de morte por malária. Esse resultado reforça a importância dos factores hidrológicos no favorecimento da transmissão e gravidade da doença.

A variância espacial ($\tau^2 = 3.04$) indica a presença de heterogeneidade residual considerável entre distritos, sugerindo que factores não incluídos no modelo, como condições socioeconômicas, acesso aos cuidados de saúde e intervenções de controle vectorial, também podem influenciar o padrão de mortalidade.

4.2.3 Evolução temporal de casos e óbitos por malária.

A figura 9 mostra os distritos com maior incidência de casos de malária. O comportamento temporal dos casos de malária apresentados no gráfico revela diferenças marcantes entre os distritos das

Tabela 8: Parâmetros estimados para casos e óbitos por malária do modelo 2.

Parâmetro	Casos de malária		Óbitos por malária	
	Estimativa(SD)	95% IC	Estimativa (SD)	95% IC
Intercepto (b_0)	-0.031(0.059)	[-0.141, 0.085]	-11.462(0.096)	[-11.659, -11.282]
Temperatura média	-0.014(0.001)	[-0.016, -0.012]	-0.651(0.242)	[-1.085, -0.139]
Runoff (escoamento)	1.325(2.51)	[2.091, 6.532]	0.826(0.247)	[0.324, 1.291]
Autocor. temporal (ρ)	0.774(0.228)	[0.221, 0.992]	0.099(0.075)	[0.004, 0.296]
Efeitos temporais (τ_δ)	0.632(2.173)	[3.621, 4.895]	—	—
Efeitos espaciais (τ_π)	0.007(0.007)	[0.006, 0.021]	—	—
Mistura (surto)	0.746(0.786)	[0.791, 2.283]	—	—
Mistura (normal)	1.516(0.902)	[0.252, 3.294]	—	—
Tau ² (modelo óbitos)	—	—	3.042(0.752)	[1.906, 4.851]

regiões Norte, Centro e Sul de Moçambique. Os distritos do norte destacam-se por apresentarem os valores mais elevados ao longo de todo o período analisado. Entre eles, a província de Nampula assume particular relevância, exibindo um crescimento contínuo desde 2017 até atingir o ponto mais alto em 2021, seguido de uma ligeira redução em 2022. Alguns distritos da região, como Mocuba, Maganja da Costa, Lugela, Mogovolas, Memba, Namarroi e Pebane, também mantêm níveis elevados de incidência, embora com variações moderadas, reforçando a ideia de que o norte permanece como a região com maior pressão epidemiológica da malária. Essa persistência está associada, a condições ambientais e sociodemográficas favoráveis à transmissão, tais como elevada humidade, pluviosidade consistente e alta densidade populacional. Nos distritos do Centro observa-se um padrão distinto, caracterizado por maior variabilidade temporal. Distritos como Barué, Buzi, Cidade de Chimoio, Ile e especialmente Gorongosa apresentam oscilações mais pronunciadas. Apesar de a zona centro manter níveis intermédios de incidência, inferiores ao norte, ainda assim, são relevantes no panorama nacional. Por fim, a zona sul apresenta situações de casos de malária muito diferente dos demais, associado a factores como menor precipitação acumulada, variações térmicas mais acentuadas e níveis mais elevados de urbanização, contribuindo para a redução do risco epidemiológico da malária.

Por outro lado, a figura 10 mostra que nem sempre os distritos de maior incidência de Malária apresentam, no mesmo sentido, maior incidência de casos de óbitos.

Os apêndices das Figuras A1, A2, A3 e A4 mostram a transmissão de casos de malária e óbitos para média e baixa transmissão. Mostram a sazonalidade da máalaria ao longo do tempo, apresentando redução de casos em meses como Novembro e Dezembro, aumento em meses como Janeiro a

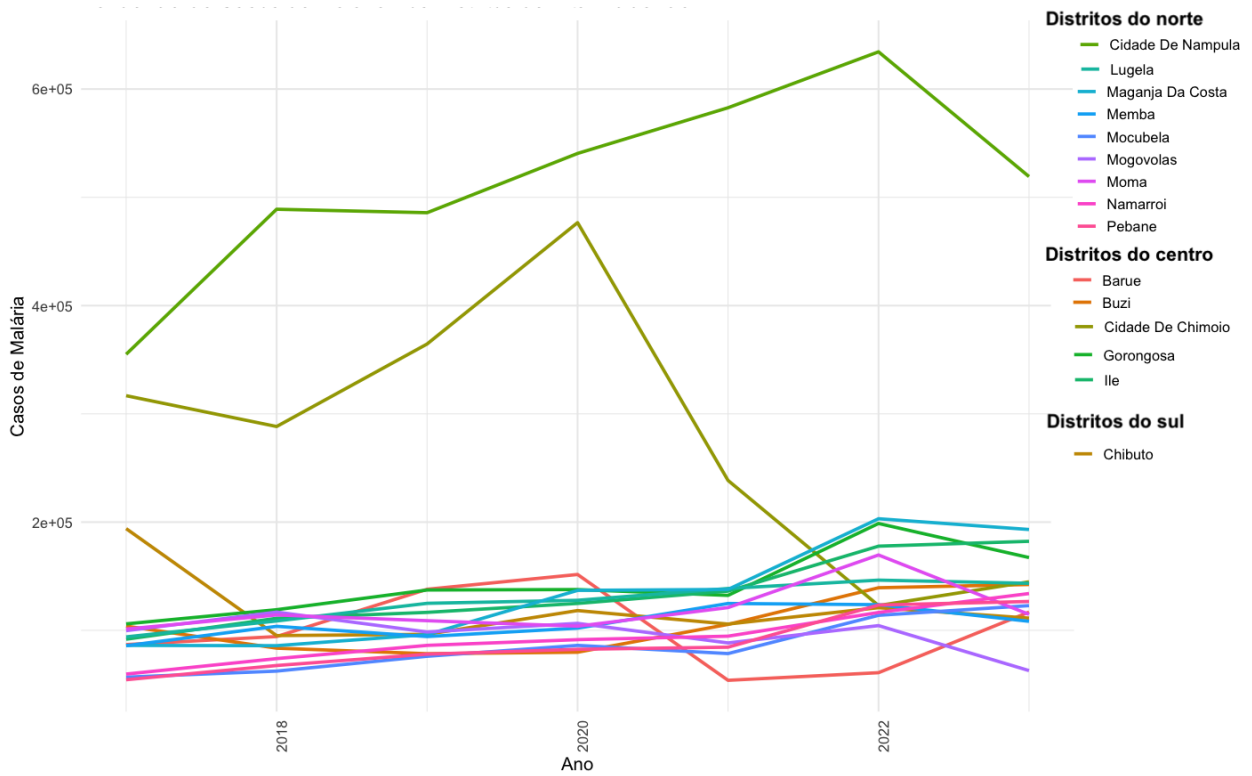


Figura 9: Linhas que mostram distritos com alta incidência de malária de 2017 a 2023 em função do tempo.

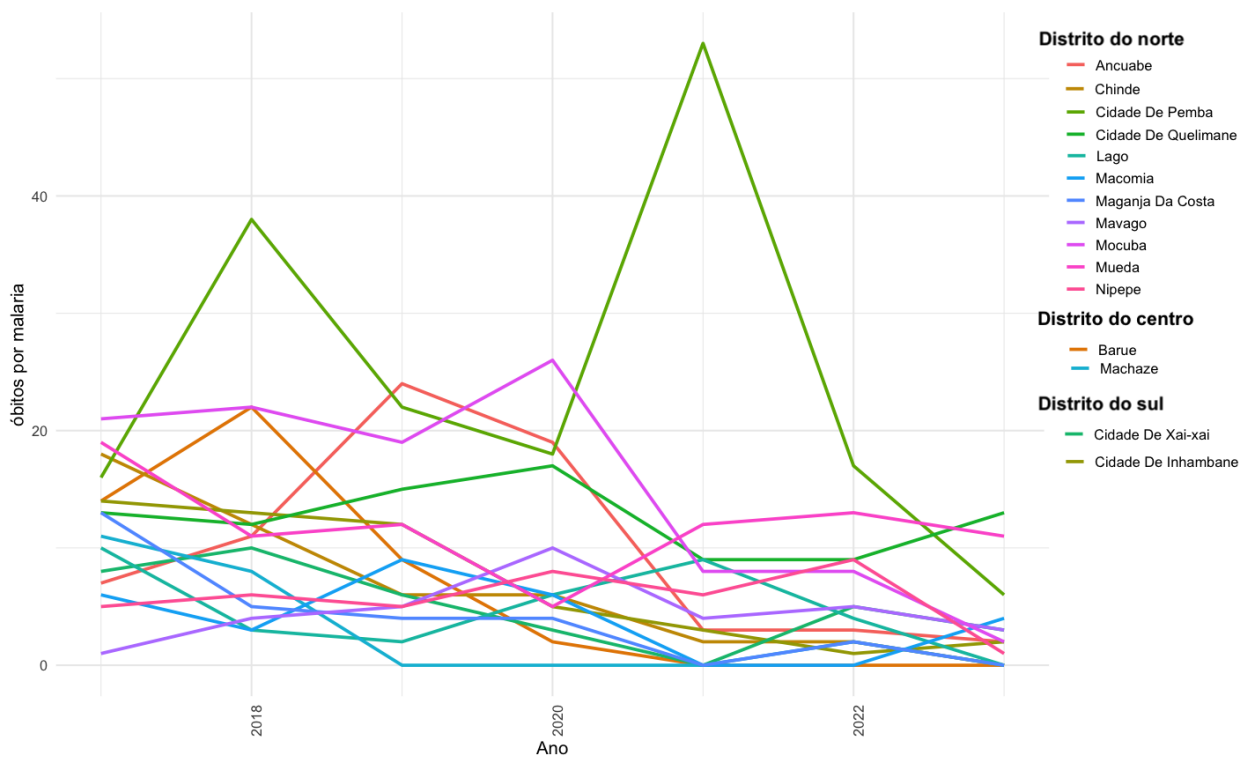


Figura 10: Linhas que mostram distritos com alta incidência de óbitos por malária de 2017 a 2023 em função do tempo.

Junho. Os distritos com baixa transmissão mostram uma queda conceptual dos casos de malária em alguns anos (2020 e 2023), ou mesmo nenhum caso de óbitos desde 2020 a 2023. Em termos regionais, a figura 12 evidencia que a mortalidade por malária segue um padrão semelhante ao observado nos casos de malária, sendo o norte com maior carga e maior variabilidade, enquanto a região centro se mantém em níveis moderados e a região sul regista os valores mais baixos. A ocorrência de picos marcados em determinados distritos reforça a necessidade de estratégias específicas de vigilância e intervenção, ajustadas às condições epidemiológicas de cada região.

4.2.4 Estimação de risco relativo

Os mapas apresentados na figura 11 evidenciam a distribuição espacial do risco relativo (RR) médio por distrito em Moçambique, considerando dois desfechos distintos: óbitos (à esquerda) e casos de malária (à direita). Observa-se que, no caso dos óbitos, os valores de RR tem maior intensidade nas regiões centrais e setentrionais do país, onde se destacam distritos que apresentam maior vulnerabilidade em termos de mortalidade associada à malária. Já no mapa referente aos casos de malária, o RR apresenta uma variação mais ampla, com distritos situados sobretudo no norte e sul, na faixa costeira, do país exibindo valores mais elevados, sugerindo maior incidência da doença nestas áreas.

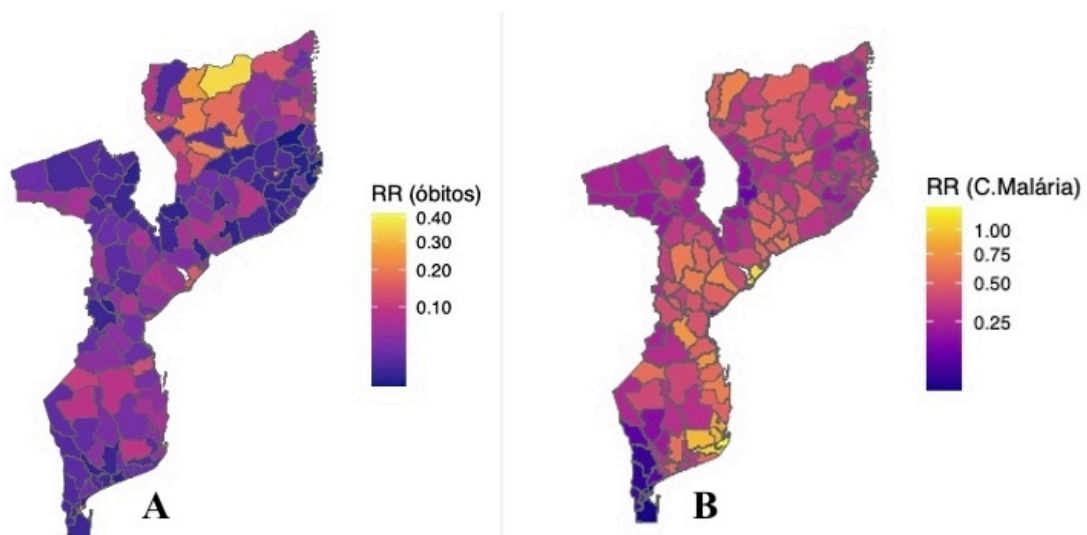


Figura 11: Risco relativo de casos de malária no lado direito e de óbitos, a esquerda da figura em Moçambique de 2017 à 2023.

O RR médio de malária (figuras 12 e apêndice A10) apresenta uma variação ao longo do tempo, com um pico claro em torno de 2020 (quando o RR sobe), seguido de uma queda e nova elevação próxima ao ano 2023, sugerindo assim, que há flutuações sazonais ou episódios epidémicos que

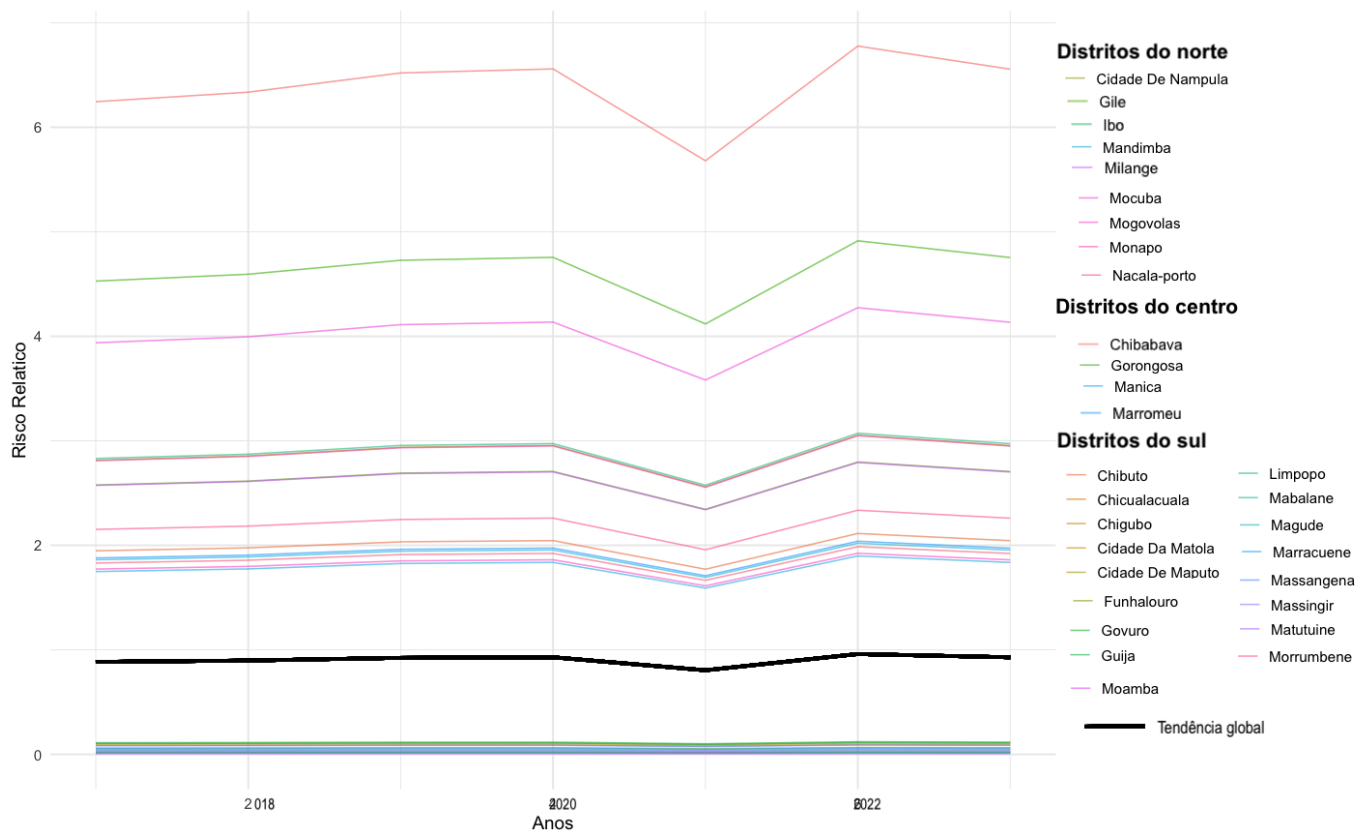


Figura 12: Tendência comum de casos estimados por malária em Moçambique no período de 2017 a 2023.

impactam todos os distritos em conjunto. Vejamos de forma regional e a nível dos distritos: A observação da Figura 12 revela que existe uma clara hierarquia regional no risco relativo de malária ao longo período em análise. A linha preta representa a tendência global, que se mantém em torno de valores próximos de 1, indicando o nível médio de risco esperado para o conjunto dos distritos. Qualquer distrito cuja linha permanece persistentemente acima dessa linha apresenta, portanto, risco superior ao esperado, reflectindo maior chance de ocorrência de casos de malária em relação à média nacional. Entre as três regiões, é evidente que o Norte concentra a maior parte dos distritos cujas curvas estão sistematicamente acima da tendência global. Distritos como Cidade de Nampula, Ibo, Mocuba, Mogovolas, Nacala-porto e Gile apresentam riscos relativos elevados, mantendo-se acima da média global, ao longo de todo o período. Isso reforça a condição epidemiológica já mencionada: o Norte constitui a região com maior intensidade de transmissão e maior persistência no risco de malária. No Centro, alguns distritos também superam consistentemente a média global. Gorongosa, em particular, destaca-se com risco relativo persistentemente acima do global, indicando elevada vulnerabilidade local. Distritos como Chibabava, Manica e Marromeu também se posicionam acima do valor global, mas com menor magnitude e menor distância

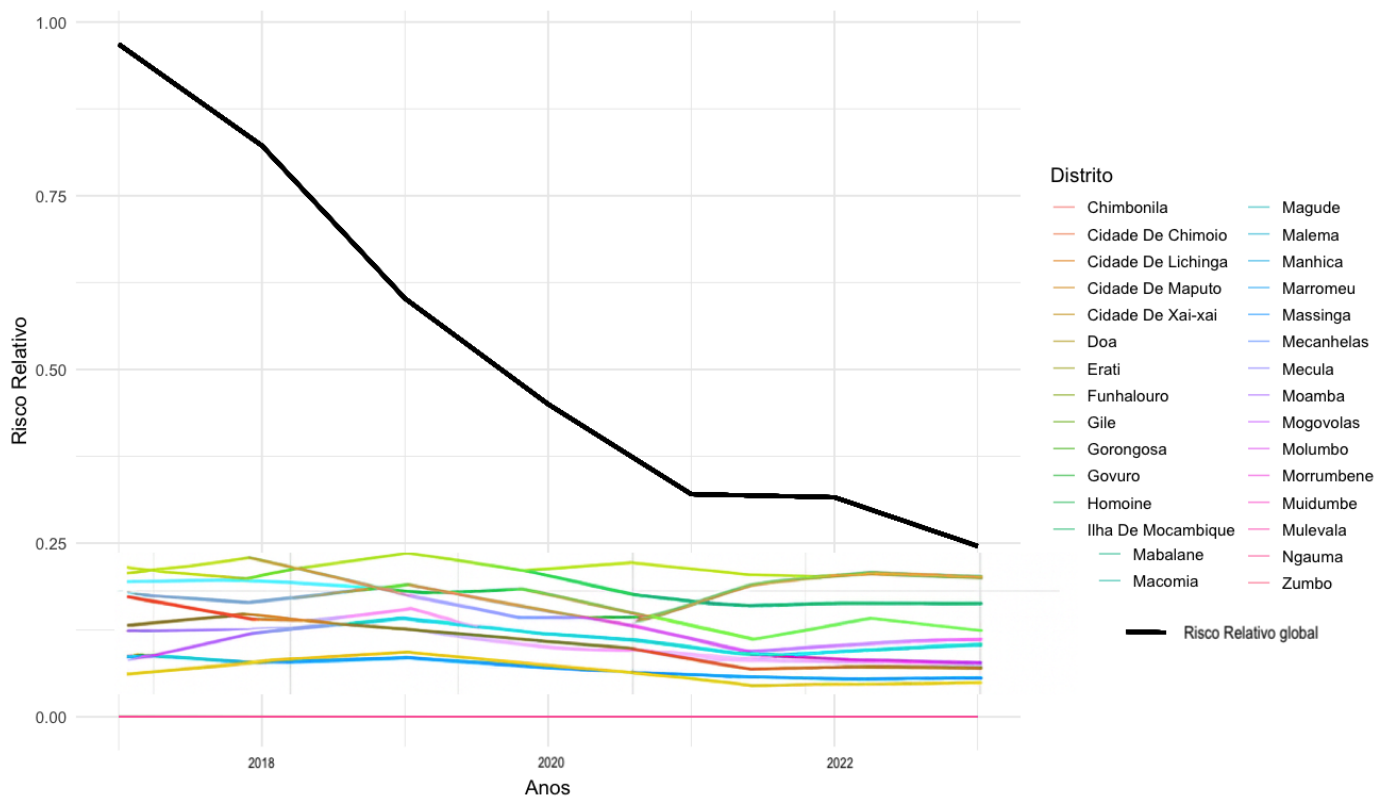


Figura 13: Tendência comum de taxa de óbito por malária em Moçambique no período de 2017 a 2023.

em relação à referência. Por outro lado, no Sul, os distritos apresentam padrões distintos. Enquanto alguns (como Limpopo, Mabalane, Magude e Massangena) se aproximam ocasionalmente da tendência global, a maioria permanece abaixo, indicando risco reduzido de transmissão em comparação com o restante do país.

O comportamento tendencial dos casos de malária não são similares á óbitos. A taxa global de óbitos (Figura 13) por malária tem reduzido ao longo dos anos e com o mesmo, as mortes se encontram por baixo da taxa global, porém há que destacar distritos como Mecalhelas, Lugela, Molumbolo e Riabué que se encontram bem próximo do risco relativo global nos últimos três anos, indicando assim, apesar de baixo, maior incidência de morte por malária.

4.2.5 Teste de convergência - Monte Carlo via Cadeias de Markov

A convergência das cadeias de Markov é uma etapa fundamental na inferência Bayesiana, pois garante que as amostras geradas pelo algoritmo MCMC (Markov Chain Monte Carlo) representem adequadamente a distribuição posterior dos parâmetros. A convergência é atingida quando as cadeias entram em regime estacionário, ou seja, quando passam a explorar de forma estável

e aleatória a região de alta probabilidade da posterior. Com recurso a amostragem de Gibbs, para garantir a reprodutibilidade e a robustez das inferências, foram consideradas as seguintes configurações do algoritmo:

- Número total de iterações: 80 000
- Período de burn-in: 20 000 iterações iniciais descartadas
- Thinning: 10 (uma em cada 10 amostras foi retida)
- Número de cadeias: 3 cadeias independentes

Após o período de burn-in e aplicação do thinning, foram obtidas amostras da distribuição posterior para todos os parâmetros do modelo. A convergência das cadeias foi avaliada utilizando métodos gráficos e critérios formais. Através de uma inspeção visual, os gráficos das cadeias apresentaram boa mistura (vide apêndice 5), comportamento estacionário e ausência de tendências sistemáticas, indicando convergência adequada. Assim, conclui-se que o processo MCMC atingiu convergência satisfatória, garantindo confiabilidade às estimativas posteriores e robustez às inferências realizadas a partir do modelo. Foi analisado também a convergência através de gráfico de Gelman-Rubin (vide apêndice 6), que evidenciou após um incremento do número de iterações, os parâmetros associados às covariáveis ambientais (β_0 , β_{runoff} , β_{tp} , β_{t2m}) e os termos temporais ρ , τ_δ , τ_π e τ_v , apresentaram valores de \hat{R} próximos de 1, indicando convergência satisfatória. Contudo, o parâmetro τ_w teve o valor de \hat{R} superiores a 1.5, mesmo após 38 000 iterações, evidenciando fraca mistura entre cadeias e ausência de convergência. Este parâmetro corresponde a componente espacial e temporal hierárquico, que são conhecidos por apresentar elevada dependência serial. No entanto não influencia nos resultados do modelo. Os valores de tamanho efectivo da amostra (ESS) foram elevados, indicando baixa autocorrelação e eficiência na amostragem. As funções de autocorrelação (vide apêndice 10) apresentaram decaimento rápido, evidenciando independência aproximada entre as amostras após o *thinning*.

4.3 Discussão

A análise espaço-temporal da malária em Moçambique no período de 2017 a 2023 evidenciou uma evolução significativa nos padrões de incidência e mortalidade, bem como uma influência de factores climáticos e espaciais sobre a disseminação da doença. Os resultados evidenciam que

a malária permanece um problema crítico de saúde pública no país, com forte heterogeneidade espacial e temporal, corroborando evidências previamente reportadas pela Organização Mundial da Saúde (OMS, 2023).

As métricas apresentadas no subcapítulo anterior permitem discutir, de forma integrada, as tendências gerais, as heterogeneidades geográficas e o desempenho dos modelos estatísticos aplicados nesse estudo.

Os resultados descritivos demonstraram um crescimento contínuo no número médio anual de casos de malária, passando de cerca de 64 mil casos em 2017 para mais de 80 mil em 2023, com uma ligeira descida no ano 2020. Tanto a média quanto a mediana apresentaram variações de aumentos consistentes, reflectindo uma intensificação flutuante da transmissão ao longo dos anos. Analisando em termos da variação, esse crescimento foi acompanhado por aumento na variabilidade (variância e desvio-padrão), o que sugere uma dispersão espacial mais acentuada dos casos e padrão desigual nos óbitos, possivelmente associada às condições ambientais heterogêneas e à desigualdade na implementação de medidas de controle vectorial entre distritos.

A análise mostrou que os modelos tradicionais de Poisson apresentam limitações importantes, especialmente na incapacidade de captar simultaneamente variações espaciais e temporais e na tendência a produzir estimativas enviesadas, conforme também discutido por Bouliéri et al. (2020). Em contraste, o modelo Bayesiano misto proposto demonstrou melhor desempenho preditivo e maior capacidade de identificação de padrões complexos, permitindo estimar com maior precisão o risco relativo global e específico por distrito. Esta evidência reforça a relevância da inferência Bayesiana em contextos epidemiológicos, conforme destacado por Gething (2011) e Escaramís et al. (2011).

O modelo simulado indica que a temperatura média (t_{2m}) possui efeito negativo estatisticamente significativo sobre os casos de malária, evidenciando que temperaturas mais altas tendem a reduzir a incidência. Em contrapartida, o escoamento superficial (runoff) apresentou efeito positivo relevante, sobretudo nos modelos de mortalidade, indicando que áreas com maior acúmulo de água estão associadas a maior risco de transmissão e gravidade da doença. Esses resultados estão em consonância com estudos anteriores de Zacarias e Andersson (2011) que apontam a temperatura e a disponibilidade hídrica como determinantes cruciais da densidade vectorial e da sobrevivência do parasita da Malária. No contexto moçambicano, onde há forte sazonalidade pluviométrica,

tais factores tornam-se determinantes para a flutuação anual dos casos. Do ponto de vista metodológico, a decomposição da tendência global e específica permitiu identificar não apenas padrões médios de evolução da doença, mas também desvios locais relevantes, superando limitações observadas em modelos anteriores, como os baseados em estruturas CAR uniformes (Armando et al., 2025). Este avanço representa uma contribuição importante para a modelação epidemiológica em Moçambique e em contextos semelhantes.

Os parâmetros estimados para os modelos espaço-temporais autorregressivos (CAR) indicaram dependência temporal moderada ($\rho_T \approx 0.15$) e dependência espacial fraca ($\rho_S \approx 0.06$), o que implica que a variação dos casos é mais explicada por factores climáticos e temporais do que pela contiguidade geográfica isolada. O parâmetro de autocorrelação temporal $\rho \approx 0.77$ no modelo de mistura confirma a persistência interanual da malária, ou seja, distritos com altas incidências tendem a manter esse padrão em anos subsequentes. Ademais, as precisões τ_π e τ_δ evidenciam heterogeneidade espacial acentuada e variabilidade temporal moderada, consistentes com a natureza não estacionária da doença.

Permitir que o parâmetro de mistura varie ao longo do tempo resultou num modelo adequado para séries temporais longas, um caso em que o modelo base demonstrou fraco desempenho. Informações sobre os momentos em que ocorrem comportamentos invulgares também foram fornecidas, facilitando a interpretação e a investigação etiológica. A priori do parâmetro de mistura é alterada para uma priori hierárquica menos informativa, que incorpora dependências espaciais e temporais, tornando o modelo mais flexível na detecção de vários padrões realistas de comportamento anómalo. O modelo de mistura Bernoulli identificou dois regimes latentes de comportamento da malária:

Um regime normal, com variações regulares associadas aos factores climáticos;

Um regime de surto, caracterizado por flutuações excepcionais na incidência; Esse resultado reforça que a malária em Moçambique não segue um padrão homogêneo, mas alterna entre períodos endêmicos e epidêmicos. A identificação desses regimes é particularmente útil para a vigilância epidemiológica, permitindo a detecção precoce de surtos e a alocação eficiente de recursos de saúde.

5 CONCLUSÕES RECOMENDAÇÕES

5.1 Conclusões

O presente estudo teve como objectivo geral modelar a distribuição espaço-temporal do risco relativo de casos e óbitos por malária em Moçambique no período de 2017 a 2023, utilizando a abordagem de modelagem Bayesiana mista. A análise foi desenvolvida com base em dados secundários provenientes do MISAU, incorporando variáveis climáticas obtidas através do Centro Europeu de Previsões a Médio Prazo, e dependências espaciais e temporais.

O objectivo geral foi plenamente alcançado, permitindo compreender de forma integrada como os factores espaciais, temporais e climáticos influenciam a incidência e a mortalidade por malária no país.

Em relação ao primeiro objectivo específico, descrever os padrões de distribuição e correlação espacial e temporal de casos e óbitos por malária em Moçambique, conclui-se que a doença apresenta uma forte variabilidade geográfica e temporal. Foram observadas diferenças marcantes entre as regiões do norte e do sul do país, sendo as províncias de Nampula, Zambézia e Cabo Delgado as mais afectadas. Confirmando-se assim que a malária não se distribui de forma uniforme, mas se concentra em áreas de maior vulnerabilidade climática e, talvez socioeconômica. Temporalmente, observou-se uma tendência crescente de casos durante o período analisado, com picos coincidentes com as estações chuvosas, o que evidencia a natureza sazonal da transmissão.

Ao modelar as variáveis climáticas, satisfaz-se o segundo objectivo da pesquisa, que singela-se em verificar a influencia da temperatura média, pressipitação e escoamento superficial das águas na sincidência da doença em moçambique. Os resultados indicaram que as variáveis climáticas exercem influência nos casos e óbitos por malária. Em particular, a temperatura média apresentou um efeito negativo estatisticamente significativo, indicando que variações nesta variável estão associadas a alterações no risco da malária, ou seja, temperaturas mais altas reduzem o risco. De igual modo, a precipitação e o escoamento superficial mostraram efeitos estatisticamente significativos, sugerindo que condições ambientais favoráveis à formação de criadouros de mosquitos contribuem para o aumento de casos de malária.

Em termos espaciais, o modelo identificou distritos com maior risco relativo de ocorrência da doença, destacando-se, entre outros, Nhamatanda, Mafambisse, Sussundenga, Alto Molócuè e

Lugela, os quais apresentaram valores consistentemente elevados ao longo do período em análise. Estes resultados evidenciam a existência de distritos de alto risco, que devem ser considerados na definição de estratégias de intervenção.

No que diz respeito ao terceiro objectivo, estimação das tendências globais e específicas de casos e óbitos, a modelagem evidenciou a coexistência de uma tendência geral de persistência da doença e de variações locais que reflectem condições ambientais e estruturais distintas. Isso reforça a importância de políticas de combate à malária que considerem as particularidades de cada distrito e ajustem as intervenções às características locais de transmissão.

Relativamente ao quarto objectivo, prever o risco relativo e a taxa de mortalidade, os modelos aplicados permitiram identificar zonas críticas com maior probabilidade de incidência e de óbitos, fornecendo subsídios importantes para o planeamento de estratégias de controlo. As previsões obtidas contribuem para a formulação de medidas preventivas baseadas em evidências, especialmente no contexto da gestão de recursos e da alocação.

Na abordagem do estudo, o objectivo principal era mostrar que a modelação Bayesiana mista é melhor forma de modelar esses tipos de dados, mas não obteve-se a convergência, não obtendo melhor WAIC, pelo que, usou-se os modelos de modelação Poisson-Lognormal, modelos esses que estão limitados no uso simultânea de efeitos espaciais, temporais e múltiplos. As estatísticas diagnósticas indicavam problemas persistentes, como valores elevados de estimativas do R e baixas amostras efectivas (ESS), sugerindo que as cadeias não exploravam adequadamente o espaço paramétrico. Além disso, a estrutura de mistura desenvolvida, particularmente sensível a problemas de identificabilidade e “label switching”, acentuou a instabilidade das cadeias e a incapacidade de o algoritmo de OpenBUGS produzir estimativas fiáveis.

Embora tenha se alcançado os objectivos, o estudo apresenta limitações dos quais destacam-se a não inclusão da sazonalidade no modelo e variáveis demográficas. Sendo assim, um campo aberto para futuras pesquisas. Ainda, reiteramos o uso de modelos para estimação simultânea de múltiplos efeitos hierárquicos (espaciais, temporais e de mistura) no entanto adaptado por província. Assim irá diminuir o número de observações na base de dados. Os registos de casos e óbitos por malária podem estar sujeitos a subnotificação, particularmente em áreas com menor acesso aos serviços de saúde, o que pode levar à subestimação dos casos e óbitos por malária. Adicionalmente, podem ocorrer erros de diagnóstico, sobretudo em contextos onde o diagnóstico clínico

não é confirmado laboratorialmente, influenciando a precisão dos dados reportados. Outro aspecto relevante refere-se à variação na cobertura dos serviços de saúde ao longo do espaço e do tempo, que pode introduzir heterogeneidade não observada nos dados.

5.2 Recomendações

Os resultados deste estudo possuem importantes implicações para as políticas públicas de saúde. A identificação das zonas de maior risco fornece subsídios para a priorização de recursos e o planeamento de campanhas de controlo vectorial e distribuição de redes mosquiteiras. A integração de dados climáticos aos sistemas de vigilância epidemiológica pode aprimorar a capacidade preditiva e permitir o desenvolvimento de sistemas de alerta precoce. Além disso, a gestão temporal dos recursos deve considerar a sazonalidade da doença, intensificando as acções de prevenção durante e após o período chuvoso.

Do ponto de vista metodológico, recomenda-se o uso contínuo da modelação climática, incorporando novas variáveis ambientais, e a integração de dados de mobilidade humana, que podem melhorar a identificação de rotas de disseminação da doença. Sugere-se ainda o uso de modelos dinâmicos espaço-temporais de mistura adaptativa, capazes de detectar mudanças no padrão de transmissão ao longo do tempo.

Finalmente, destaca-se a importância de fortalecer a capacidade técnica das instituições nacionais na aplicação de metodologias Bayesianas, promovendo formações contínuas e incentivando o uso de abordagens quantitativas avançadas na análise de dados epidemiológicos. A criação de sistemas digitais de vigilância que integrem dados de saúde, climáticos e populacionais em tempo real também é uma medida essencial para o aprimoramento da resposta nacional à malária.

Referências

- Africa Turismo. (2014). Moçambique. <http://www.africa-turismo.com/mocambique/mocambique.htm>
- Agresti, A. (2002). *Categorical Data Analysis*. 2nd ed., John Wiley & Sons, New York.
- Aitchison, J., & Ho, C. H. (1989). The multivariate Poisson-log normal distribution. *Biometrika*, 76(4), 643–653.
- Albert, J. (2009). *Bayesian Computation with R*. NY: Springer-Verlag, New York.
- Alexandre, P & Vinícios, V. (2020). *Análise Exploratória de Dados Espaciais*. Universidade Federal do Paraná, Brasil.
- Almeida, R. D. (2009). O uso de geotecnologias para o mapeamento da malária no Município de Vitória/ES. *Revista Brasileira de Geografia Médica e da Saúde*, no 5.
- Almeida, E. S. (2012). *Econometria Espacial Aplicada*. Campinas: Editora Alínea.
- Américo, J. (2023). *Experiências na utilização de perspectivas climáticas sazonais para o mapeamento de riscos na preparação do sector da Saúde em Moçambique*. INS.
- Arambepola, R., Keddie, S. H., Collins, E. L., Twohig, K. A., Amratia, P., Bertozzi-Villa, A., et al. (2020). Spatiotemporal mapping of malaria prevalence in Madagascar using routine surveillance and health survey data. *Scientific Reports*, 10(1), 18129.
- Armando, C. J., Rocklöv, J., Sidat, M., Tozan, Y., Mavume, A. F., & Sewe, M. O. (2025). Spatio-temporal modelling and prediction of malaria incidence in Mozambique using climatic indicators from 2001 to 2018. *Scientific Reports*, 15, 11971. <https://doi.org/10.1038/s41598-025-97072-6>.
- Ávila, S. L. M. (1996). In: Ferreira, A. W. *Diagnóstico laboratorial das principais doenças infecciosas e auto-imunes*. Guanabara Koogan. Rio de Janeiro/RJ.
- Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2), 192–236.
- Boulieri, A., Bennett, J. E., & Blangiardo, M. (2020). A Bayesian Mixture Modeling Approach for Public Health Surveillance. *Biostatistics*, 21(3), e369–e383. <https://doi.org/10.1093/biostatistics/kxy038>

- Carlin, B. P., Gelfand, A. E., & Smith, A. F. M. (1992). Hierarchical Bayesian analysis of change-point problems. *Applied Statistics*, 41, 389–405.
- Carter, R., & Mendis, K. N. (2002). Evolutionary and historical aspects of the burden of malaria. *Clinical Microbiology Reviews*, 15(4), 564–594.
- Casella, G., & Berger, R. L. (2011). *Inferência estatística* (2ª ed.). Cengage Learning.
- Cox, F. E. G. (2010). History of the Discovery of the Malaria Parasites and Their Vectors. *Parasites & Vectors*, 3(1), 1.
- Delfino, R. T.; Santos-Filho, O. A.; & Figueroa-Villar, J. D. (2002). Molecular docking studies of the binding of inhibitors to HIV-1 protease. *Journal of the Brazilian Chemical Society*, 13, 727–734.
- Delen, D. (2011). Predicting student attrition with data mining methods. *Journal of College Student Retention: Research, Theory & Practice*, 13(1), 17-35. Baywood Publishing Co., Inc. <https://doi.org/10.2190/CS.13.1.b>.
- Escaramís, G., Carrasco, J. L., Aponte, J. J., Nhalungo, D., Nhacolo, A., Alonso, P., & Ascaso, C. (2011). Spatio-temporal analysis of mortality among children under the age of five in Manhica (Mozambique) during the period 1997–2005. *International Journal of Health Geographics*, 10, 14. <https://doi.org/10.1186/1476-072X-10-14>
- Ferreira, M. S. (2004). In: Veronesi, R.; Focaccia, R. *Tratado de Infectologia*. Atheneu (2ª edição). Rio de Janeiro/RJ.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D., Vehtari, A., & Rubin, D. (2013). *Bayesian Data Analysis* (3rd ed.). CRC Press, Boca Raton.
- Gething, P. W., Smith, D. L., Patil, A. P., Tatem, A. J., Snow, R. W., & Hay, S. I. (2011). Climate Change and the Global Malaria Recession. *Nature*, 465(7296), 342-345. doi:10.1038/nature09098.
- Gill, J. (2008). *Bayesian Methods: A Social and Behavioral Sciences Approach*. Chapman & Hall/CRC, Boca Raton.
- Goldstein, H. (1999). *Multilevel statistical models*. Institute of Education, Multilevel Models Project – Edward Arnold.

Ghosh, M.; Song, J. J.; Miaou, S.; & Mallick, B. (2006). Bayesian multivariate spatial models for roadway traffic crash mapping. *Journal of Multivariate Analysis*, 97(1), 246–273.

Haddad, E. A.; Haddad, P. R.; Vale, A. V. (2017). *Desenvolvimento territorial no Maranhão: um diagnóstico espacial*. Banco Mundial, mimeo.

Held, L.; & Paul, M. (2011). Modeling seasonality in space-time infectious disease surveillance data. *Biometrical Journal*, 53(6), 889–908.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2ª ed.). John Wiley & Sons, Inc.

Ihaka, R., Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5, pp. 299–314.

Jana, S., Fu, S. H., Gelband, H., Brown, P., & Jha, P. (2022). Spatio-temporal modelling of malaria mortality in India from 2004 to 2013 from the Million Death Study. *Malaria Journal*, 21(1), 90.

Lesaffre, E. & Lawson, A. B. (2012). *Bayesian Biostatistics*. John Wiley & Sons.

Martens, et al. (1995). Potential impact of global climate change on malaria risk. *Environmental Health Perspectives*, nº 5.

Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37, (1), 17–23.

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 135(3), 370–384.

NIMBLE Development Team. (2023). *NIMBLE User Manual* (Version 1.1). Disponível em <https://r-nimble.org>

Organização Mundial da Saúde. (2020). *World Malaria Report 2020*. World Health Organization. Disponível em: <https://www.who.int/publications/i/item/9789240015791>.

Organização Mundial da Saúde. (2023). *World Malaria Report 2023*. World Health Organization. Disponível em: <https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023>.

Packard, R. M. (2007). *The Making of a Tropical Disease: A Short History of Malaria*. Johns

Hopkins University Press.

Plucinski, M. M., Candrinho, B., Chambe, G., Muchanga, J., Muguande, O., Matsinhe, G., et al. (2018). Multiplex serology for impact evaluation of bed net distribution on burden of lymphatic filariasis and four species of human malaria in northern Mozambique. *PLoS Neglected Tropical Diseases*, 12(2), e0006278.

Raudenbush, S. W., Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage Publications.

Rey, L. (2001). *Parasitologia*, Guanabara-Koogan Ed.: Rio de Janeiro.

Rich, S. M., Leendertz, F. H., Xu, G., LeBreton, M., Djoko, C. F., Aminake, M. N., et al. (2009). The origin of malignant malaria. *Proceedings of the National Academy of Sciences*, 106(35), 14902–14907.

Robert, A.; Benoit-Vical, F.; Dechy-Cabaret, O. & Meunier, B. (2001). From classical antimalarial drugs to new compounds based on the mechanism of action of artemisinin. *Pure and Applied Chemistry*, 73, 1173–1188.

Sequeira, A. R. (2017). *Malária em Moçambique: Políticas, provedores de cuidados, saberes e práticas de gestão de doença*. CEI, Lisboa.

Shimada, M. S. & Imakawa. (2010). Reflexões sobre as mudanças climáticas e a dinâmica da malária na Amazônia. *Congresso Brasileiro de Meteorologia*. Belém/PA.

Silva, T. H. A.; Oliveira, M. T.; Santos, H. F.; Oliveira, A. B. & Almeida, W. B. (2005). Estudos teóricos sobre derivados de artemisinina e sua atividade antimalárica. *Química Nova*, 28, 244–248.

Souza, et al. In: Leão, R. N. Q. (1997). *Doenças Infecciosas e Parasitárias: enfoque amazônico*. CEJUP. Belém do Pará/PA.

U.S. President's Malaria Initiative. (2023). *Mozambique Malaria Operational Plan FY 2023*. USAID. <https://reliefweb.int/report/mozambique/presidents-malaria-initiative-mozambique-malaria-operational-plan-fy-2023>

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11,

3571–3594.

Webb Jr., J. L. A. (2009). *Humanity's Burden: A Global History of Malaria*. Cambridge University Press.

Zacarias, O. P., & Andersson, M. (2011). Spatial and temporal patterns of malaria incidence in Mozambique. *Malaria Journal*, **10**, 189. <https://doi.org/10.1186/1475-2875-10-189>.

Zacarias, O. P., & Andersson, M. (2010). Mapping malaria incidence distribution that accounts for environmental factors in Maputo Province-Mozambique. *Malaria Journal*, *9*(1), 79.

Apêndices

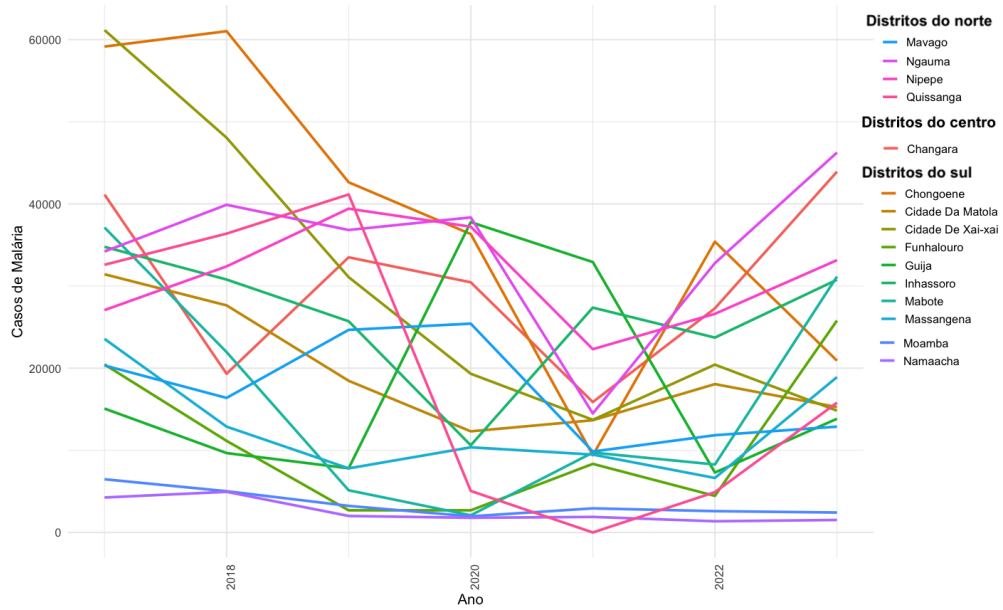


Figura A1: Tendência dos casos de malária nos distritos com baixa incidência.

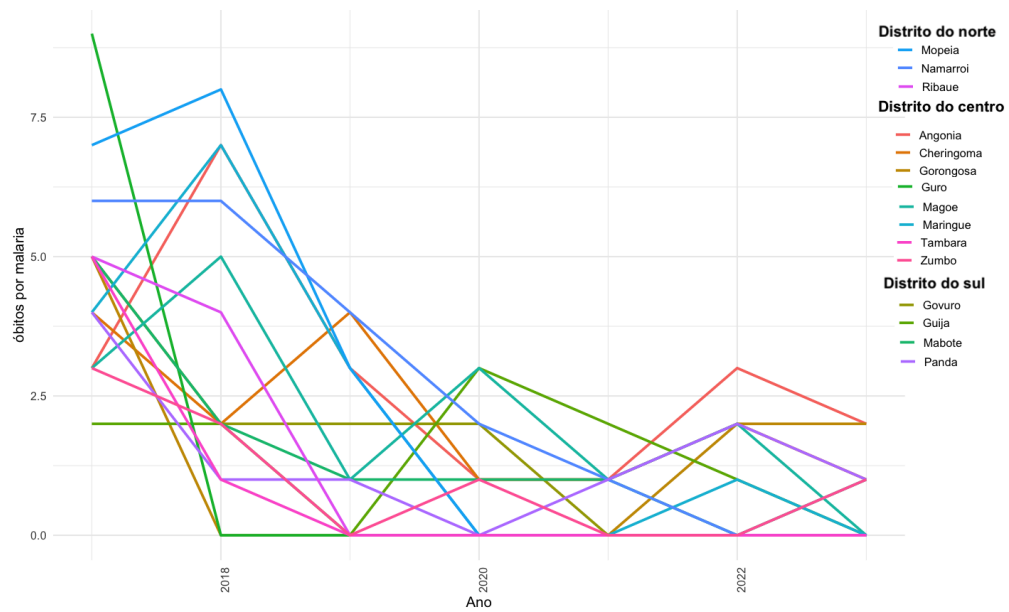


Figura A2: Tendência dos obitos nos distritos com baixa incidência.

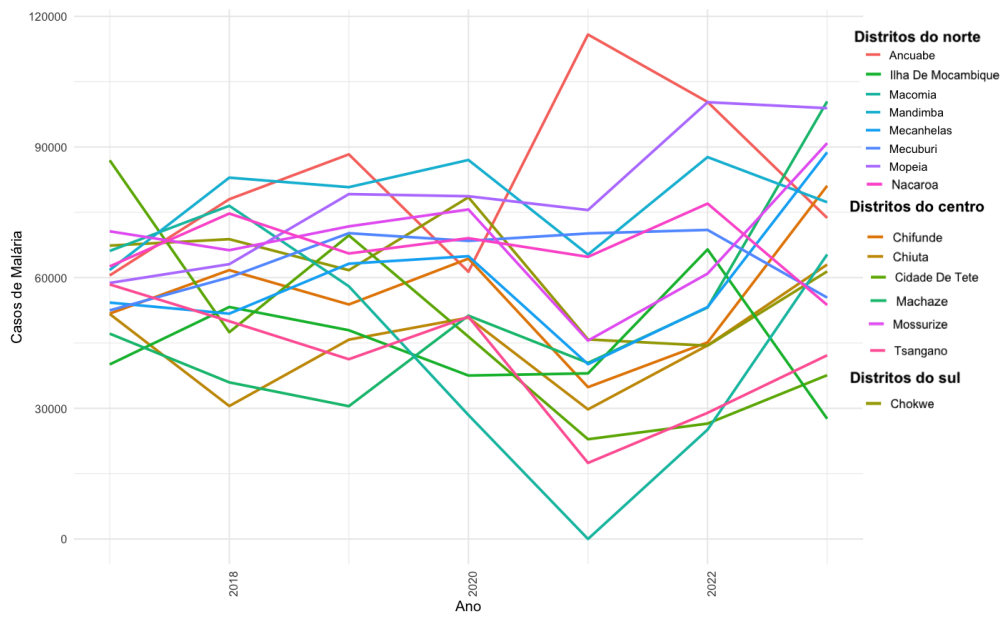


Figura A3: Tendência dos casos de malária nos distritos incidência média.

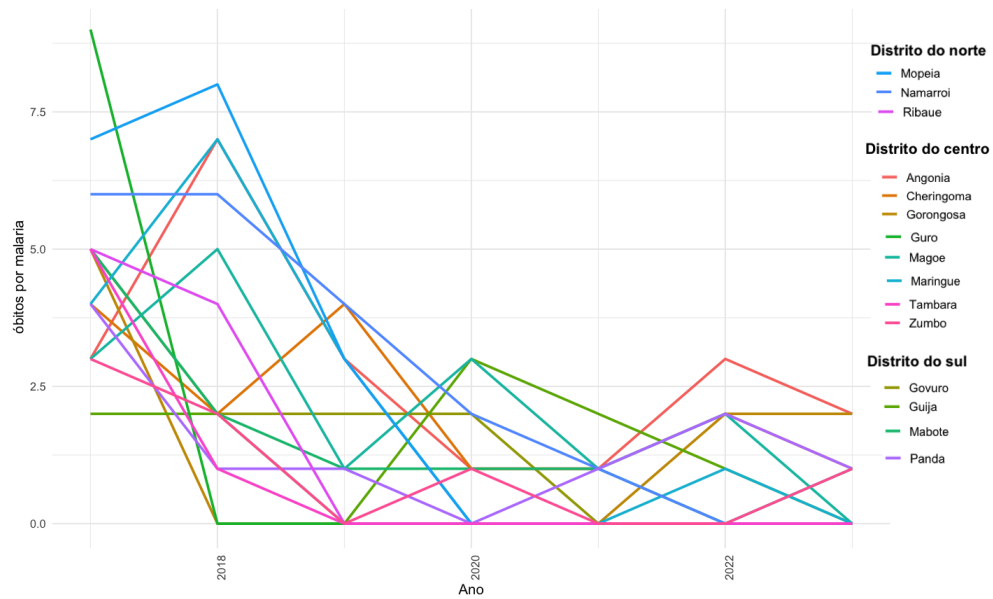


Figura A4: Tendência dos obitos por malária nos distritos com incidência média.

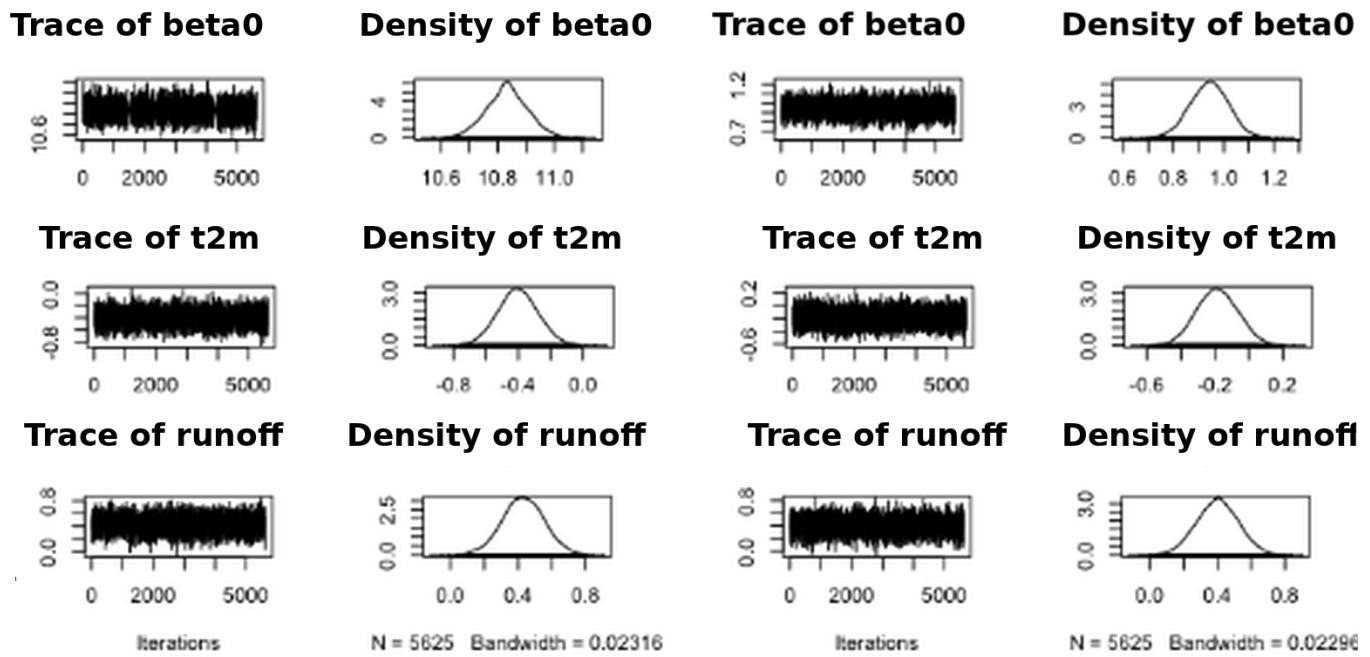


Figura A5: A esquerda, os gráficos de convergência de casos e a direita, óbitos por malária das cadeias de Markov, onde t2m e ruoff representam respectivamente a temperatura média e o escoamento superficial.

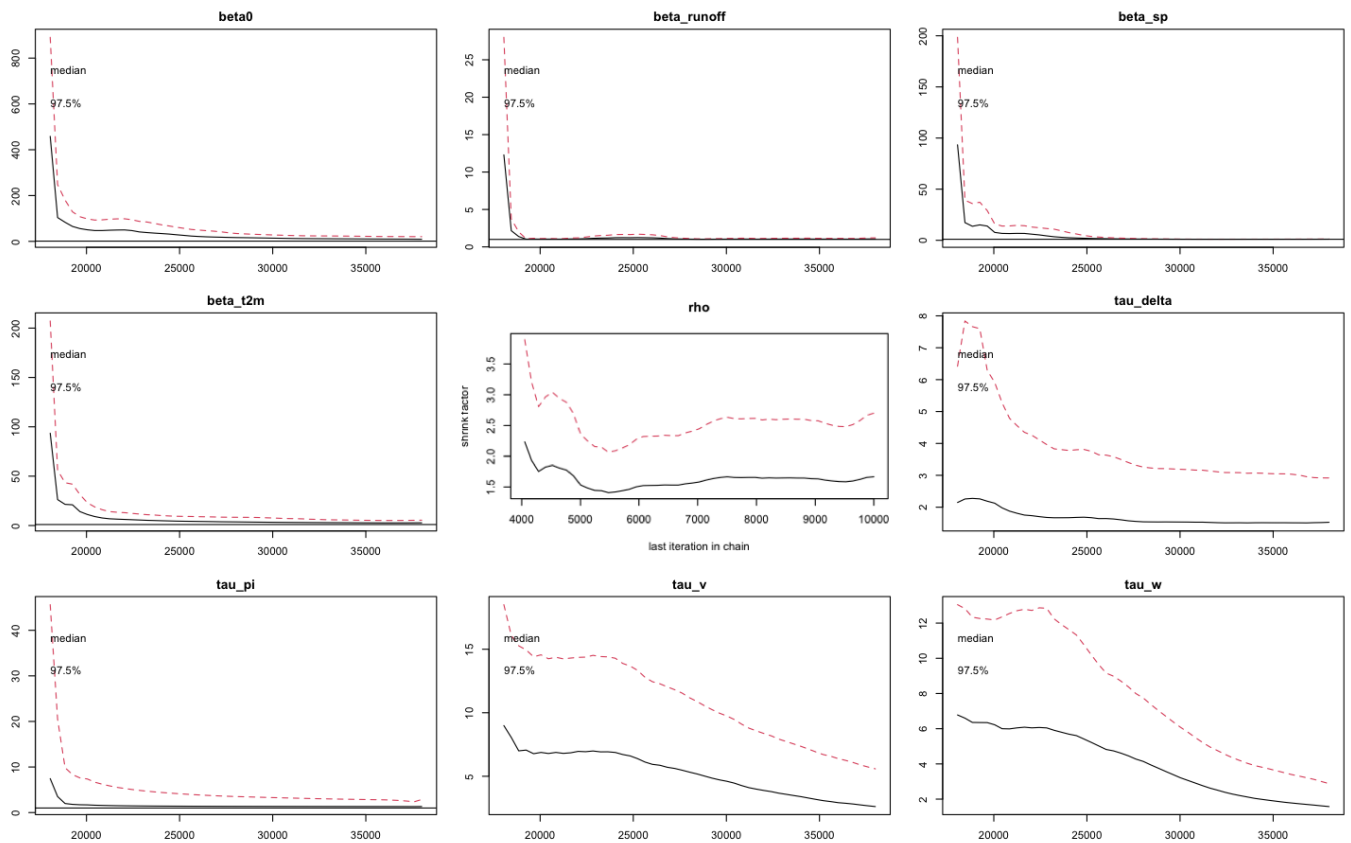


Figura A6: Convergência de parâmetros.

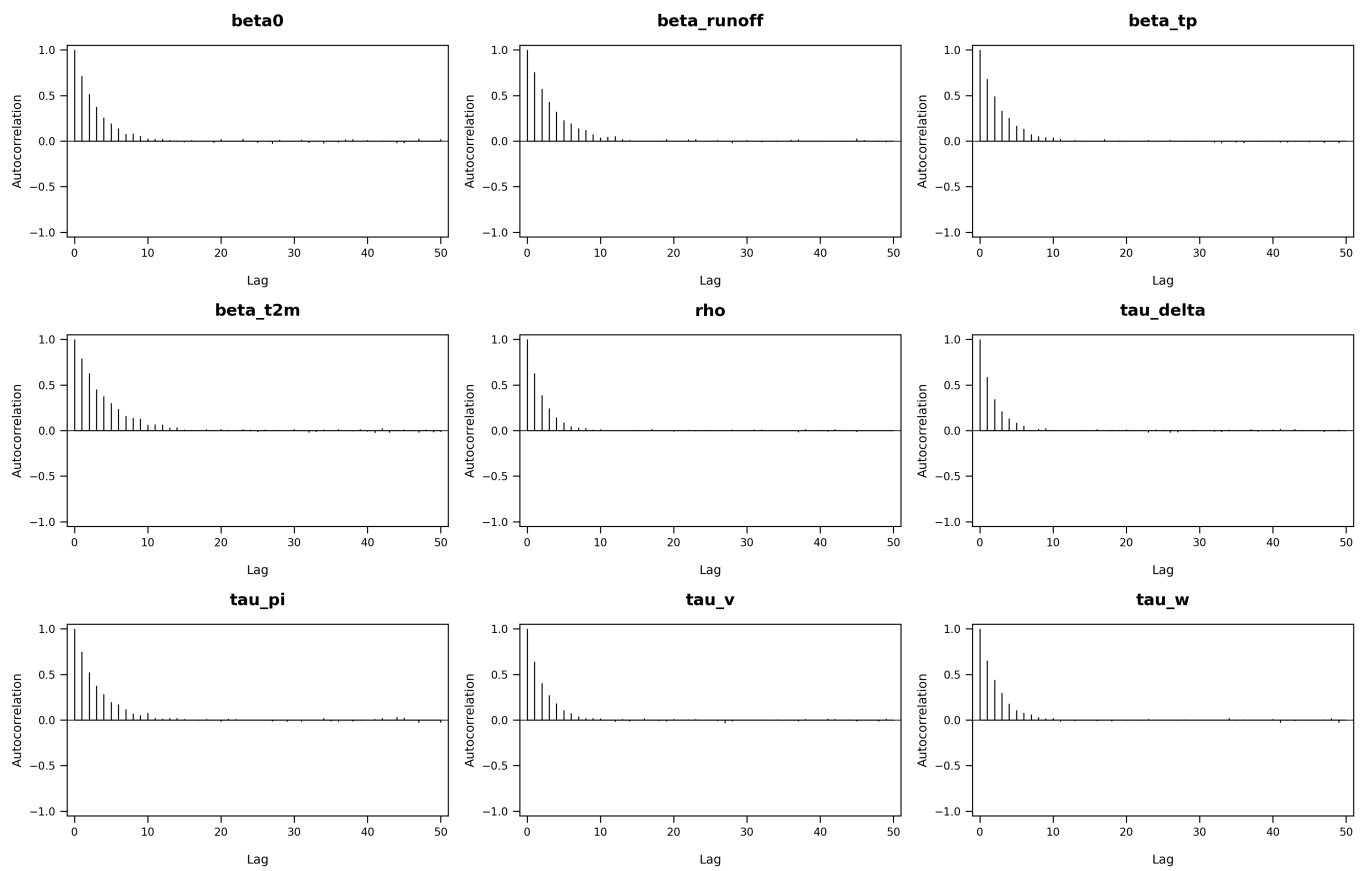


Figura A7: Funções de autocorrelação

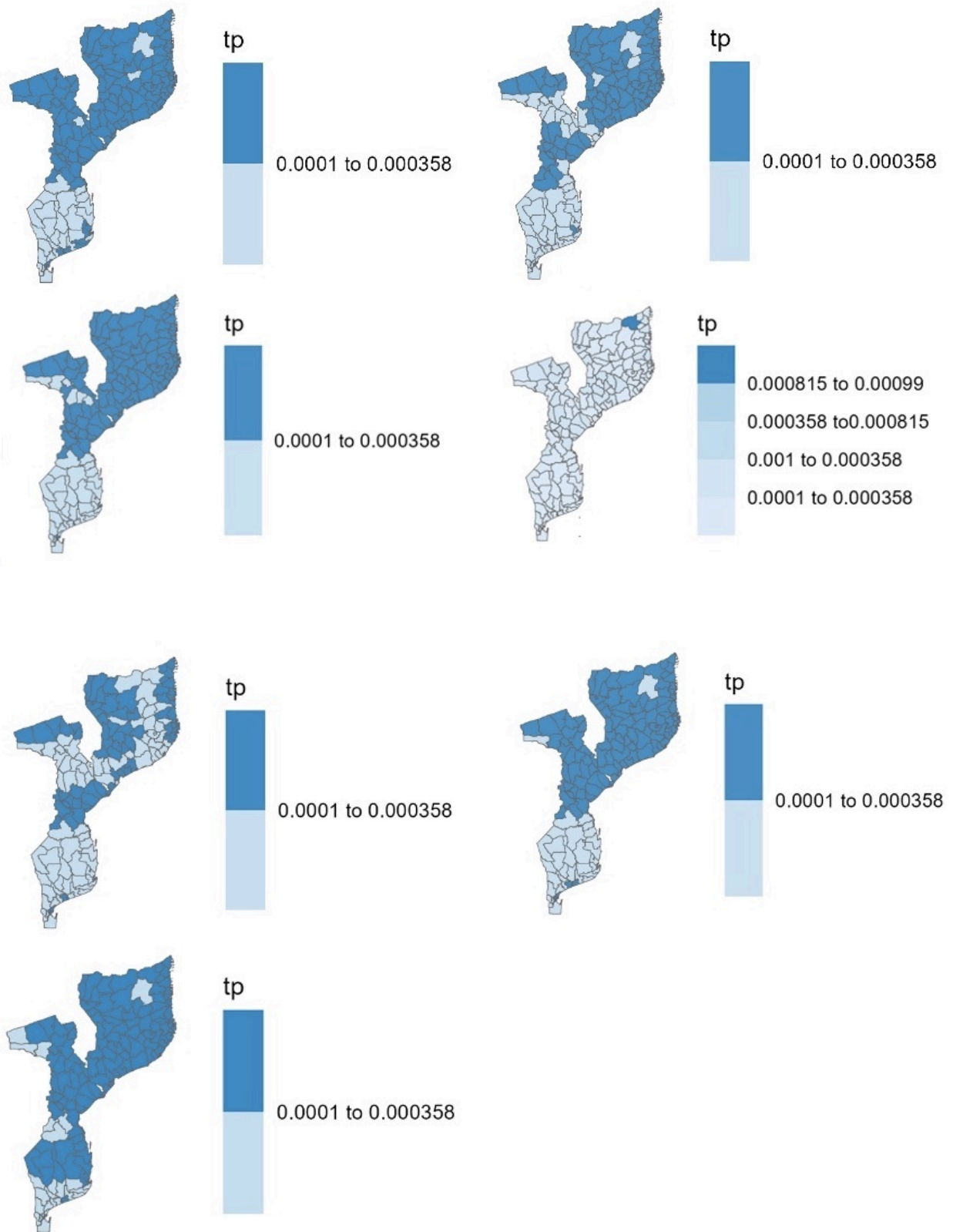


Figura A8: Os mapas ilustram a distribuição espacial da precipitação total no período de 2017 a 2022, apresentados da esquerda para a direita e de cima para baixo. Cada painel representa um ano distinto, permitindo observar a evolução temporal e a variação geográfica da influência das variáveis climáticas na propagação de casos e morbidade por malária ao longo destes seis anos.

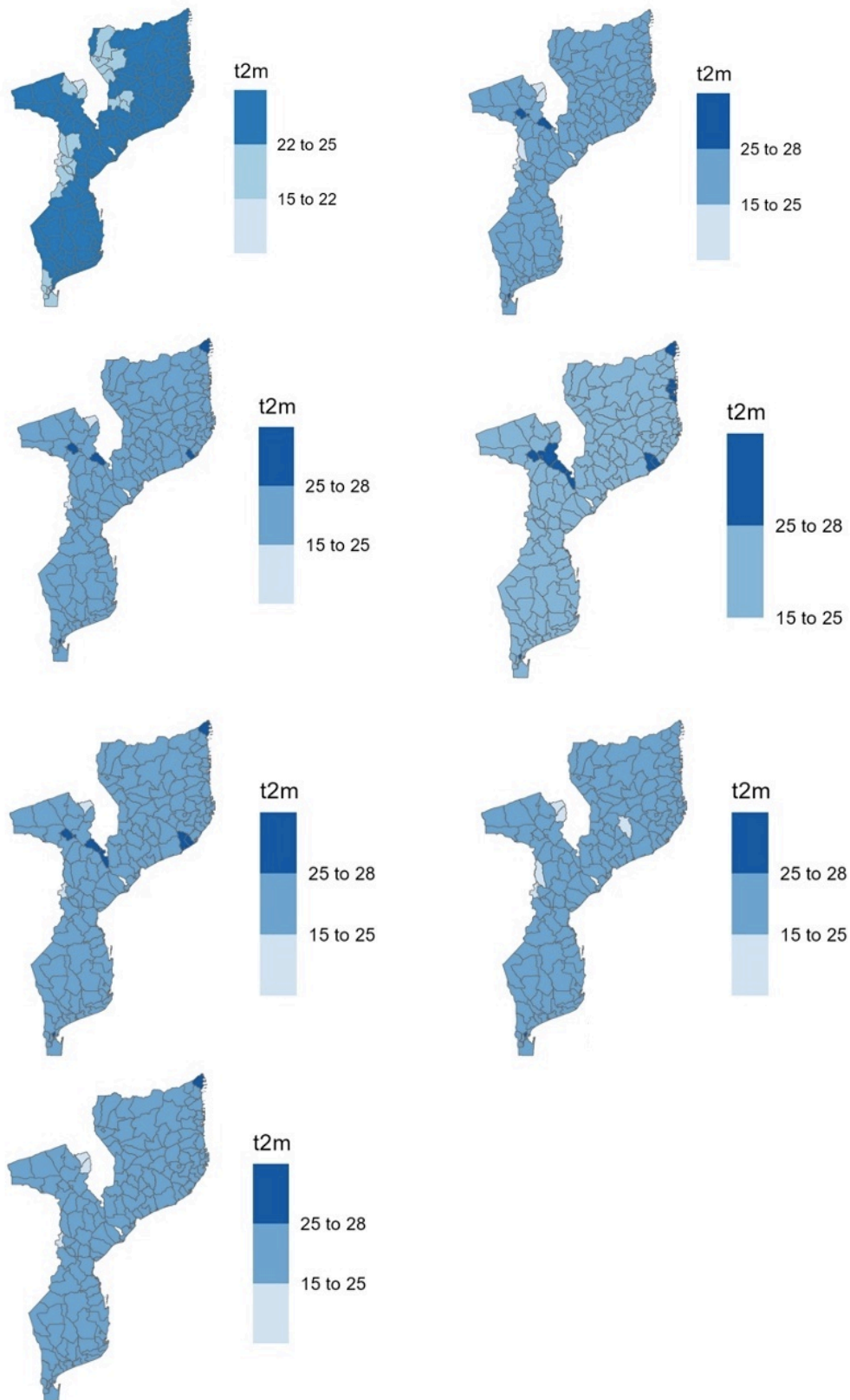


Figura A9: Os mapas ilustram a distribuição espacial da temperatura média total no período de 2017 a 2022, apresentados da esquerda para a direita e de cima para baixo. Cada painel representa um ano distinto, permitindo observar a evolução temporal e a variação geográfica da influência das variáveis climáticas na propagação de casos e morbidade por malária ao longo destes seis anos.

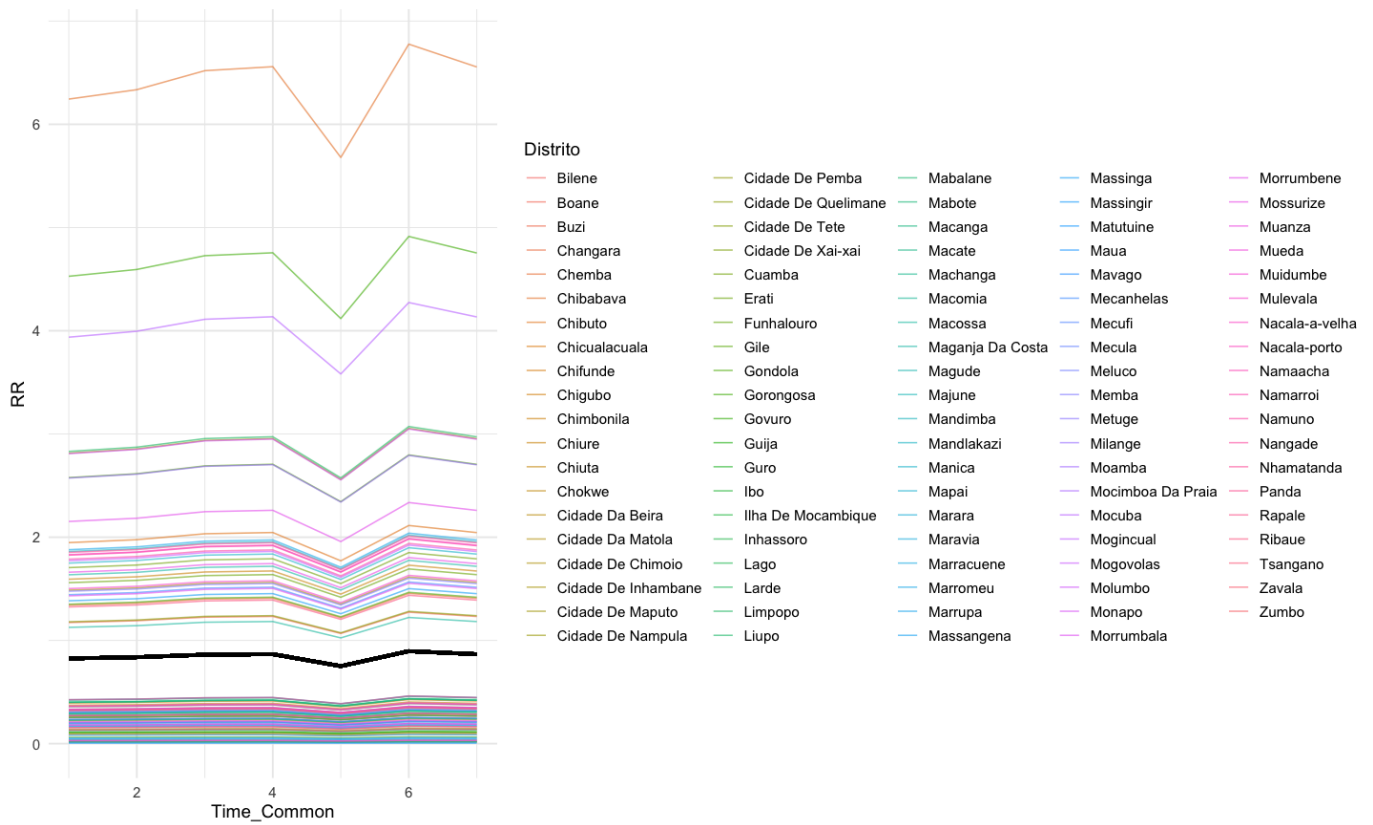


Figura A10: Tendência comum de contaminação por malária em Moçambique no período de 2017 a 2023.

<https://we.tl/t-2DFRDY1uEy>

Acessa o código mistura no link acima