

# Análise de dados de vigilância entomológica no Espírito Santo: relações com dados meteorológicos e casos de dengue

Bernardo R. A. Silva, Michele A. Brandão, Gabriel P. Oliveira

Departamento de Ciência da Computação

Universidade Federal de Minas Gerais

Belo Horizonte, Brasil

{bernardo.roberto, michele.brandao, gabrieloliveira}@dcc.ufmg.br

**Resumo**—A doença da dengue continua gerando impactos sociais e econômicos altamente negativos no Brasil e no mundo. Controlar o seu principal vetor, o mosquito *Aedes aegypti*, e entender sua dinâmica populacional são partes fundamentais do combate à doença. Utilizamos modelos de regressão para avaliar como variáveis meteorológicas influenciam a abundância dos mosquitos (estimada através de vigilância entomológica), e utilizamos análise de correlação para avaliar o relacionamento dessa abundância com a incidência de dengue, em diferentes municípios do estado do Espírito Santo. Por fim, desenvolvemos diferentes modelos epidemiológicos com os dados de dengue, de um mais simples a um mais complexo, incluindo a abundância do vetor em sua formulação, para comparar a qualidade de ajuste aos dados dos diferentes modelos. Resultados da regressão revelam maior influência da temperatura na abundância do vetor, e uma inconsistência dos modelos durante anos distintos de observação. A análise de correlação apontou um relacionamento fraco ou contraintuitivo; seguindo essa linha de raciocínio, os modelos epidemiológicos não mostraram vantagem ao incluir a infestação do vetor em sua formulação, mas alcançaram bons resultados ao incluir temperatura. O estudo destaca a importância da vigilância entomológica contínua para o entendimento da dinâmica populacional do mosquito, ao passo que o seu relacionamento com a incidência de dengue segue complexo, sugerindo que é um dado sozinho insuficiente para a avaliação de risco de dengue ou a criação de modelos de previsão.

**Index Terms**—Dengue, Monitoramento Entomológico, Análise de dados, Machine Learning, Regressão, Correlação, Modelos Epidemiológicos Compartimentais, Otimização,

## I. INTRODUÇÃO

As arboviroses, doenças transmitidas principalmente por artrópodes, como a dengue, permanecem uma grande ameaça à saúde pública do país, incorrendo em custos anuais diretos e indiretos que ultrapassam bilhões [1]. Paralelamente, a presença do vírus da dengue, antes associada principalmente às áreas tropicais e subtropicais do planeta, tem rapidamente tomado proporções globais, por conta de fatores que incluem mudanças climáticas e migrações de pessoas [2].

Medidas de prevenção e controle da dengue abrangem, além da vacinação, o controle vetorial pela aplicação de larvicidas e inseticidas e pela remoção de criadouros (controle mecânico), e também o monitoramento da população de mosquitos (monitoramento/vigilância entomológica), incluindo levantamentos

amostrais em imóveis e inspeção de armadilhas [3]. Tal medida produz indicadores da presença, distribuição e densidade dos insetos no tempo e no espaço, incluindo índices larvais e de pupas, índices de ovos e de mosquitos adultos.

Acompanhando a alta disponibilidade de dados históricos de vigilância, a modelagem matemática têm tornado-se cada vez mais relevante em diversos estudos sobre a variação populacional do mosquito *Aedes aegypti*, o principal vetor da dengue. A forte influência de variáveis meteorológicas no desenvolvimento e sobrevivência do vetor e na transmissão da dengue (influência da temperatura em processos fisiológicos importantes do mosquito e tempo de incubação do vírus; necessidade de água para ovoposição e desenvolvimento larval e pupal; necessidade de umidade para sobrevivência do adulto e do ovo; etc) [4] [5] motiva o desenvolvimento de modelos de regressão para entender o relacionamento entre o nível de infestação e os dados meteorológicos. Dessa maneira, trabalhos propõem, por exemplo, soluções para a previsão de infestações futuras e avaliação de risco da doença, de forma a habilitar medidas de controle antecipadas e eficientes [6] [7].

Por outro lado, a alta disponibilidade de dados históricos de casos de dengue possibilita o surgimento de modelos epidemiológicos compartimentais diversos [8]. Modelos compartimentais simulam a disseminação de uma doença em uma população por meio da divisão dos indivíduos em compartimentos - por exemplo, em indivíduos suscetíveis a uma doença ( $S$ ) e infectados ( $I$ ) - e da criação de fórmulas que especificam a taxa de transição entre esses compartimentos a cada intervalo de tempo. Tais modelos têm sido utilizados para diversos fins, como para a avaliação do impacto de medidas de controle e/ou vacinação na evolução da doença [9] [10], e para a avaliação de riscos de infecção de indivíduos em contexto de imigração [11].

No entanto, a relação entre os índices de vetor e os casos de dengue permanece mal compreendida pela literatura [12], apesar de que, como no caso do primeiro, trabalhos exploraram com sucesso o uso de variáveis meteorológicas para o desenvolvimento de modelos de incidência de dengue [13]. Essa barreira é provavelmente um motivo pelo qual, até onde sabemos, poucos trabalhos consideram dados de vigilância

entomológica em seus modelos epidemiológicos compartimentais, como [14].

Nesse contexto, este trabalho tem os seguintes objetivos:

- Criar modelos de regressão com dados de vigilância entomológica (índice de vetor) e meteorológicos, a partir de dados de diversas cidades do Estado do Espírito Santo. Então, avaliar o relacionamento entre as variáveis e comparar os parâmetros dos modelos de locais diferentes e de períodos (anos) distintos. Também avalia-se a consistência dos modelos durante os anos;
- Avaliar a correlação entre os casos de dengue e o índice de vetor para os mesmos locais e períodos, para verificar se a tendência descrita na literatura mantém-se;
- Criar três modelos epidemiológicos com dados de casos de dengue de algumas cidades, de um mais simples a um mais complexo que considera o índice de vetor em sua formulação, e compará-los segundo seu ajuste aos dados e parâmetros.

O restante do trabalho está organizado da seguinte forma: a seção II apresenta conceitos e parte da literatura relevantes ao trabalho; a seção III esclarece a metodologia utilizada no trabalho; a seção IV apresenta os resultados, que são discutidos na seção V; a seção VI conclui o trabalho; e a seção VII contém os agradecimentos.

## II. REFERENCIAL TEÓRICO

Modelos de regressão permitem quantificar a relação de uma variável dependente (resposta) com variáveis independentes (preditoras), por meio dos seus parâmetros (coeficientes). A Equação (1) mostra, por exemplo, um modelo de regressão linear multivariado: o modelo calcula coeficientes  $\beta_i$  para cada uma das  $x_i$  variáveis selecionadas à medida que aproxima o valor esperado da resposta ( $Y$ ) como uma função linear das variáveis, e também calcula o intercepto ( $\beta_0$ ), o valor da função quando todas as variáveis são 0. Os coeficientes tornam modelos de regressão mais interpretáveis do que outros modelos supervisionados, como árvores de decisão. Trazendo para o contexto do trabalho, cada uma das variáveis  $x_i$  pode ser um valor meteorológico, como temperatura ou nível de chuva, e a resposta ( $Y$ ), o nível de infestação do vetor naquele momento.

$$\mathbb{E}[Y] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (1)$$

Trabalhos têm utilizado modelos de regressão das mais diversas formas, para os mais diversos objetivos, tirando proveito da forte relação amplamente documentada entre o desenvolvimento e sobrevivência dos mosquitos, transmissão de dengue, e os fatores do clima, abordada na seção I. Os autores de [6], por exemplo, desenvolveram modelos com dados da cidade de Porto Alegre, Rio Grande do Sul, para descrever a dinâmica temporal da população de mosquitos e a relação com casos de dengue. Foi aplicado um modelo aditivo generalizado, no qual, em vez de coeficientes, as variáveis independentes são relacionadas à resposta através de funções não lineares ( $f(x_i)$ ), permitindo maior flexibilidade. Os fatores mais relevantes para

a predição do índice de densidade do vetor foram variáveis de temperatura mínima, umidade e o próprio índice medido na semana anterior, enquanto a precipitação, por ser bem distribuída durante o ano no clima subtropical da cidade, não foi um fator muito relevante. Utilizando um modelo logístico, também foi apontado um aumento subsequente no número de casos de dengue, quando havia aumento na densidade de vetores.

Já o trabalho [7] utilizou dados de vigilância da cidade de Vitória, Espírito Santo, mas construiu modelos lineares generalizados, nos quais a variável resposta ( $Y$ ) é relacionada às variáveis independentes por algum tipo de função de ligação ( $g(\mathbb{E}[Y])$ ). Os autores buscaram desenvolver um modelo de regressão de qualidade comparável a modelos supervisionados mais complexos, assim mantendo a vantagem da interpretabilidade, e empregaram dados meteorológicos diversos, incluindo dados de vegetação, extraídos tanto de uma zona urbana quanto de uma zona rural. O trabalho alcançou bons resultados, e concluiu uma maior influência positiva da umidade sobre a infestação de vetores, seguida por temperatura e um efeito adverso da precipitação, possivelmente pela competição das armadilhas com outros locais em época de chuva. Os autores de [15], por sua vez, compararam diversos índices vetoriais, incluindo de larvas, ovos e adultos, e tipos de armadilhas diferentes, em relação à sensibilidade dos valores às variações meteorológicas, para identificar aqueles que melhor capturam o padrão sazonal da população dos mosquitos. Foram consideradas 5 cidades de estados bem distintos do país, e foram construídos modelos para cada tipo de índice e cidade. O estudo concluiu uma grande influência da temperatura na sazonalidade da população de mosquitos e uma maior qualidade dos índices associados às armadilhas, especialmente os de densidade e positividade de adultos.

Nesse contexto, esse trabalho tem como primeiro objetivo criar modelos de regressão linear com dados de vigilância entomológica e meteorológicos de cinco municípios do estado do Espírito Santo, para avaliar o relacionamento entre as variáveis e comparar os resultados. Dada a disponibilidade de dados de diversos anos, e, até onde sabemos, o fato de que os trabalhos geralmente não consideram a consistência dos modelos durante longos períodos de tempo, também avaliamos a consistência dos resultados, ajustando assim os modelos independentemente para anos distintos. Adicionalmente, como segundo objetivo, avaliamos a correlação entre os casos de dengue e os dados de densidade de vetor, para verificar se a tendência de baixa correlação mantém-se, aproveitando a grande quantidade de dados disponíveis.

O terceiro e último objetivo, como resumido na Introdução, consiste em desenvolver modelos epidemiológicos compartimentais, e compará-los segundo sua qualidade de ajuste aos dados e os valores de seus parâmetros.

Modelos epidemiológicos são representações da epidemiologia da transmissão de doenças e seus processos associados. Entre diversas aplicações, os modelos servem para a análise dos processos específicos de doenças, a geração de hipóteses, a avaliação do impacto econômico de doenças e da efetividade

de estratégias de prevenção ou controle. Eles podem ser classificados, por exemplo, em relação à certeza, divididos entre determinísticos e estocásticos; modelos determinísticos utilizam valores fixos como parâmetros, como, por exemplo, uma taxa fixa de natalidade, e geram um resultado esperado; modelos estocásticos implementam uma variabilidade natural, gerando diversos resultados possíveis [16].

Os modelos compartmentais dividem a população de estudo em compartimentos, pelos quais indivíduos transitam a cada intervalo de tempo. Um dos mais simples nessa categoria é o modelo endêmico clássico [8] (Figura 1) também conhecido por modelo *SIR* com dinâmica vital (por incluir a dinâmica de nascimentos e mortes). Especifica-se que indivíduos podem estar suscetíveis a uma doença (compartimento *S*), infectados (e com capacidade de infectar outros - compartimento *I*) ou recuperados/imunizados (compartimento *R*). As transições, ilustradas na Figura 1, são descritas pelas Equações (2) - (4). O termo  $\beta \frac{I}{N}$  é denominado força de infecção, e corresponde à taxa com a qual indivíduos suscetíveis tornam-se infectados. O parâmetro  $\beta$  corresponde à taxa de transmissão, indivíduos recuperam-se com uma taxa de recuperação  $\gamma$ , e assume-se uma taxa de nascimentos e mortes constante  $\mu$ , de tal forma que a população total  $N = S + I + R$  permanece constante. Além disso, indivíduos infectados podem infectar indivíduos suscetíveis imediatamente, e não há perda de imunidade (transições de *R* para outros compartimentos). A Figura 2 exibe um exemplo hipotético de evolução do modelo *SIR*, dado um conjunto de parâmetros iniciais  $N, S(0), I(0), R(0), \beta, \mu$  e  $\gamma$ , em que há um pico de infecções (surto) e depois uma normalização.

Há diversos outros modelos na literatura com premissas e fórmulas mais realistas. O modelo *SEIR*, por exemplo, inclui, antes de *I*, o compartimento de indivíduos expostos (*E*), isto é, infectados mas não capazes ainda de transmitir o patógeno. Nesse caso, considera-se como um parâmetro adicional o período de latência da doença: o intervalo de tempo entre o momento que um indivíduo é infectado e o momento quando torna-se infeccioso. É importante notar que, de forma geral, a presença em *E* ou em *I* não distingue se um indivíduo possui sintomas [17]; indivíduos expostos geralmente são assintomáticos ou fracamente sintomáticos, mas podem haver indivíduos infecciosos assintomáticos; a situação é ainda mais complexa visto que indivíduos recuperados (e não infecciosos) podem permanecer sintomáticos por um tempo, como no caso do vírus da gripe. O período para a aparição de sintomas desde a infecção é denominado período de incubação.

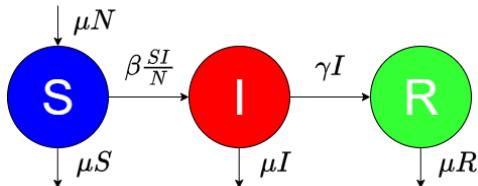


Figura 1. Modelo endêmico clássico (*SIR*).

$$\frac{dS}{dt} = \mu N - \beta \frac{SI}{N} - \mu S, \quad (2)$$

$$\frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I - \mu I, \quad (3)$$

$$\frac{dR}{dt} = \gamma I - \mu R, \quad (4)$$

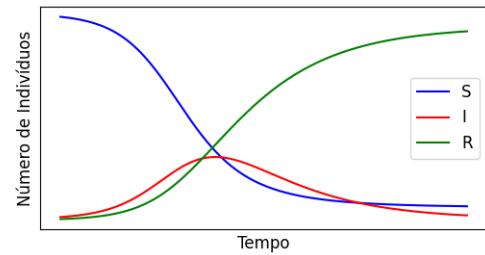


Figura 2. Exemplo de evolução do modelo *SIR*.

No caso de doenças infecciosas transmitidas por vetores, como a dengue, modelos mais simples como o *SIR* ou o *SEIR* podem não ser suficientes. Assim, adotam-se compartimentos adicionais para os vetores, e a força de infecção ocorre em dois sentidos (humano - vetor e vetor - humano). Tais modelos são baseados no famoso modelo de Ronald Ross e George Macdonald (Ross-Macdonald) [18], desenvolvido inicialmente para a malária. O modelo *SIR-SI* (Figura 3 e Equações (5) - (9)) é um exemplo deles. Tal modelo tem sido adotado em diversos trabalhos sobre dengue, como [9], [10] e [19]; por outro lado, também pode haver aplicação do modelo mais simples *SIR*, considerando de forma implícita a dinâmica vetorial, geralmente para comparação dos resultados, como em [9] e [19].

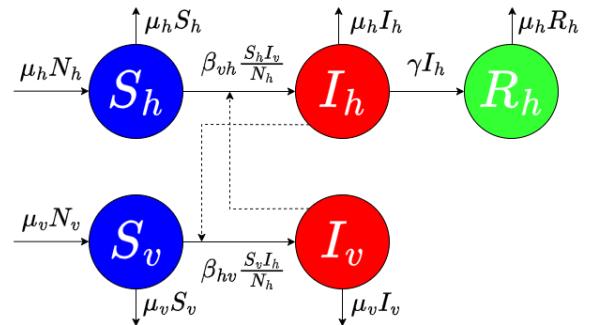


Figura 3. Modelo com dinâmica hospedeiro-vetor *SIR-SI*.

$$\frac{dS_h}{dt} = \mu_h N_h - \beta_{vh} \frac{S_h I_v}{N_h} - \mu_h S_h, \quad (5)$$

$$\frac{dI_h}{dt} = \beta_{vh} \frac{S_h I_v}{N_h} - \gamma_h I_h - \mu_h I_h, \quad (6)$$

$$\frac{dR_h}{dt} = \gamma_h I_h - \mu_h R_h, \quad (7)$$

$$\frac{dS_v}{dt} = \mu_v N_v - \beta_{hv} \frac{S_v I_h}{N_h} - \mu_v S_v, \quad (8)$$

$$\frac{dI_v}{dt} = \beta_{hv} \frac{S_v I_h}{N_h} - \mu_v I_v, \quad (9)$$

O trabalho [9] utilizou dados de casos de dengue do estado de Minas Gerais, para avaliar o impacto de métodos diversos de controle vetorial (inseticidas, larvicidas e controle mecânico) na curva de casos da doença. Foram adotados o modelo mais simples *SIR*, o modelo *SIR – SI* e uma versão modificada do *SIR – SI*, incluindo compartimentos para as fases de ovo, larva e pupa. Os autores empregaram uma taxa de transmissão ( $\beta$ ) variável, dependente de fatores meteorológicos; como será explicado na seção III-D, nós utilizamos uma abordagem similar. Em conclusão, o controle mecânico foi o mais eficaz, embora os outros métodos também obtiveram resultados satisfatórios.

O trabalho [19], por sua vez, aplicou os modelos *SIR* e *SIR – SI* em casos de dengue hemorrágica na Tailândia, implementando um fator de probabilidade da complicação, de forma a verificar o impacto das diferenças de modelagem em relação à qualidade de ajuste aos dados e aos parâmetros descobertos. Foram encontradas estimativas de parâmetros concordantes com a literatura a partir dos dois modelos e uma performance superior do modelo mais simples *SIR*. Assim, concluiu-se que a complexidade adicional do segundo modelo pode não ser necessária.

Por fim, o trabalho [14] procurou unir dados de casos de dengue com dados de vigilância entomológica, por meio de uma modelagem mais complexa (*SEIRS – SEI*), que ainda considerou a possibilidade do período de latência da doença exceder o tempo de vida do mosquito. Os autores também incluíram os quatro sorotipos da dengue como compartimentos diferentes, enquanto muitos trabalhos não fazem essa separação. O modelo foi utilizado para avaliar o efeito do controle vetorial implementado em diferentes períodos na curva da doença, e foi aplicado em dados da cidade de Vitória, Espírito Santo. Foi encontrado o maior impacto da implementação em períodos com baixo número de casos e menor quantidade de mosquitos, quando eles já representavam um gargalo para a disseminação da doença, destacando a importância de medidas de controle de longo prazo, inclusive em épocas de baixo número de casos de dengue.

Nesse contexto, recapitulando o último objetivo, exercitamos o uso de três modelos epidemiológicos compartmentais com os dados disponíveis, considerando, como feito em alguns trabalhos, modelos mais simples a mais complexos, buscando por fim incluir o índice de vetor como fator. Assim, analisamos

a diferença da qualidade de ajuste aos dados e dos valores de parâmetros descobertos.

### III. METODOLOGIA

Esta seção apresenta as etapas realizadas para atingir os objetivos do trabalho, baseadas em passos clássicos de uma análise exploratória de dados: seleção, pré-processamento e transformação dos dados (seção III-A) e modelagem (seções III-B, III-C e III-D), enquanto a apresentação e a interpretação dos resultados são reservadas, respectivamente, às seções IV e V.

#### A. Seleção, Pré-processamento e Transformação dos Dados

**Dados de Vigilância Entomológica:** Selecionamos dados de capturas semanais de armadilhas de fêmeas de *Aedes aegypti* dos anos de 2017 a 2019 de cinco municípios no Estado do Espírito Santo: Vitória, Vila Velha, Afonso Cláudio, Linhares e Ecoporanga (Figura 4). A seleção buscou um balanceamento entre altas amostragens semanais com pouca variância, diversidade geográfica, completude dos dados de capturas e completude dos dados meteorológicos dos municípios, que serão tratados à frente. Vila Velha e Vitória, apesar de sua proximidade, foram ambas selecionadas por conta da alta taxa de amostragem de armadilhas e para permitir a comparação de resultados de locais próximos.

Os dados foram-nos providos pela empresa Ecovec<sup>1</sup>, que implementa o programa de vigilância entomológica MI-Aedes, anteriormente conhecido por MI-Dengue, em muitos Estados do país desde 2005. O programa monitora semanalmente armadilhas grudentas de mosquitos (MosquiTRAP®) projetadas para fêmeas de *Aedes aegypti* grávidas, que precisam alimentar-se de sangue para desenvolver seus ovos e são os principais vetores (e o único entre os dois sexos) do vírus da dengue [20]. Calculamos, para cada município, o índice de densidade “Índice Médio de Fêmeas de *Aedes*” (IMFA) - a média de capturas por armadilha por semana - para os modelos de regressão, para a análise de correlação e para os modelos epidemiológicos, já que o indicador tem sido utilizado para avaliação de risco de transmissão de dengue<sup>2</sup>. Para evitar grandes lacunas nos dados, consideramos semanas a partir da vigésima semana de 2017. Em 2018 e 2019, poucas semanas isoladas estavam ausentes, portanto foi realizada uma interpolação linear para preencher as lacunas.

**Dados Meteorológicos:** Os dados de temperatura média diária (*temp\_media*), umidade relativa do ar média diária (*umid\_media*) e precipitação total diária (*prec\_total*) de cada cidade foram selecionados para a análise de regressão. Obtivemos os dados do Banco de Dados Meteorológicos (BDMEP) do INMET (Instituto Nacional de Meteorologia)<sup>4</sup>, começando a partir de 4 semanas antes do início do período

<sup>1</sup>Site da Ecovec: <https://ecovec.com/>

<sup>2</sup>“Onde está o Aedes? | Prefeitura de Porto Alegre.” <https://prefeitura.poa.br/sms/onde-esta-o-aedes/infestacao-de-aedes-aegypti>

<sup>3</sup>“Malha Municipal | IBGE.” <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html>

<sup>4</sup>“INMET :: BDMEP.” <https://bdmep.inmet.gov.br/>

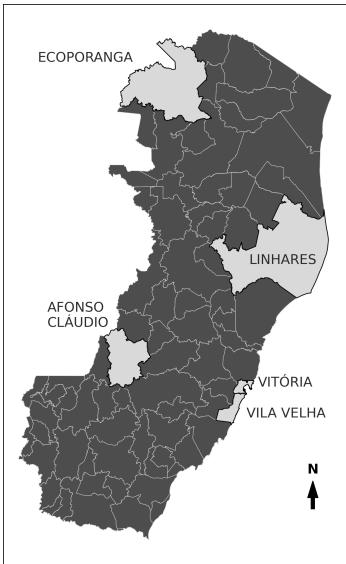


Figura 4. Municípios selecionados. Espírito Santo, Brasil.<sup>3</sup>

do índice vetorial (vigésima semana de 2017), para considerar efeitos atrasados de até 4 semanas na evolução da população do vetor. Os modelos epidemiológicos também utilizaram os dados de temperatura, mas foram desenvolvidos considerando semanas entre o final de 2018 e a maior parte de 2019, como será abordado na seção III-D. Calculamos as médias semanais para *temp\_media* e *umid\_media* e o total semanal para *prec\_total*, gerando as variáveis denotadas resumidamente por *temp<sub>i</sub>*, *umid<sub>i</sub>* e *prec<sub>i</sub>*, em que  $i = 0, \dots, 4$  indica a antecipação da variável em semanas. Para lidar com valores faltantes antes da agregação, imputamos valores utilizando as médias dos dias de outros anos. Poucos dias estavam faltantes no geral, exceto para alguns casos raros de dados ausentes por cerca de 40 dias em sequência. Antes da construção dos modelos de regressão, executamos a normalização Z-score nos dados, a qual padroniza a sua escala, tornando sua média 0 e seu desvio padrão 1. No caso dos modelos epidemiológicos, a normalização da temperatura não foi realizada, visto que a escala do valor já foi controlada por um termo.

**Casos de Dengue:** Obtivemos, para cada município, as notificações de dengue semanais confirmadas dos residentes por meio do SINAN (Sistema de Informação de Agravos de Notificação), disponibilizado pelo Ministério da Saúde<sup>5</sup>. Para a análise de correlação com o IMFA, consideramos valores adiantados de IMFA em até 10 semanas, portanto a série temporal das notificações foi extraída a partir da trigésima semana de 2017. Havia poucos dias faltantes, então realizamos interpolação linear para preencher as lacunas, exceto para o município de Ecoporanga, que possuía muitos dias faltantes e foi descartado da análise. Para Afonso Cláudio, apenas os dados de 2018 e 2019 estavam disponíveis.

A Figura 5 exibe, como ilustração, os dados coletados

<sup>5</sup>“SINAN:” <https://datasus.saude.gov.br/acesso-a-informacao/doencas-e-agravos-de-notificacao-de-2007-em-diante-sinan/>

do município de Vitória. Percebe-se a forte sazonalidade da temperatura e do índice de vetor. Além disso, há um enorme pico de notificações em 2019, período correspondente a um surto que afetou o país inteiro.

#### B. Modelos de Regressão com o índice vetorial e os dados meteorológicos

Desenvolvemos modelos parcimoniosos de regressão linear, utilizando Modelos Lineares Generalizados (MLG), com uma função de ligação logarítmica (Equação (10)). A transformação logarítmica garante que apenas valores positivos da variável resposta são considerados pelo modelo. Os modelos foram ajustados de forma independente para cada par (município, ano), para obter a influência das variáveis meteorológicas no nível de infestação do vetor (IMFA).

Assumimos primeiramente que a variável resposta (IMFA) segue uma distribuição de Poisson (Equação 10), e, para cada modelo, selecionamos os três melhores preditores rankeando-os por meio de testes F de regressão univariada, para medir o quanto bem cada uma das variáveis ( $temp_i$ ,  $umid_i$  e  $prec_i$ ) explica a resposta (MFAI).

$$\log(\mathbb{E}[Y]) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (10)$$

em que  $Y \sim \text{Poisson}(\mu)$

Em seguida, compararmos os primeiros modelos gerados com modelos que assumem que a resposta segue uma distribuição Binomial Negativa, que também é comumente utilizada para dados de contagem, como capturas, geralmente quando há uma grande dispersão nos dados. Utilizamos como métricas o Critério de Informação de Akaike ( $CIA = 2k - 2\ln(\hat{L})$ ) e o Erro Médio Absoluto ( $EMA = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$ ), e os novos modelos obtiveram um desempenho ligeiramente pior. A métrica CIA balanceia o ajuste do modelo aos dados, representado pelo termo  $\ln(\hat{L})$ , e a complexidade do modelo, representada pelo número de variáveis  $k$ , enquanto a métrica EMA avalia a diferença entre a resposta calculada pelo modelo ajustado e os valores reais da variável resposta. O ideal é que ambas medidas sejam minimizadas (não nos importamos com overfitting, já que os dados não foram divididos em conjuntos de treinamento e teste para atividades de predição).

Por fim, em comparação com a estratégia do rankeamento de variáveis, utilizamos uma seleção de modelos stepwise com a métrica CIA (inserção e remoção iterativa de variáveis para minimizar o CIA). Embora houve melhoria dos valores de CIA, a diferença se deu amplamente pela redução na complexidade dos modelos em vez do aumento na qualidade, pois apenas uma das três variáveis originais eram sempre mantidas, consequentemente reduzindo o CIA enquanto o valor do EMA aumentava. Portanto, a primeira estratégia permaneceu preferida.

#### C. Correlação entre o índice vetorial e casos de dengue

Calculamos os coeficientes de correlação de Spearman e Pearson para avaliar o relacionamento entre o índice de infestação (IMFA), considerando valores adiantados em até

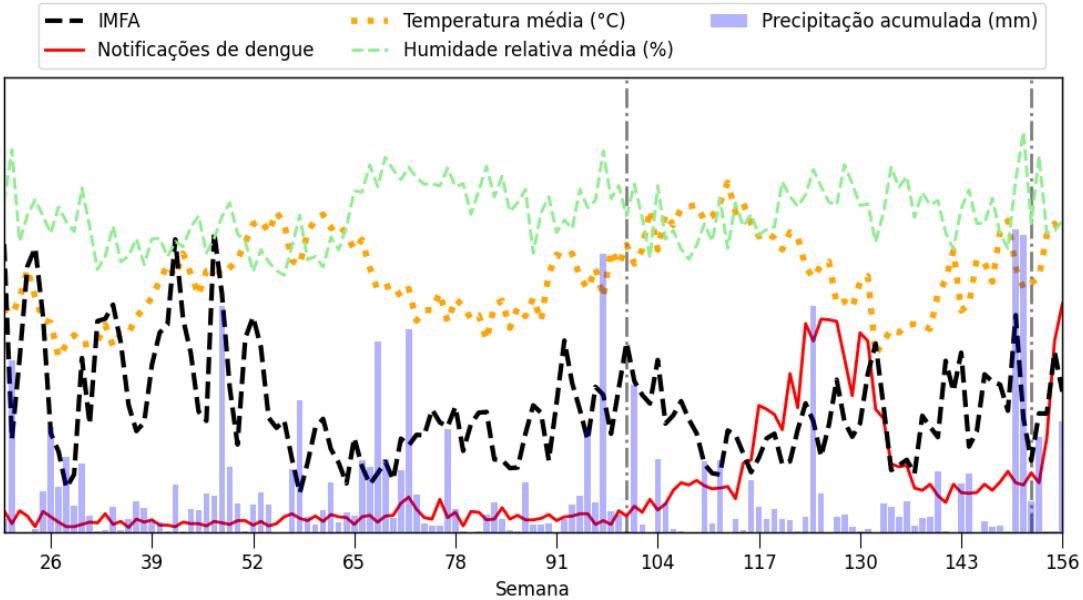


Figura 5. Dados, em escalas distintas, do município de Vitória, Espírito Santo (semanas 2017/20 a 2019/52). As linhas tracejadas verticais indicam o período escolhido para os modelos epidemiológicos.

10 semanas, e as notificações de dengue, juntamente a testes de significância baseados em p-valores.

O coeficiente de Spearman (intervalo de -1 a 1) mede a monotonicidade do relacionamento entre duas variáveis. Coeficientes próximos de -1 indicam uma forte tendência de que, se uma variável aumenta, a outra diminui; valores bem positivos indicam o oposto; e os próximos de 0 sugerem a ausência de relacionamento. O coeficiente de Pearson comporta-se similarmente, mas assume certas premissas, como um relacionamento linear entre as variáveis e uma distribuição normal bivariada delas. O p-valor indica a probabilidade de um sistema sem correlação de produzir datasets com resultados pelo menos tão extremos quanto o computado de um dado dataset; no contexto do trabalho, seria como a probabilidade dos dados de captura e de notificações possuírem a medida de correlação tão alta quanto a calculada por acaso; correlações com p-valores maiores do que 5% foram descartadas.

#### D. Modelos epidemiológicos

Como ilustrado na Figura 5, os modelos epidemiológicos foram desenvolvidos em torno do grande aumento no número de casos de dengue em 2019. Escolhemos o período da semana 48 de 2018 à semana 47 de 2019, totalizando 52 semanas, e consideramos apenas parte dos municípios: Vitória, Vila Velha e Linhares, por já possuírem curvas de infecção com características distintas e para limitar o número de ajustes de modelos, os quais requerem bastante tempo de processamento e muitas correções.

Para cada cidade, três modelos, de um mais simples a um mais complexo, foram ajustados aos dados: os já introduzidos modelos *SIR* (Equações (2) - (4)) e *SIR – SI*, nomeado *SIR – SI v1* (Equações (5) - (9)), e uma segunda versão

do modelo *SIR – SI* (*SIR – SI v2*), que inclui o índice vetorial como fator. Portanto, tomamos a abordagem mais simples de omitir o compartimento de pessoas expostas (*E*), e assumimos que novas notificações correspondem exatamente a novas infecções, as quais, nesse caso, implicam uma capacidade imediata de infeciosidade e o surgimento imediato de sintomas. O parâmetro taxa de infecção ( $\beta$ ) é variável: no caso dos primeiros dois modelos, definimos  $\beta$  como uma função linear da temperatura semanal (Equação (11)), e no caso do terceiro incluímos também o índice de vetor como um fator multiplicativo (Equação (12)). O termo  $\beta_{lag}$  é o parâmetro correspondente a quantas semanas o valor de temperatura está adiantado, para considerar efeitos atrasados na população de mosquitos, como feito no caso dos modelos de regressão.

$$\beta(t) = \beta_0 + \beta_{scale} T(t - \beta_{lag}) \quad (11)$$

$$\beta(t) = \beta_0 + \beta_{scale} T(t - \beta_{lag}) \text{IMFA}(t) \quad (12)$$

Utilizamos a redução da soma dos quadrados dos erros ( $SQE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ ) entre as notificações reais acumuladas semanalmente de dengue e o compartimento acumulado de infectados, um compartimento adicional  $C$  ( $\frac{dC}{dt} = \beta \frac{SI}{N}$ ). Utilizar os dados acumulados, por conta de sua suavidade, geralmente facilita bastante o processo de otimização, bem como o expoente no erro, que penaliza fortemente os erros grandes; entretanto, também buscou-se uma boa semelhança entre as curvas de casos reais e do modelo, o que não é garantido pela boa semelhança dos casos acumulados.

Para definir a população total de humanos ( $N$  ou  $N_h$ ), utilizamos a estimativa do IBGE (Instituto Brasileiro de Geografia e Estatística) [21] da população dos municípios de

2019 (Tabela I). Em relação aos mosquitos, assumimos a população total  $N_v = 2N_h$ . A população inicial de infectados humanos ( $I(0)$  ou  $I_h(0)$ ) foi definida como o número inicial de casos semanais de dengue, seguindo a premissa de correspondência dos novos casos com novas infecções. No caso dos modelos com dinâmica vetorial, ajustamos a proporção de mosquitos infectados inicialmente ( $I_v(0)$ , ver Tabelas III e IV). Assumimos nenhum humano recuperado inicialmente ( $R(0) = 0$ ), e o grupo de suscetíveis inicial, tanto dos humanos quanto dos mosquitos, como toda a população restante ( $S(0) = N - I(0) - R(0)$  e  $S_v(0) = N_v - I_v(0)$ ).

Os parâmetros dos modelos, seus intervalos, valores iniciais e origem são apresentados nas Tabelas II, III e IV. Utilizou-se, para cada modelo, os mesmos valores em todas as cidades. Em relação aos modelos com dinâmica de vetor explícita, assumimos uma taxa de transmissão igual nas duas direções ( $\beta = \beta_{vh} = \beta_{hv}$ ), diferenciando, portanto, a força de infecção apenas pela quantidade de infectados de cada lado (Equações (6) e (9)). Adotamos taxas de nascimento/mortes  $\mu$  constantes, de modo que a população total de humanos e de vetores permaneça constante ( $N_h = S_h + I_h + R_h = C_1$  e  $N_v = S_v + I_v = C_2$ ), bem como uma taxa de recuperação constante  $\gamma$  para os humanos.

Tabela I  
POPULAÇÃO TOTAL EM 2019 DOS MUNICÍPIOS SELECIONADOS PARA OS MODELOS EPIDEMIOLÓGICOS.

Município	Vitória	Vila Velha	Linhares
População	362.097	493.838	173.555

Tabela II  
VALORES INICIAIS E INTERVALOS DE PARÂMETROS UTILIZADOS NO MODELO SIR.

Parâmetro	Intervalo	Valor inicial	Origem
$\mu$ ( $ano^{-1}$ )	–	1/79	[22]
$\gamma$ ( $semana^{-1}$ )	–	1.4	[23]
$\beta_0$ ( $semana^{-1}$ )	(0.01, 1.0)	0.1	Experimentação
$\beta_{scale}$	(0.01, 0.2)	0.05	Experimentação
$\beta_{lag}$ (semanas)	(0, 10)	5	Experimentação

Tabela III  
VALORES INICIAIS E INTERVALOS DE PARÂMETROS UTILIZADOS NO MODELO SIR-SI v1.

Parâmetro	Intervalo	Valor inicial	Origem
$\mu_h$ ( $ano^{-1}$ )	–	1/79	[22]
$\mu_v$ ( $dia^{-1}$ )	–	1/11	[24]
$\gamma$ ( $semana^{-1}$ )	–	1.4	[23]
$\beta_0$ ( $semana^{-1}$ )	(0.1, 1.0)	0.5	Experimentação
$\beta_{scale}$	(0.01, 0.2)	0.05	Experimentação
$\beta_{lag}$ (semanas)	(0, 10)	5	Experimentação
$I_v(0)/N_v$	( $10^{-4}, 10^{-2}$ )	$10^{-4}$	Experimentação

#### IV. RESULTADOS

Esta seção apresenta os resultados da etapa de modelagem da análise exploratória dos dados de vigilância entomológica,

Tabela IV  
VALORES INICIAIS E INTERVALOS DE PARÂMETROS UTILIZADOS NO MODELO SIR-SI v2.

Parâmetro	Intervalo	Valor inicial	Origem
$\mu_h$ ( $ano^{-1}$ )	–	1/79	[22]
$\mu_v$ ( $dia^{-1}$ )	–	1/11	[24]
$\gamma$ ( $semana^{-1}$ )	–	1.4	[23]
$\beta_0$ ( $semana^{-1}$ )	(0.1, 1.0)	0.5	Experimentação
$\beta_{scale}$	(0.01, 0.2)	0.05	Experimentação
$\beta_{lag}$ (semanas)	(0, 10)	5	Experimentação
$I_v(0)/N_v$	( $10^{-4}, 10^{-2}$ )	$10^{-4}$	Experimentação

dados meteorológicos e dos casos de dengue, incluindo os modelos de regressão (seção IV-A), a análise de correlação (seção IV-B) e os modelos epidemiológicos (seção IV-C).

#### A. Modelos de Regressão

A Tabela V exibe os coeficientes dos modelos de regressão aplicados aos municípios. Variáveis de temperatura foram as mais relevantes: foram as mais selecionadas e demonstraram os maiores efeitos na densidade do vetor, o que atesta seu papel importante no desenvolvimento e sobrevivência do mosquito [25]. Mesmo assim, a temperatura nem sempre foi um dos melhores preditores em Vila Velha e Afonso Cláudio, por exemplo, embora permanecesse importante em Vitória, o município vizinho de Vila Velha. A umidade foi o segundo fator mais importante, seguido pela precipitação.

No que diz respeito à consistência do impacto das variáveis meteorológicas na variação populacional dos mosquitos, os resultados indicam que tanto o conjunto de variáveis mais relevantes quanto a magnitude de sua influência podem variar durante anos distintos para o mesmo local. Por exemplo, por 2017 e 2019, fatores de umidade e precipitação prevaleceram em Vila Velha, enquanto em 2018 a temperatura prevaleceu. Já em Ecoporanga, a influência da temperatura se deu com coeficientes bem diferentes de 2017 a 2018. Além disso, resultados de locais vizinhos podem diferir consideravelmente, como os de Vitória e Vila Velha. No entanto, ainda é possível observar casos em que não há mudanças tão consideráveis, como em Vitória (2018-2019), Afonso Cláudio (2018-2019) e Linhares (2017-2018).

#### B. Análise de Correlação

Reunimos, para cada município, as três maiores correlações entre o índice de infestação ( $IMFA$ ) e os casos de dengue (Tabela VI). Os anos 2017 de Afonso Cláudio e todos os anos de Ecoporanga não possuíam dados suficientes, por isso não estão presentes. A ausência de outros valores deu-se pela falta de uma correlação estatisticamente significante ( $p$ -valor maior do que 5%).

Levando em conta múltiplos municípios e anos distintos, o relacionamento é consistentemente contraintuitivo: quanto maior a infestação, menor o número de casos, evidenciando a conexão complexa entre as duas variáveis [12]. A única exceção desse padrão foi Afonso Cláudio. Além disso, o grande deslocamento, em semanas, um ano após o outro, das

Tabela V

COEFICIENTES DOS MODELOS DE REGRESSÃO AJUSTADOS, PARA CADA PAR MUNICÍPIO-ANO. VALORES COM MAIOR MAGNITUDE DESTACADOS, E VARIÁVEIS NÃO SELECIONADAS MARCADAS COM UM TRAÇO (-). AS SUB-COLUNAS ABAIXO DE CADA VARIÁVEL INDICAM O NÚMERO DE SEMANAS ADIANTADAS.

Ano	temp <sub>i</sub>					umid <sub>i</sub>				prec <sub>i</sub>				
	0	1	2	3	4	0	1	3	4	0	1	2	3	4
<b>Vitória</b>														
2017	-0,02	<b>0,19</b>	-	-	-	-	-	-	-	-	-0,14	-	-	-
2018	-	-	-	0,03	<b>-0,12</b>	-	-	-	-	-	-	-	0,10	-
2019	-	-	-	-0,01	-0,08	-	-	<b>0,11</b>	-	-	-	-	-	-
<b>Vila Velha</b>														
2017	-	-	-	-	-	-0,15	<b>-0,31</b>	-	-	-	-	-	-	0,20
2018	<b>0,24</b>	0,09	-	-	-	-	-0,19	-	-	-	-	-	-	-
2019	-	-	-	-	-	-	-	0,11	0,10	-	<b>-0,12</b>	-	-	-
<b>Afonso Cláudio</b>														
2017	-	-	-	0,10	<b>0,81</b>	-	-	<b>0,75</b>	-	0,50	-	0,25	-	-
2018	-	0,35	-	-	-	-	-	-	-	-	-	-	-	-
2019	-	-	0,43	<b>0,94</b>	-	-	-	-	-	-	0,31	-	-	-
<b>Linhares</b>														
2017	-	0,17	-	-	-	-0,17	<b>-0,21</b>	-	-	-	-	-	-	-
2018	0,06	<b>0,22</b>	-	-	-	-	-0,10	-	-	-	-	-	-	-
2019	-	-	-	-	-0,07	-	-	0,10	<b>0,15</b>	-	-	-	-	-
<b>Ecoporanga</b>														
2017	0,15	<b>0,32</b>	0,12	-	-	-	-	-	-	-	-	-	-	-
2018	-	-0,10	<b>0,52</b>	-	-	-	-	-	-	-	-	0,21	-	-
2019	-	-	-	-0,14	<b>-0,37</b>	-	-	0,16	-	-	-	-	-	-

Tabela VI

CORRELAÇÕES MAIS FORTES ENTRE A DENSIDADE DE MOSQUITOS (*IMFA*) E AS NOTIFICAÇÕES DE DENGUE, PARA CADA PAR MUNICÍPIO-ANO. VALORES AUSENTES MARCADOS COM UM TRAÇO (-), E P-VALORES EM PARÊNTESES. TIPOS DE CORRELAÇÃO: PEARSON (P) E SPEARMAN (S). O SUBSCRITO INDICA O VALOR DA DENSIDADE ADIANTADO EM ATÉ DEZ SEMANAS EM RELAÇÃO ÀS NOTIFICAÇÕES.

Ano	Top 1	Top 2	Top 3
<b>Vitória</b>			
2019	<i>IMFA</i> <sub>10</sub> (P): -0,39 (0,00)	<i>IMFA</i> <sub>9</sub> (P): -0,33 (0,02)	<i>IMFA</i> <sub>7</sub> (P): -0,33 (0,02)
<b>Vila Velha</b>			
2017	<i>IMFA</i> <sub>0</sub> (S): -0,43 (0,04)	-	-
2018	<i>IMFA</i> <sub>10</sub> (S): -0,41 (0,00)	<i>IMFA</i> <sub>1</sub> (P): -0,29 (0,04)	-
<b>Afonso Cláudio</b>			
2018	<i>IMFA</i> <sub>0</sub> (P): 0,54 (0,00)	<i>IMFA</i> <sub>1</sub> (P): 0,49 (0,00)	<i>IMFA</i> <sub>2</sub> (S): 0,32 (0,02)
2019	<i>IMFA</i> <sub>5</sub> (P): 0,75 (0,00)	<i>IMFA</i> <sub>4</sub> (P): 0,73 (0,00)	<i>IMFA</i> <sub>6</sub> (P): 0,68 (0,00)
<b>Linhares</b>			
2017	<i>IMFA</i> <sub>4</sub> (S): -0,51 (0,01)	-	-
2018	<i>IMFA</i> <sub>0</sub> (P): -0,45 (0,00)	<i>IMFA</i> <sub>1</sub> (P): -0,37 (0,01)	<i>IMFA</i> <sub>6</sub> (P): -0,34 (0,01)
2019	<i>IMFA</i> <sub>8</sub> (S): -0,41 (0,00)	<i>IMFA</i> <sub>7</sub> (P): -0,35 (0,01)	<i>IMFA</i> <sub>10</sub> (P): -0,33 (0,02)

variáveis com maior correlação, como em Vila Velha e em Linhares, releva uma inconsistência temporal que ressalta a falta de um padrão claro de relação entre as variáveis.

### C. Modelos Epidemiológicos

A Tabela VII apresenta os valores dos parâmetros dos modelos epidemiológicos ajustados aos municípios. Pode-se notar que o parâmetro  $\beta_{lag}$  é consistentemente próximo de 5 semanas para os modelos mais simples *SIR*, e maior, próximo de 9 semanas, para os dois tipos de modelos mais complexos *SIR-SI*. Além disso, por mais que possuam no geral um valor de  $\beta_0$  maior, os modelos mais complexos *SIR-SI* possuem sempre um  $\beta_{scaling}$  menor. Na seção V seguinte, mostraremos que os valores de  $\beta(t)$  nos modelos mais complexos tendem a ser menores. Isso pode ser explicado pela maior população de mosquitos infectados (inicialmente 0,1% de  $N_v = 2N_h$ ) se comparada com a população humana

de infectados, de modo que, para equilibrar a força de infecção  $(\beta(S_h I_v)/N_h)$  é necessário um  $\beta$  menor.

Tabela VII  
PARÂMETROS ESTIMADOS POR CADA MODELO EPIDEMIOLÓGICO, AGRUPADOS POR MUNICÍPIO.

Modelo	$\beta_0$	$\beta_{scaling}$	$\beta_{lag}$	$I_v(0)$
<b>Vitória</b>				
SIR	0,1115	0,0519	4,94	-
SIR-SI v1	0,1858	0,0199	9,00	$10^{-4}$
SIR-SI v2	0,1000	0,0254	8,67	$10^{-4}$
<b>Vila Velha</b>				
SIR	0,1118	0,0520	4,93	-
SIR-SI v1	0,1798	0,0202	9,00	$10^{-4}$
SIR-SI v2	0,1884	0,0196	8,89	$10^{-4}$
<b>Linhares</b>				
SIR	0,1105	0,0520	4,94	-
SIR-SI v1	0,1055	0,0233	9,95	$10^{-4}$
SIR-SI v2	0,1741	0,0181	9,03	$10^{-4}$

No que diz respeito à qualidade de ajuste aos dados dos modelos, a Tabela VIII compara as infecções semanais e acumuladas calculadas a partir dos modelos com os dados reais, utilizando o RMSE e o  $R^2$ . O RMSE (raiz do erro quadrático médio, Equação 13) é como uma normalização da soma dos quadrados dos erros (ver seção III-D), mantendo a unidade de medida dos dados originais e uma escala mais adequada. O  $R^2$  (coeficiente de determinação), por sua vez, indica o quanto um modelo explica a variância nos dados; Um  $R^2$  próximo de 1 é melhor, enquanto valores próximos de 0, piores; valores negativos são possíveis, e indicam um modelo totalmente não representativo, pior do que apenas uma média. Além disso, para uma comparação objetiva dos picos de casos reais e dos modelos, duas métricas foram adicionadas, as quais medem a distância da posição e do valor do pico (respectivamente,  $Dist(p)$  e  $Ratio(p)$ ).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (13)$$

O modelo *SIR* apresentou uma performance muito boa, provavelmente por sua maior simplicidade, enquanto o modelo *SIR-SI* v1, mesmo que pior, apresentou um resultado satisfatório. Por fim, o modelo mais complexo *SIR-SI* v2 mostrou uma performance muito pior consistentemente. É importante ressaltar que um bom ajuste dos casos acumulados não é garantia de um bom ajuste dos casos semanais, como a própria Tabela VIII mostra.

### V. DISCUSSÃO

Esta seção busca discutir e interpretar os resultados da etapa de modelagem, de forma a concluir a análise do trabalho.

Tabela VIII  
DESEMPENHO DOS MODELOS SIR, SIR-SI v1 E SIR-SI v2 AJUSTADOS  
PARA CADA MUNICÍPIO (2018/48 - 2019/47): ERRO (RAIZ DO ERRO  
QUADRÁTICO MÉDIO -  $RMSE$ ) E COEFICIENTE DE DETERMINAÇÃO ( $R^2$ ).  
EM RELAÇÃO AOS DADOS REAIS ACUMULADOS ( $C$ ) E SEMANAIS ( $S$ ).  
TAMBÉM, APRESENTA-SE A DISTÂNCIA SEMANAL ENTRE PICOS DE  
INFESTAÇÃO ( $Dist(p)$ ) E A RAZÃO DOS PICOS ( $Ratio(p)$ ), EM RELAÇÃO  
AOS DADOS REAIS.

Modelo	$RMSE (C/S)$	$R^2 (C/S)$	$Dist(p)$	$Ratio(p)$
<b>Vitória</b>				
SIR	209.07 / 38.35	0.99 / 0.77	3	1.01
SIR-SI	534.46 / 74.89	0.93 / 0.11	10	0.81
SIR-SI v2	833.88 / 124.15	0.84 / -1.45	17	0.89
<b>Vila Velha</b>				
SIR	146.84 / 41.56	1.00 / 0.70	1	0.88
SIR-SI v1	439.15 / 64.97	0.96 / 0.27	4	0.72
SIR-SI v2	796.82 / 119.39	0.86 / -1.48	13	0.77
<b>Linhares</b>				
SIR	122.33 / 32.40	0.99 / 0.66	1	0.66
SIR-SI v1	251.74 / 43.87	0.96 / 0.38	9	0.62
SIR-SI v2	980.21 / 104.18	0.44 / -2.50	17	0.84

#### A. Modelos de Regressão

A influência de variáveis meteorológicas na população de vetores depende fortemente do local de estudo [6] [7] [15], o que foi observado nesse trabalho até em municípios próximos, como em Vila Velha e Vitória. As inconsistências temporais nesse tipo de modelagem com índices vetoriais - como fortes mudanças nos coeficientes, deslocamentos de semanas ou mudanças de variáveis - podem ser explicadas por diversos motivos, sejam mudanças nas políticas de controle vetorial, mudanças nas frequências de amostragem semanais, características das armadilhas que conferem capacidades de captura sazonais [15], ou talvez até mudanças maiores/eventos externos, como alterações climáticas. De qualquer forma, as diferenças demonstram uma importância do monitoramento local contínuo para um melhor entendimento das dinâmicas de população do mosquito.

#### B. Análise de Correlação

Casos de dengue foram associados de forma fraca ou contraintuitiva com a infestação de vetores. Múltiplas condições podem contribuir para a evolução dos casos, como fatores climáticos e socioeconômicos [26]. Em 2019, a explosão de casos deu-se por razão de um conjunto de circunstâncias, como a circulação de diversos sorotipos do vírus, mudanças climáticas e políticas públicas ineficazes, de acordo com o infectologista José Moreira [27].

No entanto, um dos municípios, Afonso Cláudio, apresentou uma tendência de correlações bem diferente, com valores bem positivos. A Figura 6 mostra a evolução do índice de infestação e dos casos de dengue em Vitória e em Afonso Cláudio. O padrão de Vitória é similar ao dos outros municípios. A esparsidade dos dados de Afonso Cláudio, por conta da taxa de amostragem de armadilhas relativamente baixa, fortalece a coincidência dos picos de infestação com picos de dengue, de

modo que os valores observados são provavelmente excepcionais. Em suma, é necessário uma exploração mais profunda para entender o relacionamento entre as duas variáveis.

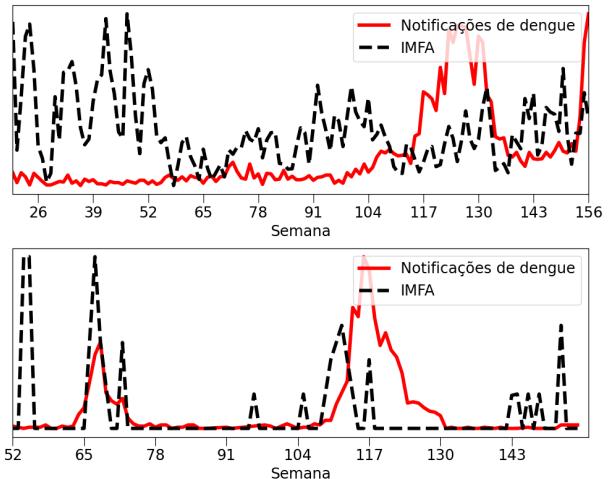


Figura 6. Evolução do índice vetorial (IMFA) e das notificações de dengue, em escalas distintas, das cidades de Vitória (acima) e Afonso Cláudio (abaixo).

#### C. Modelos Epidemiológicos

Entre os três modelos epidemiológicos, o mais simples *SIR* apresentou o melhor resultado de ajuste aos dados, inclusive aproximando as curvas de infecções reais com boa precisão, como mostram as Figuras 7, 8 e 9. Esse resultado têm sido observado em trabalhos como [19], abordado na seção II. O menor número de premissas, interações e variáveis provavelmente tornam o modelo mais fácil de ser ajustado. Por outro lado, o modelo não será suficiente se o objetivo for um estudo mais profundo das dinâmicas da doença.

Os resultados revelam um êxito ao utilizar a temperatura como um fator determinante da força de infecção da doença. Nesse caso, os modelos *SIR* identificaram uma influência mais rápida (cerca de 5 semanas) da variação da temperatura (parâmetro  $\beta_{lag}$ ), enquanto os modelos mais complexos *SIR-SI* identificaram um efeito mais tardio. Entre os fatores para o aumento da disseminação da doença, o tempo de incubação do vírus no humano e no vetor é fortemente impactado pela temperatura [28], cujas flutuações naturalmente poderão resultar em um efeito atrasado na população.

Em relação aos valores da taxa de transmissão, os modelos mais complexos *SIR-SI* tendem a possuir valores de  $\beta(t)$  menores, como abordado na seção IV-C. As Figuras 11, 12 e 13 exibem exatamente a evolução dos valores de  $\beta$  de cada modelo e cada município. Considerando um mesmo modelo, as curvas de transmissão de cada município são bem similares entre si, com um pequeno deslocamento vertical por conta das variações nos números de casos reais. Além disso, as curvas dos primeiros dois modelos possuem uma aparência similar de pico às curvas de casos, mas são deslocadas à esquerda, o que não é uma surpresa, pois o ápice da taxa de transmissão é bem antes do pico, e começa a decair logo em seguida. Ao

observar as Figuras 14, 15 e 16, percebe-se (de forma mais clara com os dados de 2019) um padrão claro entre dengue e temperatura, justificando essa aparência.

Por fim, o terceiro modelo, que considera o índice de vetor em sua formulação, obteve resultados consideravelmente piores. A Figura 10 mostra a comparação das curvas de casos reais com as curvas do melhor resultado do modelo (Vila Velha). Observando-se novamente as Figuras 14, 15 e 16, percebe-se que o índice de vetor (*IMFA*) é altamente sazonal, assim como a temperatura. Por outro lado, visualmente não há um padrão claro de ligação do *IMFA* com os casos de dengue, o que vai ao encontro da análise de correlação.

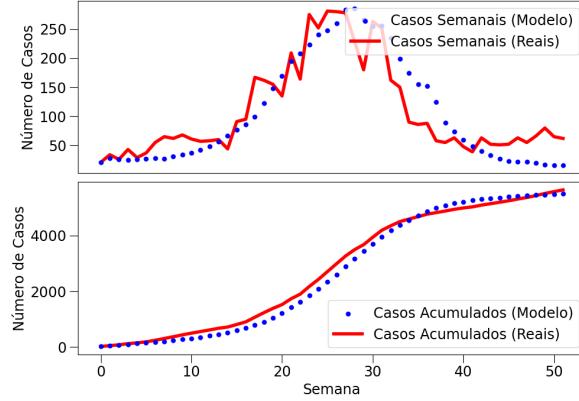


Figura 7. Modelo *SIR* vs casos reais - Vitória (2018/48 - 2019/47).

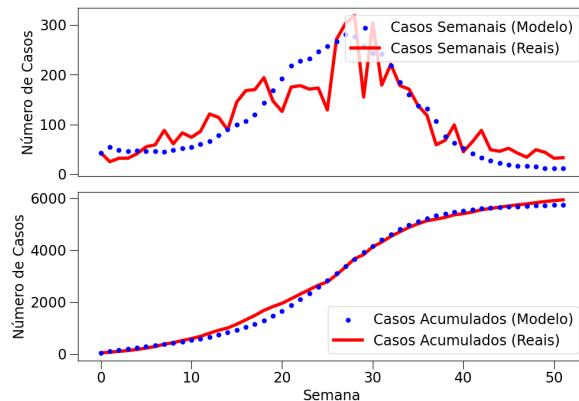


Figura 8. Modelo *SIR* vs casos reais - Vila Velha (2018/48 - 2019/47).

Um dos efeitos disso nas curvas de taxa de transmissão do terceiro modelo (Figura 13) é a criação de um vale por volta das primeiras semanas, quando o valor deveria ser mais alto. Isso se deu pela baixa infestação nesse período inicial (mais perceptível em Vitória ou em Linhares, Figuras 14 e 16), e o pelo aumento da infestação apenas conforme os casos encaminhavam-se para uma normalização.

Dessa maneira, o uso dos índices no modelo *SIR* – *SI* v2 de forma geral assemelhou-se mais à introdução de um ruído. Portanto, concluímos que há necessidade de um estudo

mais aprofundado e multifatorial para a associação do índice vetorial com os casos de dengue em modelos epidemiológicos.

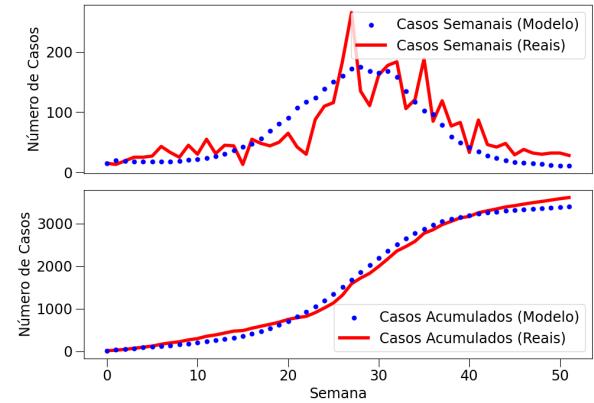


Figura 9. Modelo *SIR* vs casos reais - Linhares (2018/48 - 2019/47).

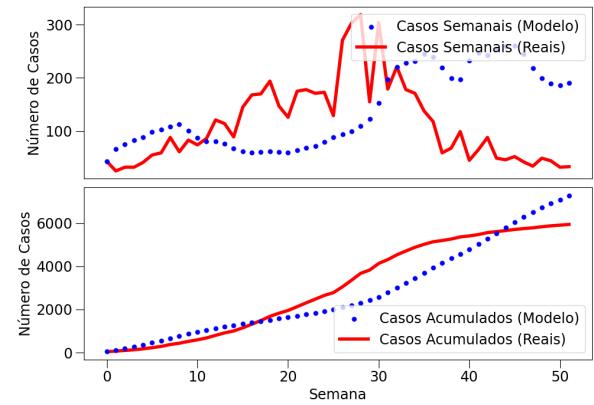


Figura 10. Modelo *SIR* – *SI* v2 vs casos reais - Vila Velha (2018/48 - 2019/47).

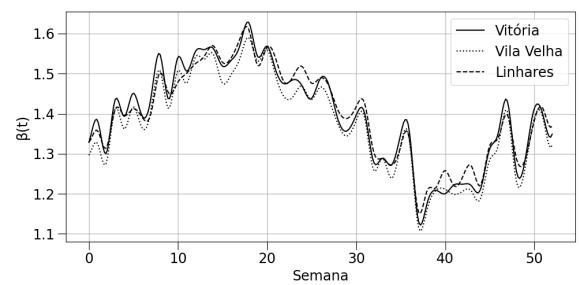


Figura 11. Modelo *SIR* - Evolução da taxa de transmissão em cada município.

## VI. CONCLUSÃO

Este trabalho teve por objetivo analisar o relacionamento de dados de vigilância entomológica (índice de vetor) com dados meteorológicos e com casos de dengue de alguns municípios do estado do Espírito Santo. Desenvolvemos modelos de regressão e efetuamos análises de correlação, de forma a,

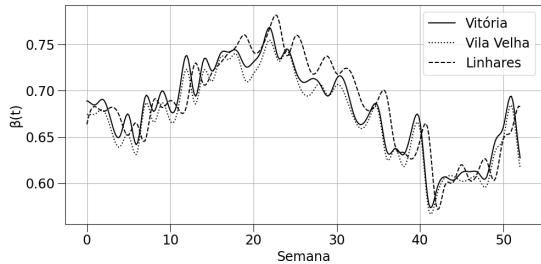


Figura 12. Modelo  $SIR - SI$  v1 - Evolução da taxa de transmissão em cada município.

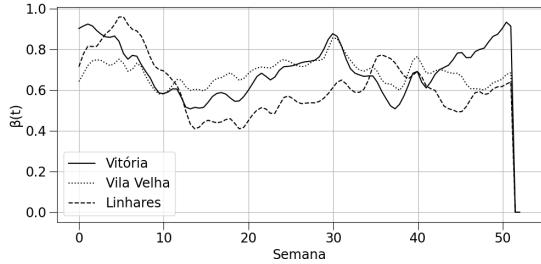


Figura 13. Modelo  $SIR - SI$  v2 - Evolução da taxa de transmissão em cada município.

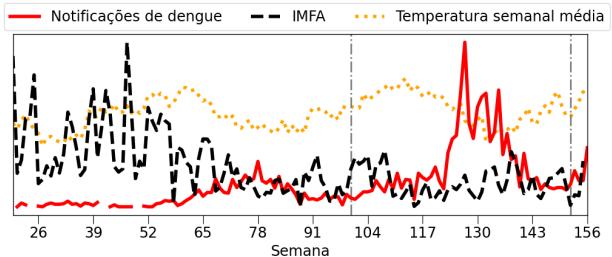


Figura 16. Em escalas distintas, notificações de dengue, índice de infestação (IMFA) e temperatura semanal média - Linhares (2017/20 - 2019/52). As linhas tracejadas verticais indicam o período escolhido para os modelos epidemiológicos.

pacto da inclusão do índice de vetor na formulação.

É bem estabelecido que há uma forte relação entre fatores ambientais, como climáticos, e índices de mosquitos, sejam larvais, pupais ou de adultos [4] [5] [28]. Um dos fatores mais relevantes é a temperatura, como encontrado pelos modelos de regressão desenvolvidos. Além disso, as diferenças consideráveis nos modelos, ano após ano nos mesmos locais e até em relação a locais vizinhos, demonstra a importância da vigilância contínua e localizada.

Em relação à análise de correlação, os casos de dengue exibiram uma correlação fraca ou contraintuitiva com a abundância do vetor, destacando a característica multifatorial da doença. O mecanismo que descreve o relacionamento entre as duas variáveis não é bem compreendido pela literatura [12]. Nesse sentido, a utilização do índice de vetor não apresentou vantagem nos modelos epidemiológicos de incidência de dengue, enquanto a temperatura mostrou-se um ótimo fator.

Limitações na etapa de regressão incluem não considerar outras variáveis relevantes, como vegetação [7], não considerar a possível influência do ambiente na efetividade das capturas das armadilhas ou possíveis mudanças de regimes de controle vetorial nos municípios. No caso da análise de correlação, consideramos correlações apenas temporais e não espaciais. Por fim, a modelagem epidemiológica, ao passo que demonstra um aspecto importante dos dados, pode ter sido simplória, sendo necessário um desenvolvimento de um modelo multifatorial para integrar os dados de vigilância de forma satisfatória.

Sugestões de trabalhos futuros incluem: analisar diferenças e a consistência de modelos de previsão de casos de dengue com variáveis meteorológicas, aplicadas em locais diferentes e por longos períodos de tempo distintos (como anos); integrar dados reais de infecção de dengue em mosquitos ou integrar dados espaciais, para análises de correlação com casos em humanos; e desenvolver e comparar modelos epidemiológicos de dengue com processos/mecanismos mais complexos, que considerem o índice de vetor em sua formulação.

## VII. AGRADECIMENTOS

Agradecemos à BH-TEC<sup>6</sup> pelo projeto TCCLAB, que trouxe a conexão com a empresa Ecovec para a realização do

<sup>6</sup>Site da BHTEC: <https://bhtec.org.br/>

respectivamente, avaliar o relacionamento dos índices de vetor com as variáveis meteorológicas e com a incidência de dengue. Finalmente, exercitamos o uso de modelos epidemiológicos compartimentais com os casos de dengue, verificando o im-

trabalho; agradecemos à empresa Ecovec pela disponibilização dos dados; e agradecemos à Secretaria Estadual de Saúde do Espírito Santo (SESA-ES) e ao Núcleo Especial de Vigilância Ambiental (NEVA) pela autorização do uso e divulgação dos dados de vigilância entomológica dos municípios.

## REFERÊNCIAS

- [1] V. Teich, R. Arinelli, and L. Fahham, “Aedes aegypti e sociedade: O impacto econômico das arboviroses no Brasil,” *Jornal Brasileiro de Economia da Saúde*, vol. 9, no. 3, p. 267–276, Dec 2017.
- [2] W.-X. Zhang, T.-Y. Zhao, C.-C. Wang, Y. He, H.-Z. Lu, H.-T. Zhang, L.-M. Wang, M. Zhang, C.-X. Li, and S.-Q. Deng, “Assessing the global dengue burden: Incidence, mortality, and disability trends over three decades,” *PLOS Neglected Tropical Diseases*, vol. 19, no. 3, Mar 2025.
- [3] “Vigilância e Controle do Votor – Ministério da Saúde,” Ministério da Saúde, 2025. [Online]. Available: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/aedes-aegypti/vigilancia-entomologica>
- [4] L. Lambrechts, K. P. Paaijmans, T. Fansiri, L. B. Carrington, L. D. Kramer, M. B. Thomas, and T. W. Scott, “Impact of daily temperature fluctuations on dengue virus transmission by Aedes aegypti,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 18, p. 7460–7465, Apr 2011.
- [5] J. P. Messina, M. U. Kraemer, O. J. Brady, D. M. Pigott, F. M. Shearer, D. J. Weiss, N. Golding, C. W. Ruktanonchai, P. W. Gething, E. Cohn, and et al., “Mapping global environmental suitability for zika virus,” *eLife*, vol. 5, Apr 2016.
- [6] D. A. d. Cruz Ferreira and et al., “Meteorological variables and mosquito monitoring are good predictors for infestation trends of Aedes aegypti, the vector of dengue, chikungunya and zika,” *Parasites & Vectors*, vol. 10, no. 1, Feb 2017.
- [7] O. Mudele and et al., “Modeling dengue vector population with earth observation data and a generalized linear model,” *Acta Tropica*, vol. 215, p. 105809, Mar 2021.
- [8] H. W. Hethcote, “The mathematics of Infectious Diseases,” *SIAM Review*, vol. 42, no. 4, p. 599–653, Jan 2000.
- [9] G. C. Silva, A. G. Utsumi, and N. G. Sousa, “Dengue transmission dynamics analysis and simulation in Minas Gerais - Brazil,” *IEEE Latin America Transactions*, vol. 20, no. 6, p. 1012–1017, Jun 2022.
- [10] M. Derouich, A. Boutayeb, and E. Twizell, “A model of dengue fever,” *BioMedical Engineering OnLine*, vol. 2, no. 1, Feb 2003.
- [11] R. Ximenes, M. Amaku, L. F. Lopez, F. A. Coutinho, M. N. Burattini, D. Greenhalgh, A. Wilder-Smith, C. J. Struchiner, and E. Massad, “The risk of dengue for non-immune foreign visitors to the 2016 Summer Olympic Games in Rio de Janeiro, Brazil,” *BMC Infectious Diseases*, vol. 16, no. 1, Apr 2016.
- [12] L. R. Bowman, S. Runge-Ranzinger, and P. J. McCall, “Assessing the relationship between vector indices and dengue transmission: A systematic review of the evidence,” *PLOS Neglected Tropical Diseases*, vol. 8, no. 5, May 2014.
- [13] E. Descloux, M. Mangeas, C. E. Menkes, M. Lengaigne, A. Leroy, T. Tehei, L. Guillaumot, M. Teurlai, A.-C. Gourinat, J. Benzler, and et al., “Climate-based models for understanding and forecasting dengue epidemics,” *PLOS Neglected Tropical Diseases*, vol. 6, no. 2, Feb 2012.
- [14] C. B. Leach, J. A. Hoeting, K. M. Pepin, A. E. Eiras, M. B. Hooten, and C. T. Webb, “Linking mosquito surveillance to dengue fever through Bayesian mechanistic modeling,” *PLOS Neglected Tropical Diseases*, vol. 14, no. 11, Nov 2020.
- [15] C. T. Codeço, A. W. Lima, S. C. Araújo, J. B. Lima, R. Maciel-de Freitas, N. A. Honório, A. K. Galardo, I. A. Braga, G. E. Coelho, and D. Valle, “Surveillance of Aedes aegypti: Comparison of house index with four alternative traps,” *PLOS Neglected Tropical Diseases*, vol. 9, no. 2, Feb 2015.
- [16] M. GARNER and S. HAMILTON, “Principles of epidemiological modelling,” *Revue Scientifique et Technique de l’OIE*, vol. 30, no. 2, p. 407–416, Aug 2011.
- [17] O. N. Bjørnstad, K. Shea, M. Krzywinski, and N. Altman, “The SEIRS model for infectious disease dynamics,” *Nature Methods*, vol. 17, no. 6, p. 557–558, Jun 2020.
- [18] D. L. Smith, K. E. Battle, S. I. Hay, C. M. Barker, T. W. Scott, and F. E. McKenzie, “Ross, macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens,” *PLoS Pathogens*, vol. 8, no. 4, Apr 2012.
- [19] A. Pandey, A. Mubayi, and J. Medlock, “Comparing vector-host and SIR models for dengue transmission,” *Mathematical Biosciences*, vol. 246, no. 2, p. 252–259, Dec 2013.
- [20] “Aedes aegypti – ministério da saúde,” Ministério da Saúde, 2025. [Online]. Available: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/aedes-aegypti>
- [21] “IBGE divulga as estimativas da população dos municípios para 2019,” IBGE, Aug 2019. [Online]. Available: <https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/25278-ibge-divulga-as-estimativas-da-populacao-dos-municipios-para-2019>
- [22] “Tábua completa de mortalidade para o Brasil – 2019,” IBGE, Nov 2020. [Online]. Available: [https://agenciadenoticias.ibge.gov.br/media/com\\_mediaibge/arquivos/65c3023462edaabf0d7318c1a0f80ca4.pdf](https://agenciadenoticias.ibge.gov.br/media/com_mediaibge/arquivos/65c3023462edaabf0d7318c1a0f80ca4.pdf)
- [23] C. P. Simmons, J. J. Farrar, N. van Vinh Chau, and B. Wills, “Dengue,” *New England Journal of Medicine*, vol. 366, no. 15, p. 1423–1432, Apr 2012.
- [24] M. Otero, H. G. Solari, and N. Schweigmann, “A stochastic population dynamics model for aedes aegypti: Formulation and application to a city with temperate climate,” *Bulletin of Mathematical Biology*, vol. 68, no. 8, p. 1945–1974, Jul 2006.
- [25] J. Couret and M. Q. Benedict, “A meta-analysis of the factors influencing development rate variation in aedes aegypti (diptera: Culicidae),” *BMC Ecology*, vol. 14, no. 1, p. 3, 2014.
- [26] S. C. de Sousa, M. Carneiro, E. Eiras, J. M. Bezerra, and D. S. Barbosa, “Factors associated with the occurrence of dengue epidemics in Brazil: A systematic review,” *Revista Panamericana de Salud Pública*, vol. 45, p. 1, Aug 2021.
- [27] A. Tavares, “Por que os casos de dengue aumentaram no Brasil? Circulação dos quatro vírus e mudanças climáticas ajudam a explicar crescimento,” Instituto Butantan, Aug 2024. [Online]. Available: <https://butantan.gov.br/noticias/por-que-os-casos-de-dengue-aumentaram-no-brasil-circulacao-dos-quatro-virus-e-mu>
- [28] M. Chan and M. A. Johansson, “The incubation periods of dengue viruses,” *PLoS ONE*, vol. 7, no. 11, Nov 2012.