

Mathematical and Statistical models as tools for the control of mosquito-borne diseases: The experience of Costa Rica

Fabio Sanchez, Luis A. Barboza, Paola Vásquez, Yury E. García,
Juan G. Calvo, Shu-Wei Chou, Gustavo Mery

June 24, 2021

Abstract

The modeling of infectious diseases provides valuable input in the development of mitigating strategies and implementation of public health interventions. We highlight results and current research conducted in Costa Rica using mathematical and statistical tools to develop optimal strategies for mosquito control and mosquito-borne disease prevention/control methods in the country.

1 Introduction

Currently, despite strong control efforts and the increased knowledge acquired throughout history, timely and effective management of mosquito-borne pathogens continue to elude public health authorities worldwide. The past decades have been especially challenging as increased urbanization, population mobility, deforestation, climate change, insecticide resistance, and deficient vector control programs have created suitable conditions for both, pathogens and vectors, to emerge into new areas and to reemerge in regions from which previously successful vector control campaigns had led to their eradication [1, 2].

As a result, more than half the world's population now live in areas at risk of dengue infections [3], an estimated 229 million cases and 409,000 deaths related to malaria were notified in 2019 by the World Health Organization [4], and previously obscured viruses such as Chikungunya and Zika reached the Americas bringing with them unexpected health complications to the population [5, 6].

Under this scenario, public health officials around the world have highlighted the need to explore innovative and cost-effective tools to strengthen the prevention and mitigation programs currently in place in affected countries. However, the complexity involved in the transmission dynamics, coupled with newly recognized pathogenic mechanisms and modes of transmission [7], urgently demands a more interdisciplinary approach to reach this goal [8]. In this effort, the use of mathematical and statistical modeling techniques have been increasingly recognized as important and plausible tools for policymakers.

In essence, these models provide a simplified representation of a complex system, which involves a variety of underlying factors, interactions, heterogeneity, random variations and fluctuations. They allow to, either improve the understanding of how a certain pathogen spreads or to provide predictive models to develop preventive methods and reduce disease incidence [9, 10]. For their development, multiple techniques [11] and variables have been historically used, including climatological, environmental, entomological, demographic and socioeconomic data [12, 13, 48]. The results obtained, have shown the potential of such models to highlight key factors for public health interventions and to guide public policy, thus helping allocate human and economic resources more efficiently for the fight against mosquito-borne pathogens [14].

However, as every model has limitations, both intrinsic to the model itself and user or data dependent. There are several challenges that have to be overcome when attempting to use them for public health purposes [9]. Nevertheless, the increased knowledge in mosquito population biology, advances in technology, access to enormous amount of real-time epidemiological, demographic, socioeconomic and climate data have opened up a new array of possibilities for these models to reinvigorate disease control [8].

In Costa Rica, according to data from the Ministry of Health, since 1993, more than 392,000 suspected and confirmed dengue cases have been reported by the Ministry of Health [16]. Although in 2013 the country achieved the non-report of autochthonous malaria patients, since 2016 the local transmission has slowly resumed its upwards trend [16]. In 2014, the first cases of chikungunya were reported, and by February 2016, the first two indigenous cases of Zika virus were reported in the Pacific coast [16]. Therefore,

the introduction and development of new methods and tools is at the forefront of public health priorities in the country.

Here, we review the results, limitations and public health implications of research conducted at Universidad de Costa Rica using different mathematical and statistical techniques to model mosquito-borne diseases in the country. The models were used to better understand the dynamics involved in the 2016-2017 Zika outbreak in Costa Rica, and to retrospectively predict the relative risk of acquiring dengue in five climatically diverse municipalities during the 2016-2017 outbreak. The models were developed using suspected and confirmed cases data from the Ministry of Health and other public health institutions.

The article is organized as follows: In Section 2, we provide details of the data and statistical methodology applied to estimate parameters, as well as the description of the models used. In Section 3, we discuss the results obtained and their public health implications and, in Section 4, we give our conclusions and final thoughts.

2 Materials and Methods

Data. Publicly available data was used for the development of the models. The Ministry of Health of Costa Rica provided weekly epidemiological information of suspected and confirmed dengue and Zika cases. Demographic information of the study areas was provided by the National Institute of Statistics and Census and the National Meteorological Institute provided the daily temperature, precipitation and relative humidity from meteorological stations located in the selected study areas. In order to include further climate information, data of the daily Sea Surface Temperature Anomaly (SSTA) from El Niño 3.4 region was obtained from the Climate Prediction Center (CPC) of the NOAA.

Methods. Throughout history different mathematical models have been used to study the dynamics and behavior of vector-borne diseases [17, 18, 19, 20, 21, 22, 23, 24] among others. These models address questions related to population disease dynamics. The range of mathematical models of this type is quite broad: systems of ordinary and partial differential equations, graph theory, agent-based modeling, among others are used to study and understand vector-borne disease dynamics.

A different question of interest in terms of modeling is the precise prediction of the relative risk of the disease using historical data. This type of problem has been investigated using predictive statistical models, from classical generalized linear models to statistical learning methods, see for example: [25, 26, 27, 28].

In order to fit the models with observed information, different statistical methods have been used to infer the underlying parameters and state variables of each model. In this section, we explain three different estimation approaches employed in the modeling of vector-borne diseases in Costa Rica.

2.1 Classical Methods

In [29], the authors explore a single-outbreak deterministic model with state variables describing the host and vector dynamics of dengue and chikungunya. The estimated parameters were: the transmission rate, the diagnosis rate, the average vector infectious period and the initial value of the susceptible population. The main assumption behind is that the observed number of cases at week k follows a Poisson distribution whose rate is determined as a function of the expected number of cases within the k -th week. In this case the authors used a least-squares procedure based on the normalized differences between observed and expected weekly cases in order to fit the whole set of parameters and initial populations.

2.2 Approximate Bayesian Computation

The Bayesian statistical approach was also explored in [30] by means of a single-outbreak model with sexual transmission mathematical model that describes the overall dynamics of Zika during 2016-2017 in Costa Rica. The authors define as parameters to estimate the average life-span of mosquitoes, the mosquito biting rate and the per-capita diagnosis rate. From a Bayesian point of view, the parameters are treated as random variables whose posterior distribution can be determined once we adopt a prior distribution for each of them. The main problem with this approach is that any usual Bayesian estimation technique requires a well-defined likelihood function in order to express the posterior distribution of the parameters and due to the complexity of the assumed dynamical system in this case it was impossible to obtain it. The authors

were able to overcome that limitation by using an Approximate Bayesian computation approach, see [31] with a rejection sampling scheme.

2.3 Statistical Learning

A different question of interest is the prediction of dengue incidence in Costa Rica based on past information from different sources. This problem was tackled in [32] using two different approaches: A Generalized Additive Model and Random Forest. Both approaches belong to a larger set of models and procedures called Statistical learning methods or Machine Learning techniques.

Machine Learning models represent an extensive set of algorithms with a high statistical component that allow solving different tasks, for example see [33] for more details): prediction of dependent variables based on some covariate behavior (Supervised Learning models), pattern recognition over spaces of observations or variables (Non-supervised Learning models). Machine learning models are designed under the premise that they must be adjusted through a learning process on a set of test or training data. The increasing availability of new technologies, such as machine learning models, have come to enter as a technique increasingly used in the modeling of infectious diseases. Through the use of computational algorithms, this method allows more refined prediction models, as it learns from past trends and transmission dynamics. Algorithms under this approach, process data in sufficient quantity and without explicit external instructions (and therefore potentially biased) provided by the researcher [34]. Given these characteristics, several studies have evidenced its usefulness and predictive capacity, using for this purpose, a diversity of variables [35, 36, 37]. In Costa Rica, the use of these techniques is still in its early stages, however, the study conducted with dengue and different climatic variables in five municipalities of the country has shown promising results [32]. In the current stage, we are developing statistical models including 32 municipalities in the country and involving climatic variables along with geographic and temporal information, in order to improve the monthly dengue predictability per municipalities.

2.4 Network model

Complex systems can be modeled via network models given that real interactions occur in complicated patterns and are for the most part not deterministic. This type of modeling allows for inclusion of complex heterogeneity, such as including different attributes for each node (individual) and its interactions among contacts (connected nodes in the graph). It could also include multilayers for different type of social groups in order to include particular parameters. For vector-borne diseases, it allows to include one layer for the vectors where we keep track of its population and interactions among them and between humans, per county. It includes an underlying SIR-type epidemiological model for which we mimic the evolution of the disease, allowing a generalization of these types of models with less restrictive assumptions and more complex dynamics.

3 Results

To better understand the transmission dynamics of mosquito-borne diseases in Costa Rica, the application of four different modelling techniques each one aimed to answer different research questions has provided information and showed the potential to be used as an additional tool for public health authorities in the country. With the introduction of the Chikungunya virus in 2014, health authorities faced, for the first time, the circulation of two pathogens with similar clinical manifestations, temporal and spatial distribution. Under this scenario, the use of classical methods of mathematical modelling allowed to study and compare the trend of dengue and chikungunya cases in Costa Rica during the 2015-2016 outbreak. Based on available information, results from the model showed that during the study period it was likely that dengue cases were misdiagnosed and chikungunya cases were under reported. These results, highlighted the importance of continued laboratory-based surveillance. By January 2016, with the introduction of another pathogen transmitted by the Aedes mosquito, the Zika virus, the country faced yet another public health challenge, this time with the added complexity of new ways of transmission and the birth defects and pregnancy complications related to the zika virus. To understand what population was most at risk, the use of a Bayesian statistical approach allowed to demonstrate that people most at risk were those that spent most of their time inside their homes, which according to the time use survey reported by the national institute of statistics and census, women were the part of the population that spent most of their time inside their

homes. Given the characteristics of zika infection, the information brought an extra layer of information and showed the need to increased the prevention and control interventions adjusted to each specific region, interventions that need a more active involvement of the communities members.

While the first two methods allowed to better comprehend the transmission dynamics of past outbreaks, the use of statistical learning algorithms and climate information showed that, by using historical epidemiological and meteorological data, as well as by incorporating lag periods, both the generalized additive model and the random forest model presented adequate performance in predicting the temporal patterns of the relative risk of dengue during 2017, the year that was used as the test period. Periodic fluctuations are frequent in the dynamics of disease transmission.

In an increasingly interconnected and ever-changing world, public health systems of countries with endemic mosquito-borne infections, such as Costa Rica, must be prepared to detect and respond quickly and effectively to changes in the transmission dynamics, potential drivers, geographical distribution and disease risk [38]. The successful implementation of this response requires not only evolution from traditional control approaches, but also, the introduction of novel tools that can be leveraged to further optimize the allocation of available resources [8].

Given the multi-factorial nature and complexity of the problem, collaboration among sectors, institutions and disciplines across the country is certainly needed [39]. In this sense, the use of mathematical and statistical models can serve as platforms to integrate large collections of data from different sectors, helping to systematize and integrate experience and expertise from different areas of knowledge and disciplines. This includes techniques that a priori had not been extensively used for these purposes [40]. For instance, estimates of transmission risk can be used to design tailored interventions to specific scenarios [41, 44]. These models can support large quantities of information, such as demographic, climatological and epidemiological data [32]. Modeling can also clarify the geographical expansion and spread of diseases, therefore allowing to better invest the already scarce resources into specific high-risk areas [42].

However, a better understanding of past transmission dynamics or the availability to predict future transmission risk by the use of these models alone will not be sufficient to achieve sustainable impact in long-term case reduction. It is necessary that public health authorities start a more widely use of these tools in the active decision-making process, always taking into consideration the limitations that every model has [43]. At this stage, it is also necessary, that modelers and health authorities involve knowledge translation activities that will allow the general community to better comprehend their part in the fight against these diseases [45]. These education strategies have the potential to be more targeted thanks to the information that the different models can provide about the unique characteristics of each specific region or community in the country, for example the increased risk of Zika infections among those people that spent most of the time at home inside their homes as opposed to those who spent a substantial amount of time outside due to work or other activities, demonstrate in [30]. Results that go in hand with mosquito behaviour.

In diseases where a wide range of epidemic drivers are involved, much of which are generated by the behavior of individuals, the active involvement of the community previously mentioned, needs to be at the forefront of research [46], thereby, letting the models to have a greater impact in the daily fight against vectors mosquitoes. Currently, even though much limitations are still in place, there is no question about the good results that multiple modeling techniques have had, both nationally and internationally, in better elucidating and predicting disease outbreaks. Therefore, the involvement of authorities, the community and researchers of different disciplines is crucial to move modeling results from an academic exercise to the public health scenario [40].

4 Discussion

With recent increases in resources, both in technology and knowledge, there is a growing awareness in the value of including mathematical and statistical models into the public health arsenal of active available tools to fight the transmission of mosquito-borne diseases. By using the available data, the appropriate modeling techniques, and considering the limitations present in every model, public health decision makers, as part of an interdisciplinary work, have the possibility to design and implement tailored control and prevention strategies that take into consideration the unique characteristics of endemic regions [47].

In this article, we analyzed research conducted in Costa Rica, involving three different mathematical and statistical techniques. Taking into account the specific characteristics of the country, the amount of epidemiological data available, the multiple micro-climates present throughout the national territory and different demographic variables, the authors were able to implement models which allowed both: a better

understating of past transmission dynamics of Zika virus in the country and to retrospectively predict transmission risk of dengue.

However, before these models can be widely implemented in Costa Rica, channels between modelers and the different actors involved must first be established. Several limitations in the developments of the models need to also be targeted. As shown in [12], the wide range of approaches and variables available, where climate information has been one of the most common approach, the process of choosing the right set of data is a challenge that needs to be further investigated. In an effort to face this challenge, future work is in process to include other sources of information, such as those provided by satellite imaging (like the vegetation index), and social data, which can provide better results to account for the complexity in the transmission patterns of these diseases. This would allow that the interpretation, use and dissemination of the results can be properly implemented. Furthermore, channels developed need to be permanent open in order to implement continuing improves to the techniques as well as the information used in the development of the models.

It is also crucial to translate the results into actions involving the community, that would allow a more active and empowering participation of the key members that can disseminate the strategies needed to bring possible breeding areas to a minimum. This collaborative endeavor aims to add to the current actions being held in the country a perspective that has not been used before, and that has the potential of becoming a powerful strategy towards an increasing evidence-based public health actions.

Acknowledgements

The authors would like to thank the Research Center in Pure and Applied Mathematics and the School of Mathematics at Universidad de Costa Rica for their support during the preparation of this manuscript.

References

- [1] Gubler DJ. The Global Emergence/Resurgence of Arboviral Diseases as Public Health Problems. *Arch Med Res.* 2002;33(4):330-342.
- [2] Gould E, Pettersson J, Higgs S, Charrel R, de Lamballerie X. Emerging arboviruses: Why today? *One Health.* 2017;4:1-13.
- [3] World Health Organization. Global strategy for dengue prevention and control 2012-2020. 2012. Available in: <https://www.who.int/denguecontrol/9789241504034/en/>.
- [4] World Health Organization. World Malaria Report 2020. 2020. Available in: <https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2020>.
- [5] Yacaty S, Staples JE, Millot V, Cibrelus L, Ramon-Pardo P. Epidemiology of Chikungunya in the Americas. *J Infect Dis.* 2016;214(5):441-445.
- [6] Song BH, Yun SI, Woolley M, Lee YM. Zika virus: History, epidemiology, transmission, and clinical presentation. *J Neuroimmunol.* 2017;308:50-64.
- [7] Gutiérrez-Bugallo G, Piedra LA, Rodríguez M, Bisset JA, Lourenco de Oliveira R, Weaver SC, et al. Vector-borne transmission and evolution of Zika virus. *Nat Ecol Evol.* 2019;3:561-569.
- [8] World Health Organization. Global vector control response 2017-2030. 2017. Available in: <https://www.who.int/vector-control/publications/global-control-response/en/>.
- [9] Keeling MJ, Pejman R. *Modeling Infectious Diseases in Humans and Animals.* New Jersey. Princeton University Press. 2008:5-14.
- [10] Brauer F, Kribs C. *Dynamical Systems for Biological Modelling: An Introduction.* 1 ed. Florida: CRC Press. 2016:3-17.
- [11] Reiner RC, Perkins TA, Barker CM, Niu T, Chaves LF, Ellis AM et al. A systemic review of mathematical models of mosquito-borne pathogen transmission: 1970-2010. *J R Soc Interface.* 2013;10(81):1-13.

- [12] Johansson MA, Apfeldorf KM, Dobson S, Devita S, Buczak AL, Baugher B, et al. An open challenge to advance probabilistic forecasting for dengue epidemics. *Proc Natl Acad Sci USA*. 2019;1-7.
- [13] Massaro E, Kondor D, Ratti C. Assessing the interplay between human mobility and mosquito borne diseases in urban environments. *Sci Rep*. 2019;9(1):1-13.
- [14] Egger M, Johnson L, Althaus C, Schöni A, Salanti G, Low N, et al. Developing WHO guidelines: Time to formally include evidence from mathematical modelling studies. *F1000Res*. 2018;6:1-24.
- [15] Ministerio de Salud. Lineamientos Nacionales para el control del Dengue. 2010. Available in: http://www.solucionesss.com/descargas/G%20Leyes/LINEAMIENTOS_NACIONALES_PARA_EL_CONTROL_DEL_DENGUE.pdf. (Accessed September 27, 2019).
- [16] Ministerio de Salud Análisis de Situación de Salud, 2019. Available in: <https://www.ministeriodesalud.go.cr/index.php/vigilancia-de-la-salud/analisis-de-situacion-de-salud>. (Accessed September 25, 2019).
- [17] Feng Z, Velasco-Hernández J. Competitive exclusion in a vector-host model for the dengue fever. *J Math Biol*. 1997;35(5):523-544.
- [18] Esteba L, Vargas C. Analysis of a dengue disease transmission model. *Math Biosci*. 1998;150(2):131-151.
- [19] Sanchez F, Engman M, Harrington L, Castillo-Chavez C. Models for dengue transmission and control. *Contemp Math*. 2006;410:311—326.
- [20] Murillo D, Holechek SA, Murillo AL, Sanchez F, Castillo-Chavez C. Vertical transmission in a two-strain model of dengue fever. *Lett Biomath*. 2014;1(2):249-271.
- [21] Manore CA, Hickman KS, Xu S, Wearing HJ, Hyman JM. Comparing dengue and chikungunya emergence and endemic transmission in *A. aegypti* and *A. albopictus*. *J Theor Biol*. 2014;356:174-191.
- [22] Lee S, Castillo-Chavez C. The role of residence times in two-patch dengue transmission dynamics and optimal strategies. *J Theor Biol*. 2015;374:152-164.
- [23] Brauer F, Castillo-Chavez C, Mubayi A, Towers S. Some models for epidemics of vector-transmitted diseases. *Infect Dis Model*. 2016;1(1):79-87.
- [24] Olawoyin O, Kribs C. Effects of multiple transmission pathways on zika dynamics. *Infect Dis Model*. 2018;3:331-344.
- [25] Chen SC, Liao CM, Chio CP, Chou HH, You SH, Cheng YH. Lagged temperature effect with mosquito transmission potential explains dengue variability in southern Taiwan: insights from a statistical analysis. *Sci Total Environ*. 2010;408(19):4069-4075.
- [26] Pinto E, Coelho M, Oliver L, Massad E. The influence of climate variables on dengue in Singapore. *Int J Environ Health Res*. 2011;21(6):415-426.
- [27] Wang C, Jiang B, Fan J, Wang F, Liu Q. A study of the dengue epidemic and meteorological factors in Guangzhou, China, by using a zero-inflated poisson regression model. *Asia Pac J Public Health*. 2014;26(1):48-57.
- [28] Struchiner CJ, Rocklöv J, Wilder-Smith A, Massad, E. Increasing dengue incidence in Singapore over the past 40 years: population growth, climate and mobility. *PLoS One*. 2015;10(8):1-14.
- [29] Sanchez F, Barboza LA, Burton D, Cintrón-Arias A. Comparative analysis of dengue versus chikungunya outbreaks in Costa Rica. *Ricerche di Mat.*, 2018;7(1):163–17.
- [30] Sanchez F, Barboza L, Vásquez P. Parameter estimates of the 2016-2017 Zika outbreak in Costa Rica: An Approximate Bayesian Computation (ABC) approach. *Math Biosci Eng*. 2019;16(4):2738–2755.
- [31] Sisson SA, Fan Y, Beaumont M. Handbook of approximate Bayesian computation. Florida: Chapman and Hall/CRC. 2018;87-123.

- [32] Vásquez P, Loría A, Sanchez F, Barboza LA. Climate-driven Statistical Models as effective predictors of local Dengue incidence in Costa Rica: A Generalized Additive Model and Random Forest Approach. *Revista de Matemática Teoría y Aplicaciones Mathematical Modeling: A New Frontier*. 2020;27(1):1-21.
- [33] Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2 ed. Florida: Springer Series in Statistics. Springer; 2009;745.
- [34] Holsinger A. Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Inform*. 2016;3(2):119-131.
- [35] Guo P, Zhang Q, Xia J, Zhang Q, Luo G, Li Z et al. Developing a dengue forecast model using machine learning: A case study in China. *PLoS Negl Trop Dis*. 2017;11(10):1-22.
- [36] Ong J, Liu X, Rajarethinam J, Kok S, Liang S, Tang C et al. Mapping dengue risk in Singapore using Random Forest. *PLoS Negl Trop Dis*. 2018;12(6):1-12.
- [37] Stolerman L, Maia P, Kutz N. Forecasting dengue fever in Brazil: An assessment of climate conditions. *PLoS One*. 2018;14(8):1-16.
- [38] Kading RC, Golnar AJ, Hamer SA, Hamer GL. Advanced surveillance and preparedness to meet a new era of invasive vectors and emerging vector-borne diseases. *PLoS Negl Trop Dis*. 2018;12(10):
- [39] Moore CG. Interdisciplinary research in the ecology of vector-borne diseases: Opportunities and needs. *J of Vector Ecology*. 2008;33(2):218-224.
- [40] Garnett GP, Cousens S, Hallet TB, Steketee R, Walker N. Mathematical models in the evaluation of health programmes. *Lancet*. 2011;378:525-525.
- [41] Moss R, Hickson RI, Mc Vernon J, McCaw JM, Black J, Madden JR et al. Model-Informed Risk Assessment and Decision Making for an Emerging Infectious Disease in the Asia-Pacific Region. *PLoS Negl Trop Dis*. 2016;10(9):1-25.
- [42] Walters CE, Meslé MM, Hall, IA. Modelling the global spread of diseases: A review of current practice and capability. *Epidemics*. 2018;25:1-8.
- [43] Luz PM, Struchiner CJ, Galvani AP. Modeling Transmission Dynamics and Control of Vector-Borne Neglected Tropical Diseases. *PLoS Negl Trop Dis*. 2010;4(10):1-6.
- [44] Cohen JM, Le Menach A, Pothin E, Eisle TP, Gething PW, Eckhoff PA et al. Mapping multiple components of malaria risk for improved targeting of elimination interventions. *Malar J*. 2017;16(459):1-12.
- [45] Star L, Moghadas SM. The role of mathematical modeling in public health planning and decision making. *National Collaboration Center for Infectious Diseases*. 2010;2:1-6.
- [46] Winch PJ, Kendall C, Gubler D. Effectiveness of community participation in vector-borne disease control. *Health Policy Plan*. 1992;7(4):342-351.
- [47] Struchiner CJ, Luz PM, Codeco CT, Coelho FC, Massad, E. Current research issues in mosquito-borne diseases modeling. *Contemp Math.*, 2006;410:349-366.
- [48] Aswi A, Cramb S, Moraga P, Mengersen K. (2019). Bayesian spatial and spatio-temporal approaches to modelling dengue fever: A systematic review. *Epidemiology and Infection*, 147:E33. doi:10.1017/S0950268818002807.