



MÁRCIO ROGER PIAGIO

**Modelo Espacial de Captura com Remoção
para Estimação de Casos Não Notificados em
Epidemiologia**

Maringá - Paraná
2025

MÁRCIO ROGER PIAGIO

Modelo Espacial de Captura com Remoção para Estimação de Casos Não Notificados em Epidemiologia

Dissertação apresentada ao Programa de Pós-graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá como requisito parcial para obtenção do título de mestre em Bioestatística.

Orientador: Prof. Dr. Diogo Francisco Rossoni

Membro do PBE: Prof. Dr. Josmar Mazucheli

Membro Externo: Prof. Dr. George Lucas Moraes Pezzott

Universidade Estadual de Maringá - UEM

Departamento de Estatística - DES

Programa de Pós-Graduação em Bioestatística

Maringá - Paraná

2025

SUMÁRIO

1	Introdução	3
2	Metodologia	6
2.1	Modelo de Remoção	6
2.2	Modelo N-Mistura	8
2.2.1	Modelo para população fechada	8
2.2.2	Modelo dinâmico para população aberta	9
2.3	Modelo proposto	10
2.3.1	Modelo <i>N-mistura</i> Dinâmico com Remoção	11
3	Análise Exploratória dos Dados	14
3.1	Análise dos Padrões de Notificação	14
3.1.1	Análise Temporal do Atraso na Notificação	14
3.1.2	Heterogeneidade Espacial	16
3.1.3	Dinâmica Espaço-Temporal da Dengue	18
3.2	Padrões e Componentes do Modelo Proposto	19
4	Cronograma das atividades do projeto de pesquisa	21
	Referências	23

INTRODUÇÃO

As arboviroses configuram um dos mais significativos desafios virais em saúde pública, impondo obstáculos persistentes devido à sua complexa transmissão entre artrópodes hematófagos e hospedeiros vertebrados (VIGLIETTA et al., 2021). Dentre elas, a dengue destaca-se como uma das doenças mais negligenciadas, apesar de sua crescente carga epidemiológica, somando-se a Zika, febre amarela e chikungunya, em um cenário marcado pelo desafio recorrente no controle vetorial (ARAÚJO et al., 2022). Causada por vírus do gênero *Flavivirus*, a dengue é perpetuada principalmente pelo *Aedes aegypti*, vetor de notória resistência a intervenções urbanas, cuja expansão é agravada por desigualdades socioambientais e falhas nas políticas de saneamento (TORRES-FLORES; REYES-SANDOVAL; SALAZAR, 2022). A existência de quatro sorotipos (DENV-1 a DENV-4), imunologicamente distintos, não apenas dificulta a obtenção de imunidade cruzada duradoura, como também eleva o risco de reinfecções, culminando em formas hemorrágicas e óbitos evitáveis (ALAGARASU et al., 2021). A identificação de um quinto sorotipo (DENV-5) na Malásia em 2007 (MUSTAFA et al., 2015; NORMILE, 2013), ainda pouco estudado e sem medidas de controle específicas após mais de uma década, evidencia a limitação dos sistemas globais de vigilância para antecipar ameaças emergentes, perpetuando estratégias reativas ineficazes diante da evolução viral.

A dengue, anteriormente considerada uma doença tropical com distribuição geográfica limitada, emergiu como um dos maiores desafios de saúde pública em escala global nas últimas décadas. A sua trajetória epidemiológica, antes concentrada no Sudeste Asiático, sofreu uma drástica transformação a partir da década de 1990, com uma rápida dispersão para as Américas, sendo o Brasil um dos países mais afetados por esta nova dinâmica de transmissão (PAIXÃO; TEIXEIRA; RODRIGUES, 2018; BARRETO; TEIXEIRA, 2008). Esta expansão geográfica sem precedentes, agora documentada em todos os continentes habitáveis, incluindo surtos inesperados em regiões da Europa e Austrália (GWEE; CHUA; PANG, 2021), demonstra os limites das atuais estratégias de controle de doenças frente às mudanças ambientais e padrões de urbanização acelerada. Os dados epidemiológicos recentes revelam um quadro preocupante: as epidemias não apenas se expandiram territorialmente, mas também apresentam

maior gravidade clínica, com aumento significativo nas taxas de mortalidade, especialmente a infantil, associada às formas graves da doença (BUCHS et al., 2022; BERBERIAN et al., 2022; MANNA et al., 2022). Este agravamento reflete a complexa interação entre a circulação simultânea de múltiplos sorotipos virais e os desafios nos sistemas de vigilância e atenção à saúde. Os padrões de transmissão, intensificados por períodos de chuvas intensas que favorecem a proliferação de criadouros artificiais em áreas urbanas, somados ao aumento das temperaturas médias que aceleram o ciclo do vetor e a replicação viral (NOR et al., 2021; MORIN; COMRIE; ERNST, 2013; LIU et al., 2023), configuram um cenário epidemiológico cada vez mais complexo, exigindo abordagens inovadoras para seu controle.

Apesar dos avanços em vigilância e diagnóstico, os sistemas de saúde pública continuam enfrentando o desafio crônico da subnotificação. Um número significativo de casos de dengue permanece não registrado, particularmente nas apresentações oligossintomáticas ou assintomáticas, que frequentemente não motivam a busca por atendimento médico (HALSTEAD, 1988). Embora técnicas diagnósticas precisas, como RT-PCR e testes sorológicos, estejam disponíveis, sua utilização permanece desigual entre diferentes regiões e serviços de saúde (ORGANIZATION et al., 2009). Esta lacuna na notificação gera um retrato incompleto da endemia, subestimando sistematicamente a incidência real da doença e comprometendo a eficácia das estratégias de controle. Diante disso, emerge a questão central de pesquisa: como estimar com precisão a carga real da dengue em uma população, considerando a conhecida discrepância entre os casos ocorridos e os efetivamente notificados?

Modelos estatísticos tradicionais, como a técnica de captura-recaptura (LINCOLN, 1930; OTIS et al., 1978), vêm sendo utilizados para estimar o tamanho de populações ocultas, inclusive em estudos epidemiológicos. No entanto, esses modelos baseiam-se em pressupostos difíceis de serem satisfeitos na prática, como a independência entre as fontes de dados e a possibilidade de recaptura do mesmo indivíduo, o que raramente se aplica a sistemas de notificação de base passiva, como o utilizado para a dengue (CARLE; STRUB, 1978). Alternativas metodológicas oriundas da Ecologia têm ganhado espaço nesse contexto. Os modelos de remoção, por exemplo, estimam a abundância com base em contagens sucessivas de indivíduos removidos (ou notificados), assumindo que a população observada vai diminuindo a cada tentativa (ZIPPIN, 1956). Já os modelos *N-mistura*, propostos por Royle (2004), utilizam uma estrutura hierárquica para estimar separadamente a abundância latente de indivíduos e a probabilidade de sua detecção. Essa abordagem foi ampliada por Dail e Madsen (2011) para incorporar dinâmicas populacionais ao longo do tempo, resultando em modelos dinâmicos aplicáveis a contextos com variação temporal na incidência.

Essas ferramentas, ainda que originalmente aplicadas à Ecologia, têm se mostrado promissoras para lidar com os desafios da subnotificação em Epidemiologia. Aplicações recentes incluem a estimação da prevalência real de clamídia (BRINTZ; FUENTES; MADSEN, 2018), a avaliação de subnotificação da COVID-19 (PARKER et al., 2024) e a sensibilidade de testes

rápidos para SARS-CoV-2 (SANDERLIN et al., 2021). A principal vantagem de tais modelos reside na sua capacidade de lidar com dados fragmentados e contagens agregadas, permitindo a estimação de parâmetros epidemiológicos mesmo em situações de observação imperfeita.

No presente trabalho, propõe-se uma extensão do modelo *N-mistura* dinâmico de Dail e Madsen (2011), que incorpora um esquema de remoção sequencial ao processo de observação. A lógica adotada representa a notificação como uma cadeia de tentativas sucessivas de detecção e remoção, refletindo a realidade dos sistemas de vigilância passiva. Essa estrutura busca capturar a ocorrência de notificações tardias, múltiplos canais diagnósticos e a permanência de casos ocultos ao longo do tempo. Espera-se, com isso, desenvolver um modelo estatístico mais realista e flexível, capaz de estimar de forma robusta a verdadeira carga de doenças infecciosas como a dengue em cenários marcados por subnotificação crônica. Ao integrar essa abordagem hierárquica ao contexto epidemiológico brasileiro, o presente estudo pretende contribuir com novas ferramentas analíticas para o aprimoramento da vigilância em saúde pública.

O trabalho está organizado da seguinte forma. No Capítulo 2, descrevem-se os fundamentos metodológicos, abrangendo os modelos de remoção, *N-mistura* e a proposta de modelo estendido com remoção sequencial. O Capítulo 3 realiza a Análise Exploratória dos dados, caracterizando empiricamente os padrões de notificação de dengue através de análises gráficas temporais, estatísticas descritivas e identificação de padrões espaço-temporais. Esta seção evidencia a pertinência do novo modelo proposto, ao demonstrar a sua necessidade a partir de dados empíricos. Finalmente, o Capítulo 4 apresenta as Propostas e Cronograma, estabelecendo o planejamento para implementação prática do modelo, as etapas de validação empírica, um cronograma detalhado para conclusão e defesa do trabalho, além das perspectivas para a aplicação do método em outros contextos epidemiológicos.

METODOLOGIA

Este capítulo descreve alguns dos principais modelos propostos na literatura para estimação de tamanhos de populações não marcadas. Inicialmente, discutimos o modelo de captura com remoção (RIVERA; MCCREA, 2021) e, em seguida, os Modelos da classe N-mistura (MADSEN; ROYLE, 2023), especificamente a formulação original de Royle (2004) para população fechada e sua extensão para populações abertas seguindo o modelo N-mistura dinâmico de Dail e Madsen (2011). Cabe ressaltar que, embora os modelos possam ser escritos covariáveis espaço-temporais, apresentamos as formas sem heterogeneidade nos parâmetros para fins de ilustração. Mantivemos a terminologia clássica dos modelos para população animal. No final do capítulo, interpretamos a modelagem e parâmetros sobre o contexto epidemiológico e apresentamos a formulação do modelo estatístico proposto neste trabalho.

2.1 MODELO DE REMOÇÃO

O modelo de remoção (do inglês, *removal model*) é amplamente utilizado para estimar o número de elementos de uma população a partir de sucessivas tentativas de captura (épocas/períodos de amostragem), nas quais os indivíduos capturados são removidos da população, isto é, eles não são devolvidos após a captura até o final do estudo. A ideia central reside na redução progressiva da população observável, gerando um padrão decrescente de contagens.

A técnica de captura com remoção é aplicada em diversos contextos, principalmente em populações onde o processo tradicional de marcação nos modelos de captura-recaptura apresenta alto custo financeiro (por exemplo, DNA, câmera-trap ou tags para tigres e ursos) ou impraticável (populações que trocam de pele e perdem a marcação, por exemplo). Um outro exemplo de captura com remoção é muito utilizado na engenharia de qualidade, onde um número de itens defeituosos são inspecionados por diversos revisores e, a cada item defeituoso encontrado, ele é corrigido, isto é, não volta para a população de 'item defeituosos'.

Considere uma população de N indivíduos que é submetida a T ocasiões de amostragem. Considere que a população é fechada, isto é, existem os mesmos N indivíduos ao longo

das T ocasiões de amostragem, sem nascimentos, mortes e/ou migrações. Em cada ocasião de amostragem, todos os indivíduos capturados são removidos da população antes da próxima ocasião. Suponha que cada indivíduo tem probabilidade p de ser capturado em uma determinada ocasião de amostragem, dado que não foi previamente capturado. Logo, considerando que $0 < p < 1$, temos que $1 - p$ dos indivíduos ainda presentes na população permanecerão não observados após cada ocasião.

Para o processo observacional, seja y_t o número de indivíduos capturados na t -ésima ocasião de captura. Supondo independência entre indivíduos, temos a estrutura hierárquica via captura com remoção definida por

$$\begin{aligned} y_1 | N, p &\sim \text{Binomial}(N, p) \\ y_t | N, p, n_t &\sim \text{Binomial}(N - n_t, p) \quad \text{para } t = 2, 3, \dots, T, \end{aligned} \quad (2.1)$$

onde $n_t = y_1 + \dots + y_{t-1}$ é o número de indivíduos capturados (e, conseqüentemente, removidos) até a ocasião $t - 1$. Observe que a probabilidade do indivíduo ser capturado na ocasião t é $(1 - p)^{t-1}p$ e a probabilidade dele não ser capturado em ocasião alguma é $(1 - p)^T$. De fato, o processo hierárquico acima pode ser representado pela distribuição conjunta multinomial na forma

$$(y_1, \dots, y_T) | N \sim \text{Multinomial} \left\{ N; (p, (1 - p)p, (1 - p)^2p, \dots, (1 - p)^{T-1}p, (1 - p)^T) \right\}$$

com $y_t \geq 0$ e $y_1 + \dots + y_T \leq N$. Com base nos dados observados (y_1, \dots, y_T) , utiliza-se o modelo Multinomial acima para estimar os parâmetros N e p .

Suponha agora que o processo de amostragem acima é replicado em R regiões disjuntas. Logo, para cada região com N_r indivíduos e probabilidade de captura p_r , as frequências de observação $\mathbf{y}_r = (y_{r,1}, \dots, y_{r,T})$ segue o modelo Multinomial na forma

$$(y_{r,1}, \dots, y_{r,T}) | N_r \sim \text{Multinomial} \left\{ N_r; (p_r, (1 - p_r)p_r, (1 - p_r)^2p_r, \dots, (1 - p_r)^{T-1}p_r, (1 - p_r)^T) \right\}$$

com $y_{t,r} \geq 0$ e $y_{r,1} + \dots + y_{r,T} \leq N_r$. Pode-se realizar inferências para N_r e p_r separadamente para cada região utilizando a função de verossimilhança local dada por

$$\begin{aligned} L_r(N_r, p_r | \mathbf{y}_r) &= p(y_{r,1}, \dots, y_{r,T} | N_r, p_r) \\ &\propto \frac{N_r!}{(N_r - n_r)!} \prod_{t=1}^T [(1 - p_r)p_r]^{y_{r,t}} [(1 - p_r)^T]^{N_r - n_r} \end{aligned}$$

onde $n_r = y_{r,1} + \dots + y_{r,T}$ é o número total de animais capturados na região. Todavia, é natural que informações entre as regiões sejam compartilhadas no processo inferencial. Neste caso, a função de verossimilhança pode ser escrita como

$$L(N_1, \dots, N_R, p_1, \dots, p_R | \mathbf{y}) = \prod_{r=1}^R L_r(N_r, p_r | \mathbf{y}_r).$$

Dorazio, Jelks e Jordan (2005) adotam uma estrutura hierarquica $N_r \sim \text{Poisson}(\lambda)$ e $p_r \sim \text{Beta}(a, b)$. Neste caso, a função de verossimilhança marginal para os parâmetros (λ, a, b) é dada por

$$L(\lambda, a, b | \mathbf{y}) = \prod_{r=1}^R \left\{ \sum_{N_r=n_r}^{\infty} \int L_r(N_r, p_r | \mathbf{y}) f(p_r | a, b) dp_r P(N_r = n_r) \right\}$$

Para modelar heterogeneidade de abundâncias locais, Royle e Dorazio (2006) adotam uma abordagem Bayesiana não paramétrica supondo $N_r | \lambda_r \sim \text{Poisson}(\lambda_r)$, $\phi_r = \log \lambda_r \sim G$, sendo G seguindo um Processo Dirichlet (FERGUSON, 1973).

Por outro lado, a modelagem conjunta entre regiões torna-se necessária quando se obtém covariáveis regionais. Por exemplo, Royle e Dorazio (2006) adotam uma estrutura hierarquica $N_r | \lambda_r \sim \text{Poisson}(\lambda_r)$, e utiliza-se uma formulação de regressão Poisson $\log \lambda_r = \mathbf{w}_r' \alpha$, em que \mathbf{w}_r é um vetor de covariáveis e α é o vetor de parâmetros. Fontes estocásticas de variação na abundância média (implicando variação extra-Poisson) podem ser especificadas via distribuições adicionais, como um prior normal multivariado:

$$\phi \sim \text{Normal}(\mathbf{W}\alpha, \Sigma_\phi),$$

com $\phi_r = \log \lambda_r$ e $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n)'$.

A matriz de covariância Σ_ϕ pode assumir diferentes formas, como:

$$\Sigma_\phi = \sigma_\phi^2 \mathbf{I} \quad \text{ou} \quad \Sigma_\phi = \sigma_\phi^2 \mathbf{R}_\phi,$$

onde \mathbf{I} é a matriz identidade e \mathbf{R}_ϕ incorpora dependência espacial. Adicionalmente, a variação espacial na probabilidade de detecção pode ser modelada de forma análoga à abundância por uma regressão logística $\text{logit } p_r = \mathbf{v}_r' \beta$ para modelar a variação sistemática da detecção em função de covariáveis \mathbf{v}_r e parâmetros β .

2.2 MODELO N-MISTURA

2.2.1 MODELO PARA POPULAÇÃO FECHADA

O modelo N-mistura de Royle (2004) foi originalmente proposto para estimar a abundância de populações animais em múltiplas regiões a partir de contagens repetidas com detecção imperfeita. A diferença do modelo N-mistura para Modelos com Remoção é que o processo de captura não remove os animais capturados como é feito nos modelos com remoção, isto é, é um processo puramente observacional sem intervenção na população pois realiza-se apenas a contagem de animais.

Considerando a notação previamente adotada, temos que no modelo N-mistura proposto por Royle (2004), o número de animais observados na região r no tempo t tem distribuição

$$y_{r,t} | N_r \sim \text{Binomial}(N_r, p), \quad r = 1, 2, \dots, R, \quad t = 1, 2, \dots, T.$$

Este modelo considera população fechada, isto é, o número de animais N_r na região r é constante durante todas as T ocasiões de observação, assim como o modelo de remoção descrito anteriormente.

O modelo considera um processo latente para N_r supondo

$$N_r \sim \text{Poisson}(\lambda), \quad i = 1, \dots, R,$$

onde λ representa a média populacional de indivíduos por região. Como os valores de N_r são desconhecidos, a verossimilhança marginal para os parâmetros (λ, p) é obtida somando sobre todos os valores possíveis para as variáveis latentes N_1, \dots, N_R .

Considerando independência condicional entre as regiões, temos que a função de verossimilhança é escrita na forma:

$$L(\lambda, p | \mathbf{y}) = \prod_{r=1}^R L_r(\lambda, p | \mathbf{y}_r)$$

onde

$$L_r(\lambda, p | \mathbf{y}_r) = \sum_{N_r=y_{i+}}^{\infty} \left[\prod_{t=1}^T \binom{N_r}{y_{r,t}} p^{y_{r,t}} (1-p)^{N_r-y_{r,t}} \right] \cdot \frac{e^{-\lambda} \lambda^{N_r}}{N_r!} \quad (2.2)$$

com $y_{r+} = \max_t \{y_{r,t}\}$ representando o máximo observado de animais na região r .

2.2.2 MODELO DINÂMICO PARA POPULAÇÃO ABERTA

O modelo proposto por Dail e Madsen (2011) estende a estrutura original de Royle (2004) para incorporar dinâmica populacional ao longo do tempo, permitindo estimar a abundância latente de uma população sujeita a processos de recrutamentos (nascimentos, imigração, etc) e perdas (mortes, emigração, etc) entre períodos consecutivos.

A modelagem considera que as abundância iniciais $N_{r,1}$ nas regiões seguem o modelo Poisson, assim como o modelo N-mistura clássico de (ROYLE, 2004):

$$N_{r,1} \sim \text{Poisson}(\lambda), \quad r = 1, 2, \dots, R. \quad (2.3)$$

Nos períodos seguintes, o número de indivíduos $N_{r,t+1}$ decorre da soma de dois processos:

$$N_{r,t+1} = S_{r,t+1} + R_{r,t+1} \quad (2.4)$$

onde $S_{r,t+1}$ é o número de indivíduos presentes na população no tempo t que sobreviveram até o tempo $t+1$, e $R_{r,t+1}$ é o número de indivíduos recrutados no tempo $t+1$, isto é, eles não estavam presentes no tempo t mas foram recrutados (nascimentos, imigração) para o tempo $t+1$.

Consequente, adotam-se as seguintes distribuições:

$$S_{r,t+1} | N_{r,t} \sim \text{Binomial}(N_{r,t}, \phi) \quad (2.5)$$

e

$$R_{r,t+1} \mid N_{r,t} \sim \text{Poisson}(N_{r,t} \gamma) \quad (2.6)$$

onde ϕ representa a taxa de sobrevivência e os novos indivíduos surgem no sistema com taxa proporcional γ à abundância anterior.

Assumindo que $S_{r,t}$ e $R_{r,t}$ são independentes, a probabilidade de transição em um passo do estado $N_{r,t-1} = j$ para o estado $N_{r,t} = k$, denotada por P_{jk} , é obtida usando a convolução discreta:

$$P_{jk} = \sum_{c=0}^{\min\{j,k\}} \binom{j}{c} \omega^j (1-\omega)^{j-c} \frac{e^{-\gamma j} (\gamma j)^{k-c}}{(k-c)!}$$

Supondo independência condicional entre as regiões, temos a função de verossimilhança para os parâmetros $(p, \lambda, \gamma, \phi)$ é dada por

$$L(\lambda, p | \mathbf{y}) = \prod_{r=1}^R L_r(\lambda, p, \gamma, \phi | \mathbf{y}_r)$$

onde

$$L_r(p, \lambda, \gamma, \omega \mid \{\mathbf{y}_r\}) = \sum_{N_{r,1}=y_{r,1}}^{\infty} \cdots \sum_{N_{r,T}=y_{r,T}}^{\infty} \left\{ \left(\prod_{t=1}^T \binom{N_{r,t}}{y_{r,t}} p^{y_{r,t}} (1-p)^{N_{r,t}-y_{r,t}} \right) \times \right. \quad (2.7)$$

$$\left. \times \frac{e^{-\lambda} \lambda^{N_{r,1}}}{N_{r,1}!} \cdot \prod_{t=2}^T P_{N_{r,t-1}, N_{r,t}} \right\}. \quad (2.8)$$

2.3 MODELO PROPOSTO

Nesta seção, os modelos para estimação de abundância populacional são reinterpretados sob a perspectiva epidemiológica com relação ao número de indivíduos contaminados por uma determinada característica 'A' de interesse (por exemplo, Dengue, SarCov-2 ou Influenza) com base em casos confirmados nas bases do sistema de saúde. Novamente, suponha que a região de estudo é particionada em R região (por exemplo, estados da federação, cidades ou regionais de saúde). A evolução dos casos contaminados pela característica 'A' nos R região serão acompanhados por T períodos (por exemplo, cada período é um dia, um mês ou uma semana epidemiológica).

Considere $N_{r,t}$ o número de casos em uma determinada região r em um determinado tempo t . Cada indivíduo destes entrará no sistema como caso notificado com probabilidade p no tempo t . Interpretamos esse parâmetro como a taxa de notificação que condiz aos indivíduos que buscaram atendimento de saúde, fizeram o teste para a característica, o teste deu resultado positivo e seu nome foi registrado no sistema. Neste caso, como o teste deu positivo, o indivíduo é considerado 'capturado' na terminologia anteriormente adotada e, uma vez positivado, ele fará o tratamento específico e não mais precisará refazer o teste, ou seja, ele é 'removido' da população.

Observe que se essa taxa p se mantiver ao longo dos períodos seguintes e os $N_{r,t}$ ficarem contaminados indefinidamente, temos o modelo clássico de captura com remoção apresentado na Seção 2.1. Contudo, a hipótese de população fechada dos modelos de remoção é extremamente forte para estudos epidemiológicos. Para exemplificar esta afirmação, suponha que cada período t seja uma semana epidemiológica, e que nesta semana tenhamos $y_{r,t}$ casos notificados. Logo, $N_{r,t} - y_{r,t}$ indivíduos não foram notificados e estariam aptos a serem notificados no tempo $t + 1$. Contudo, é natural que os sintomas comecem a diminuir ou até mesmo desaparecer em uma ou duas semanas, fazendo com que uma parcela dos casos não observados ‘desapareçam’ da população entre períodos, o que pode então ser caracterizado como modelagem de população aberta.

Para população aberta, apresentamos na Seção 2.2.2 o modelo de Dail e Madsen (2011) que modela a taxa ϕ de sobreviventes entre duas ocasiões consecutivas por $S_{r,t+1} \sim \text{Binomial}(N_{r,t} - y_{r,t}, \phi)$. Contudo, este modelo é observacional (sem marcação e remoção), e a modelagem considera que os sobreviventes se somam aos novos casos $R_{r,t+1}$ no tempo $t + 1$ por $N_{t+1,r} = S_{r,t+1} + R_{r,t+1}$ como definido na equação (2.4).

Por outro lado, os dados de notificação de dengue possuem uma característica adicional em relação aos modelos observacionais: a data de início dos sintomas é registada num sistema oficial, como o SINAN (Sistema de Informação de Agravos de Notificação) no Brasil, e, portanto, é possível saber em que período t o indivíduo foi infetado. Consequentemente, torna-se possível distinguir os ‘sobreviventes’ dos ‘novos recrutados’. Por exemplo, suponha que, numa região r na semana 3 ($t = 3$), dois indivíduos testem positivo. O indivíduo 1 informa que os sintomas iniciaram na semana 3, enquanto o indivíduo 2 relata que os seus sintomas começaram na semana 1. Logo, o indivíduo 1 deve ser contabilizado em $N_{r,3}$ (como um novo ‘recrutado’), e o indivíduo 2 é contabilizado em $N_{r,1}$ (pois é um ‘sobrevivente’ desde o primeiro período). Neste cenário, a distinção entre sobreviventes e recrutados é explícita. Por isso, ajustar um modelo como o de Dail e Madsen (2011) levaria à perda de precisão nas estimativas, pois este não foi desenhado para incorporar tal informação.

Diante do exposto, este trabalho propõe na próxima seção um novo modelo N -mixture com remoção para dados epidemiológicos com registro da data da notificação e data do contágio.

2.3.1 MODELO N -MISTURA DINÂMICO COM REMOÇÃO

Considere $y_{r,t}^k$ o número de casos notificados na região r no tempo k com início dos sintomas no tempo t , para $k \geq t$. Em outras palavras, o índice t informa a data de contágio e o índice k informa o período que foi registrado no sistema. Observe que o índice k diferencia os sobreviventes dos recrutados, como previamente discutido. Além disso, a utilização deste índice adicional gera um novo conjunto de dados, não suportado pelos modelos N -mixture de Royle (2004) e Dail e Madsen (2011).

Seja $N_{r,t}$ o número de indivíduos da região r que contraíram a característica no período t . Vale ressaltar que $N_{r,t}$ não contempla os indivíduos contaminados em tempos t' ($t' < t$) que permaneceram contaminados até o tempo t . Adotamos a seguinte dinâmica latente do número de casos ao longo do tempo, com possibilidade de crescimento ou declínio em função de parâmetros epidemiológicos:

$$N_{r,1} \sim \text{Poisson}(\lambda_r) \quad (2.9)$$

$$N_{r,t} \mid N_{r,t-1} \sim \text{Poisson}(\gamma \cdot N_{r,t-1}), \quad t \geq 2 \quad (2.10)$$

onde

- λ_r : valor esperado do número inicial de casos na localidade i ;
- γ : taxa per capita de crescimento ou declínio do número de casos entre períodos sucessivos.

Consequente, adotamos uma nova estrutura de captura com remoção para população aberta. Inicialmente, para o período que aconteceu o contágio, temos a contagem de número de casos notificados modelado por

$$y_{r,t}^t \sim \text{Binomial}(N_{r,t}, p)$$

O parâmetro p informa a taxa de notificação dos casos ativos.

Devido ao atraso na notificação, para os indivíduos contaminados no tempo t , o registro pode ocorrer em um tempo k ($k > t$). Contudo, adicionamos ao modelo de captura com remoção uma estrutura de ‘cura’ dos indivíduos entre consecutivos períodos k e $k + 1$ através do parâmetro ϕ : probabilidade de um indivíduo com a doença no tempo k permanecer contaminado (e passível de ser capturado) no tempo $k + 1$. A terminologia ‘passível’ indica, por exemplo, os casos de indivíduos com sintomas ou suspeitas de contágio, que ainda podem realizar um teste para a característica. Interpretamos $1 - \phi$ como a fração de ‘cura’, isto é, a proporção de indivíduos passíveis de serem capturados no tempo k mas que não foram capturados e se ‘curaram’ no tempo $k + 1$ (desistiram de fazer um teste, por exemplo).

Consequente, assumimos que se um indivíduo apresenta teste positivo em um determinado período k , o mesmo não irá refazer o teste futuramente, isto é, ele é capturado e removido da população. Logo, temos que, para os indivíduos que contraíram a característica no tempo t , eles serão observados no tempo k com probabilidade θ_k^t , com

- $\theta_k^t = p$ se $k = t$ (capturado no mesmo período do contágio) e

$$\blacksquare \theta_k^t = \phi^{k-t}(1-p)^{k-t}p, \text{ para } k = t+1, t+2, \dots, T$$

Observe que, para o indivíduo contaminado no tempo t ser capturado no tempo k é necessário que ele permaneça contaminado (ϕ^{k-t}) até o tempo k e que não seja capturado ($[1-p]^{k-t}$) antes da ocasião k , para então ser efetivamente capturado (p) na ocasião k . É possível observar o vetor

$$(\theta_t^t, \theta_{t+1}^t, \dots, \theta_T^t) = (p, \phi(1-p)p, \phi^2(1-p)^2p, \dots, \phi^{T-t}(1-p)^{T-t}p)$$

Adicionalmente, a probabilidade de um indivíduo contaminado no tempo t nunca ser capturado é dado por

$$1 - \sum_{k=t}^T [\phi(1-p)]^{k-t}p = \frac{p([\phi(1-p)]^{T-t+1} - 1)}{\phi(1-p) - 1}$$

Finalmente, para uma região r e um tempo inicial de contágio t , os registros $\mathbf{y}_{r,t} = (y_{r,t}^t, \dots, y_{r,t}^T)$ seguem um modelo Multinomial definido por

$$\mathbf{y}_{r,t} \sim \text{Multinomial}\left(N_{r,t}, (\theta_t, \dots, \theta_T^t)\right)$$

com $n_{r,t} = y_{r,t}^t + \dots + y_{r,t}^T \leq N_{r,t}$ é o número total de registros para os indivíduos da região r com contágio no tempo t .

ANÁLISE EXPLORATÓRIA DOS DADOS

3.1 ANÁLISE DOS PADRÕES DE NOTIFICAÇÃO

A presente análise exploratória utiliza como base de dados as 452.650 notificações de dengue registradas no Sistema de Informação de Agravos de Notificação (SINAN) para o estado do Paraná, abrangendo o período de 2021 a 2023. O conjunto de dados original, composto por 121 variáveis, foi submetido a um processo de seleção e tratamento, resultando em três informações essenciais para o estudo: (i) a data de início dos primeiros sintomas, (ii) a data de notificação do caso e (iii) a Regional de Saúde (RS) de ocorrência – totalizando 22 RS no estado. Essa seleção permitiu calcular o intervalo de atraso na notificação, métrica central desta pesquisa, além de viabilizar a análise dos padrões temporais e da heterogeneidade espacial entre as diferentes regionais de saúde. Tais resultados constituem a base empírica para o modelo proposto.

3.1.1 ANÁLISE TEMPORAL DO ATRASO NA NOTIFICAÇÃO

As Figuras 1 e 2 revelam que, embora a maioria das notificações ocorra nos primeiros dias após o início dos sintomas, existe uma cauda longa e persistente na distribuição dos atrasos. A Figura 1 mostra que, apesar do pico de notificações se concentrar nos primeiros 10 dias, um número expressivo de casos ainda é registrado mais de 60 dias após o início dos sintomas. Em escala semanal, a Figura 2 evidencia que o processo de notificação se prolonga por mais de 24 semanas (6 meses). Esse padrão sugere que a probabilidade de detecção decai lentamente ao longo do tempo, o que faz com que uma fração considerável dos casos permaneça invisível ao sistema por múltiplos períodos antes de ser, eventualmente, notificada.

Do ponto de vista clínico, muitos casos de dengue são assintomáticos ou apresentam sintomas leves e inespecíficos, o que pode retardar a procura por atendimento médico. Além disso, em contextos de co-circulação de diferentes arboviroses (como Zika e Chikungunya), o diagnóstico diferencial se torna mais complexo, contribuindo para atrasos na confirmação e notificação. Do ponto de vista social, barreiras no acesso aos serviços de saúde, especialmente

em regiões periféricas ou com infraestrutura precária, podem postergar o contato inicial com o sistema de vigilância. Fatores como desinformação, baixa percepção de risco, medo de estigmatização ou desconfiança nas instituições públicas também podem contribuir para o adiamento da busca por atendimento. Do ponto de vista institucional, atrasos operacionais dentro do próprio sistema de vigilância epidemiológica também são determinantes. Processos burocráticos, escassez de profissionais treinados, sobrecarga durante surtos e inconsistências nos fluxos de informação entre unidades de saúde e bases de dados centralizadas (como o SINAN) afetam diretamente o tempo entre o início dos sintomas e a efetiva notificação.

Figura 1 – Distribuição Diária do Tempo até Notificação

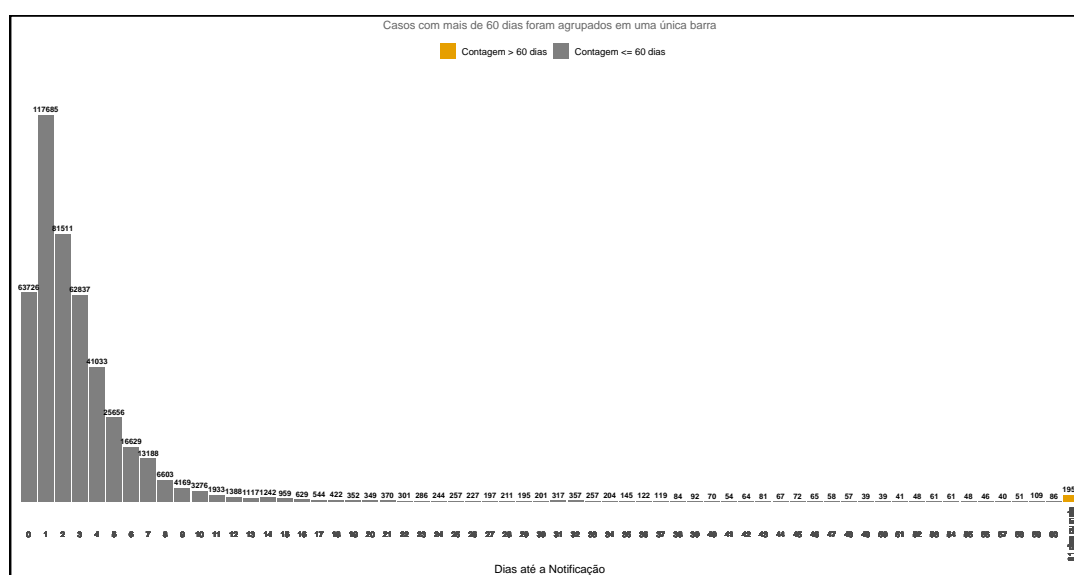
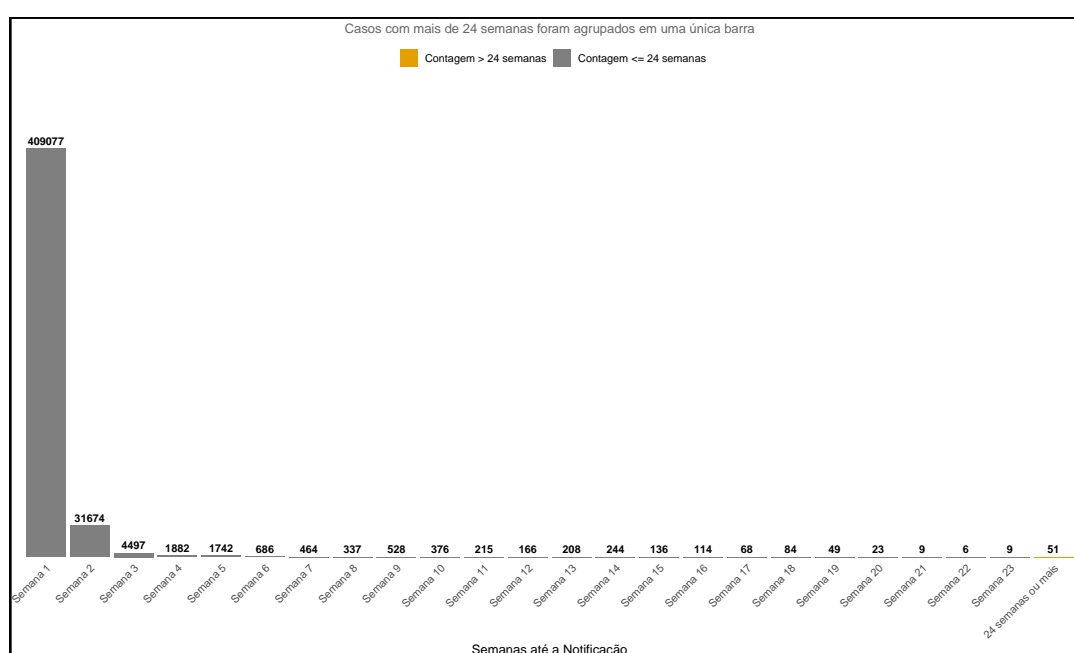


Figura 2 – Distribuição Semanal do Tempo até Notificação



3.1.2 HETEROGENEIDADE ESPACIAL

A Figura 3 evidencia uma desigualdade geográfica acentuada na distribuição dos casos de dengue. Observam-se “hotspots” de alta incidência em determinadas Regionais de Saúde (RS), como Maringá (15ª RS), Londrina (17ª RS) e Cianorte (13ª RS), enquanto outras regiões apresentam uma carga de doença significativamente menor. Essa heterogeneidade espacial indica que a epidemia de dengue não se distribui de forma uniforme no território, o que pode estar relacionado a fatores ambientais, demográficos e sociais, bem como à presença e eficácia de estratégias locais de controle vetorial.

As Figuras 5 e 4 complementam essa análise ao mostrar que o desempenho do sistema de vigilância epidemiológica também varia substancialmente entre regiões. Em Paranaguá (1ª RS), por exemplo, o tempo médio entre o início dos sintomas e a notificação é de aproximadamente 3,5 dias, enquanto na Região Metropolitana (2ª RS) esse valor sobe para cerca de 6,5 dias. Essa diferença sugere que a agilidade do sistema de detecção e registro não é homogênea e é influenciada por fatores locais que vão além da carga de doença em si.

Essa dissociação entre a incidência e a rapidez do sistema de notificação aponta para a existência de fatores estruturais que afetam a capacidade de resposta local. Do ponto de vista institucional, a capacidade instalada do sistema de saúde, a efetividade da atenção primária e a organização dos fluxos de informação são decisivas: regiões com maior infraestrutura e processos informatizados respondem mais rápido, enquanto áreas com sistemas fragmentados enfrentam atrasos operacionais. Paralelamente, o contexto social e político mostra-se determinante, pois a atuação de gestores comprometidos e a mobilização social fortalecem a vigilância, ao passo que a falta de priorização política resulta em sistemas passivos. Por fim, do ponto de vista clínico, a própria dinâmica dos surtos agrava o problema, uma vez que a sobrecarga nos serviços de saúde leva à priorização do atendimento ao paciente em detrimento do registro formal e imediato dos casos.

Figura 3 – Total de casos de Dengue no Paraná por Regional de Saúde

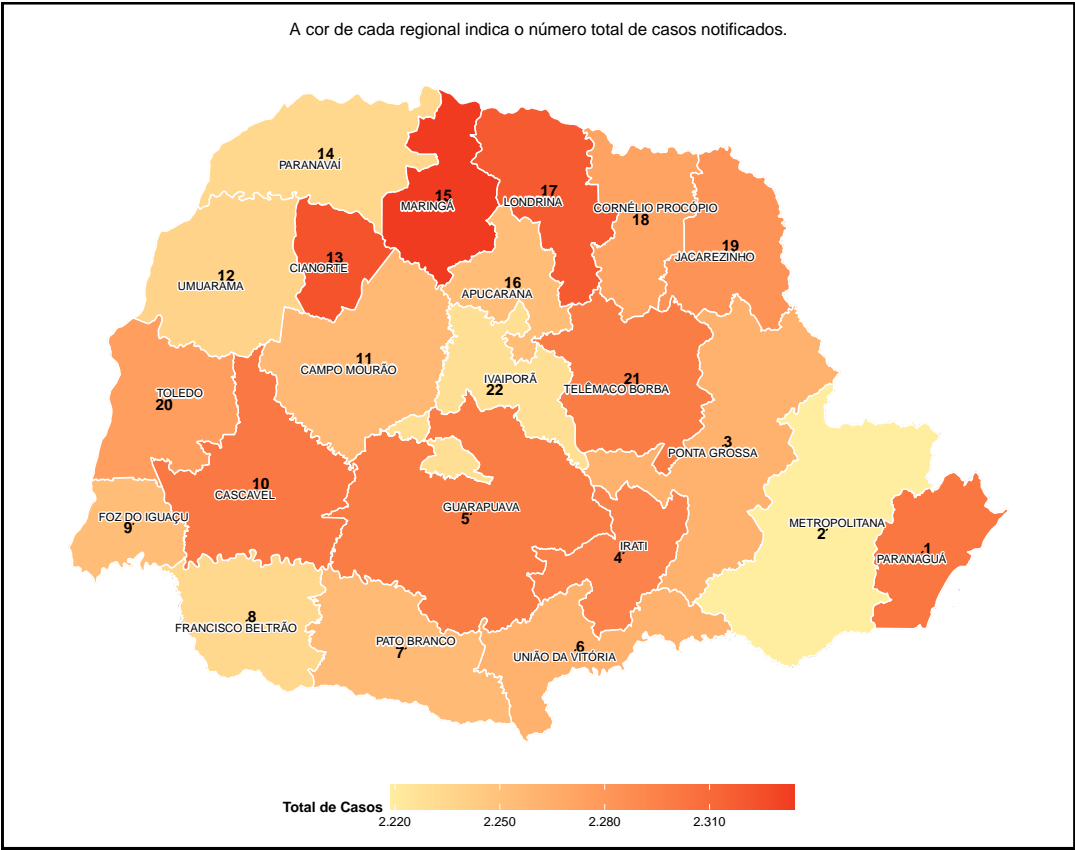


Figura 4 – Tempo Médio (em dias) entre Primeiros Sintomas e Notificação por Regional de Saúde

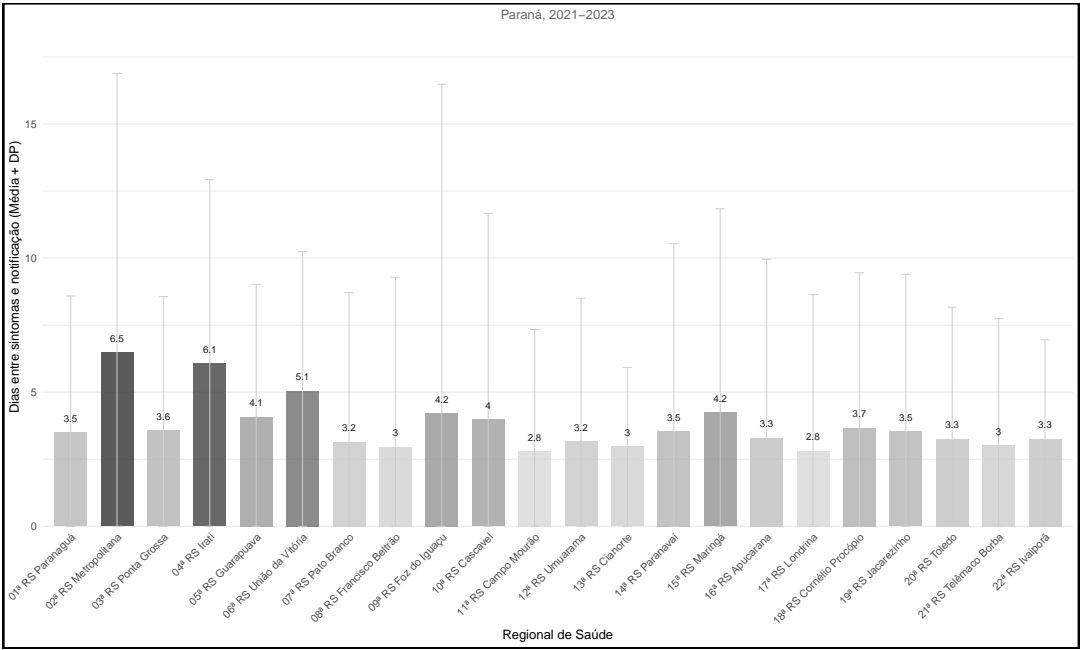
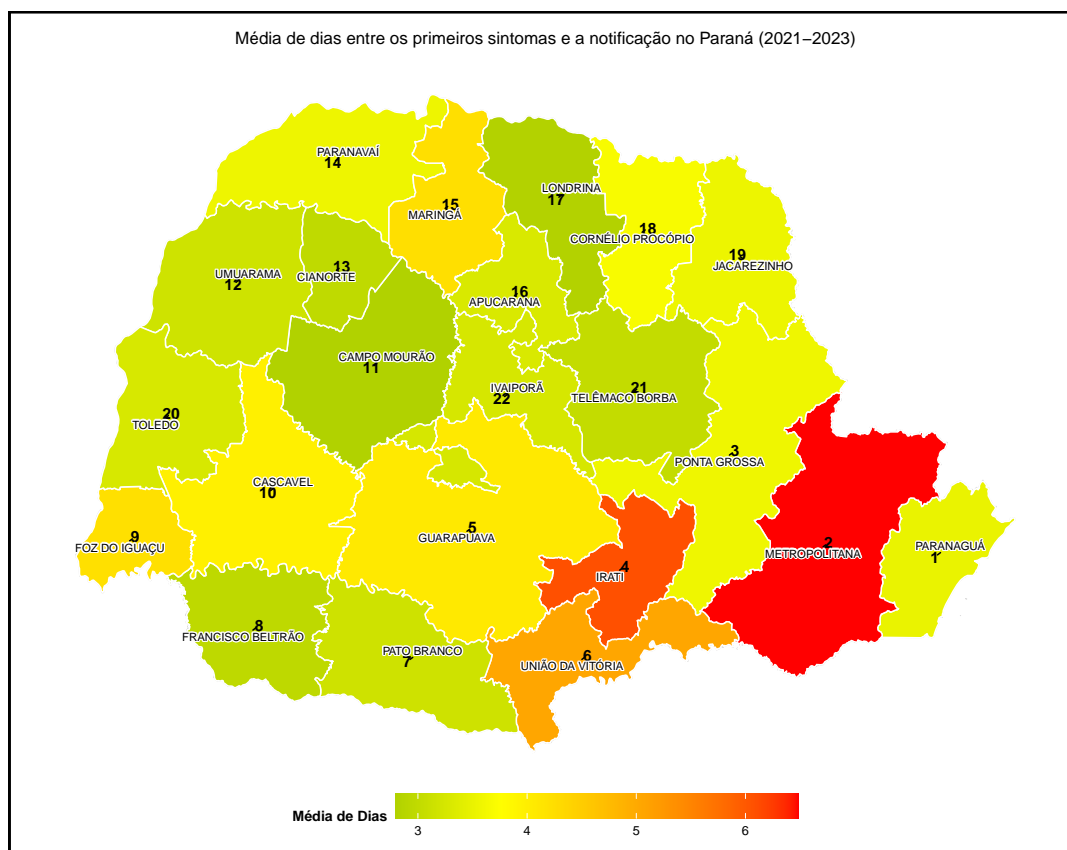


Figura 5 – Tempo Médio pra Notificação de Dengue por Regional de Saúde

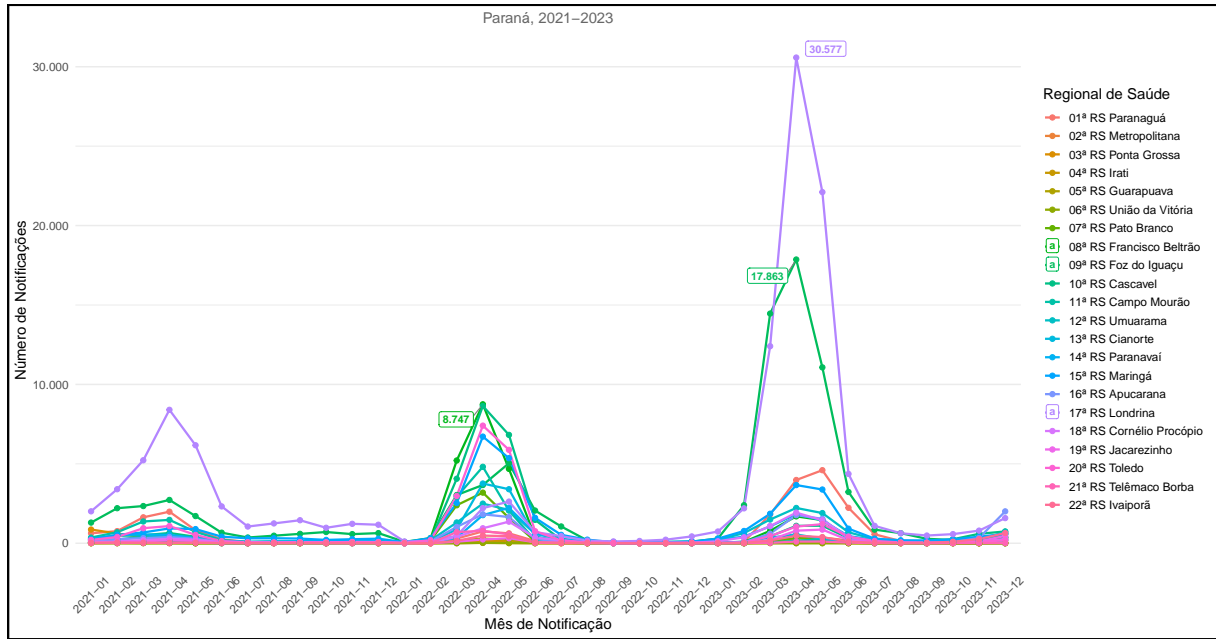


3.1.3 DINÂMICA ESPAÇO-TEMPORAL DA DENGUE

A Figura 6 apresenta as séries temporais mensais de casos de dengue por Regional de Saúde no estado do Paraná, revelando que a dinâmica da doença é marcadamente não estacionária. Observam-se padrões de sazonalidade, variações interanuais significativas e uma notável heterogeneidade na ocorrência entre as diferentes regiões. Essa complexa dinâmica pode ser compreendida através da interação de múltiplos fatores que se sobrepõem.

Primeiramente, a dinâmica é impulsionada por fatores ambientais e climáticos, com a forte sazonalidade a ser o padrão mais evidente: os picos de incidência coincidem com os meses mais quentes e chuvosos, que aceleram o ciclo de vida do vetor *Aedes aegypti*. Sobre essa base climática, atuam os fatores sociais e demográficos, como a densidade em ambientes urbanos, que facilita a disseminação viral, e a mobilidade populacional, que ajuda a dispersar o vírus entre diferentes localidades. Finalmente, a magnitude dos surtos é modulada pela resposta institucional, pois a capacidade das autoridades sanitárias, a continuidade das ações de controle vetorial e a eficiência da vigilância epidemiológica são cruciais para mitigar a transmissão e controlar a epidemia localmente.

Figura 6 – Evolução Mensal de Casos de Dengue por Regional de Saúde



3.2 PADRÕES E COMPONENTES DO MODELO PROPOSTO

Justificativa para a Remoção Sequencial: O padrão de cauda longa observado nas Figuras 1 e 2 fundamenta a adoção de um componente de remoção sequencial. A evidência de que casos permanecem não detectados por múltiplos períodos invalida a premissa de que a detecção ocorre em uma única janela temporal. No modelo proposto, isso é capturado pela estrutura hierárquica que permite que um indivíduo infectado, se não notificado em um período t , "sobreviva" no conjunto de casos latentes e seja passível de detecção em períodos futuros $(t + 1, t + 2, \dots)$. Isso replica empiricamente o processo de atraso.

Justificativa para Efeitos Regionais: A expressiva variabilidade espacial na incidência (Figura 3.3) e na eficiência da notificação (Fig. 3 e 5) torna indispensável a modelagem explícita da variação entre as Regionais de Saúde. Para capturar essa heterogeneidade, o modelo incorpora parâmetros específicos por região, como a abundância inicial de casos latentes (λ_r). Essa estrutura hierárquica (via efeitos aleatórios ou fixos por regional) permite que tanto a intensidade da epidemia quanto, implicitamente, a probabilidade de detecção (p) variem entre as regiões, acomodando as diferenças observadas nos dados.

Dinâmica sazonal: O padrão sazonal e de crescimento abrupto de surto (Figura 6) justifica o componente dinâmico do modelo (Eqs. 2.9–2.10). Especificamente, a evolução latente do número de casos $N_{r,t}$ é modelada por um processo Poisson dinâmico com taxa de crescimento/declínio γ . Esse mecanismo captura o aumento populacional de casos ao longo do tempo (que pode refletir sazonalidade implícita e efeitos epidêmicos anuais). Em essência, a fórmula $N_{r,t} \mid N_{r,t-1} \sim \text{Poisson}(\gamma N_{r,t-1})$ permite que $N_{r,t}$ cresça de forma consistente com

os múltiplos surtos observados. Se necessário, variações sazonais poderiam ser incluídas em γ ou em λ_r , mas a estrutura dinâmica básica já acomoda grande parte da sazonalidade e variabilidade interanual dos dados.

Justificativa para o Componente Dinâmico: A forte sazonalidade e a variação interanual, visíveis nas séries temporais da Figura 6, exigem a utilização de um componente dinâmico. Um modelo estático seria incapaz de capturar os picos epidêmicos e as tendências observadas. A formulação proposta modela a evolução do número de casos latentes $N_{r,t}$ por meio de um processo que depende do estado anterior $N_{r,t} \mid N_{r,t-1} \sim \text{Poisson}(\gamma N_{r,t-1})$. Este mecanismo permite que o número de casos cresça ou decaia ao longo do tempo, refletindo a dinâmica de surtos observada empiricamente.

Seção 4

CRONOGRAMA DAS ATIVIDADES DO PROJETO DE PESQUISA

Tabela 1 – Cronograma das atividades do projeto de pesquisa

Etapa principal	Atividades detalhadas	Início	Término	Observações
Implementação do novo modelo	Levantamento de requisitos; codificação do modelo estendido em R/C++ (Rcpp); verificação de convergência (MCMC/MV)	Ago 2025	Set 2025	Scripts reprodutíveis
Estudo de simulação	Definição de cenários (variação de λ , ϕ , γ , tamanho amostral, grau de subnotificação); dados sintéticos; avaliação de viés, RMSE, cobertura e tempo computacional	Out 2025	Nov 2025	Resultados alimentarão ajustes finos
Comparação com o modelo original	Ajuste do modelo N-Mistura dinâmico de Dail & Madsen (2011) aos mesmos dados simulados; métricas (AIC, WAIC, LOO-CV, acurácia de parâmetros, tempo); análise estatística das diferenças	Nov 2025	Nov 2025	Evidência empírica da vantagem da extensão proposta
Análise com dados reais (SINAN)	Identificação e limpeza de bases públicas (p. ex. séries COVID-19, IST, arboviroses); ajuste dos dois modelos; interpretação epidemiológica dos parâmetros (prevalência real, fração não detectada)	Dez 2025	Dez 2025	Dados de dengue do Sistema de Informação de Agravos de Notificação (SINAN) serão utilizados.
Outros	(a) Revisão bibliográfica contínua (Jul 2025–Fev 2026); (b) Redação de capítulos (Out 2025–Fev 2026); (c) Revisão final e formatação ABNT (Fev 2026); (d) Depósito e defesa (Fev 2026)	Jul 2025	Fev 2026	1 mês de folga para ajustes e trâmites administrativos

Nota. Cada macro-etapa será revisada mensalmente com a orientação; as datas indicam prazos máximos e podem ser antecipadas caso a etapa anterior termine antes do previsto.

REFERÊNCIAS

- ALAGARASU, K. et al. Serotype and genotype diversity of dengue viruses circulating in india: a multi-centre retrospective study involving the virus research diagnostic laboratory network in 2018. *International Journal of Infectious Diseases*, Elsevier, v. 111, p. 242–252, 2021.
- ARAÚJO, N. J. S. et al. Control of arboviruses vectors using biological control by wolbachia pipientis: a short review. *Archives of Microbiology*, Springer, v. 204, n. 7, p. 376, 2022.
- BARRETO, M. L.; TEIXEIRA, M. G. Dengue no brasil: situação epidemiológica e contribuições para uma agenda de pesquisa. *Estudos avançados*, SciELO Brasil, v. 22, p. 53–72, 2008.
- BERBERIAN, G. V. et al. Dengue beyond the tropics: a time-series study comprising 2015–2016 versus 2019–2020 at a children's hospital in the city of buenos aires. *Sociedad Argentina de Pediatría*, 2022.
- BRINTZ, B.; FUENTES, C.; MADSEN, L. An asymptotic approximation to the n-mixture model for the estimation of disease prevalence. *Biometrics*, Oxford University Press, v. 74, n. 4, p. 1512–1518, 2018.
- BUCHS, A. et al. The threat of dengue in europe. *New Microbes and New Infections*, v. 49, p. 101061, 2022.
- CARLE, F. L.; STRUB, M. R. A new method for estimating population size from removal data. *Biometrics*, JSTOR, p. 621–630, 1978.
- DAIL, D.; MADSEN, L. Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics*, Oxford University Press, v. 67, n. 2, p. 577–587, 2011.
- DORAZIO, R. M.; JELKS, H. L.; JORDAN, F. Improving removal-based estimates of abundance by sampling a population of spatially distinct subpopulations. *Biometrics*, Oxford University Press, v. 61, n. 4, p. 1093–1101, 2005.
- FERGUSON, T. S. A bayesian analysis of some nonparametric problems. *The annals of statistics*, JSTOR, p. 209–230, 1973.
- GWEE, X. W. S.; CHUA, P. E. Y.; PANG, J. Global dengue importation: a systematic review. *BMC infectious diseases*, Springer, v. 21, n. 1, p. 1078, 2021.

- HALSTEAD, S. B. Pathogenesis of dengue: challenges to molecular biology. *Science*, American Association for the Advancement of Science, v. 239, n. 4839, p. 476–481, 1988.
- LINCOLN, F. C. Calculating waterfowl abundance on the basis of banding returns. *US Department of Agriculture*, n. 118, 1930.
- LIU, Z. et al. The effect of temperature on dengue virus transmission by aedes mosquitoes. *Frontiers in cellular and infection microbiology*, Frontiers Media SA, v. 13, p. 1242173, 2023.
- MADSEN, L.; ROYLE, J. A. A review of n-mixture models. *Wiley Interdisciplinary Reviews: Computational Statistics*, Wiley Online Library, v. 15, n. 6, p. e1625, 2023.
- MANNA, S. et al. Dengue outbreaks in south asia amid covid-19: Epidemiology, transmission, and mitigation strategies. *Frontiers in Public Health*, Frontiers Media SA, v. 10, p. 1060043, 2022.
- MORIN, C. W.; COMRIE, A. C.; ERNST, K. Climate and dengue transmission: evidence and implications. *Environmental health perspectives*, National Institute of Environmental Health Sciences, v. 121, n. 11-12, p. 1264–1272, 2013.
- MUSTAFA, M. et al. Discovery of fifth serotype of dengue virus (denv-5): A new public health dilemma in dengue control. *Medical journal armed forces India*, Elsevier, v. 71, n. 1, p. 67–70, 2015.
- NOR, M. K. M. et al. A systematic analysis of the effects of climate and environmental exposures on dengue cases from 2010 to 2020. *Malaysian Journal of Medicine & Health Sciences*, v. 17, 2021.
- NORMILE, D. *Surprising new dengue virus throws a spanner in disease control efforts*. [S.l.]: American Association for the Advancement of Science, 2013.
- ORGANIZATION, W. H. et al. *Dengue: guidelines for diagnosis, treatment, prevention and control*. [S.l.]: World Health Organization, 2009.
- OTIS, D. L. et al. Statistical inference from capture data on closed animal populations. *Wildlife monographs*, JSTOR, n. 62, p. 3–135, 1978.
- PAIXÃO, E. S.; TEIXEIRA, M. G.; RODRIGUES, L. C. Zika, chikungunya and dengue: the causes and threats of new and re-emerging arboviral diseases. *BMJ global health*, BMJ Publishing Group Ltd, v. 3, n. Suppl 1, 2018.
- PARKER, M. et al. Faster asymptotic solutions for n-mixtures on large populations. *Journal of Agricultural, Biological and Environmental Statistics*, Springer, p. 1–16, 2024.
- RIVERA, O. Rodriguez de; MCCREA, R. Removal modelling in ecology: A systematic review. *PLoS One*, Public Library of Science San Francisco, CA USA, v. 16, n. 3, p. e0229965, 2021.
- ROYLE, J. A. N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, Oxford University Press, v. 60, n. 1, p. 108–115, 2004.
- ROYLE, J. A.; DORAZIO, R. M. Hierarchical models of animal abundance and occurrence. *Journal of Agricultural, Biological, and Environmental Statistics*, Springer, v. 11, n. 3, p. 249–263, 2006.

SANDERLIN, J. S. et al. Occupancy modeling and resampling overcomes low test sensitivity to produce accurate sars-cov-2 prevalence estimates. *BMC public health*, Springer, v. 21, p. 1–10, 2021.

TORRES-FLORES, J. M.; REYES-SANDOVAL, A.; SALAZAR, M. I. Dengue vaccines: an update. *BioDrugs*, Springer, v. 36, n. 3, p. 325–336, 2022.

VIGLIETTA, M. et al. Vector specificity of arbovirus transmission. *Frontiers in microbiology*, Frontiers Media SA, v. 12, p. 773211, 2021.

ZIPPIN, C. An evaluation of the removal method of estimating animal populations. *Biometrics*, JSTOR, v. 12, n. 2, p. 163–189, 1956.