

# Decomposição em modos dinâmicos em um episódio de propagação geográfica da dengue

Lucy T. Takahashi<sup>1</sup>

DM/UFJF, Juiz de Fora, MG

Luis A. D'Afonseca<sup>2</sup>

DM/CEFET-MG, Belo Horizonte, MG

Wilson C. Ferreira Jr.<sup>3</sup>

IMECC/UNICAMP, Campinas, SP

Daniel M. Barbosa<sup>4</sup>

IMECC/UNICAMP, Campinas, SP

**Resumo** Neste trabalho, o método de Decomposição em Modos Dinâmicos é apresentado e aplicado a um episódio epidemiológico geográfico de propagação da dengue no Brasil referente ao período de 2014 a 2019, utilizando dados empíricos extraídos do Departamento de Informática do Sistema Único de Saúde do Ministério da Saúde do Brasil. E pela análise dos modos associados aos dados da dengue as principais frequências de ocorrência da doença são determinadas, assim como suas características espaciais.

**Palavras-chave.** Big Data, Modelagem, Aproximação Linear

## 1 Introdução

A contínua e crescente capacidade de armazenamento de dados deu lugar, em tecnologia da informação, ao que chamamos de Big Data. Big Data é um grande conjunto de dados que podem ser gerados, ou capturados, por fontes como transações (ex. compras pela internet), comunicações e mensagens (ex. celulares), imagens, buscas, etc. e se caracteriza por ter um grande volume, uma grande variedade e uma grande velocidade que, em geral, não são planejados, nem estruturados e são dados orgânicos (“não pagos”).

Na biomatemática pesquisadores têm apresentado trabalhos utilizando informações do Big Data, entre outros, Proctor e Eckhoff (2015), em [2], analisaram a dinâmica da gripe nos Estados Unidos utilizando dados do Google Flu Trends; e as dinâmicas do sarampo no Reino Unido e da poliomelite na Nigéria utilizando dados da pré-vacinação, por meio do método matemático Decomposição em Modos Dinâmicos (DMD).

O DMD foi recentemente desenvolvido por Schmid, [3], para modelar fenômenos dinâmicos temporais complexos, com distribuição baseada em grande quantidade de dados. Esse método é caracterizado por ser uma descoberta de conhecimento analítico que visa construir um modelo dinâmico a partir de uma massa de dados de alta dimensão, seguindo uma projeção em espaços de baixa dimensão que mantém apenas informações significativas. O método é linear, mas pode ser aplicado a fenômenos não-lineares e funciona concentrando-se nos dados de captura instantânea.

---

<sup>1</sup>ltiemi@gmail.com

<sup>2</sup>luis.dafonseca@cefetmg.br

<sup>3</sup>wilson@ime.unicamp.br

<sup>4</sup>daniel.barbosa@engenharia.ufjf.br

Neste trabalho aplicamos o método DMD num conjunto de dados empíricos, de número de casos de dengue em cada município e distrito brasileiro, no período de 2014 a 2019, disponíveis no Departamento de Informática do Sistema Único de Saúde (DATASUS) [1].

## 2 Episódio de Dengue

O método DMD é utilizado para analisar um episódio epidemiológico geográfico de propagação da dengue em território brasileiro, por meio dos dados gerados pelo Sistema Único de Saúde (SUS), programa de saúde público brasileiro que disponibiliza o número de casos da dengue, seus agravamentos, entre outros, com data e municípios de notificação, por meio do programa de informações à saúde (TABNET), DATASUS [1]. O que nos dá um bom diagnóstico da dengue em cada município e distrito brasileiro. Os dados utilizados aqui correspondem ao período de 2014 a 2019, em semanas epidemiológicas que é um tempo padrão em eventos epidemiológicos, sendo o ano então dividido em 52 ou 53 semanas. A partir da presença de picos anuais e do período no ano em que ocorrem tais picos, vide Figura 1, fica nítido o comportamento sazonal da dengue em território nacional.

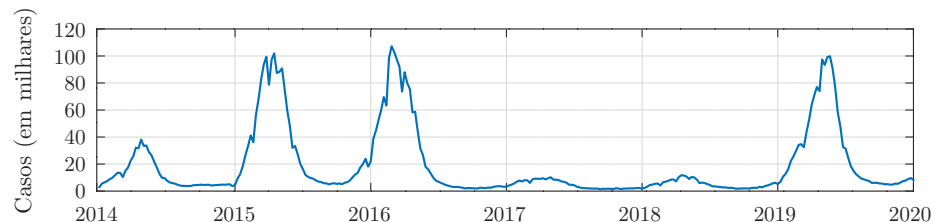


Figura 1: Número de casos totais (em milhares) da dengue no Brasil no período de 2014 a 2019. Dados extraídos do DATASUS [1].

## 3 Decomposição em Modos Dinâmicos

Nesta seção, descrevemos o método DMD num cenário epidemiológico considerando  $x \in \mathbb{R}^n$  e  $k \in \mathbb{N}$ , tais que

$$x_k = (x_k^1, x_k^2, \dots, x_k^n), \quad (1)$$

representam os dados reais, ou observados, espaço-temporais de uma dinâmica, sendo que os valores  $x_k^i$  representam o número detectado de indivíduos infectados na  $i$ -ésima localização,  $i = 1, \dots, n$ , no tempo  $k$ ,  $k = 1, \dots, m$ . Dizemos que  $x_k$  é o dado espacial instantâneo no tempo  $k$ . Já o vetor  $x_{k+1}$  representa o dado espaço-temporal em um cenário seguinte, ou seja, resultante do estado anterior  $x_k$ . Pelo método supõe-se que os dados observados podem ser reproduzidos aproximadamente pela relação de recorrência

$$x_{k+1} = A x_k, \quad (2)$$

sendo  $A \in M_{n \times n}(\mathbb{R})$  o melhor ajuste (no sentido dos quadrados mínimos) obtido levando em consideração todos os pares de dados instantâneos. Uma relação funcional entre os  $m$  sucessivos dados  $x_k$  e  $x_{k+1}$  é descrita aproximadamente pela expressão linear

$$X' \approx AX, \quad (3)$$

sendo

$$X = \begin{bmatrix} | & | & \cdots & | \\ x_1 & x_2 & \cdots & x_{m-1} \\ | & | & \cdots & | \end{bmatrix} \text{ e } X' = \begin{bmatrix} | & | & \cdots & | \\ x_2 & x_3 & \cdots & x_m \\ | & | & \cdots & | \end{bmatrix}.$$

A DMD para o par de conjuntos de dados  $X$  e  $X'$  é a decomposição espectral da matriz  $A_{n \times n}$  da equação (3),

$$A = X'X^\dagger = X'V\Sigma^{-1}U^*,$$

sendo que  $\dagger$  indica a inversa de Moore-Penrose e  $U\Sigma V^*$  é a Decomposição em Valores Singulares (SVD) de  $X$ ; e  $*$  indica a transposta conjugada. A dimensão,  $r$ , do espaço de projeção pode ser escolhida ou depender da escolha da porcentagem de energia,  $E_r$ , neste último caso podemos determinar  $r$  resolvendo

$$E_r = \left( \sum_{i=1}^r \sigma_i \right) \left( \sum_{i=1}^n \sigma_i \right)^{-1}.$$

Determinado  $r$  consideramos

$$A \approx \bar{A} = X'\tilde{V}\tilde{\Sigma}^{-1}\tilde{U}^*, \tag{4}$$

tal que  $\tilde{\Sigma} \in M_{r \times r}(\mathbb{R})$ , e  $r$  é o limiar de truncamento para os valores singulares da decomposição SVD de  $X$ . As características dinâmicas do modelo são representadas pela decomposição

$$\tilde{A}W = W\Lambda,$$

sendo que as matrizes  $W$  e  $\Lambda$  contém os autovetores e os autovalores, respectivamente, da matriz

$$\tilde{A} = \tilde{U}^*X'\tilde{V}\tilde{\Sigma}^{-1}. \tag{5}$$

Os modos dinâmicos

$$\phi = X'\tilde{V}\tilde{\Sigma}^{-1}w \in \mathbb{C}^n \tag{6}$$

relacionam a baixa dimensão  $\tilde{A}_{r \times r}$  com a alta dimensão  $\bar{A}_{n \times n}$ . Os modos dinâmicos descrevem como as localizações espaciais estão relacionadas, pois cada componente de  $\phi$  contém duas importantes informações: o seu valor absoluto que é a medida da sua participação na dinâmica; e o valor de fase que dá o atraso do comportamento oscilatório com respeito às outras componentes, além disso cada autovalor  $\lambda \in \Lambda$  carrega consigo a informação da frequência do comportamento oscilatório.

## 4 Aplicação

O episódio epidêmico da dengue foi discretamente descrito pela população infectada durante 313 unidades de tempo (semana epidemiológica) nos 5568 municípios e nos 2 distritos brasileiros. Logo a matriz  $X$  da equação (3) possui dimensões  $5570 \times 312$ . Foi feita uma análise de erro máximo obtido ao utilizarmos 157 valores singulares que ficou na ordem de  $e^{-11}$ , o que foi considerado adequado. Assim, a matriz  $\tilde{A}$  possui dimensão  $157 \times 157$ , ou seja, a dinâmica que está num espaço de dimensão 313 é projetada num espaço de dimensão 157.

Introduzimos uma medida de relevância,  $\gamma$ , aos modos obtidos, de forma a sermos capazes de identificar os modos mais relevantes na dinâmica obtida, essas medidas dependem dos autovalores de  $\tilde{A}$ . Na Figura 2 dispomos  $\gamma$  sobre o eixo temporal dado em anos. Ao analisarmos os períodos, ou seja, o inverso das frequências, em anos e compararmos os valores de  $\gamma$  obtemos picos mais relevantes em 1 e 3 anos, o que indica a sazonalidade da dengue.

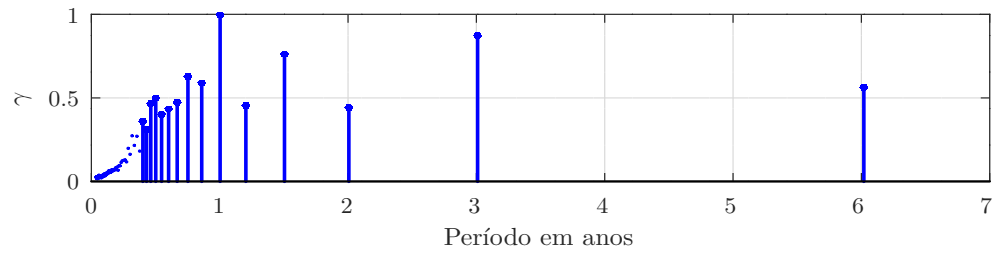


Figura 2: Relevância normalizada por período associado a cada modo,  $\phi_j$ ,  $j = 1, \dots, 157$ , obtido.

Estamos interessados em uma visualização da distribuição de fase espacial do modo mais relevante. Assim, para uma visualização de quais municípios e distritos encontram-se na mesma fase, ou seja, que apresentam casos de dengue no mesmo período do ano, sobrepomos em um mapa os municípios e os distritos, com mais de 50 mil habitantes, vide Figura 3, de forma a sermos capazes de extrair algumas informações sobre a propagação da doença.

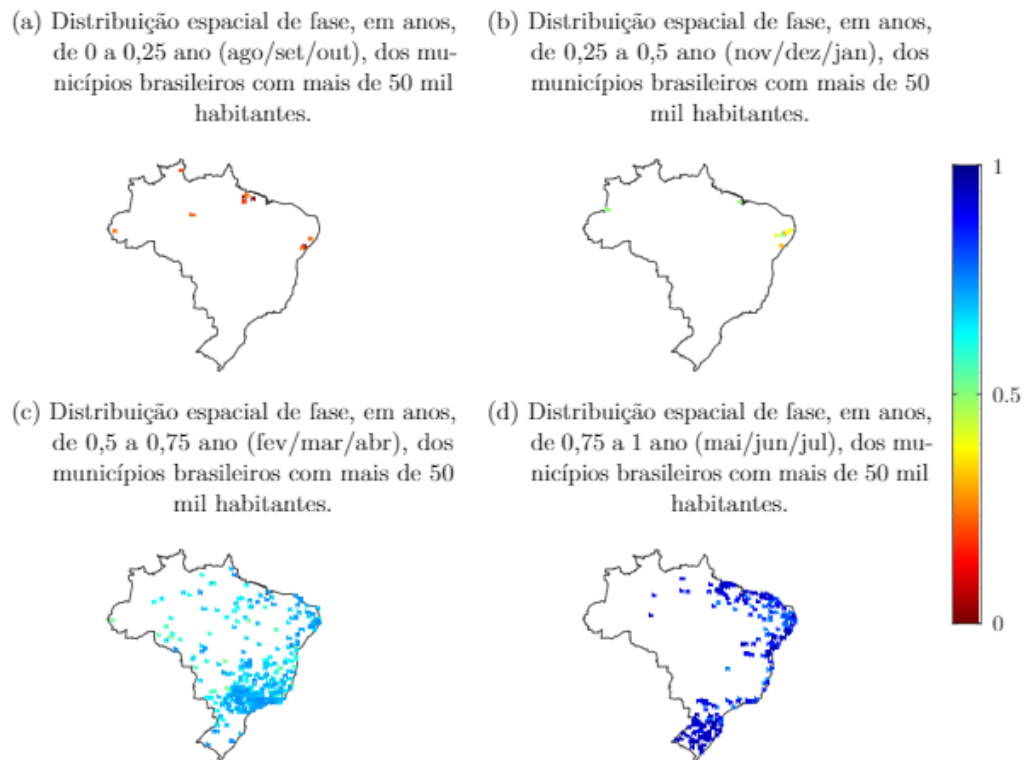


Figura 3: Distribuição espacial das fases, em anos, dos municípios e distritos brasileiros, com mais de 50 mil habitantes, separados em 4 faixas de fase, iniciando em agosto e terminando em julho. Em cada subfigura (a), (b), (c) e (d) temos na fase correspondente os municípios e distritos com mais de 50 mil habitantes, que notificaram casos de dengue, a cor representa a densidade de casos.

Nas simulações apresentadas na Figura 3 temos a distribuição espacial dos municípios, com mais

de 50 mil habitantes, que estão na mesma faixa de fase em relação a dengue num período de um ano. Nas regiões norte e nordeste cerca de 84% e 90,8%, respectivamente, dos municípios não chegam a ter 50 mil habitantes, logo sobre os municípios que aparecem nessa simulação podemos dizer que são centros: comerciais, educacionais e, ou, de saúde. Logo estes viabilizam o espalhamento da dengue para o interior. Na Figura 3(a) é apresentado os municípios que possuem casos da dengue, no período de agosto a outubro, período do pós inverno no Brasil que não possui condições favoráveis ao mosquito transmissor, com exceção das regiões norte e litoral nordestino. Por meio da Figura 3(b), que representa o período de novembro a janeiro, verifica-se que os municípios do entorno dos municípios da Figura 3(a) passam a ter casos de dengue, evidenciando um importante papel dos municípios da Figura 3(a) para a manutenção da dengue e da sua dispersão no seu entorno. Com o calor e as chuvas no período de fevereiro a abril, há uma explosão de casos de dengue no país, vide Figura 3(c). Com a chegada do inverno, período de maio a julho, a dengue desaparece novamente da região centro sul e com o retorno das chuvas nas regiões norte e nordeste verificamos a interiorização da dengue nessas regiões, vide Figura 3(d). Destacamos que as cidades da região sul que aparecem na Figura 3(d), na verdade estão com aproximadamente zero casos de dengue.

## 5 Conclusões

O método DMD apresenta-se como uma forma simples, porém efetiva de obter um conjunto de vetores, os modos dinâmicos, que atuam como uma “base”, não propriamente de um espaço vetorial, mas para a dinâmica e representa o fenômeno com precisão adequada, isto é, tomamos fenômenos com dinâmicas complexas e projetamos essas dinâmicas, em um espaço gerado pelos modos, onde os modos gerados contém as principais informações da dinâmica tanto de noção espacial, intensidade, quanto temporal, fase e frequência. No episódio da dengue fomos capazes de observar isto, tanto as principais frequências de ocorrência da doença, quanto as suas características espaciais que aparecem nitidamente quando observados os modos e as frequências associadas. Pelas fases apresentadas, outro destaque é que devido a diferença de clima entre as regiões do país ao longo do ano é um fator que favorece a permanência da dengue no país, pois quando uma região não tem condições favoráveis à doença uma outra tem e, assim, se mantém o vírus circulando. Podemos conjecturar também rotas de propagação através da distribuição de fase espacial dos modos associados as principais frequências.

Algumas informações relevantes ainda podem ser extraídos no caso da dengue, como a identificação de cidades chave que atuam na permanência da dengue no Brasil, destacamos que dos mais de 5000 municípios, identificamos cerca de 200 apenas que possuem alta atividade da doença em cerca de metade do ano. Seria possível também identificar melhor a propagação em uma região menor, conhecendo as principais vias de transporte da doença.

## Agradecimentos

Este trabalho foi financiado em parte pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

## Referências

- [1] DATASUS, Banco de dados do Sistema Único de Saúde, Ministério da Saúde do Brasil. Disponível em: <https://tabnet.datasus.gov.br/cgi/sim/obtmmap.htm#>.

- [2] J. L. Proctor and P. A. Eckhoff. Discovering dynamic patterns from infectious disease data using DMD, *Int. Health*, 7: 139-145, 2015. DOI: 10.1093/inthealth/ihv009.
- [3] P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.*, 656:5-28, 2010. DOI:10.1017/S0022112010001217.