



DETERMINANDO PADRÕES DINÂMICOS PARA A DENGUE

Lucy Tiemi Takahashi*

Universidade Federal de Juiz de Fora, Juiz de Fora, Brazil

Resumo

O método matemático de Decomposição em Modos Dinâmicos é um procedimento desenvolvido recentemente para modelar fenômenos dinâmicos temporais complexos com distribuição baseada em grande quantidade de dados. Este método caracteriza-se 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 pode ser aplicado a fenômenos altamente não-lineares e funciona focando em dados instantâneos. Neste trabalho em particular, a Decomposição em Modos Dinâmicos é aplicada a um episódio epidemiológico geográfico de propagação da dengue, gerado por simulações.

Trabajo realizado en conjunto con:

Luis Alberto D'Afonseca¹, Centro Federal de Edu. Tec. de Minas Gerais, Belo Horizonte, Brasil.

Wilson Castro Ferreira Jr.², Universidade Estadual de Campinas, Campinas, Brasil.



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. O Big Data se caracteriza por ter um grande volume, uma grande variedade e uma grande velocidade. E estes dados, em geral, não são planejados, nem estruturados e são dados orgânicos (“não pagos”).

*Parcialmente financiado por CAPES, e-mail: ltiemi@gmail.com

¹e-mail: luis.dafonseca@gmail.com

²e-mail: wilson@ime.unicamp.br

Ao se utilizar do Big Data temos que estar cientes de que este conjunto “pode responder” algumas perguntas, mas nem sempre as perguntas que queremos respostas sobre um problema real. Mas, trata-se de uma fonte alternativa para substituir ou complementar fontes tradicionais. No Big Data como num levantamento estatístico devemos estar atentos a variabilidade e a veracidade, e é ainda uma tarefa complexa combinar dados de multiplas fontes. Logo mais dados não é necessariamente bom. Outro ponto importante é acessibilidade, buscar dados que estão e continuarão disponíveis.

Na biomatemática pesquisadores têm apresentado trabalhos utilizando informações do Big Data. Entre outros, Wesolowski *et al.* (2015), em [7], usa dados de telefones móveis para analisar o fluxo da rubéola no Kênia, Yang *et al.* (2015), em [8], utiliza dados de pesquisas feitas no Google para rastrear a ocorrência da gripe.

Proctor and Eckhoff (2015), em [3], utilizaram o Big Data para estudar a dinâmica da gripe nos Estados Unidos utilizando dados do Google Flu Trends; e para a dinâmica do sarampo no Reino Unido e da poliomelite na Nigéria utilizam dados da pré-vacinação. Mas, para analisar estes dados eles lançam mão ao método matemático chamado Decomposição em Modos Dinâmicos (DMD), que reduz a dimensão do espaço de estado do fenômeno, mantendo suas características principais.

O DMD foi desenvolvido por Peter J. Schmid em 2008 para estudar campos de fluxo na hidrodinâmica, em [4] Schmid apresenta o DMD como sendo um método capaz de extrair informações dinâmicas do fluxo. O DMD determina um conjunto formado por modos dinâmicos que estão associados a uma frequência de oscilação fixa e a taxa de crescimento/decaimento. Diferentemente do PCA o DMD determina modos não ortogonais o que pode ser mais trabalhoso de se interpretar. Mas, a vantagem é que cada modo está associado a um comportamento sinusoidal amortecido no tempo.

Segundo Proctor em [3] e Kutz em [1] o DMD é adequado para modelar fenômenos dinâmicos temporais e complexos, com uma distribuição baseada em uma grande quantidade de dados. E que o método é caracterizado por usar um conhecimento analítico que visa descobrir um modelo dinâmico a partir de uma massa de dados de alta dimensão, por meio de uma projeção em espaços de baixa dimensão que mantém apenas informações significativas. O método pode ser aplicado a fenômenos altamente não-lineares e funciona, concentrando-se em dados instantâneos, os *snapshots*.

Neste trabalho em particular, o DMD é apresentado e aplicado a um episódio epidemiológico geográfico de propagação da dengue num cenário formado por cidades do estado de São Paulo, no Brasil. Neste trabalho apresentamos como é feita a DMD, passo-a-passo, utilizando termos no contexto de epidemiologia e utilizaremos o método para analisar dados gerados por um modelo dinâmico para a propagação da dengue, proposto por Takahashi em [5].

2 Decomposição em Modos Dinâmicos

Na Decomposição em Modos Dinâmicos (DMD) para um cenário epidemiológico consideramos $x \in \mathbb{R}^n$ e $k \in \mathbb{N}$, tais que x_{k+1} representa os dados espaço-temporais de uma dinâmica epidêmica em um cenário seguinte a um estado anterior x_k . Aqui

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

é chamado de *snapshots*, onde x_k^i representa o número de indivíduos infectados detectados na i -ésima localização no tempo k , com $i = 1, \dots, n$, $k = 1, \dots, m$ e $n < m$. Definidos os *snapshots*, equação (1), buscamos a seguinte recursão

$$x_{k+1} \approx Ax_k, \quad (2)$$

onde $A \in M_{n \times n}(\mathbb{R})$ é o melhor ajuste (no sentido de quadrados mínimos) obtido levando em conta todos os pares de *snapshots*, ou seja, $k = 1, \dots, m - 1$. A relação funcional entre dados sucessivos x_k e x_{k+1} é aproximadamente descrita pela expressão linear

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

onde

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

assim, tomamos o operador

$$A = X'X^\dagger, \quad (4)$$

com \dagger indicando a pseudoinversa de Moore-Penrose, que minimiza o erro

$$\|X' - AX\|_F$$

onde $\|\cdot\|_F$ denota a norma de Frobenius.

Para construir a DMD consideramos a decomposição em valores singulares (SVD) de X e que ela seja dada por

$$X = U_{n \times n} \cdot \Sigma_{n \times m-1} \cdot V_{m-1 \times m-1}^* \quad (5)$$

onde $*$ indica a matriz transposta conjugada.

Daí, de (4) temos

$$A = X'X^\dagger = X'V\Sigma^{-1}U^*. \quad (6)$$

Na SVD da matriz X , equação (5), temos no máximo $(m-1)$ valores singulares. Uma forma de se determinar o valor de truncamento, r , para se obter a *low-dimensional* do espaço de aspecto é escolher uma certa porcentagem de energia, entre os r maiores primeiros modulos dos valores singulares, definida por

$$E_r = \frac{\sum_{i=1}^r \sigma_i}{\sum_{i=1}^m \sigma_i}. \quad (7)$$

Neste trabalho apresentaremos resultados considerando 90% de energia para o truncamento [3, 6, 4].

Assim, uma vez escolhido r tomamos

$$A \approx \bar{A} = X' \tilde{V} \tilde{\Sigma}^{-1} \tilde{U}^* \quad (8)$$

onde $\tilde{\Sigma} \in M_{r \times r}(\mathbb{R})$. Tomando

$$\tilde{A}_{r \times r} = \tilde{U}^* X' \tilde{V} \tilde{\Sigma}^{-1}. \quad (9)$$

temos

$$\tilde{A} \approx \tilde{U}^* A \tilde{U} \quad (10)$$

e assim os autovalores de \tilde{A} são autovalores de A . Desta forma, as características dinâmicas do modelo são representadas pelos autovalores e autovetores da decomposição

$$\tilde{A}W = W\Lambda, \quad (11)$$

onde W é matriz com os autovetores relacionados com os autovalores na matriz diagonal Λ .

A relação entre a dinâmica do modelo com ordem reduzida (*low-dimensional*), \tilde{A} , e o modelo (*high-dimensional*), A , é obtida pelos chamados **modos dinâmicos**:

$$\phi = X' \tilde{V} \tilde{\Sigma}^{-1} w \quad (12)$$

onde, ϕ está na dimensão n e w é autovetor de \tilde{A} . Note que se $\lambda \neq 0$ este é o modo dinâmico para λ . Agora se $\lambda = 0$ então o modo dinâmico é $\phi = \tilde{U}w$.

Os modos dinâmicos, ϕ , descrevem como as localizações espaciais estão relacionadas. Dentro de um único modo dinâmico, cada entrada (componente) tem duas importantes informações: a magnetude (valor absoluto) que dá uma medida da participação daquela posição espacial na dinâmica; e o ângulo entre a parte real e imaginária de cada entrada que fornece a fase de oscilação daquela entrada em relação as outras, ou seja, dá a frequência dos modos.

3 DMD para a dengue

Neste trabalho aplicamos o DMD em dados gerados por um modelo dinâmico desenvolvido por Takahashi [2004] para descrever a propagação da dengue em uma região restrita do sudeste do Brasil, o estado de São Paulo, veja Figura 1. Este modelo reproduz de modo razoável dados esparsos disponíveis para esta doença nas últimas décadas, veja Figura 4.

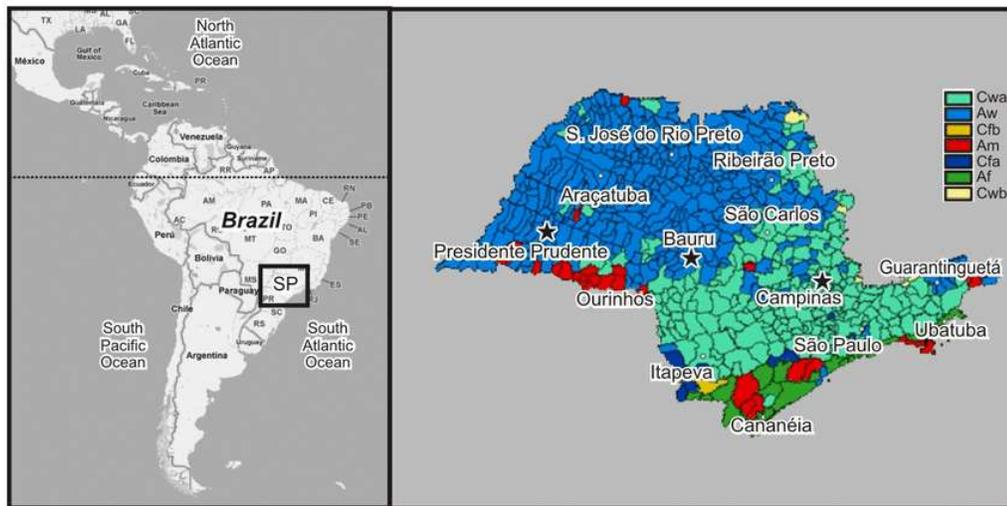


Figura 1: Figura extraída de [2]. À esquerda temos o mapa da América do Sul com destaque o Estado de São Paulo e à direita mapa do Estado de São Paulo com classificação climática de Köppen–Geiger.

No estado de São Paulo existem regiões com climas mais favoráveis ao mosquito que outras, e isto foi considerado para gerar os *snapshots*. Na Figura 1 à esquerda temos o mapa da América do Sul com destaque o Estado de São Paulo e à direita mapa do estado de São Paulo com classificação climática de Köppen–Geiger, indicando os municípios e algumas cidades importantes do Estado. Veja [2] para mais detalhes.

Nas simulações consideramos uma rede discreta formada por 60 municípios no estado de São Paulo, conectadas por rodovias, conforme apresentado na Figura 2. E dentro de cada município (nó da rede), foi considerada uma dinâmica epidemiológica local simples modelando as populações de mosquitos e de humanos ambos com infectados e suscetíveis. A dispersão espacial da doença foi descrita pela conexão das localidades com um fluxo medido pelo respectivo tráfego rodoviário. Esses 60 municípios foram escolhidos entre os 645 existentes no estado por serem centros regionais e, ou, cidades localizadas na fronteira do estado. A região externa ao estado é considerada nula, ou seja, não consideramos a influência de outros municípios do país fora do estado de São Paulo na dinâmica.

Segundo a Sucen - Superintendência de Controles de Endemias, da Secretaria de Estado da Saúde do Governo do Estado de São Paulo a propagação da dengue, em 1985, teve início em três cidades:

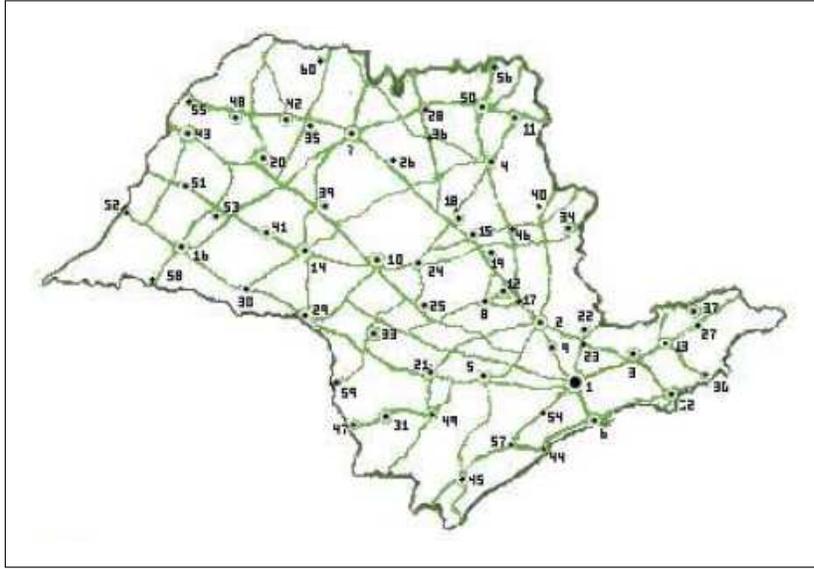


Figura 2: 60 municípios do Estado de São Paulo com as rodovias consideradas interligando-os [5].

7 - São José do Rio Preto, 20 - Araçatuba e 16 - Presidente Prudente, dado que foi utilizado como condições iniciais para as simulações [5]. Na Figura 3 temos ao longo de 1710 dias o comportamento das populações: de Mosquito Suscetível, Infectado e Total; e a de Humano Suscetível, Infectado e Removido. O tempo 1710, correspondendo a 5 anos, onde cada ano possui um período de 270 dias corridos que, nesta região do Brasil, são favoráveis ao mosquito devido as temperaturas elevadas com bastante humidade, e mais 90 dias corridos que representam um período no inverno que é desfavorável ao mosquito, por ser seco e frio. O último período desfavorável foi descartado nestes 5 anos.

As entradas do *snapshot* x_{1710} de Mosquito Total são representadas por cores no mapa do estado, estas cores representam a densidade de *Aedes aegypti* no ano 1990 em cada município, veja Figura 4 à esquerda. Já os municípios infestados por *Aedes aegypti* do Estado de São Paulo em 1990, após 5 anos de invasão, conforme dados da Sucec é exibido na Figura 4 à direita, [5].

Para o modelo DMD das simulações, exibidas na Figura 3, consideramos apenas as de Humanos Infectados. Assim, construímos os *snapshots* e montamos a matriz X da equação (3). A matriz X então possui dimensão 60×1709 .

Para realizar a DMD neste conjunto de *snapshots* com uma precisão de 90% foram necessários os primeiros 28 valores singulares entre os 60. O maior autovalor em módulo de \tilde{A} encontrado é real. Estas informações são exibidas na Figura 5.

Para cada um dos 60 municípios do Modo 2 determinamos a magnetude Figura 6 à esquerda e a fase Figura 6 à direita.

4 Conclusões

Neste trabalho geramos uma propagação da dengue pela simulação do modelo epidêmico de [5]. Utilizamos o DMD para analisar o comportamento desta epidemia pela dinâmica dos humanos infectados via os modos dinâmicos.

O maior autovalor é real logo o Modo 1 correspondente terá entradas reais, o que implica que todas as cidades estão na mesma fase. Entendemos que isso implique que a dengue ao chegar em qualquer um dos municípios seu comportamento mais forte seja de crescer exponencialmente sem

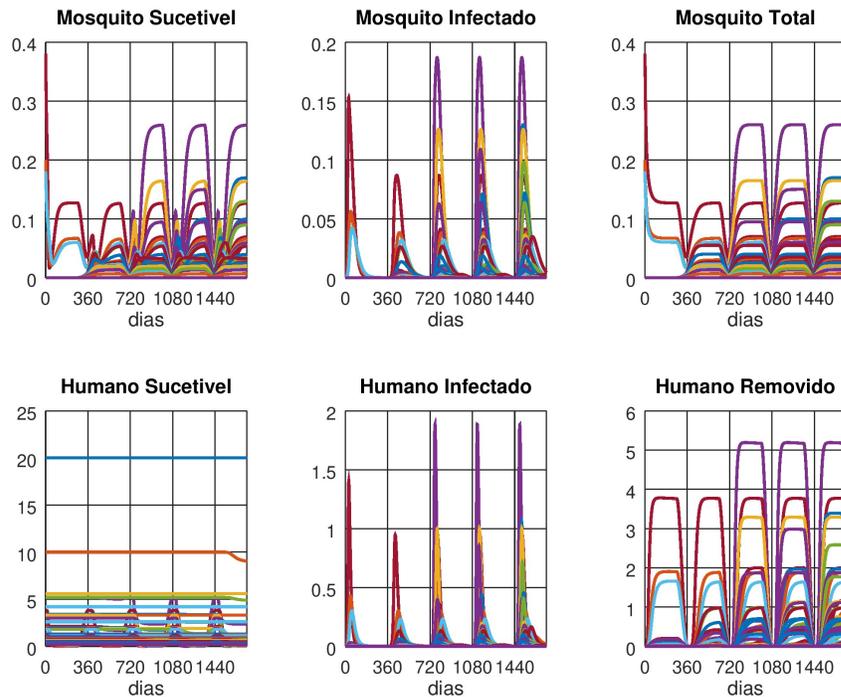


Figura 3: Na figura temos o compartimento das populações, utilizando o sistema (21) e considerando a influência negativa do inverno (90 dias por ano), em relação ao tempo de observação 1710 dias. Os parâmetros são os das Tabelas 3 e 4. Da Tabela 2 apenas as populações das cidades de São José do Rio Preto, Araçatuba e Prudente foram alteradas inserindo 2 infectados em cada uma e alguns mosquitos saudáveis.

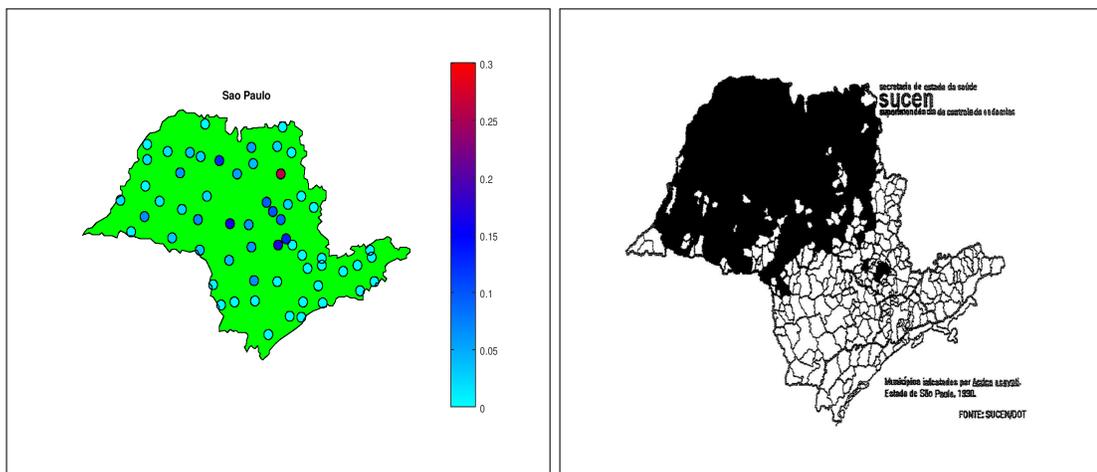


Figura 4: Na figura à esquerda o *snapshot* x_{1710} de Mosquito Total correspondente ao ano 1990 gerado utilizando o sistema proposto por Takahashi, com os 60 municípios em destaque, onde as cores representam a densidade da população de mosquitos. E na figura à direita temos os municípios infestados por *Aedes aegypti* do Estado de São Paulo em 1990, após 5 anos de invasão [5].

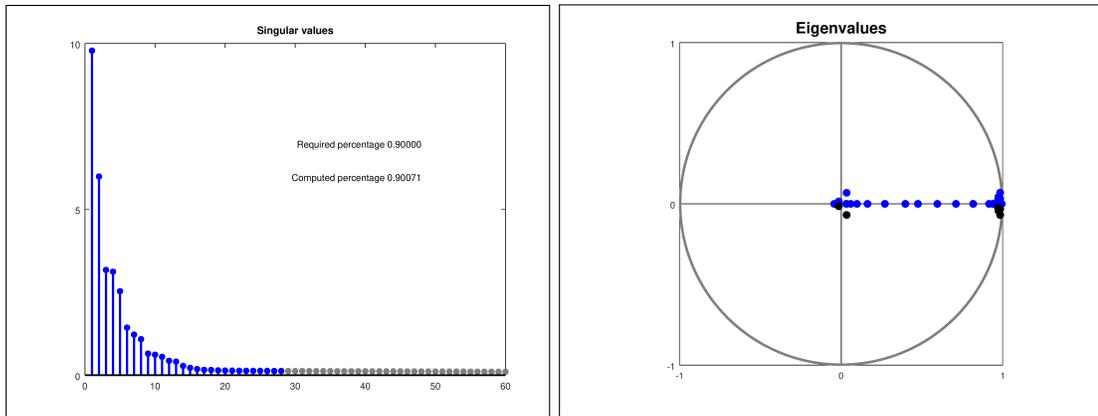


Figura 5: À esquerda temos os valores singulares, em azul são os 28 utilizados para se obter uma energia de 90%. À direita temos os autovalores de \tilde{A} plotados no plano complexo.

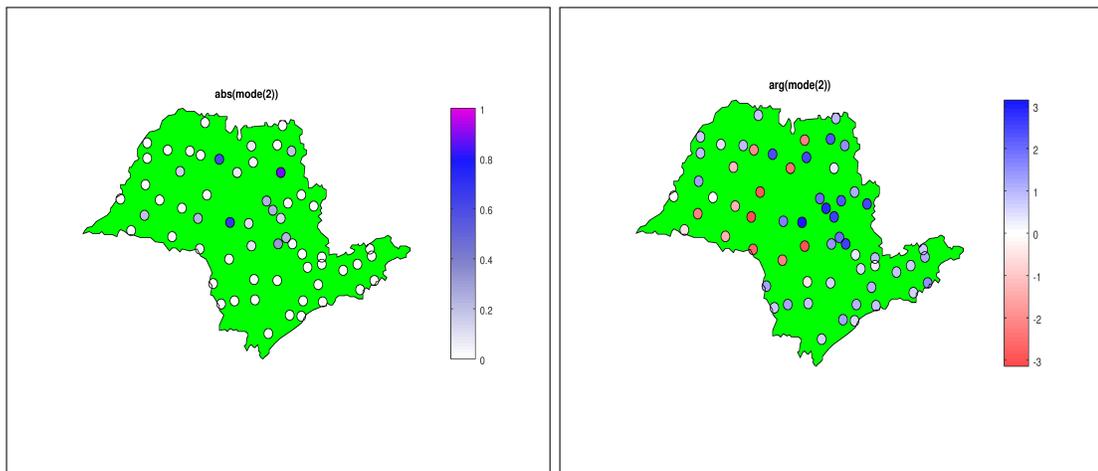


Figura 6: À esquerda magnitude das entradas do Modo 2. À direita fase em cada município obtida do Modo 2.

oscilações, pois não estamos considerando que os humanos mudem seu comportamento frente a chegada da dengue.

Passamos então para a análise do Modo 2, referente ao segundo maior autovalor em módulo, que agora não é real. Do Modo 2 construímos uma representação da magnetude e da fase em cada município, o que é exibido na Figura 6. Na Figura 6 à direita as cores azul escuro e vermelho representam a mesma fase e elas se concentram em municípios com populações de tamanhos diferentes, mas industrializados. E estes municípios são interligados por rodovias com pista dupla o que facilita o fluxo. E suas cidades “satélites” encontram-se em fases um pouco atrasadas. Entendemos então que a dengue ao chegar numa cidade economicamente importante ela espalha a dengue para as cidades satélites, mas isso demora um tempo para ocorrer. Temos também cidades com a mesma fase sem estarem diretamente ligadas pelas rodovias, o que demonstra que elas possuem características parecidas e importantes para a doença. Outra informação importante a ser extraída da Figura 6 é que dos três municípios que possuem maior influência na propagação da dengue, os com cores azuis na figura à esquerda, apenas uma delas foi considerada na condição inicial contendo humanos infectados. Logo nem sempre o local que inicia a propagação é o que exerce maior influência na propagação ao longo do tempo. E destes resultados obtidos do Modo 2 podemos propor estratégias de controle da propagação. Uma vez surgindo a dengue no estado de São Paulo, não seria num primeiro momento necessário agir em todos os municípios, mas nestes três municípios os com maior magnetude no Modo 2 deveriam ter ações mais intensas de controle do mosquito, e assim da dengue.

Destas análises concluímos também que: O DMD é uma ferramenta capaz de representar um conjunto de dados espaço temporais complexos de forma sistemática e simplificada; Dependendo dos dados, o ajuste por mínimos quadrados usado para calcular a matriz A pode introduzir grandes erros; Os dados modelados reproduzem a dinâmica observada do episódio epidemiológico da dengue; e para 90% requerido o DMD usou 38 valores singulares dos 60 e retornou mais que os 90% requeridos.

Referências

- [1] J. N. Kutz, S. L. Brunton, B. W. Brunton and J. L. Proctor, *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*, SIAM, Philadelphia, 2016.
- [2] L. C. Labaki *et al.*, Conforto térmico em espaços públicos de passagem: estudos em ruas de pedestres no estado de São Paulo, *Ambiente Construído*, Porto Alegre, v. 12, n. 1, p. 167-183, 2012.
- [3] PROCTOR, J. L.; ECKHOFF, P. A. Discovering dynamic patterns from infectious disease data using dynamic mode decomposition, *Int. Health*, 7: 139-145, 2015. DOI:10.1093/inthealth/ihv009.
- [4] P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.*, 656:5-28, 2010. DOI:10.1017/S0022112010001217.
- [5] L. T. Takahashi. Modelos Matemáticos de Epidemiologia com Vetores: Simulação da Propagação Urbana e Geográfica da Dengue. PhD Thesis, IMECC, Unicamp, Brazil, 2004. Online: <https://www.ime.unicamp.br/pos-graduacao/modelos-matematicos-epidemiologias-com-vetores-simulacao-propagacao-urbana-geografica>.
- [6] J. H. Tu, C. W. Rowley, and D. M. Luchtenburg, On dynamic mode decomposition theory and applications, submetido para *Journal of Computational Dynamics*, arXiv:1312.0041v1, 2013.
- [7] A. Wesolowski *et al.*, Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data, *PNAS*, 112:11114-11119, 2015.

- [8] S. Yang, M. Santillana, and S. C. Kou, Accurate estimation of influenza epidemics using Google search data via ARGO, *PNAS*, 112:14473-14478, 2015.