

Inversion of the SIR-SI System for Estimation of Human-Vector Contact Rate and Prediction of Dengue Cases

T. Silva, and J. Montalvão

Abstract—A method for simulating future scenario of dengue cases is proposed in this article. Its main feature is the use of disease notification history to retrieve information about the SIR-SI type, such as: human-vector contact rate history, value related to the mosquito population; and the present state of the system, which allows the evolution of the system to a next stages. This feature is important because it enables the responsible health agency to assess dengue outbreak risk based on its history. For this purpose, a SIR-SI dynamic model was adapted, with human-vector contact rate as input and fraction of new infected individuals as output. Furthermore, a procedure was suggested to estimate the input and the state of the system in such a way that they explain the history of disease notifications, using as criterion the average quadratic error between simulated and real data.

Index Terms—Dengue, Model SIR-SI, Input and State System Retrieve, Prediction of Dengue Cases.

I. INTRODUÇÃO

A dengue é uma doença transmitida pela fêmea do *Aedes Aegypti*, sendo comum em países de clima tropicais devido às condições ideais para o desenvolvimento do mosquito. No Brasil, surtos de dengue costumam aparecer em períodos chuvosos do ano, meses com maiores índices de infestação do mosquito, o que provoca alta procura por serviços de saúde com suspeita da doença. Para se ter noção quantitativa, no mês de fevereiro de 2007 em Mato Grosso do sul foram registrados 23762 notificações de suspeitas de casos da dengue, isso corresponde a mais de 1% da população residente [1]. Esse não foi o único período grave do estado, tampouco é a única região brasileira a sofrer tais problemas.

Não por menos, a dengue é considerada um problema de saúde pública grave, demandando esforços de diferentes setores que atuam, essencialmente, na prevenção, controle dos vetores e cuidados dos infectados. Essa doença, além de possuir afincado direto com a questão da saúde nacional, tem sido constantemente revisitado em estudos científicos, em especial, quando relacionados à aplicações de teorias matemáticas de processamento de sinais. Estes, por exemplo, podem se preocupar em variados objetivos, como: entender o comportamento da propagação da dengue por meio de modelos dinâmicos [2]; estimativa das populações de acordo com as condições de saúde (suscetíveis, infectadas e recuperadas) [3]; previsão de novos casos da doença [4].

A possibilidade de prever os casos de dengue pode ser bastante útil aos órgãos responsáveis para prevenção e precaução da doença. Nesta linha, modelos matemáticos podem ser usados para esta tarefa. De fato, [4] realizaram isto utilizando métodos estocásticos, os quais, segundo os autores, mostraram resultados coerentes, exceto para um período em que houve racionamento de água, distúrbio que provocou um aumento na população de vetores devido ao maior número de criadouros. Isto não tinha sido considerado no modelo usado pelos autores. Deste modo, o número de mosquito tem importância nas previsões, variável que é incluída no modelo SIR (Suscetíveis, Infectados e Recuperados) em [5], mas não com o objetivo de predição. De fato, o modelo SIR é versátil, primeiramente aplicado no campo da saúde [6], mas adaptável a fenômenos análogos artificiais, como em [7]. Do mesmo modo, os modelos dinâmicos do tipo SIR para dengue apresentam versatilidade na inclusão da variável número de mosquitos, sendo bastante úteis também em simulações futuras, mostrando-se um campo a ser explorado.

Para que modelos dinâmicos sejam úteis em simulações futuras, condizentes com a realidade, deve-se conhecer suas constantes, além dos valores atuais das variáveis de estado¹ e da entrada. Nesse sentido, há dois problemas nos modelos para a dengue: seu estado atual não é conhecido, tampouco a entrada é de fácil obtenção. No entanto, tem-se disponível o histórico de notificações de casos da dengue, dados esses que ‘contam a história’ da evolução do sistema, assim, estes podem conter informações úteis do passado de todos os elementos, inclusive os faltantes. Diante disso, abre-se espaço para a possibilidade de simular situações futuras embasadas no que ocorreu. De fato, essas limitações são focos deste trabalho, concentrando os esforços em recuperar a ‘história’ do sistema.

Os resultados desses interesses são dispostos neste trabalho da seguinte forma: Seção II, é apresentado o modelo dinâmico SIR-SI (humanos Suscetíveis, Infectados, Recuperados e vetores Suscetíveis, Infectados); Seção III, expõe-se o problema de forma matemática; Seção IV, divulgam-se alguns detalhes pertinentes na obtenção dos dados históricos de notificação dos casos de dengue e de seu uso; Seção V e VI, mostra-se a resolução do problema proposto junto das considerações feitas; Finalmente, na Seção VII é apresentado o potencial de

T. Silva, Departamento de Engenharia Elétrica, Universidade Federal de Sergipe, São Cristóvão, Sergipe, Brasil, thiagosil143@gmail.com.

J. Montalvão, Departamento de Engenharia Elétrica, Universidade Federal de Sergipe, São Cristóvão, Sergipe, Brasil, jmontalvao@ufs.br.

¹As variáveis de estado de um sistema dinâmico determinístico correspondem ao menor conjunto de variáveis, tais que seu conhecimento no instante inicial em conjunto com a entrada do sistema, determina completamente o comportamento do sistema no futuro, ver [8] para mais detalhes.

simulação das situações futuros, possibilitadas pelos resultados subsequentes.

II. O MODELO SIR-SI

Os modelos dinâmicos para a propagação da dengue do tipo SIR-SI são encontrados em diferentes formas na literatura, com pequenas mudanças de um trabalho para outro, em vista do que se pretende estudar. Por conseguinte, é possível encontrar sistemas SIR-SI autônomos, como em [2] tanto quanto sistemas SIR-SI variantes no tempo, como em [3]. Na mesma linha, o modelo passou por algumas adaptações para ser usado aqui, sendo este apresentado a seguir como uma extensão do SIR.

O modelo SIR clássico foi analisado em [6] e sua expressão é dada por:

$$\begin{cases} \dot{S}(t) = \mu(N - S) - \beta IS/N \\ \dot{I}(t) = \beta IS/N - (\gamma - u)I \\ R(t) = N - S(t) - I(t) \\ S(0) = S_0 > 0, \quad I(0) = I_0 > 0, \quad R(0) \geq 0, \end{cases} \quad (1)$$

Em que (S, I, R) são as partes da população de tamanho N suscetíveis, infectadas e recuperadas, respectivamente. μ é expectativa de vida, β é a taxa de contatos médios com transmissão para cada indivíduo humano e γ a taxa de recuperação. Algumas considerações foram tomadas na modelagem desse sistema, delimitadas para simplificação da análise, são elas: população constante, fazendo com que a taxa de mortalidade seja igual à de natalidade; a população está uniformemente distribuída; a transmissão é feita pelo contato entre indivíduos infectados com os sucessíveis; indivíduos recuperados adquirem imunidade.

Deve-se atentar ao mecanismo de transmissão da doença, expresso em (1) como $\beta IS/N$, sendo β a taxa de contatos de um indivíduo passível de propagação da infecção num período curto (instantâneo, no modelo) de tempo, então βI é o número de pessoas com as quais os infectados interagiram de modo capaz de transmitir a doença. Do modelo, foi assumido população uniformemente distribuída, implicando que $\beta IS/N$, portanto, é a taxa de novos infectados.

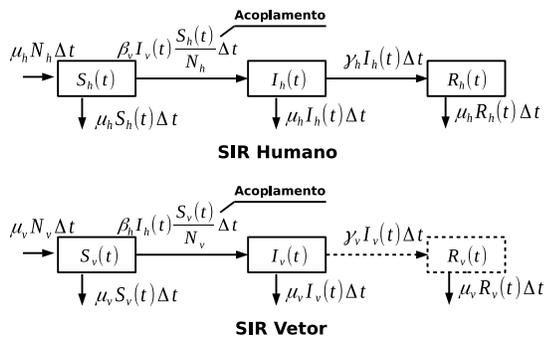


Fig. 1. Diagrama de fluxo do modelo SIR-SI, formado pelo acoplamento de dois modelos SIR. Detalhe na classe recuperada do vetor, pois pode ser desprezada para o caso da dengue [2].

O modelo SIR-SI busca modelar a propagação de uma doença com a interação de duas espécies diferentes, homem-mosquito. Considerando que as interações entre estas não

alterem suas relações vitais individuais, μ e γ , o modelo SIR-SI é formado por dois modelos SIR acoplados pelos termos de contato homem-vetor e vetor-homem, ver Figura 1. Em equações, usando os sub índices h e v para indicar a característica de humanos e vetores, respectivamente, obtêm-se:

$$\begin{cases} \dot{S}_h(t) = \mu_h(1 - S_h) - \beta_v I_v S_h / N_h \\ \dot{I}_h(t) = \beta_v I_v S_h / N_h - (\gamma_h + u_h) I_h \\ \dot{S}_v(t) = \mu_v(1 - S_v) - \beta_h I_h S_v \\ \dot{I}_v(t) = \beta_h I_h S_v / N_v - (\gamma_v + u_v) I_v \\ R_{h,v}(t) = N_{h,v} - S_{h,v}(t) - I_{h,v}(t) \end{cases} \quad (2)$$

em $\Delta = \{(S_{h,v}, I_{h,v}) | S_{h,v} \geq 0, I_{h,v} \geq 0 \text{ e } S_{h,v} + I_{h,v} \leq N_{h,v}\}$.

É pertinente observar que β_v e β_h são as taxas de contato vetor-humano e humano-vetor, respectivamente, e seus valores geralmente não são iguais, pois o número de encontros homem-mosquito é comparado a duas populações diferentes, com N_h e N_v indivíduos cada. Entretanto, elas estão relacionadas – sendo essa relação a primeira adaptação do modelo proposta por este trabalho – pelo número total de contatos. Supondo que as duas populações, de humanos e mosquitos, interagem somente uma com a outra, o número médio de contatos humano-vetor ($\beta_h N_h$) e vetor-humano ($\beta_v N_v$) são iguais, ou seja

$$\beta_v N_v = \beta_h N_h. \quad (3)$$

Da equação em (3), cabe a seguinte observação: assumindo que β_v , valor relacionado com os hábitos alimentares dos mosquitos, e N_h variem muito pouco, estas podem ser consideradas constantes, havendo, portanto, uma relação linear entre N_v e β_h .

Torna-se interessante para usos subsequentes, simplificar o sistema em (2). Para tal, considera-se o tempo de vida do mosquito da mesma ordem e/ou inferior ao tempo de infeccioso dele, não havendo, portanto, tempo hábil para recuperação destes ($\gamma = 0$). Assim, a classe de recuperados é desprezada para a população de vetores [2]. Juntando isto com o que foi visto anteriormente, de posse da equação (3), escrevendo $(i, s)_{h,v} = (I, S)_{h,v} / N_{h,v}$ e após algumas manipulações algébricas, chega-se ao modelo SIR-SI:

$$\begin{cases} \dot{s}_h(t) = \mu_h(1 - s_h) - \beta_h i_v s_h \\ \dot{i}_h(t) = \beta_h i_v s_h - (\gamma_h + u_h) i_h \\ \dot{i}_v(t) = \beta_v(1 - i_v) i_h - \mu_v i_v, \end{cases} \quad (4)$$

definido em $\Delta = \{(s_h, i_h, i_v) | s_h \geq 0, i_h \geq 0, i_v \geq 0, s_h + i_h \leq 1 \text{ e } i_v \leq 1\}$.

Algumas das propriedades deste modelo podem ser encontradas em [2]. Dentre estas, destacam-se a de estabilidade que está ligada a taxa básica de reprodução, definida por $\bar{R}_0 = \sqrt{\frac{\beta_v \beta_h}{\mu_v (\mu_h + \gamma_h)}}$. O sistema em (4) possui o ponto de equilíbrio não epidêmico localizado em $(1, 0, 0)$, sendo este globalmente assintoticamente estável para $\bar{R}_0 \leq 1$ e instável para $\bar{R}_0 > 1$, fazendo com que o sistema tenda ao ponto epidêmico localizado dentro de Δ . Essa informação é importante, uma vez que \bar{R}_0 depende de β_h , este da população de vetores. Ou

seja, manter a quantidade de mosquitos suficientemente baixa poderia levar o sistema ao ponto não epidêmico.

Em vista do que se pretende aqui, a simulação de situações de novos casos da doença, propõe-se a adição de uma entrada variante no tempo e uma saída do sistema. Pelo problema, é natural que a saída do sistema seja a taxa de novos casos de dengue $\beta_v I_v S_h / N_h$, no entanto por comodidade será tomada como saída a fração desses casos $\beta_h i_v s_h$. Quanto à entrada, escolhe-se a taxa de contatos humano-vetor (β_h). A razão para isto é sua relação direta com a população de vetores, o qual é sabido apresentar comportamento sazonal, além disso, os órgãos responsáveis utilizarem medidas de infestações de mosquitos para avaliar o risco de epidemias. Feitas as devidas considerações – segunda adaptação proposta para o modelo aqui –, o sistema do tipo SIR-SI não autônomo pode ser escrito como

$$\begin{cases} \dot{s}_h(t) = \mu_h(1 - s_h) - \beta_h(t)i_v s_h \\ \dot{i}_h(t) = \beta_h(t)i_v s_h - (\gamma_h + u_h)i_h \\ \dot{i}_v(t) = \beta_v(1 - i_v)i_h - \mu_v i_v, \\ y(t) = \beta_h(t)i_v s_h. \end{cases} \quad (5)$$

Consideram-se constantes os valores de μ_h , μ_v e para todo o tempo. Quanto ao valor de N_v , no uso do restante do trabalho, é tido como constante por partes com a finalidade de manter válidas as suposições feitas na modelagem.

III. O PROBLEMA

Supondo que o modelo SIR-SI, dado em (5), represente bem a realidade, e conhecidas as constantes do modelo, seu estado atual e o valor da entrada $\beta_h(t)$, é possível simular as saídas futuras. Isto pode ser útil para os órgãos responsáveis pela prevenção e controle da dengue, uma vez que, idealmente, se pode calcular o número de infectados em determinado período em posse da taxa da fração de novos infectados ($y(t)$). Entretanto, no que diz respeito ao modelo, apenas as constantes estão disponíveis, desconhecendo-se o estado atual. Além disto, a entrada $\beta_h(t)$, taxa de contatos humano-vetor, é de difícil medição, já que se sabe apenas que ela tem uma relação linear com N_v , população de vetores, e essa, por sua vez, apresenta comportamento sazonal. Posto isto, para que as simulações de situações futuras sejam possíveis, deseja-se, portanto, encontrar o estado atual e, suficientemente, devido a sazonalidade, o histórico de β_h . Não por menos, para que isto seja minimamente praticável, é necessário dados que contenham a ‘história’ do sistema, tais quais os usados aqui: o histórico do número de notificações de casos de dengue por mês (C_k), que representa o mesmo significado físico da média de $N_h y(t)$, saída do sistema multiplicado pela população humana, no mês ou $N_h \bar{y}(k) = N_h \int_{k-1}^k y(t) dt$.

Sabe-se que o modelo SIR-SI responde a uma entrada de forma única. Em outras palavras, conhecido o estado inicial e a entrada, determina-se a saída. Existe, por isso, uma relação entre esses três elementos pertencentes ao sistema. Destes, dispõe-se do histórico da saída ($N_h \bar{y}(k) \equiv C_k$), desejando-se recuperar os demais. Este problema não possui solução única, sendo que para cada estado inicial existe um $\beta_h(t)$

correspondente à saída desejada. Dito isto, busca-se aqui resolver os seguintes problemas (ver Fig. 1): Como encontrar β_h conhecidos a saída e o estado inicial (inversão do sistema)? e apresentar um critério de escolha do estado inicial. Destaca-se, ainda, que com a resolução destes, encontra-se, conjuntamente, o último estado do sistema, referente a saída mais recente.

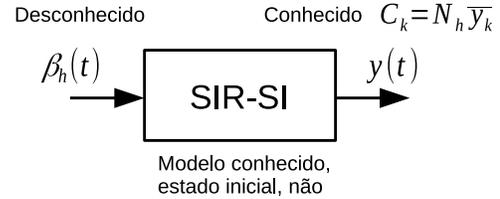


Fig. 2. Esquema do problema, resumindo o que se têm disponível e o que não.

Os problemas apresentados são resolvidos separadamente nas próximas seções, cabendo, antes, algumas informações sobre a coleta dos dados. Seus resultados são posteriormente usados nas simulações de situações para alguns estados brasileiros, na seção de resultados e discussões.

IV. OS DADOS DOS CASOS DE DENGUE

A coleta de dados foi realizada, principalmente, no site do Sistema Único de Saúde (SUS) brasileiro [1]. O SUS possui uma base com dados de diferentes doenças e o da Dengue é constituído de agravos e notificações no período de 2001 a 2017 (acessado em: 24 out. 2018), podendo ser coletados de diferentes maneiras. Para o que se pretende aqui, foram coletados dados de notificações por mês de alguns estados brasileiros, sendo mencionados ao longo do texto os estados da Bahia, Goiás, Mato Grosso, Mato Grosso do Sul e São Paulo. Vale ressaltar que as constantes do modelo utilizado são obtidas de diferentes fontes, podendo ser vistas na Tabela I. A utilização do número de notificações por mês da Dengue está alinhado ao número de habitantes na região no período, uma vez que o modelo se usa da fração da taxa de novos infectados ($y(t)$), por isso aqui os termos fração e número são usados indistintamente, sendo especificado quando necessário.

A base utilizada com os estados mencionados está disponível no site do Grupo Biochaves da UFS em formato txt, constituído de duas colunas numéricas: a primeira contendo o número de notificação e a segunda a população residente naquela data. Além dos dados, está disponível também o método implementado (descrito neste trabalho) em uma linguagem de programação para uso livre².

V. O MÉTODO

Para resolução do primeiro problema, encontrar o passado de β_h , se considera o estado atual conhecido – a escolha desse é tratado em seguida. Além disto, os dados de saída são amostrados mensalmente. Sendo assim, por simplicidade, β_h é tido como constante em períodos de um mês. O intuito disso é utilizar algumas informações do comportamento do

²Os links são: <http://www.biochaves.com>, site do biochaves; <http://www.biochaves.com/download/1557/>, base; <http://www.biochaves.com/download/1554/>, código.

sistema atuando de forma autônoma, são elas: a evolução dos estados do sistema possui um comportamento oscilatório tendendo a um ponto de equilíbrio; a região do espaço de estados – conjunto de todos os estados possíveis de sistema dinâmico – em que o sistema oscila é limitada, ficando em Δ , o que implica em variações dos estados limitadas – para mais informações ver [2]. Estas características mostram uma evolução no tempo previsível, o que abre espaço para um primeiro algoritmo de busca por tentativa e erro. Basicamente, para um conjunto de valores para β_h em um intervalo de tempo, simula-se o sistema do estado atual até o próximo para todos os valores do conjunto, escolhendo, como melhor representante para β_h do intervalo, aquele em que a média da saída possuir menor erro quadrático em relação ao dado real, $(N_h \bar{y}_k - C_k)^2$.

O método relatado é lento, uma vez que se trata de uma busca ‘cega’. Ainda, o valor escolhido para β_h no intervalo possui o menor erro quadrático dentre os testados, mas não há garantias de que esse erro seja pequeno. Posto isto, o primeiro método rudimentar foi melhorado com a inclusão do método iterativo da bisseção, havendo um aumento considerável da velocidade de execução (menor custo computacional), além de erros menores. Em passos, o método de inversão utilizado neste trabalho pode ser descrito do seguinte modo (ver Fig. 2):

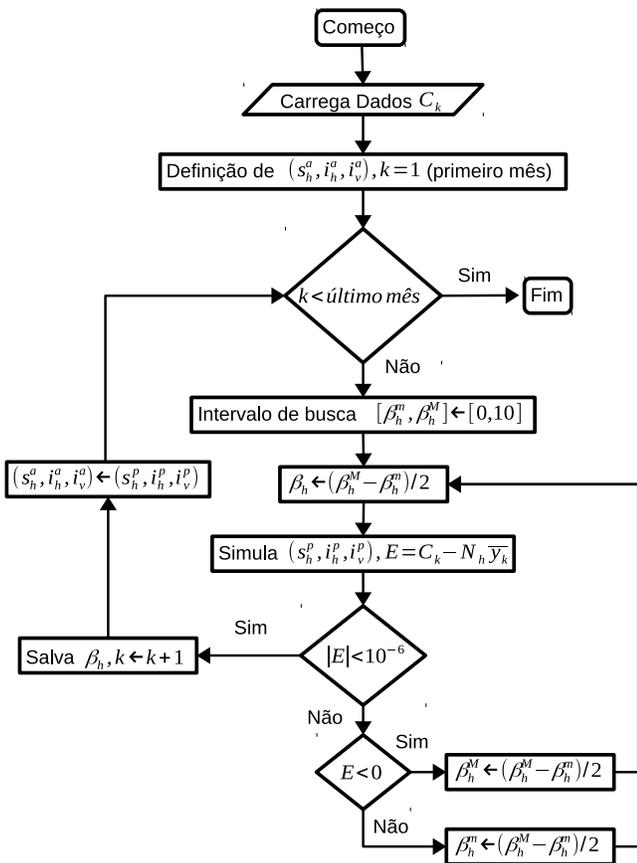


Fig. 3. Fluxograma do método de inversão.

- 1) Carregar dados das frações de novos infectados $y(t)$ [1]. Além das constantes do modelo μ_h, μ_v, γ e β_v . Os valores aqui usados são mostrados na Tabela 1.

TABELA I
VALORES DOS PARÂMETROS ASSUMIDOS CONSTANTES

Parâmetro	Valor	Fonte
μ_h	$1,2 \times 10^{-3}$ mês ⁻¹	[9]
μ_v	$6,7 \times 10^{-1}$ mês ⁻¹	[10]
γ_h	5 mês ⁻¹	[10]
β_v	100 (mês.mosquito) ⁻¹	Nota ³

- 2) Define-se o estado inicial/atual $(s_h^a, i_h^a, i_v^a) = (s_h(0), i_h(0), i_v(0))$, sua escolha é discutida na próxima seção.
- 3) Aqui começa a busca pelo melhor β_h do intervalo k , iniciado em $k = 1$ sendo incrementado à medida que os β_h forem encontrados. Primeiro define-se o intervalo de procura do β_h , normalmente $[\beta_h^m, \beta_h^M] = [0, 10]$.
- 4) Simula-se a entrada do sistema que está no estado atual (s_h^a, i_h^a, i_v^a) , com o β_h igual ao meio do intervalo $(\beta_h^M - \beta_h^m)/2$, determinando o próximo estado do sistema (s_h^p, i_h^p, i_v^p) pelo método de Runge Kutta de ordem quatro com passo de 0.01 (um centésimo do intervalo de mês).
- 5) Compara-se a saída com os valores reais, encontrando-se o erro $E = C_k - N_h \bar{y}(k)$.
- 6) Caso o módulo do erro seja menor que 10^{-6} , salva-se o valor atual de β_h para este intervalo, e faz-se a atribuição $(s_h^a, i_h^a, i_v^a) \leftarrow (s_h^p, i_h^p, i_v^p)$. Após isto, volte ao passo 3, incrementando $k \leftarrow k + 1$ para o próximo mês.
- 7) Caso o módulo do erro seja maior que 10^{-6} , atualiza-se o intervalo de busca de β_h de acordo com o sinal do erro

- Para $E < 0$,

$$[\beta_h^m, \beta_h^M] \leftarrow [\beta_h^m, (\beta_h^M - \beta_h^m)/2],$$

- Para $E > 0$

$$[\beta_h^m, \beta_h^M] \leftarrow [(\beta_h^M - \beta_h^m)/2, \beta_h^M],$$

- 8) O algoritmo fica preso nos passos 4 e 5 até encontrar β_h com erro inferior a 10^{-6} . Para que ele não entre em um *loop* infinito, é aconselhável encerrá-lo com o valor atual de β_h após 12 interações do mesmo. Após todos os dados serem avaliados o algoritmo encerra.

O processo tem o custo máximo de 1998 multiplicações por mês simulado, o que permite a sua execução em computadores pessoais com simulação de meses em apenas poucos segundos (ou mesmo frações de segundo). Quanto a seu resultado, a estimação de $\beta_h(t)$ abre espaço para algumas discussões, mas antes é interessante discutir a escolha do estado inicial.

³Um único mosquito fêmea pode transmitir a doença em média a 300 pessoas no seu tempo de vida [11] apud [12]. As unidades usadas são em meses, e considerando um tempo de vida adulta do mosquito de 45 dias, logo 200 pessoas por mês. Além disso, o modelo leva em conta a população de vetores e, como os machos não transmitem a doença, chega-se a $\beta_v = 100$ mês⁻¹.

VI. ESCOLHA DO ESTADO INICIAL

A escolha do estado inicial é arbitrária. No entanto, uma vez que todas as soluções encontradas explicam os dados, isto é, são todas excitações/perturbações que poderiam explicar satisfatoriamente as observações – o método foi construído com esse intuito –, então uma boa prática para simulações futuras de novos casos é utilizar as soluções mais simples. Deste modo, o próprio modelo com os dados da saída podem sugerir estados iniciais que proporcionem soluções mais simples para $\beta_h(t)$ em relação às outras.

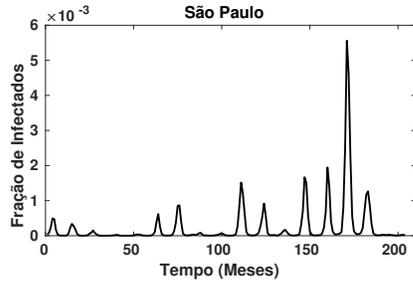


Fig. 4. Fração de novos infectados do estado de São Paulo - Brasil entre os anos de 2001 a 2017, saída real do sistema.

Na Fig. 4 se pode ver os dados da fração de novos infectados para cada mês do estado de São Paulo, do qual é possível ver uma periodicidade aproximada nos picos. Estes dados correspondem às somas mensais do único termo não linear $\beta_h(t)s_h i_v$ no sistema em (5), o que possibilita uma simplificação interessante. Assumindo $y(t) = C_k$ constante em cada período $(k-1, k)$, então se pode entendê-lo como uma ‘entrada’ no tempo para os sistemas de ordem 1 desacoplados, como segue:

$$\dot{s}_h = \mu_h - y(t) - \mu_h s_h \quad (6)$$

$$\dot{i}_h = y(t) - (\gamma + \mu_h) i_h. \quad (7)$$

É proveitoso mencionar que a estimação de $\beta_h(t)$ está relacionada à evolução dos estados do sistema, assim, quanto mais variabilidade estes tiverem, mais detalhes $\beta_h(t)$ pode ter. Nesse contexto, respostas de sistemas lineares de ordem 1 podem ser separadas em resposta transitória e estacionária, na qual a transitória possui uma exponencial decrescente, que se extingue depois de um tempo, somada a resposta forçada, que se mantém todo o tempo. Deste modo, entende-se aqui como soluções mais simples àquelas que possuem menos coisas a serem representadas. Logo, entre uma solução com resposta transitória somada a forçada e somente a forçada, esta última é tida como a mais simples. Isto posto, de (6) e (7), considerando a componente constante da ‘entrada’, $y(t)$, a resposta transitória estará eliminada/atenuada quando não houver variação dos estados, isto é, $\dot{s}_h = 0$ e $\dot{i}_h = 0$, portanto

$$s_h = \frac{\mu_h - \bar{y}(t)}{\mu_h}, \quad (8)$$

$$i_h = \frac{\bar{y}(t)}{\gamma_h + \mu_h}. \quad (9)$$

Estas são as escolhas dos valores iniciais da fração de suscetíveis e infectados humanos para se ter menor influência

do transitório na estimação de $\beta_h(t)$, uma vez que o sistema é iniciado no estacionário.

Quanto a fração de vetores infectados, o problema não pode ser simplificado ao linear de imediato como os já feitos. Entretanto, a ideia é semelhante, de modo que, primeiro, $i_h(t)$ pode ser encontrado da equação em (7). Agora, $i_h(t)$ está ‘disponível’, o que permite a ele ser tratado como uma entrada na equação

$$\dot{i}_v(t) = \beta_v(1 - i_v)i_h - \mu_v i_v,$$

a qual ainda não é linear. Vale lembrar que a fração de i_v é limitada ao intervalo de 0 a 1, o que torna verdadeira a inequação

$$\dot{i}_v(t) \leq \beta_v i_h - \mu_v i_v,$$

e dela, pode-se usar o lema da comparação (ver [13])⁴ para concluir que $i_v(t)$ estará limitada pela resposta do sistema linear dado na igualdade superior da inequação. Por fim, usando os mesmos argumentos anteriores para s_h e i_h em i_v , o estado inicial que foi utilizado no método para o restante do trabalho é

$$\left(\frac{\mu_h - \bar{y}(t)}{\mu_h}, \frac{\bar{y}(t)}{\gamma_h + \mu_h}, \frac{\beta_v \bar{y}(t)}{(\gamma_h + \mu_h)\mu_v} \right). \quad (10)$$

Neste ponto, torna-se proveitoso mencionar que os sistemas lineares apresentados para as frações de infectados humanos e vetores possuem transitório da ordem de alguns meses. Portanto, o estado inicial só afeta a estimação de $\beta_h(t)$, significativamente, nos meses iniciais. Por esta razão, as escolhas de i_h e i_v não são tão cruciais, no sentido de não alterar os valores obtidos de β_h no método de inversão (exceto os dos primeiros meses). Quanto a fração de humanos suscetíveis, seu transitório é da ordem de séculos, o que torna sua escolha inicial relevante. Para elucidar isto, na Fig. 4 foi utilizado o método para estimação de $\beta_h(t)$ com três estados iniciais diferentes, sendo alterado apenas o valor da fração de suscetíveis. No estado inicial sugerido em (10), $s_h(t)$ apresenta um menor intervalo de abrangência em relação aos outros, em vista da menor influência do transitório. Já para o valor mais acima, pode ser visto uma tendência decrescente em na fração de humanos suscetíveis, acontecendo o oposto para o valor mais abaixo. No que diz respeito a $\beta_h(t)$, ela apresenta, além da sazonalidade esperada, módulos maiores para valores iniciais de s_h menores, sendo que para o menor destes, a taxa de contatos humano-vetor possui, ainda, um decréscimo causado pela, anteriormente mencionada, resposta livre de $s_h(t)$. Isto exemplifica, de maneira prática, os critérios adotados nesta seção, sendo proveitoso citar o trabalho de [14] por apresentar argumentos semelhantes com aspectos estocásticos em outro sistema epidemiológico.

⁴Sob determinadas condições, que o sistema aqui atende, o lema garante que para uma inequação diferencial escalar

$$v(t) \leq g(t, v), \quad v(t_0) \leq u_0,$$

$v(t)$ é limitada pela resposta da equação diferencial

$$u(t) = g(t, u) \quad u(t_0) = u_0$$

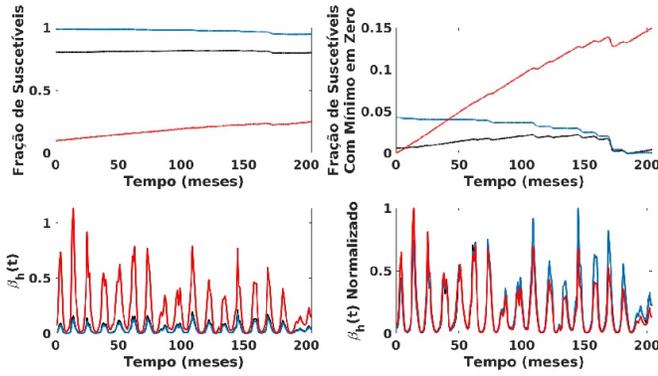


Fig. 5. Resultados do método de inversão obtidos utilizando os estados iniciais ($s_h(0), i_h(0), i_v(0)$) iguais a $(1 \times 10^{-1}, 4 \times 10^{-5}, 6 \times 10^{-3})$ (vermelho), $(99 \times 10^{-2}, 4 \times 10^{-5}, 6 \times 10^{-3})$ (azul), $(8 \times 10^{-1}, 4 \times 10^{-5}, 6 \times 10^{-3})$ (preto). Foram utilizados os dados de novos infectados do estado de são paulo.

VII. RESULTADOS E DISCUSSÕES

Exposto o método de estimação da entrada e o critério de escolha do estado inicial, nesta seção cabem algumas discussões acerca dos valores encontrados para β_h , além de seu uso para simulação de possíveis cenários de surto da dengue. Nessa ordem, é proveitoso lembrar a dificuldade prática de medir a taxa média de contatos humano-vetor, o que torna comparações com dados reais irrealizáveis, devido à inexistência dos dados. Entretanto, existe algumas correspondências intrigantes, como: os vales de $\beta_h(t)$ ocorrerem em períodos de campanhas contra o mosquito transmissor.

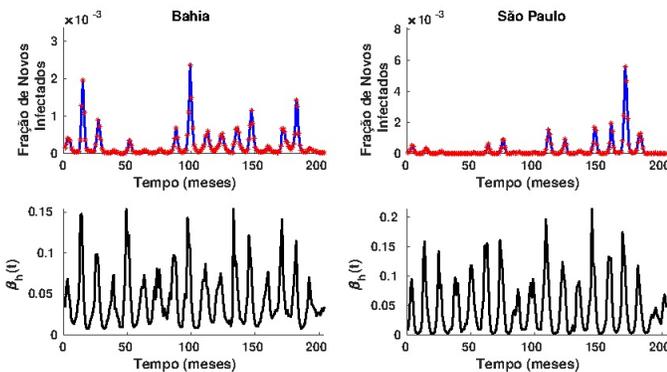


Fig. 6. Resultados do método de inversão com o melhor estado inicial quando aplicados à Bahia e a São Paulo: saídas simuladas em azul, dados reais (saída real) em vermelho, histórico recuperado da entrada em preto.

Na Fig. 6 é exemplificado a solução dos problemas de inversão e escolha do estado inicial para dois estados brasileiros, obtendo como resultado a recuperação da entrada, a evolução da taxa de contatos humano-vetor. Dito isto, é interessante observar que alguns vales de $\beta_h(t)$ ocorrem depois de um decréscimo acentuado, o que pode indicar alguma campanha de combate ao *Aedes Aegypti* nas condições alarmantes do surgimento de novos casos. De fato, se for usado um agregador de notícias⁵, como o **Google** na aba notícias, para realizar uma

⁵Agregadores são programas que juntam/organizam informações que possam ser vistas por usuários finais.

busca por ‘Combate a dengue em São Paulo’ em períodos de epidemias acentuadas, têm-se com resposta mais notícias sobre o assunto que comparado a anos sem epidemias. No ano 2015 (de 169 a 180, meses), por exemplo, destacam-se as seguintes notícias: [15] informa sobre o número de caso recorde em São Paulo; Já em [16], [17], [18] são mostrados medidas de combate ao mosquito, sendo elas: a tomada providências sanitárias, combate genético do vetor e sancionamento de leis que facilitam o trabalho dos agentes de endemias, respectivamente. O mesmo pôde ser feito para Bahia, com resultados semelhantes. Correspondências como estas podem ser vistas também nos picos, como exemplo no ano de 2012 (133 a 144, meses) $\beta_h(t)$ apresenta um dos maiores valores para a Bahia, fato esse que coincide com a notícia, alarmante, do aumento do índice de infestação predial (IIP) em salvador mostrada em [19].

De fato, a taxa de contato humano-vetor e o IIP são diretamente proporcionais à população de mosquitos, portanto as mesmas estão relacionadas por uma constantes de proporcionalidade, $\beta_h(t) = \sigma \cdot IIP$. Isto é relevante, uma vez que conhecida tal constante se pode utilizar as estimações de IIP, medição comumente realizada nas campanhas contra a dengue, no sistema SIR-SI para simular novos casos da doença. Além disso, essa relação pôde ser empregada para comprovar a validade do método proposto aqui. Para tanto, foram utilizados os dados de infestação predial encontrados em [20] referentes aos anos de 2001 a 2005 do estado de Goiás. Infelizmente as estimativas de IIP não são divulgadas em nenhuma base de dados pública pelos órgãos que as coletam, o que dificulta o acesso a essas informações. Por isso, os valores foram coletados visualmente da Figura 1-A em [20]. Dito isto, na Fig. 7 pode ser comparado o histórico da entrada encontrado pelo método proposto aqui com o IIP, onde se pode observar como eles possuem as mesmas tendências, sendo $\sigma = 0,11$ determinado por ajuste do mínimo erro quadrático. Isto é de suma importância, pois valida toda a metodologia usada na recuperação do histórico do sistema. Vale ressaltar que o primeiro ano do histórico recuperado de β_h é desprezado na determinação de σ por conter influência da escolha do estado inicial, como visto na seção VI.

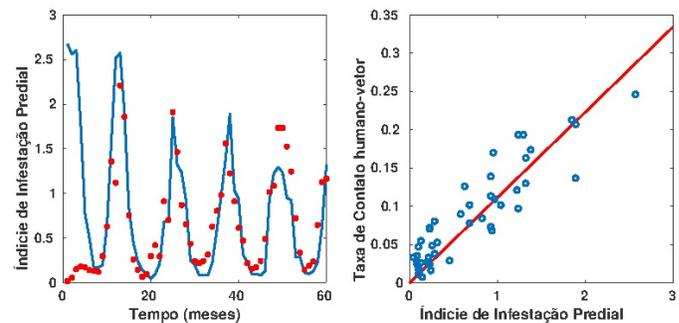


Fig. 7. À esquerda, o índice de infestação predial mostrado em [20] (azul) e o estimado pelo método proposto aqui ($\beta_h(t)/\sigma$) (vermelho) de 2001 a 2005, com $\sigma = 0,11$. À direita, relação $IPP \times \beta_h$ para os anos de 2002 a 2005 (azul) e a reta com o σ que melhor ajusta os pontos na forma $\beta_h(t) = \sigma \cdot IIP$ (vermelho).

Mencionada a dificuldade em ter acesso ao índice de

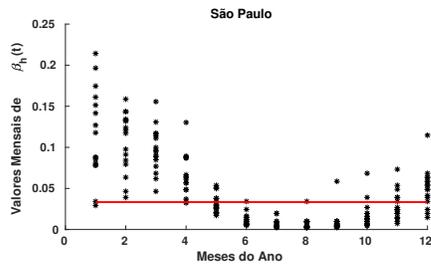


Fig. 8. Valores da taxa de contato humano-vetor distribuídos para cada mês do estado de São Paulo. Dados usados de 2001 a 2017. Linha vermelha indica estabilidade não endêmica para valores inferiores a ela $\bar{R}_0 \leq 1$.

infestação predial, uma alternativa, de interesse principal aqui, é usar o histórico de β_h recuperado como base para simular novas situações dos casos de dengue. Nesse sentido, os dados recuperados estão disponíveis e podem ser usados para simulações de futuros casos da doença de diferentes formas. Assim, uma maneira simples de se fazer isso é utilizar cada ano como entrada do sistema para o estado atual. Ao fazer isso, levando em conta a quantidade de dado disponível (17 anos), têm-se um total de, no máximo, 17 situações diferentes. O que pode ser insuficiente em razão da liberdade do sistema. Em vista disso, uma maneira de extrapolar a quantidade de situações simuladas, devido a sazonalidade pressuposta, é tratar cada mês do ano como sendo uma distribuição de probabilidade. Deste modo, pode-se simular situações diversas compreendendo as já ocorridas.

De fato, utilizar uma distribuição de probabilidade para extrapolar as informações disponíveis faz sentido, ao olhar, por exemplo, os valores mensais de $\beta_h(t)$, eles estão dispostos em uma determinada região (ver Fig. 8). Nesse sentido, para determinar tal distribuição, com tão poucas amostras, atenta-se ao fenômeno que está por trás da taxa de contato humano-vetor. Primeiramente, o contato se dá através das picadas do mosquito fêmea em uma pessoa, que por vezes pode ocorrer sem que haja infecção, mesmo entre um infectado e um suscetível, sendo, devido a isso, associado uma probabilidade de transmissão p por picada – vale ressaltar que no modelo dinâmico determinístico só foram contabilizados os contatos passíveis de transmissão. Se forem realizadas N_p picadas por mês, sendo diferente para cada mês do ano, com probabilidade de contágio igual p , então a variável aleatória X , que conta o número de sucessos em um mês, tem distribuição do tipo binomial. E para obter a taxa de contatos humano-vetor basta dividir X pela população de humanos, então $\beta_h \sim \text{Binomial}(N_p, p)/N_h$. Neste ponto, torna-se interessante mencionar que N_p é da ordem de $\beta_h N_h$, o que tornaria bastante lenta a simulação de uma instância da distribuição binomial, por essa razão é aconselhável a aproximação da binomial pela normal. Por fim, utilizou-se nas simulações de situações dos surtos de dengue β_h como sendo do tipo normal, ignorando as eventuais instâncias negativas por não fazerem sentido físico.

Para testar a usabilidade da abordagem foram supostos desconhecidos os dados de 2016, ou seja, deseja-se simular diferentes situações para a quantidade de novos infectados por

mês em 2016, conhecendo o histórico destes nos anos de 2001 a 2015. O que, em resumo, é feito do seguinte modo:

- 1) Recupera-se $\beta_h(t)$ pela aplicação do método mostrado na seção V, utilizando o estado inicial dado em (10), por conseguinte, o estado atual é também encontrado.
- 2) Supõe-se que $\beta_h(t)$ possui uma distribuição de probabilidade normal específica para cada mês, em que seus parâmetros são estimados pelas médias e desvios a partir dos dados mensais recuperados de cada ano. Vale salientar que o primeiro ano de $\beta_h(t)$ é ignorado, por conter influência da escolha do estado inicial, como visto na seção VI.
- 3) Simula-se, via Monte Carlo, 1000 vezes a evolução do sistema, novas situações, utilizando as instâncias da distribuição de $\beta_h(t)$ encontrada anteriormente.

Ao aplicar essa abordagem, torna-se disponível 1000 situações da propagação da dengue obtidas pela extrapolação dos dados históricos de $\beta_h(t)$. Os resultados podem ser visualizados de diferentes maneiras, aqui optou-se por duas: uma para dar ideia da evolução dos casos no ano e a outra para as possibilidades de cenários por mês. Nessa ordem, na Fig. 9 é mostrado as médias e desvios laterais dos novos casos da doença por mês simulados feitas para quatro estados brasileiros conjuntamente com os dados reais. É interessante observar que as médias estão próximas aos dados reais para os primeiros meses, afastando-se nos meses subsequentes. Isso se deve, em partes, pela propagação do erro. Os meses acima de um não utilizam o estado atual obtido pelo método com o uso do mês anterior para o ano em questão – pode-se melhorar esta aproximação utilizando o método até o mês anterior ao que se quer realizar as simulações. Além disto, em vista do surto, os órgãos responsáveis podem ter exercido um controle dos vetores, fazendo os casos reais diminuírem abaixo da média simulada, situação presente/possível nas simulações, mas não como ente mais provável.

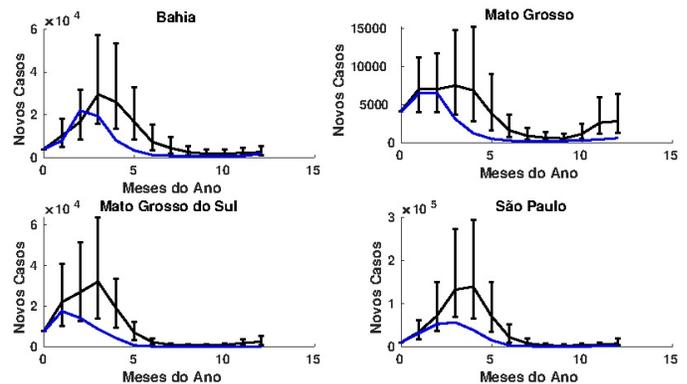


Fig. 9. Em azul os dados reais obtidos em cada estado brasileiro, em preto os dados as médias simuladas e seus desvios superior (calculado com os valores acima da média) e inferior (calculado com os valores abaixo da média) para o ano de 2016.

Outro modo de visualizar os dados é agrupar os valores simulados em caixas/bins, de acordo com o intervalo a que eles pertencem, contando quantas aparições ocorreram em cada uma, em resumo, cria-se um histograma. Isto dá ideia

de quão provável é determinada situação ao custo de ter a visualização reduzida a um mês por vez. Dessa forma, na Fig. 10 é possível ver o histograma do mês de janeiro para os mesmos quatro estados brasileiros no ano em questão. Neste, observa-se que as situações são mais dispersas para valores acima da média, ao passo que os valores extremos são pouco prováveis. Além disso, pode ser observado que cada valor real medido recai próximo ao valor médio correspondente, estes últimos localizados em *bins* dos histogramas associados a altas frequências de ocorrência.

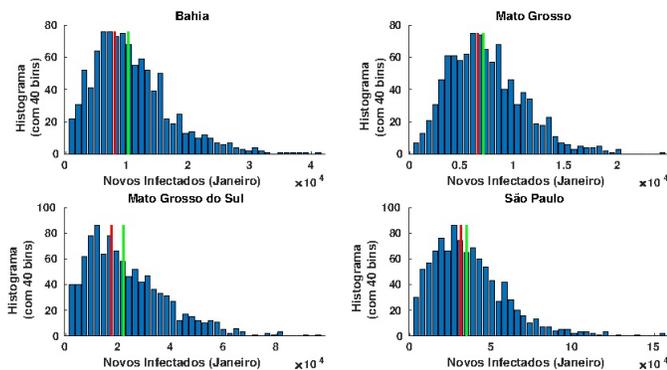


Fig. 10. Histogramas obtidos via Monte Carlo com 1000 simulações de janeiro de 2016 para os estados: Bahia, Mato Grosso, Mato Grosso do Sul e São Paulo. As barras em verde e vermelho indicam média das simulações e o valor real dos novos casos registrados, respectivamente.

Dos resultados mostrados, deve-se atentar ao objetivo do trabalho, que é simular situações de surto de dengue com base no histórico de notificações. Nesse contexto, foram mostrados um modelo SIR-SI, com uma entrada e uma saída; ferramentas para recuperar o par histórico da entrada e estado inicial mais simples sob determinado critério; além de se obter, como subproduto desse processo, o estado atual do sistema. Feito isto, as simulações futuras tornam-se possíveis, sendo extrapolados os dados históricos dos contatos humano-vetor por uma distribuição gaussiana com intuito de um maior número de situações simuladas. Entretanto, para previsões mais precisas desses novos casos, $\beta_h(t)$ deve ser melhor explorado em termos da dinâmica populacional do mosquito sem e com os efeitos de controle humano sobre os vetores, incluindo também consequências de eventos atípicos que podem elevar a população de mosquito, como o racionamento de água – relatado em [4] por explicar, devido a anormalidade causada no modelo de séries temporais usado pelos autores, o afastamento dos casos previstos e reais. Ou ainda, como alternativa, utilizar as informações e estimativas dos índices de infestação do mosquito para encontrar $\beta_h(t)$ e, com isso, estimar os novos casos.

VIII. CONCLUSÕES

O modelo SIR-SI foi adaptado neste trabalho ao fenômeno da propagação da dengue em cinco regiões do Brasil. Nesta adaptação, o problema dos dados faltantes, mais especificamente, o desconhecimento dos sinais de entrada e dos estados dinâmicos do sistema, se tornou um objetivo deste trabalho, levando assim a uma reformulação do problema como inversão

cega de um sistema não-linear. A solução desenvolvida de forma original neste trabalho para cinco regiões do Brasil pode ser usada em quaisquer outras regiões, independente das condições ambientais e intervenções humanas, mediante as restrições típicas do modelo SIR-SI. Com os dados recuperados, o objetivo inicial de simular cenários futuros baseados nos passados se tornou possível, tendo como impasse adicional a quantidade pequena de situações ocorridas, o que culminou em um problema de extrapolação. A solução proposta se mostrou proveitosa às regiões testadas ao estabelecer cenários associados a probabilidades de ocorrência em função das últimas observações, pois esses cenários podem servir como base a políticas de prevenção e/ou mitigação de dengue.

IX. CONTINUIDADE DO TRABALHO

Os problemas e a metodologia apresentados estão sujeitos às condições de modelagem do modelo SIR-SI, a exemplo: população constante e uniformemente distribuída. Nesta linha, as soluções propostas estão limitadas a estas condições. Dito isto, um estudo em andamento é o impacto das restrições geográficas – tais como rios e montanhas – sobre o método proposto, uma vez que estas podem dividir a população total em centros populacionais ligados por migrações recorrentes, de forma que a condição de uniformidade pode não ser atendida. Uma possível solução é a utilização do modelo proposto aqui em uma rede de modelos SIR-SI acoplados como um só modelo mais complexo.

AGRADECIMENTOS

Agradecemos ao programa de mestrado de engenharia elétrica da Universidade Federal de Sergipe pelo espaço dedicado a pesquisa. A FAPITEC/SE pela bolsa de estudos do primeiro autor.

REFERÊNCIAS

- [1] M. da Saúde, “Datasus,” Disponível em: <http://tabnet.datasus.gov.br>, s.d., acessado em: 24 out. 2018.
- [2] L. Esteva and C. Vargas, “Analysis of a dengue disease transmission model,” *Mathematical biosciences*, vol. 150, no. 2, pp. 131–151, 1998.
- [3] A. Iggidr and M. O. Souza, “State estimators for some epidemiological systems,” *Journal of mathematical biology*, vol. 78, no. 1-2, pp. 225–256, 2019.
- [4] D. M. CORDEIRO and G. M. CORDEIRO, “Modelo de previsão combinada: Uma aplicac aoa série mensal das notificac oes do “dengue” no estado de pernambuco,” *Rev. Mat. Estat.*, vol. 22, no. 3, pp. 57–80, 2004.
- [5] P.-A. Bliman and B. D. Barros, “Interval observers for sir epidemic models subject to uncertain seasonality,” in *International Symposium on Positive Systems*. Springer, 2016, pp. 31–39.
- [6] H. W. Hethcote, “Qualitative analyses of communicable disease models,” *Mathematical Biosciences*, vol. 28, no. 3-4, pp. 335–356, 1976.
- [7] J. S. V. Martinez, P. G. Lopez, J. A. Gonzalez, and G. D. Reyes, “Numerical approaching of sir epidemic model for propagation of computer worms,” *IEEE Latin America Transactions*, vol. 13, no. 10, pp. 3452–3460, 2015.
- [8] H. K. Khalil, “Nonlinear systems,” *Prentice-Hall, New Jersey*, vol. 2, no. 5, pp. 5–1, 1996.
- [9] M. da Saúde, “A.11 esperança de vida ao nascer,” Disponível em: <http://tabnet.datasus.gov.br/cgi/tdb2006/a11t.htm>, s.d., acessado em: 24 out. 2018.

- [10] S. da Saúde do Paraná, “Dengue - características clínicas e epidemiológicas - combate à dengue,” Disponível em: <http://www.dengue.pr.gov.br/modules/conteudo/conteudo.php?conteudo=11>, s.d., acessado em: 24 out. 2018.
- [11] W. J. McBride and H. Bielefeldt-Ohmann, “Dengue viral infections; pathogenesis and epidemiology,” *Microbes and infection*, vol. 2, no. 9, pp. 1041–1050, 2000.
- [12] I. M. d. A. Silva, “Mortalidade e comportamento de aedes aegypti (diptera: Culicidae) resistente a piretróides exposto ao óleo essencial de aristolochia trilobata l.(piperales: Aristolochiaceae),” 2017.
- [13] H. K. Khalil, “Nonlinear systems. 2002,” *ISBN*, vol. 130673897, no. 9780130673, p. 893, 2002.
- [14] S. Cauchemez and N. M. Ferguson, “Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in london,” *Journal of the Royal Society Interface*, vol. 5, no. 25, pp. 885–897, 2008.
- [15] G. S. Paulo, “G1 - com 222 mil casos, estado de sp bate recorde histórico de dengue,” Disponível em: <http://g1.globo.com/sao-paulo/noticia/2015/04/estado-de-sao-paulo-registra-222-mil-casos-de-dengue.html>, 2015, acessado em: 26 Abr. 2019.
- [16] J. de Brasília, “Secretaria de saúde anuncia medidas para combater a dengue em são paulo,” Disponível em <http://www.jornaldebrasil.com.br/brasil/secretaria-de-saude-anuncia-medidas-para-combater-a-dengue-em-sao-paulo/>, 2015, acessado em: 26 Abr. 2019.
- [17] G. P. e Região, “Projeto usa ‘aedes do bem’ para combater o mosquito da dengue,” Disponível em <http://g1.globo.com/sp/piracicaba-regiao/noticia/2015/03/projeto-usa-aedes-do-bem-para-combater-o-mosquito-da-dengue.html>, 2015, acessado em: 26 Abr. 2019.
- [18] G. S. Paulo, “Sp sanciona lei que autoriza entrada à força em casas para combater dengue,” Disponível em <http://g1.globo.com/sao-paulo/noticia/2015/10/sp-sanciona-lei-que-autoriza-entrada-forca-em-casas-para-combater-dengue-.html>, 2015, acessado em: 26 Abr. 2019.
- [19] G. Bahia, “Salvador tem alerta para epidemia de dengue,” Disponível em <http://g1.globo.com/bahia/noticia/2012/06/salvador-tem-alerta-para-epidemia-de-dengue.html>, 2012, acessado em: 6 Jul. 2019.
- [20] S. S. d. Souza, I. G. d. Silva, and H. H. G. d. Silva, “Associação entre incidência de dengue, pluviosidade e densidade larvária de aedes aegypti, no estado de goiás,” *Rev Soc Bras Med Trop*, pp. 152–155, 2010.



Thiago Silva Nasceu em Poço Verde, Sergipe, Brasil, em 1994. Recebeu o título de Engenheiro Eletricista (2017) pela Universidade Federal de Sergipe. Atualmente é aluno de mestrado do programa de pós graduação em engenharia elétrica da Universidade Federal de Sergipe. Seus principais interesses de pesquisa são: processamento de sinais e sistemas não lineares.



Jugurta Montalvão Nasceu em Aracaju, Brasil, em 1968. Recebeu o título de Engenheiro Eletricista (1992) pela Universidade de Campina Grande (UFPB II), Mestre em Engenharia Elétrica (1995) pela Universidade de Campinas (UNICAMP) e Doutor em “Automatique et traitement du signal”(2000) da University Paris-Sud XI. Ingressou no Departamento de Engenharia Elétrica da Universidade Federal de Sergipe (UFS) em 2005. Seus principais interesses de pesquisa são: reconhecimento de padrões e processamento de sinais.