

**LARGE-SCALE VECTOR-BORNE DISEASE
AGENT-BASED MODEL, WITH APPLICATION TO
CHIKUNGUNYA IN COLOMBIA**

GUIDO FELIPE CAMARGO ESPAÑA

Tesis presentada como requisito parcial para obtener el título de
DOCTOR OF PHILOSOPHY.
ELECTRICAL ENGINEERING

Director:
Hernando Díaz Morales, Ph. D.
Profesor Titular

UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE INGENIERÍA
BOGOTÁ D. C.
2015

Aprobada por la Facultad de INGENIERÍA,
en cumplimiento de los requisitos exigidos
para otorgar el título de:

**Doctor of Philosophy. — Electrical
Engineering**

Hernando Díaz Morales, Ph. D.
Director de la Tesis

Fernando Pío de la hoz Restrepo
Jurado

John Grefenstette
Jurado

Willem van Panhuis
Jurado

Universidad Nacional de Colombia
Bogotá D. C., Septiembre de 2015

RESUMEN

MODELO BASADO EN AGENTES DE GRANDE ESCALA PARA REPRESENTAR LA DINÁMICA DE ENFERMEDADES CONTAGIADAS POR MOSQUITOS, CASO CHIKUNGUNYA EN COLOMBIA

por

GUIDO FELIPE CAMARGO ESPAÑA

Doctor of Philosophy. en Electrical Engineering

UNIVERSIDAD NACIONAL DE COLOMBIA

Director: Hernando Díaz Morales, Ph. D.

En esta tesis se presenta el desarrollo de un modelo basado en agentes para representar la dinámica de enfermedades transmitidas por vectores. En específico, en Colombia para chikungunya, con posibles aplicaciones para dengue. Con el modelo se busca estimar la carga de casos de chikungunya en el país y evaluar el impacto de control vectorial para controlar su expansión en el país.

El chikungunya es una enfermedad transmisible por mosquitos del tipo aedes, en especial aedes aegypti y albopictus. Su sintomatología es similar al dengue, con una diferencia de dolor en las articulaciones que puede prolongarse por años, dependiendo del paciente. En Colombia, el mosquito aegypti se encuentra en mayor proporción que el albopictus. Este es considerado un mosquito residencial, debido a que se alimenta principalmente de humanos, consecuentemente, se encuentra en su mayoría en cercanías a los hogares.

El modelo representa estados de salud en humanos, basados en un modelo SEIR (Susceptible-Expuesto-Infectado-Recuperado). Mientras que para mosquitos se basa en la estructura SEI. La transmisión del virus en el modelo ocurre en lugares específicos, tales como, hogares, lugares de trabajo o colegios. Dentro de estos establecimientos, se encuentra un número determinado de mosquitos que depende de la cantidad de humanos y de la temperatura promedio anual. Estos mosquitos pueden infectarse con el virus con una probabilidad de infección determinada en el modelo, para luego transmitirlo a los humanos susceptibles que visitan ese lugar. Por su parte, cada humano tiene asignadas una cantidad de actividades dependiendo de su situación (estudiante, trabajado, ama de casa, etc), estas actividades se realizan con prioridad, mientras que hay una lista extra de actividades opcionales como viajar, visitar a un vecino, etc. De esta manera, los agentes pueden transmitir o infectarse con el virus.

Con el objetivo de representar una población cercana a la realidad, hubo la necesidad de desarrollar una población sintética que represente estadísticamente la población de Colombia. Además, que represente las actividades principales de cada agente, e.g. estudiar, trabajar, etc. La población sintética representa los 1122 municipios del país. Además, el modelo requiere de grillas de temperatura que fueron obtenidos de bases de datos de libre acceso. Finalmente, el modelo incluye una estimación de los viajes interdepartamentales, basado en datos de flujo entre aeropuertos del país.

El modelo fue sintonizado utilizando reportes de casos de chikungunya del 2014-2015. Usando un municipio como muestra y estimando el desempeño del modelo con municipios no sintonizados. Este modelo sintonizado fue utilizado para evaluar el impacto de las campañas de control en el municipio de Santa Marta (Magdalena), donde se registró un caso exitoso de prevención de la enfermedad utilizando control vectorial. Finalmente, el efecto del control vectorial fue estimado, simulando una epidemia en todo el país con diferentes estrategias vectoriales.

ABSTRACT

LARGE-SCALE VECTOR-BORNE DISEASE AGENT-BASED MODEL, WITH APPLICATION TO CHIKUNGUNYA IN COLOMBIA

by

GUIDO FELIPE CAMARGO ESPAÑA

Doctor of Philosophy. in Electrical Engineering

UNIVERSIDAD NACIONAL DE COLOMBIA

Advisor: Hernando Díaz Morales, Ph. D.

This document presents the development of a large-scale agent-based model to represent vector-borne disease transmission dynamics. Specifically the model represents the transmission of chikungunya in Colombia. Due to their similarities, the model can also be applied to simulate dengue epidemics. The aim of this model is to contribute to the knowledge of chikungunya, to reproduce realistic epidemics, and to quantify the impact of vector control programs to halt the spread of the disease.

Chikungunya is a disease transmitted by the *aedes* mosquitoes, particularly *aedes aegypti* and *Aedes albopictus*. Chikungunya symptoms are similar to dengue, but it is characterized by acute joint-pain that can last for years. In Colombia, the *aedes aegypti* is found in larger proportions than the *Aedes albopictus*. The *aedes aegypti* mainly obtains its food from humans, hence it is often considered a residential mosquito.

The model proposed in this thesis represents humans and mosquitoes. Humans are represented by agents whose health status can be classified in a S-E-I-R structure (Susceptible, Exposed, Infectious, Recovered). Whereas mosquitoes are represented by a homogeneous meta population model with the S-E-I compartments. In the model, the virus transmission occurs in specific locations such as households, workplaces, or schools. In each location, the number of mosquitoes are computed based on temperature and the human density. Mosquitoes and humans can transmit the infection to each other with specific probabilities determined in the model. Transmission occurs when an infectious agent visits a place with susceptible vectors, or when a susceptible agent visits a place with infectious mosquitoes. These visits are determined by each agent's activities that are assigned in a synthetic population, these activities include: household visits, school attendance, work attendance, and travel.

A synthetic population was developed to represent a realistic population of Colombia. The synthetic population represents the population of the 1122 municipalities and 33 departments of the country. Additionally, the synthetic population reproduces daily activities for each individual based on the census data. Human mobility was also represented in the model implementing a calibrated gravity model to represent air travel.

The model's parameters were calibrated to represent chikungunya dynamics reported in the Riohacha, Guajira. Some of the parameter values were obtained from the literature while others were adjusted using an optimization algorithm. This calibrated model was used to estimate the impact of vector control strategies in the city of Santa Marta, Magdalena. The control parameters in the model were modified to determine improvements to design optimal vector control strategies. Lastly, the model was simulated in a national-scale to evaluate the burden of the chikungunya with and without vector control strategies.

ACKNOWLEDGEMENTS

The author would like to express his acknowledgements to:

- To Wilbert, John, and Fernando for having accepted to be part of my committee.
- To Hernando, and Fredy for all their support and advise.
- To the Benter Foundation for their financial support.
- To Donald Burke for funding my research and for giving me the opportunity to work at the PHDL.
- To the Colombian National Institute of Health for sharing their valuable data and for their collaboration.
- To David Galloway for his constant technical help with FRED.
- To Molly and Sharon for all their patience and help (Even in English questions).
- To Jeannette for her helpful English proof reading of the document.
- To all the Public Health Dynamics Laboratory of the Graduate School of Public Health; it has been such a wonderful experience.
- To Megan for all her patience, English proof reading and love.

DEDICATORIA

To Megan, to whom my heart belongs,

and to Philip J. Fry.

Table of Contents

Table of Contents	x
List of Tables	xii
List of Figures	xiv
1 Introduction	1
1.1 Dengue and Chikungunya	3
1.2 Modeling of Infectious Diseases	6
1.3 Models' Calibration and Optimization	10
1.4 Document Content	14
2 Dengue and Chikungunya Epidemiology in Colombia	15
2.1 Dengue and Chikungunya Surveillance Reports	15
2.2 Department Classification Analysis Using Wavelet Analysis	17
2.3 Recent dengue and chikungunya data classification analysis	23
2.4 Summary	31
3 Synthetic Population of Colombia	33
3.1 Generation of the Persons and Households Attributes of the Synthetic Population	34
3.2 Geographical Location of Households	38
3.3 Schools' location and assignment	40
3.4 Workplaces location and assignment	41
4 Large-Scale Vector-borne Agent-Based Model	45
4.1 Vector Layer	45
4.2 Vector-borne transmission model	49
4.3 Model Verification	51
4.4 Model Assumptions and Limitations	54
4.5 Summary	55
5 Model Calibration to Represent Chikungunya Epidemics in Colombia	57
5.1 The Data	57
5.2 Calibration Strategy	57
5.3 Calibration Results	64
5.4 Comparison at National Scale Epidemic	72
5.5 Summary	81
6 Vector Control Strategies	83
6.1 Santa Marta Case	83
6.2 Vector Control Strategies in the Model	84
6.3 Calibration of Vector Control to Surveillance Data	85
6.4 Impact of Parameters' Values in the Vector Control Strategies	86
6.5 Vector Control Strategies at National Level	91

6.6	Summary	99
7	Conclusions and Future Work	101
7.1	Original Contributions	104
7.2	Future Work	105
	References	107

List of Tables

2.1	Municipalities with epidemics of Dengue and Chikungunya	24
2.2	Dengue clusters sorted by peak-time difference. Average climate variables are included as well as mosquito infestation indexes and population size. The R_0 was calculated using the initial growth rate proposed by Wallinga [1].	27
2.3	Chikungunya clusters sorted by peak-time difference. Average climate variables are included as well as mosquito infestation indexes and population size. The R_0 was calculated using the initial growth rate proposed by Wallinga [1].	27
2.4	Clusters of dengue and chikungunya	27
3.1	IPUMS household records included in the synthetic population.	34
3.2	Schools datasets from the Ministry of Education of Colombia	40
3.3	Schools grades equivalent in age	40
3.4	Workplace attributes from the National Census of 2005.	41
4.1	Agents' characteristics in FRED	45
4.2	Adjusted parameters of the gravity model.	48
4.3	Parameters of the vector-borne disease transmission model.	51
4.4	List of municipalities to explore the effect of temperature in the disease transmission model.	52
5.1	Model's parameter values adjusted to reported values in the literature. A closed population was assumed for the <i>aedes aegypti</i> , hence the mortality and birth rate are assigned to the same value. Latency and infectious period were obtained from dengue distributions.	59
5.2	Errors' estimations using different goodness-of-fit measures for the model's output presented in figure ???.	60
5.3	Parameter values fitted using the local optimization algorithm.	61
5.4	Parameter values fitted using global optimization algorithm.	62
5.5	Cumulative cases of chikungunya reported by the surveillance system and estimated by the model.	64
5.6	16 Municipalities with good performance. Gray background color indicates Riohacha (44001) which is the municipality used to calibrate the model. 15 municipalities' data are closely represented by the model without fitting. Light gray background indicates the four municipalities with good fit except for extreme values of the data.	67
5.7	Seven Municipalities with poor fit performance. Light Gray background indicates that the model predictions are larger than the incidence reports.	67
5.8	Three municipalities with initial good fit but sudden decay of cases in the data.	67
5.9	Cumulative diagnosed cases in the model and in the surveillance reports.	73
5.10	Departments with cumulative incidence > 10 people per 10,000 inhabitants	73
5.11	Errors of prediction of cumulative incidence rates (per 10000 inhabitants)and the list of covariates by department.	79
6.1	Parameter values of the vector control programs adjusted from the literature and using the <i>Simplex Nelder-Mead</i> optimization algorithm.	85

6.2	Total count of cases reported by the surveillance system in Santa Marta and predicted by the model.	88
6.3	Multiple scenarios of vector control. Error = (Model - Data)/Data * 100.	91

List of Figures

1.1	Reconstruction of dengue incidence in Colombia since 1970 to 2012	4
2.1	Dengue weekly reports by department. Each pixel in X axis represents a week and each pixel in the Y axis represents a department. The number of cases per 10,000 persons per week are represented by a color code. Representing high incidence with red and low incidence with blue. Black represents non-reported weeks.	16
2.2	Dengue and Chikungunya weekly incidence rates (cases per 10,000 inhabitants) by municipality. Left graphic corresponds to dengue. Right graphic corresponds to chikungunya.	16
2.3	Wavelet function scaling and translation properties. Left figure corresponds to scaling the Morlet wavelet by a factor of 0.5 (magenta) and 2 (green). At the right, the wavelet is translated by a factor of 2.	18
2.4	Wavelet transform of Norte de Santander's weekly incidence reports. The hot colors represent high power spectrum.	19
2.5	Wavelet transform of Santander's weekly incidence reports. The hot colors represent high power spectrum.	19
2.6	Wavelet transform of Barranquilla's weekly incidence reports. The hot colors represent high power spectrum.	19
2.7	Wavelet transform of Bogotá's weekly incidence reports. The hot colors represent high power spectrum.	19
2.8	Department clusters in a dendrogram. A 'ward' distance measure was used. The threshold was set to 14 in order to obtain four clusters.	21
2.9	Wavelet spectra in cluster 1.	22
2.10	Wavelet spectra in cluster 2.	22
2.11	Wavelet spectra in cluster 3.	22
2.12	Wavelet spectra in cluster 4.	22
2.13	Department clusters based on wavelet spectrum. Each color represents a different cluster that agrees with the color code in figure 2.8.	23
2.14	Examples of two municipalities when the selection criteria has been applied.	24
2.15	Municipalities' peak time of chikungunya epidemic	25
2.16	Municipalities' peak time of dengue epidemic	25
2.17	Clusters of dengue and chikungunya epidemics by municipality. Top graph shows chikungunya clusters. Bottom graph shows dengue clusters.	28
2.18	Multivariate Analysis of Variance for chikungunya clusters. Precipitation, peak time, and reproductive number variate within the difference clusters.	29
2.19	Multivariate Analysis of Variance for dengue clusters. Temperature, house index, Breutau index, peak time, and reproductive number variate within the difference clusters.	30
3.1	Model's inputs and outputs. The synthetic population and places locations are inputs of a facsimile ABM.	33
3.2	Population pyramid of Colombia in 2010. Blue bars represent the males proportion in each age group. Beige bars represent the females proportion in each age group.	35
3.3	Regression analysis of the synthetic population estimates compared to the census data for each municipality.	36

3.4	Histogram of χ^2 values for the age-distribution of the 1122 municipalities of Colombia	37
3.5	Examples of age distributions of the synthetic population compared to the census. The left graph shows a municipality with an acceptable χ^2 . The right graph shows a municipality with a high χ^2 that represents a poor fitting.	37
3.6	Regression analysis of the synthetic population estimates of students, workers, and females compared to the census, by municipality.	37
3.7	The territory of Colombia divided in $1Km^2$ cells. Each cell contains information about the municipality, department, urban status, and population density.	38
3.8	Graphical representation of the population density for the synthetic population (Left) and WorldPop (Right).	39
3.9	Spatial correlation between WorldPop and the synthetic population by department.	39
3.10	Synthetic school capacities by grade and total capacity by school.	41
3.11	Graphical representation of the workplace assignment based on commuting distances.	42
3.12	Regression analysis of the number of workplaces by each capacity category. From left to right: small, medium, large.	43
3.13	Histogram of the number of workplaces by χ^2 goodness-of-fit measure. 70 % of the workplaces are drawn from the same distribution than the census reports. 20 % are highly dissimilar.	43
4.1	Mosquito density based on annual temperature values. Left, annual temperature in Celsius for a grid of 30 arc seconds of resolution, obtained from WorldClim project. Right, mosquito resultant grid of 30 arc seconds of resolution based on temperature and population density, calculated using equation 4.1.	46
4.2	Passengers flow between city airports. The green line is shown to represent the linear trend of the data.	48
4.3	Gravity model comparison with data. The coefficient of determination is $R^2 = 0,46$	49
4.4	Model Structure: SEIR for Humans and SEI for Mosquitoes.	50
4.5	Susceptible proportion for each dengue serotype in the departments of Bogotá and Norte de Santander. Computations are made on historic data reports and estimations of under-reporting ratios.	52
4.6	Impact of temperature in the model. Four municipalities with different climate conditions were simulated. The temperature influences the model's response as assumed, i.e. high temperatures implies high incidence.	53
4.7	Impact of probabilities of contagion in the model's response. Small values of probability of contagion produce unsustainable epidemics.	54
4.8	Impact of previous exposure to the virus in the model's response. Previous immunity to the virus reduce the susceptible population, hence the impact of the epidemic is reduced as well. The age-structure incidence of the disease moves toward the young population when the population has been previously exposed to the virus.	55
5.1	Weekly incidence chikungunya reports in Riohacha, Guajira. Obtained from the National Institute of Health.	58
5.2	Weekly incidence chikungunya reports filtered to remove extreme points while maintaining the cumulative number of cases.	58
5.3	Two simulations of the model contrasted with the weekly incidence reports of chikungunya in Riohacha. Left, the model's response is characterized by a wide curve that fits the peak of the data but misrepresent the initial growth rate. Right, the model's response is characterized by a narrow curve with that fits the initial growth rate of the data but overestimates the peak.	60
5.4	Model calibrated to Riohacha reports using two parameters and two different goodness of fit measures. In both cases the Nelder-mead optimization algorithm was used with two parameters and a maximum number of iterations of 100.	61
5.5	Model calibrated to Riohacha reports using three parameters and two different goodness of fit measures. In both cases the Nelder-mead optimization algorithm was used with three parameters and a maximum number of iterations of 100.	62

5.7	Calibration convergence points from the global strategy 2. Random initial values to calibrate three parameters. The circle denotes the minimum point found by the optimization algorithm. The color represents the error between the model and the reports. 40 initial points were tested with maximum 100 iterations for each calibration step.	63
5.6	Calibration convergence points from the global strategy 1. Random initial values to calibrate two parameters. The circle denotes the minimum point found by the optimization algorithm. 40 initial points were tested with maximum 100 iterations of each calibration step.	63
5.8	Model fitted to chikungunya weekly reports in Riohacha (Guajira). The top figure shows the mean response of 100 simulations of the model compared to the filtered data. The bottom figure shows the model compared to the original data from the National Institute of Health in Colombia. In both graphs, the bars represent the standard deviation.	65
5.9	Mean attack rate of 100 simulations of the model fitted to chikungunya data. 60 % of the population is infected at some point in the epidemic.	66
5.10	Exposed and Infectious variables of 100 simulations of the model fitted to chikungunya data. The green curve represents exposed individuals per week, whereas the red curve represents infectious individuals per week. There is a peak of 120000 exposed individuals and 8000 infectious individuals.	66
5.11	Municipalities that showed good fit performance when simulated with the calibrated parameter values to the Riohacha epidemic curve.	68
5.12	Municipalities that showed good fit performance when simulated with the calibrated parameter values to the Riohacha epidemic curve.	69
5.13	Municipalities that showed good fit performance when simulated with the calibrated parameter values to the Riohacha epidemic curve. The reports present larger impact than the model's predictions.	70
5.14	Municipalities that showed good fit performance when simulated with the calibrated parameter values to the Riohacha epidemic curve. The model's predictions present larger impact than the reports.	70
5.15	Three municipalities that showed initial good fit performance and sudden decay of cases in the middle of the epidemic, compared to simulations of the calibrated parameter values to the Riohacha epidemic curve.	71
5.16	Municipality of Turbaco, Bolívar. The blue color represents the location of the municipality. The epidemic of chikungunya initiated in this municipality.	72
5.17	Weekly incidence of Chikunguya cases in the model and surveillance reports.	73
5.18	Regions of Colombia. 1. Andina, 2. Amazonía, 3. Caribe, 4. Insular, 5. Orinoquía, 6. Pacífico.	74
5.19	Error of reports (Individuals per 10000 inhabitants) - Model predictions (Individuals per 10000 inhabitants) by region.	75
5.20	Model and reports incidence curves by region in Colombia.	76
5.21	Error of reports (Individuals per 10000 inhabitants) - Model predictions (Individuals per 10000 inhabitants) by region.	77
5.22	Error of reports (Individuals per 10000 inhabitants) - Model predictions (Individuals per 10000 inhabitants) by region.	77
5.23	Residuals of regression analysis of the model's prediction errors.	78
5.24	Incidence cases by municipality reported by the surveillance system (Left) and predicted by the model (Right).	80
5.25	Error of reports (Individuals per 10000 inhabitants)- Model predictions (Individuals per 10000 inhabitants) by municipality.	80
6.1	100 simulations of the model for Santa Marta without implementing vector control programs contrasted to the surveillance reports. The model fits the initial growth rate but the data shows a sudden decay that the model misrepresents.	83

6.2	100 simulations of the model for Santa Marta implementing vector control programs contrasted with the surveillance reports. The model fits the data better compared to the simulations without vector control.	85
6.3	Model simulation of Santa Marta with and without vector control strategies. The red shade represents the cases overestimated by the model without including vector control. The blue shade corresponds to the additional case counts estimated by the model with vector control.	86
6.4	Mosquito population when applying vector control. The red dashed line represents the introduction time of the interventions.	87
6.5	Estimated Attack Rate of the chikungunya epidemic in Santa Marta with and without vector control.	87
6.6	Variation of the threshold of cases to start the intervention programs. The figure at the left represents the percentage of reduction in symptomatic cases when varying the threshold to initiate vector control interventions. The figure at the right shows the different curves produced with the multiple threshold values.	88
6.7	Variation of the recruitment rate of the intervention. The figure at the left represents the percentage of reduction in symptomatic cases when varying the recruitment rate. The figure at the right shows the various curves produced with the multiple recruitment rate values.	89
6.8	Variation of the recruitment rate of the intervention. The figure at the left represents the percentage of reduction in symptomatic cases when varying the participation rate. The figure at the right shows the different curves produced with the multiple participation rate values.	90
6.9	Variation of the recruitment rate of the intervention. The figure at the left represents the percentage of reduction in symptomatic cases when varying the efficacy of the interventions. The figure at the right shows the different curves produced with the multiple efficacy values.	90
6.10	Variation of the stop time of the intervention programs. The figure at the left represents the percentage of reduction in symptomatic cases when varying the stop time. The figure at the right shows the different curves produced with the multiple parameter values.	91
6.11	Cumulative symptomatic incidence in Colombia.	92
6.12	Symptomatic Incidence curves of different control scenarios in Colombia.	92
6.13	Model and reports incidence curves for the Andean region in Colombia. Top left figure compares the model and the data in the data period. Top right figure displays the data and the model for the whole simulation period. Bottom figure presents the model compared to the data delayed 9 weeks.	94
6.14	Model and reports incidence curves for the Amazon region in Colombia. Top left figure compares the model and the data in the data period. Top right figure displays the data and the model for the whole simulation period. Bottom figure presents the model compared to the data delayed 9 weeks.	95
6.15	Model and reports incidence curves for the Caribbean region in Colombia. Top left figure compares the model and the data in the data period. Top right figure displays the data and the model for the whole simulation period. Bottom figure presents the model compared to the data delayed 9 weeks.	96
6.16	Model and reports incidence curves for the Insular region in Colombia. Top left figure compares the model and the data in the data period. Top right figure displays the data and the model for the whole simulation period.	96
6.17	Model and reports incidence curves for the Orinoquía region in Colombia. Top left figure compares the model and the data in the data period. Top right figure displays the data and the model for the whole simulation period.	97
6.18	Model and reports incidence curves for the Pacific region in Colombia. Top left figure compares the model and the data in the data period. Top right figure displays the data and the model for the whole simulation period.	97
6.19	Heatmaps of incidence rates by municipalities for seven different scenarios.	98

Chapter 1

Introduction

Vector-borne diseases have a serious impact around the world with approximately 3 billion people living in risk zones [2]. Dengue virus is known as one of the most critical vector-borne diseases in the world. It is estimated that 390 million cases occur each year. While 96 millions cases are reported, the remainder are attributed to asymptomatic presentations of the disease. According to Bhatt *et al.*, the geographic distribution of cases around the world can be described as: Africa (48.8 million) Asia (204.4 millions), the Americas (40.5 millions), and Oceania (550.000) [3]. However, these numbers could be higher than predicted, as the World Health Organization (WHO) stated in the 2012 dengue report [4].

Dengue infection can lead to two different types of clinical presentations: dengue fever (DF) and severe dengue (SD). While DF usually produces febrile symptoms, SD can produce severe symptoms, including plasma leakage, that could lead to death [5]. The presentations of SD are often related to secondary infections, although there are additional factors that determine the SD occurrence, for instance: gender, race, chronic diseases or age [6]. Although SD constitutes a risk to the patient's life, death can be prevented with the appropriate treatment [5].

Dengue presents itself in four varieties or serotypes (DENV1-2-3-4). Each serotype generates permanent immunity to itself but only temporary immunity to the other serotypes. Consequently, the same person can be exposed to multiple infections. Also, the number of cases of SD are often related with secondary infections, although there are additional factors that determine the SD occurrence, for instance: gender, race, chronic diseases or age [5].

Another major vector-borne disease, chikungunya, is present in Africa, Asia, Indian Ocean and, recently, in the Americas, where the suspected cases are more than seven hundred thousands as for October of 2014 [7, 8]. In contrast with dengue, chikungunya disease occurs by one serotype of the virus (CHIKV).

Chikungunya disease is characterized by an incubation period that varies from 1 to 12 days, followed by an acute illness, and a late stage of illness. The acute illness stage can last from 4 to 7 days and it is characterized by fever, polyarthralgia, headache, among others. After the acute stage, some symptoms such as arthralgia can remain present in a late stage of illness having a life quality reduction in the individual. Owing to the clinical representations, the asymptomatic rate is considered to be smaller than in the case of dengue [8].

The transmission of chikungunya and dengue to humans is accredited to the *Aedes* mosquitoes. The main vector is the mosquito *aedes aegypti*; however, the *aedes albopictus* is also able to transmit the virus to humans. *Aedes aegypti* lives mostly in warm regions with temperatures above 18°C. The *aedes aegypti* is mostly found in urban habitats, where non mature stages of the mosquito can develop in man-made water reservoirs such as barrels, trash cans or water containers. Due to the fact that the *aedes aegypti* lives near humans, human blood is considered its main food source [3].

The concern about an introduction of the chikungunya virus to developed countries has increased, considering that *aedes albopictus* can survive in lower temperatures than the *aedes aegypti* [8]. Chikungunya was introduced for the first time in the Americas in 2013, thus there is a large susceptible population that could be infected, since there is no previous exposure to the virus. Hence, the impact

of the chikungunya epidemic could be massive across the continent [7].

Currently, there is no vaccine available for chikungunya. In contrast, there are multiple vaccine candidates for dengue. At this moment none of the vaccines have been implemented in a population to control the epidemic. Hence, control measures are focused on vector control. These measures have shown little impact when irregularly applied and inefficiently planned [9]. The possibility of the implementation of a dengue vaccine, would shift the horizon of the disease [10]. In the case of the vaccine's success and approval, it would be necessary to design optimal distribution strategies that guarantee the eradication of the disease. In order to accomplish that, WHO suggests the use of mathematical or computational models to design control programs with the inclusion of vaccines and vector control as a combined strategy [4].

Whereas multiple models have been constructed to represent dengue dynamics, just a few have been proposed for chikungunya. These models can be classified as: compartmental models, based on Ordinary Differential Equations (ODE); stochastic models and Agent-Based Models (ABM). In the case of dengue, they can also be classified by the number of serotypes included in the model. For instance, compartmental models of one serotype of dengue have been proposed by multiple authors. Regardless of the simplicity of these models, they can be employed to do useful analysis such as determination of the reproductive number of the disease [11, 12, 13].

Furthermore, two or more serotypes have been included in dengue compartmental models [14, 15, 16, 17, 18, 19, 20, 21]. In these models it is possible to represent the effects of multiple infections by different serotypes. For example, Esteva *et al.* [16] formulated a compartmental model where the mosquito and the human population can be infected by two serotypes of the virus.

Some mathematical models of chikungunya have been developed to compute the reproduction number for past epidemics [22, 23]. Also, some models evaluated the implementation of vector control strategies and proposed thresholds to ensure the success of these programs [23, 24]. Additionally, the spread of the disease to different regions has been studied for small regions to large populations such as the European Union or the United States of America [25, 26, 27].

Although ODE models exhibit a mathematical formality that allow formal analysis, it is not possible to represent the heterogeneity of a population or the individual characteristics. Meanwhile, ABM allow the inclusion of individual characteristics and behaviors. In particular, Otero *et al.* created a model for dengue in Buenos Aires, Argentina where the human population is represented as an ABM, while the mosquitoes are represented by grids. Furthermore, Chao *et al.* assessed the impact of a dengue vaccine using an ABM for a rural-area in Thailand [28]. Finally, in the case of chikungunya, one model was proposed by Dommar *et al.* [29] to evaluate the impact of human mobility in the spread of the epidemic.

In Colombia, no model has been proposed for chikungunya, whereas some have studied dengue dynamics. For instance, Castañeda-Orjuela *et al.* assessed the burden of dengue disease in Colombia using a two-serotypes compartmental model based on ODE [30]. The model was tuned to represent the overall dynamics of dengue in the country from 1997 to 2011. Torres *et al.* applied *fuzzy* techniques to represent adequate weekly dengue reports in Colombia [31]. Velez *et al.* [32] used the model developed by Focks [33] to study the dengue dynamics with climate effects in the municipality of Bello, Antioquia. On the other hand, Padmanabha *et al.* [34] built an ABM of dengue dynamics for a neighborhood in the city of Armenia, Quindío., where the authors represent the population characteristics by means of a synthesized population of the area.

These approximations of dengue Agent-Based Models are focused on small areas, and this constitutes a limitation in the possible recommendations or conclusions. Due to the similarity of the transmission of chikungunya and dengue viruses, the construction of an ABM would be suitable to represent both diseases. Consequently, the aim of this project is to develop a large-scale vector-borne disease ABM, specifically to represent the transmission of chikungunya within the entire population of Colombia, with application to dengue dynamics in the country. Hence, the model would allow testing the design of control strategies that are applicable on a national basis.

The transmission dynamics of dengue and chikungunya occur within similar conditions. However, dengue dynamics is considerably more complex than chikungunya. Nevertheless, since the vector for them is the same, several componentes of the model are common to both diseases. Dengue has been present in the Americas for some decades, whereas chikungunya was introduced in the region in 2013.

Consequently, the amount of data concerning dengue is significantly more abundant than the data available for chikungunya. Thus, a model for chikungunya would share significant components that are also required for a model of dengue transmission dynamics. Therefore, the construction of a model for chikungunya would be a fundamental step for developing a dengue model. This is the reason why there will be a considerable amount of space devoted to the discussion of dengue epidemiology.

1.1 Dengue and Chikungunya

Progression of Dengue

Although, the origin of the dengue virus has not been determined with confidence, it is considered that it developed one thousand years ago in Asia or Africa. The dengue disease has been formally described since 1979-80 [6, 35]. In Central America and India, the virus appeared in the 17th century, whereas in the United States of America, dengue was first registered in the 18th century [35]. SD cases have been reported since 1953 [35].

According to the WHO, dengue virus cases have increased more than 30 fold in the last five decades [4]. This quick spread has been attributed to climate changes and fast unplanned urbanization. Furthermore, population migration is an additional factor that contributes to the geographical spread of the virus all around the world [4, 6].

This growth in the transmission of dengue started during World War II, probably due to the massive transportation of potential water containers such as tires. Also, the weather conditions in Southeast Asia contributed to the proliferation of the virus. After the war, governments in the Americas implemented an international eradication plan that eliminated most of the *aedes aegypti* in the region. The success of these programs resulted in a significant reduction of the disease. However in most of the countries in Central and South America the campaigns were abandoned because of the lack of resources or for political reasons. As a consequence, the region experienced a re-infestation of the virus in the 1970's. Since then the dengue virus has been spreading throughout the region. Currently, the four serotypes of the virus are circulating through most countries in Central and South America [35, 6].

Dengue in the Americas

In the Americas, the presence of the disease can be divided into four different stages, according to Dick *et al.*[9]. In 1600 the virus was first introduced in the region producing outbreaks in multiple countries. Then, the above mentioned eradication plan of the *aedes aegypti* took place from 1947 to 1970, and as a consequence most of the countries were free of dengue during that period. However, an *aedes aegypti* reinfestation occurred around 1971 and since then the virus has been present mostly in South and Central America. Recently, starting in 2000, dengue dynamics has been characterized by a spread of the four serotypes through most of the countries in South and Central America.

The first reports of a dengue-like disease in the Americas occurred in Martinica, Guadelupe and Panama. In South-America multiple cases were reported during the 19th century. In Peru, around 50 thousand cases were reported in 1818. In the 19th century and the early the 20th century, several clinical reports similar to dengue occurred in Cuba, Brazil, Colombia, Venezuela, Argentina and even in the United States of America. These cases have been attributed to dengue owing to their clinical presentations, however they could be confused with similar disease such as Chikungunya or yellow fever, taking into account that the dengue virus was not isolated until 1943.

The eradication plan started as an initiative from William Gorgas in Cuba aiming to eliminate the yellow fever vector. This initiative was adopted in Brazil, where the Rockefeller foundation and the Brazilian Government worked together to completely eliminate the mosquito from the region. As a consequence, in 1947, the Pan American Health Organization (PAHO) approved the *aedes aegypti* eradication plan for all the affected countries. Due to the countries' efforts, the mosquito was apparently eradicated from the region. However, owing to the lack of political interest in sustaining these campaigns, the vector control was significantly reduced after 1962 [36, 37, 9].

The decline in the vector control campaigns led to a reinfestation of the virus all around the region. Between 1971-1977, serotypes DENV2-3 were reported in Puerto Rico and Colombia. An epidemic was reported between 1977-1980, where DENV1 was the predominant serotype. In the 1980's the incidence of dengue increased significantly, DHF epidemics were reported in the region, mostly in Cuba and Venezuela. In 1981, DENV4 was introduced in the region generating more DHF cases. As a result, in 1996 PAHO proposed an intensification of the *aedes aegypti* control, however it was not implemented broadly. As a consequence, the virus continued to spread in most South and Central American countries.

Between 2000-2010, dengue reports displayed a drastic increase. In the Americas, two outbreaks were reported in 2001-2002 and 2010. Cases were reported in multiple countries such as Ecuador, Paraguay, Costa Rica, Brazil, Colombia and Peru. For instance, in 2001, Peru experienced its biggest dengue outbreak reported to date with approximately 24,000 cases. The biggest outbreak in the whole region occurred in 2010 with more than 1700 million cases. During the same year, Colombia contributed the highest number of cases in the continent.

Within Colombia, the presence of dengue is similar to the rest of the countries in the Americas, mostly in South and Central America. During the 1900-1952 period, dengue was present in some regions of the country. The virus was partially eradicated in 1952-1970. However, the virus has been present in the country since 1971 due to the *aedes aegypti* reinfestation that took place throughout most of the countries in the Americas [38].

Figure 1.1 shows all dengue outbreaks reported in the country. Outbreaks reported in 1971, 1975 and 1977 were caused by the DENV2, DENV3 and DENV1, respectively [38, 9]. Those outbreaks numbered over 400,000 cases a year. After those outbreaks, dengue has been present as an endemic disease. The DENV4 serotype was first reported in 1982 in the country. Also, in 1985 it was registered the first DHF case and multiple serotype outbreaks have occurred periodically since [37, 39].

Currently, the National Health Institute (INS) is in charge of the epidemiological surveillance in the country, through the National System of Public Health Surveillance (SIVIGILA). Dengue cases are reported weekly by municipalities. In 2010, the largest epidemic occurred since the reporting of dengue cases became mandatory; approximately 150,000 cases of DF and 5,000 cases of DHF were

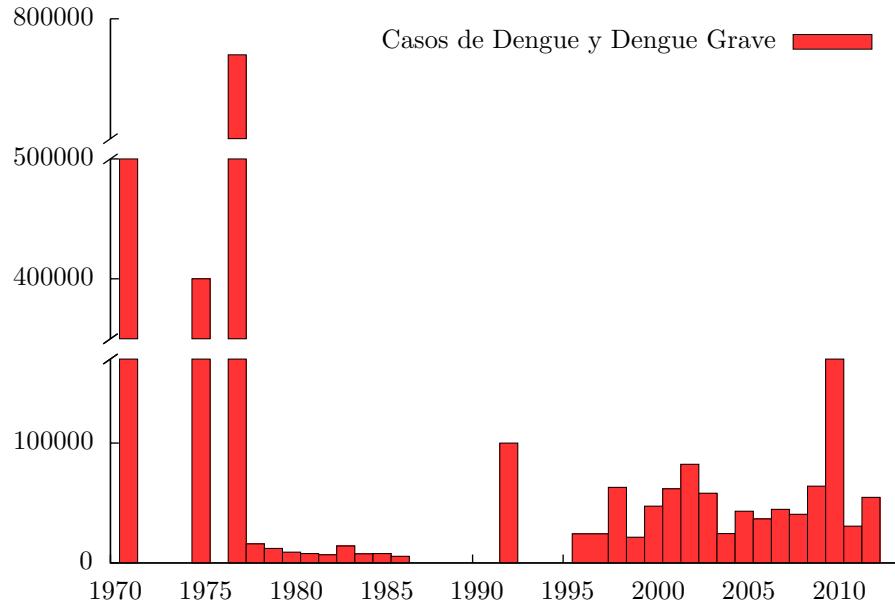


Figure 1.1: Reconstruction of dengue incidence in Colombia since 1970 to 2012

reported during that year [40].

Progression of Chikungunya

The first reported outbreak of chikungunya took place in Africa in the 1950s. The virus was first isolated in 1952 in Tanzania. Since then, outbreaks have been reported most often in some developing countries of Asia and Africa. Moreover, in 2005 an outbreak was reported in the Indian Ocean in La Réunion Island, France. Although no autochthonous transmission has been documented for developed countries, some imported cases have been recorded in some regions in Europe [8].

Concern is generating about the impact of the outbreak in tropical countries in the Americas and the possible introduction to some regions of the United States. In 2013 an outbreak occurred in Saint Martin Island and, owing to the frequency of travels between the Americas, the disease has spread to the Caribbean, Central America, South America and even North America.

In October of 2015, more than 1'500,000 suspected and more than 40,000 confirmed cases had been reported across the Americas. In Colombia, more than 425,000 suspected cases and almost 2,000 confirmed cases had been recorded as of October of 2015. The impact of the chikungunya in the tropical countries of the Americas could be significant, because of the climate and social conditions that allow for the transmission of the virus [41, 42].

Vector Control

Vector control strategies aiming to stop the transmission of dengue and chikungunya disease are focused on the elimination of the *aedes* mosquitoes. These methods are focused on: decreasing the number of water containers, killing immature mosquitoes and eliminating the adult mosquito. Elimination of the adult mosquito is performed mainly during outbreak seasons [35, 43]. Unfortunately, these interventions are often short term efforts that are seldom evaluated consistently by the public health institutions. According to the World Health Organization (WHO), any interventions should be sustained with constant evaluation and monitoring [4].

According to WHO [44], vector strategies to control chikungunya or dengue can be divided into: preventing man-mosquito bites and controlling the vector population. The mosquito bites can be avoided using long-sleeve clothes, applying repellents or using mosquito nets or screens. The vector-control measures include indoor spraying, outdoor space spraying, thermal fogging and larva control measures. The larva control programs basically consist on eliminating any possible source of breeding sites. These actions can be performed by the inhabitants of a household, eliminating trash that can hold water, emptying water containers or covering them.

WHO recommends that the vector control should be integrated with vector surveillance and include social mobilization besides the use of pesticides. The organization also defines the key elements of Integrate Vector Management as: advocacy, social mobilization and legislation, collaboration with the health sector, integrated approach, evidence-based decision-making, and capacity-building. An issue to effectively implement integrated vector management programs is that its benefits are not visible to the community. Hence, measuring or providing with estimates of reduction of cases owing to the community efforts, could incentive the community to participate in vector control programs.

In regards to the introduction of the chikungunya in the Americas, PAHO manifests as well that the participation of the community is vital to halt the spread of the virus in new areas [45]. PAHO recommends to prioritize regions where dengue has been present with high impact, since the chikungunya virus is expected to appear in those regions. PAHO recommends that the interventions are made mainly in the immature stages of the mosquito. In order to stop the epidemic in new areas, first confirmed cases should be communicated to the Integrated Vector Management programs in a local and national level. The programs should focus on the areas where the initial cases are detected to prevent the spread, if these programs fails, PAHO recommends to apply an intensified vector control program in a large area.

Vaccines

Historically, vaccination programs have proved to be a useful way to control epidemics. In particular, vaccination has decreased notably the burden of contagious diseases such as smallpox, polio, measles, pertussis, among others [46]. In the case of dengue, there is no vaccine that protects the population against the four serotypes of the virus. Nonetheless, numerous vaccine candidates have been developed around the globe [47, 10, 48, 49].

The development of a dengue vaccine represents a big challenge due to the complex immunity response of the four heterologous dengue serotypes. Nonetheless, tetravalent dengue vaccine candidates are currently in different stages of development. In fact, six vaccine candidates are in different phases of clinical development and one of them is presently in phase three of clinical trials [47].

Moreover, there are several vaccine candidates in pre-clinical stage of development that are based on different technological approaches. Pre-clinical candidates show potential advantages that constitute an improvement step in the continuous generation of dengue vaccines [48].

The most advanced vaccine candidate is the Sanofi Pasteur tetravalent dengue vaccine that has shown to be safe for humans [47]. In the phase 2b of trials, the vaccine did not meet the efficacy goals, particularly for the serotype 2. Nevertheless, recently Sanofi Pasteur published successful results of the phase 3 trials conducted in Asia-Pacific. These trials involved more than 10,000 children between the ages of 2 and 14 [50]. However, the vaccine candidate showed less efficacy against the serotype 2 compared to the other three. Another drawback of the vaccine candidate is the vaccination schedule that is longer than other current commercial vaccines.

In general, the Sanofi Pasteur vaccine candidate has shown significant advances in the dengue vaccine development. This could lead to an early commercialization and implementation of the vaccine in endemic countries that would aid in accomplishing the WHO goals for 2020 [4, 50]. Nonetheless, it is important to consider multiple factors in order to design successful vaccination programs that contribute to the eradication of dengue around the world [46].

Concerning chikungunya, currently there are no available vaccines to prevent the disease. However, multiple efforts have been made in the subject. Some vaccine candidates have been developed using different technologies. Unfortunately, all the candidates are in the early stages of development, without any human trials, hence they are not likely to be implemented in the population in the short term [8].

1.2 Modeling of Infectious Diseases

Epidemiological models can be classified as static or dynamic. Static models exclude changes in time. Dynamic models include the variation of the system over time. Dynamic epidemic models are represented mainly by two different approaches: Differential Equation Based models or Computational models. The appropriate model is implemented depending on the characteristics to be represented.

Mathematical Models

Mathematical models in epidemiology are usually described by Ordinary Differential Equations (ODE) that represent the epidemiological state in a population by means of compartments and the rates of change among these compartments. In the most common structure found in these models, a homogeneous population is split in three epidemic states: Susceptible, Infectious, Recovered (SIR). In the case of vector borne diseases, the epidemiological state of the vectors is also represented.

Regarding dengue models, complexity of the ODE models could vary depending on the characteristics included in the model. For instance, there are simple models that represent the dengue dynamics in a population in risk of one serotype of the virus [13, 12]. Others include two serotype to increase the complexity [16, 14, 15, 51]. Four-serotype models are used to describe more specific dynamics for regions where all the serotypes circulate [17, 18, 19, 20, 21, 52, 53, 54].

Moreover, age-structured models can be represented using ODE models, allowing the inclusion of the risk difference by age. Supriatna *et al* represented that difference assigning a greater contagion probability to the children population [51]. More complex models include climate impact on transmission [55]. Climate and socioeconomic conditions influence the *aedes aegypti* life cycle, affecting dengue

and chikungunya transmission. Chen and Hsieh formulated a mathematical model that includes the impact of climate in a mosquito model with temperature dependent parameters [56].

Climate factors have been represented using statistical models. Fuller *et al.* represents dengue cases in Costa Rica where temperature observations and vegetation index are included [57]. Also, Wu *et al.* found out that humidity and temperature are statistically related to dengue incidence [58]. More complex characteristics can be represented in ODE models such as GIS (Geographic Information Systems) to represent socioeconomic factors.

On the other hand, multiple models have included the vaccination effects owing to the expectation of a dengue tetravalent-vaccine [59, 52, 53]. These models can assess the effectiveness of different vaccination strategies, the economic impact of the vaccine and the cost-effectiveness ratio.

Few mathematical models have been developed to represent chikungunya transmission. Due to the possibility of spread in the recent years, one particular modeling interest is the evaluation of the impact of travelers on the spread of the epidemic across a region. Other models have evaluated the possible invasions of chikungunya in fully susceptible populations. The majority of these models are focused on the 2005 outbreak in La Réunion Island.

For instance, Moulay *et al.* developed a metapopulation model that represents chikungunya transmission in the Réunion Island, specifically the 2005-2006 epidemic. For that purpose, a network was constructed using population density grids and road networks. Then realistic human and mosquito mobility patterns were included in the model. Each node of the network is characterized by a compartmental model which includes individual import/export. This model shows the importance of human and mosquito mobility in the spread of the chikungunya virus [25].

Seyler *et al.* used a monte carlo model to estimate the risk of dengue and chikungunya introduction in European union by travelers from endemic countries. This model used real data to tune the model regarding flights and duration of the stay in the European union. Of all the countries, Italy showed a higher risk of chikungunya introduction. Nevertheless, the lack of incidence data from some endemic countries limits the capabilities of the model analysis [27].

Ruiz-Moreno *et al.* assessed the probability of occurrence of an outbreak of chikungunya across the United States of America using monte carlo simulations. The authors used this model to study the possible scenarios of chikungunya outbreaks in different regions of the country at different times of the year. Also, the model allows the simulation of control interventions, such as larvae removal, that reduce the mosquito-human ratio. In general, the model suggests that preventive vector control could prevent the virus introduction in the U.S. [26].

Some studies have found estimates of the reproduction number for a chikungunya epidemic. For instance, Yakob et al. developed a compartmental model to reproduce the epidemic in La Réunion Island, and then to calculate the reproduction number. This estimate allows the authors to find a critical vector mortality rate that ensures the success of the control programs (i.e. $R_0 < 1$). Additionally, this model suggest that vector control is more effective than quarantine of infectious symptomatic individuals [22].

Another compartmental model for La Réunion epidemic of 2005 was proposed by Dummont *et al.* In this study, the reproduction number was computed for each city of the island. In addition, thresholds were found for different strategies of vector control to ensure an R_0 below 1. According to the simulation results, vector strategies focused on egg removal are more effective than the use of insecticides to halt the spread of the epidemic across the island [23].

Furthermore, Moulay et al. evaluated optimal control strategies in a compartmental model that represents mosquito and human population dynamics. Three different control interventions were modeled: reduction of human and mosquito contact, diminution of breeding sites and treatment of infectious individuals. The optimal control strategies suggest that extensive larvae reduction is needed to minimize the effect of the epidemic [24].

Although multiple characteristics can be represented in ODE models, these models cannot describe individual characteristics in a population. For that reason, other models may need to be used such as Agent-Based Models.

Agent-Based Models (ABM)

ABM are computational models that represent dynamics using social interactions. ABM are comprised of agents and an environment. An agent represents an individual entity such as a person, an animal or even an organization. Agents interact with other agents and with the environment. Also, agents have goals that determine the activities assigned to each one in order to reach their goals. The environment is an abstract representation of the real world where the agents interact. Although the environment can be seen as a geographical space where the agents interact, it could also illustrate complex features, for instance in a network model, the environment could represent links between agents [60].

In contrast with the compartmental models mentioned above, ABM are capable of representing heterogeneous populations with multiple individual characteristics. When defining the environment of an ABM as a geographical representation of the space, it is possible to include several characteristics such as: geography, climate, socio-economic status, etc.

In short, agent's characteristics can be summarized as [61]:

- Adaptation: the kind of decisions that an agent makes to adapt to the environment and other agents.
- Goals: Any decisions made by an agent must be consistent with its own goals. These goals can be adapted to the environment or the model evolution.
- Learning: how an agent makes new decisions based on past experiences.
- Sensing: the ability of an agent to measure variables from the environment or other agents.
- Interaction: how an agent shares information with other agents. According to their own goals, these interactions could lead to cooperation or competition.
- Collectives: different social relationships can be attributed to agents, such as family, friendship, etc. Also, agents can form collectives based on their mutual goals.

Multiple ABM have been used in epidemiology to represent diseases such as malaria, influenza or Chagas. For instance, Linard *et al.* proposed a model to study malaria dynamics in France. In this model, humans' and mosquitoes' interactions depend on human activities, like farming, urban dwelling, or tourism [62]. In Peru, Lloyd *et al.* developed a contagion model for malaria [63] where socio-economic factors were included. These models allow the inclusion of control strategies such as vaccination or vector control. For example, Gu *et al.* evaluated the effect of larvae reduction in water containers near human settlements [64].

Chagas disease has also been modeled using ABM. In Cameroon, Muller proposed an ABM for Chagas disease that incorporates human activities, mosquito density and mosquito migration. In that model, other hosts for the disease different than humans are considered, such as cattle or wild animals [65].

Regarding dengue, Focks *et al.* [66] developed two different models to represent dengue dynamics: CENSiM and DENSiM. CENSiM represents *aedes aegypti*'s different maturation age characteristics that are affected by geographic effects. DENSiM uses the CENSiM output as input to simulate dengue dynamics in a human population. This model has been used in Iquitos, Peru to estimate parameters of dengue disease in the region [67].

Otero *et al.* represented the *aedes aegypti*'s dynamics in a stochastic model. In this model, the geographical region is represented as a grid, where each patch contains the information about the different stages of the *aedes aegypti*. Mosquito spread is represented by a diffusion coefficient as a traveling wave. The model is parameterized to represent the field data from a small region in Buenos Aires, Argentina [68]. This *aedes aegypti* model is used as an input for an ABM to represent dengue outbreaks; in which, humans are modeled as agents and mosquitoes are represented by a compartmental stochastic model within each patch [69].

Isidoro *et al.* proposed an ABM to represent dengue epidemics including mosquito and human populations. In that model, the following characteristics are included: mating, female mosquitoes' biting

behavior, virus transmission (both ways: infected mosquito transmitting the virus to a susceptible human or an infected human transmitting the virus to a susceptible mosquito) and oviposition behavior [70].

Chao *et al.* assessed the possibility of vaccination for a small region in Thailand using a synthetic population that matched the census data. In the model, humans are represented as agents while mosquitoes are modeled as follows: in every place, such as workplace school or household, the number of susceptible mosquitoes is recorded. When an infected human visits a place, a number of mosquitoes become infected. Then, this group of mosquitoes are modeled as agents and can bite humans and infect them until they die. The authors concluded that at least 50 % of vaccination coverage would be needed to decrease dengue local transmission [28].

For chikungunya, only one ABM was found, proposed by Dommar *et al.* [29]. In this model, villages are represented as a network. Agents are defined as travelers or non travelers in order to evaluate the effect of human mobility in the spread of the epidemic. The model suggests that travel restrictions in symptomatic individuals are not enough to halt the spread of the epidemic, because of the possibility of asymptomatic travelers. Therefore, the authors propose to implement control interventions that would trace symptomatic infections to identify exposed humans. However, this model does not represent a real population. Also, it cannot be implemented for large-scale populations because of the computational power needed.

Models in South America

Diverse studies have presented models to represent dengue dynamics in South America, primarily in Brazil, Argentina and Peru [71, 72, 73, 68, 74, 69, 33, 67, 75, 76]. However, no model was found to analyze the impact of a chikungunya invasion in South America.

In Colombia, Velez *et al.* [32] used the models CIMSIM and DENSiM proposed by Focks *et al.* [33] to study dengue dynamics in a small municipality in Antioquia, Colombia. The model includes climate factors calibrated with the climate data of the region. In 2012, Padmanabha *et al.* [34] built an ABM to represent dengue dynamics in an urban neighborhood located in Armenia, Colombia. The authors created a synthetic population with the correct distribution of the households, age structure of the population and mosquito density by household. In that study, the model was used to simulate several scenarios of the spread of dengue in the region, including vector control in pupae stages. Also, the effects of immigration of infectious humans were simulated in the model. Furthermore, the authors found minimum thresholds of pupae density to achieve the successful dengue eradication in the region.

Castañeda-Orjuela *et al.* [30] developed a two-serotype age-structure deterministic model to evaluate the burden of the dengue disease in Colombia in 2011-2014. Torres *et al.* represented dengue dynamics in Colombia for the period 1995-2011 using fuzzy modeling techniques, and made predictions for a three years horizon [31].

Nonetheless, epidemiological models have not evaluated a possible chikungunya outbreak in the Americas. Also, the scale of Agent-Based Models of vector-borne diseases is restricted to small populations. Due to the recent innovation of chikungunya across the region, it is crucial to develop tools that can assess the impact of such an epidemic in the region. In this dissertation, a large-scale Agent-Based Model is developed to evaluate the impact of possible outbreaks in a country scale.

ABM Platforms

According to Railsback *et al.* [77] there is no ideal platform to program every kind of ABM. Nevertheless, there are several comparisons between the most common platforms. Arunachalam *et al.* made a comparison between MASO, NetLogo, Ascape, RePastS and DIVAs. Based on the results, the best tool for non-programmers is DIVAs. While, RePast shows a friendly environment complexity and NetLogo appears to be the best for distribution [78].

Another platform is the Framework for Reconstructing Epidemic Dynamics (FRED) [79], developed at the University of Pittsburgh. “FRED is a free-available epidemic modeling system that uses census-based synthetic populations to capture the demographic and geographic heterogeneities of the population” [79]. It has been used in multiple studies to address research questions in public health.

Owing to its characteristics, it could be an appropriate platform to implement a large-scale dengue agent-based model.

FRED is a highly modular, object-oriented program written in C++ that was released in 2013 under the BSD 3-Clause Open Source License. FRED exhibits fast performance to simulate epidemics over large populations. For instance, a FRED simulation of influenza within a population of 1 million completes in around two minutes on a conventional laptop. The same FRED simulation when implemented on a supercomputer using data for the whole U.S. population of over 321 million, takes approximately 4 hours to run [79].

FRED requires a synthetic population with a specific format to load agents' demographic information (i.e. sex, age, race, household location,etc.). Currently, FRED has the synthetic population of the United States and some other selected countries. However, more countries can be simulated if an appropriate synthetic population is available [79].

Disease transmission in FRED is location specific and the schedule is updated daily, which could be an issue when simulating diseases with short latency or infectious periods. However, it is not a problem with diseases with a latency period of several days such as influenza or dengue. Also, multiple strains can be set up to run in FRED. At present, FRED supports only diseases that can be expressed as agent to agent transmission. Nonetheless, due to FRED's flexibility, vector transmission diseases such as dengue can be implemented by adding a vector-transmission class in the source code [79].

Several public health concerns have been addressed using FRED. For example, Lee et al. evaluated the school closure impact in an influenza outbreak in Allegheny County, PA, finding the time needed to produce a significant reduction in the overall attack rate of the epidemic (School closure to mitigate 2010). Moreover, Brown et al. used FRED to exhibit the high cost of school closure, concluding that as a sole control strategy, school closure would be too expensive compared to the epidemic cost itself (School closure 2011).

Work policies have also been studied in FRED. Kumar et al. used FRED to assess the effect of the stay at home policy and the importance of access to paid sick days(PSDs). Kumar et al. showed that providing access to PSDs and staying at home would decrease significantly the attack rate of an influenza epidemic (Policies to reduce influenza 2013). Another study in FRED showed that a vaccination program focused on low income communities would decrease the overall impact of an influenza outbreak in a population (Poor communities 2011). Furthermore, FRED was used to analyze the effect of public transportation in an influenza epidemic in New York City (SUBWAY 2011).

These studies display the capability of FRED to address various questions regarding epidemics over large populations. Hence, as a synthetic population is created for this research project pertaining to dengue in Colombia, it can be used to address similar question. Although vector transmission is not included in FRED currently, its implementation is planned for future versions. Consequently, a dengue model for Colombia fits in the FRED development aims.

1.3 Models' Calibration and Optimization

Dynamical models, such as those described in the previous section, attempt to represent a system whose behavior is predetermined by the physical laws governing the model and a set of parameters. Several approaches to represent the dynamics of epidemiological systems have been mentioned in the last section. Validation of the model consists of assessing that the model is consistent with the phenomenon of interest. The validation of models can be achieved with the adjustment of the parameter values to obtain coherent outcomes. This calibration process depends on the model's nature, capabilities and limitations. In general, models can be classified in three different categories: abstract, specific and measurable processes [60].

- **Abstract models.** These models attempt to represent basic social patterns that are not comparable with any kind of data. These models can be validated qualitatively from the macro-level patterns that are expected to appear. Another way to calibrate these models is to systematically vary parameters in the model and observe the patterns obtained. In order to have a valid model, the changes in the patterns obtained should be interpretable.

- **Middle Range Models.** The aim of middle range models is to describe characteristics of a particular social phenomenon in a general way, hence the conclusions can be applied widely. The outcomes of these models are general and not comparable with specific data. Therefore, the validation can be done qualitatively with expected statistic distributions or shapes.
- **Facsimile Models.** The objective of these kinds of models is to represent a specific target phenomenon. The intention of the model could be to predict a future stage or to evaluate the response of the target to any kind of intervention. The validation of these models is performed using quantifiable data that is compared to the model's output. However, there are random factors in the model and in reality that makes it challenging to reproduce an exact scenario.

The model developed in this thesis is a facsimile model that intends to obtain quantifiable conclusions. Hence, the model needs to be compared to empirical data. The calibration steps for this kind of model can be summarized in the following steps [80].

- Selection of the parameters.
- Selection of the targets to compare the model's outcome.
- Definition of a Goodness-of-Fit measure.
- Choice of a parameter search strategy
- Definition of acceptable goodness of fit criteria

Parameters to Calibrate and Selection of Targets

A facsimile model is composed of observable and unobservable parameters. For instance, in epidemiological models, observable parameters can be the incidence cases, incubation period, age of incidence, etc. Whereas, the probability of transmission or the asymptomatic level are difficult to measure. In general, the parameters used in the calibration process are the unobservable parameters or those parameters that are difficult to measure for a specific situation. Some observable parameter values can vary from source to source, in which case it could be advisable to include those parameters in the calibration process [80].

The selection of the targets depends on multiple factors such as the nature of the model, the availability of quality data, and the purpose of the model. In the case of epidemiological studies, it is preferable to use local data, since the population specifics can influence the outcome of the targets.

The calibration process consists of minimizing the difference between the outcomes of the data and the empirical targets chosen. Multiple techniques can be applied to fit the parameters. Mathematically, the parameter calibration can be expressed as:

$$\begin{aligned}\hat{\theta} &= \min_{\theta} \|F(\theta) - Y\| \\ F(\theta) &= (f(t_1; \theta), f(t_2; \theta), \dots, f(t_m; \theta)) \\ Y &= (y_1, y_2, \dots, y_m)\end{aligned}$$

Where $F(\theta)$ represents the outcome of the model for a specific time (t_n), and Y represents the observations for each time step (t_n). The expression $\|F(\theta) - Y\|$ represents the measure of the difference between the empirical data and the outcome of the model. Hence, the estimated parameters are the set of parameters that minimizes the difference between the output of the model and the expected observations of the data. There are several measures that quantify this difference and are called goodness of fit measures.

Goodness-of-Fit Measures

In order to minimize the difference between the model and the experimental data, this difference needs to be measured quantitatively. There are several goodness-of-fit measures that are used to compare observed and modeled data. The importance of these measures is to avoid the human subjective perception about a 'good' or 'bad' fit. Depending on the field, different measures are used. In epidemiological models, some of the measures commonly used are: least squares, χ^2 and likelihood [80].

Least squares (equation 1.1) calculates the sum of the square of errors for each observation. The square of the errors increases the weight of the larger data points. The χ^2 (equation 3.1) is similar to the least squares, however each error is divided by the standard deviation of the data. Hence, the measure puts more weight on the more precise data observations. Another measure is likelihood, which estimates how likely a specific set of parameters is to match the empirical data. The objective of the calibration is to maximize this value rather than minimize it. This measure requires more data than the other calculations. Variations of these measures can be created to assign more weight to different aspects of the data. For example, in Error expression in 1.4, there is more weight on the larger data points.

$$Error_{lsq}(\theta) = \sum_t (f(t; \theta) - y(t))^2 \quad (1.1)$$

$$Error_{\chi^2}(\theta) = \sum_t \left(\frac{f(t; \theta) - y(t)}{\sigma_t} \right)^2 \quad (1.2)$$

$$Error_L(\theta) = \prod_t p(t|\theta)^{y(t)} (1 - p(t|\theta))^{n(t)-y(t)} \quad (1.3)$$

$$Error(\theta) = \sum_t y(t) \cdot (f(t; \theta) - y(t))^2 \quad (1.4)$$

Models can also be calibrated to multiple data. Depending on the data, multiple measures can be used to estimate the errors between the model and the empirical observations. The total error is expressed as a weighted sum of the error measures. In any case, with one or multiple targets, the purpose of the calibration of the model is to minimize the total error. This problem can be seen as an optimization problem that can be solved using multiple strategies with different levels of complexity.

Model Calibration and Optimization

Basic parameterization strategies involve a grid search of all the possible parameters' values. The error is computed for each and the point with the optimum performance is selected. and computing the error for each point, then selecting the point with better performance. This strategy could be effective for small number of parameters within small boundaries. The advantage of this strategy is that its implementation is straightforward and does not require any assumptions of the data. However, as the number of parameters and their possible values increase, so does the combinations of these parameters, and the search could become computationally inefficient [81].

Mathematical models with analytical expressions can be parameterized using gradient methods that evaluate the derivatives of the system of equations. Minima values of the system can be found moving along these gradients in the direction of the steepest possible curve. This method can be implemented for linear and nonlinear systems. Nonetheless, for a nonlinear system, the optimal value can be a local solution. Therefore, several initial points should be tested in order to find a global minimum. This method cannot be used for models without explicit mathematical formulation of the system. Other iterative optimization methods are Simplex Nelder-Mead, Trust-Region, Lavengerg-Marquardt, and genetic algorithms.

The simplex method proposed by Nelder and Mead in 1965 is used to minimize a function outcome without information about the derivatives of the system [82]. Therefore, it can be implemented in

computational models that are not mathematically formulated, such as an ABM. Basically, the Nelder-Mead simplex method generates non degenerate simplex that in n dimensions is denoted by $n+1$ vertices, i.e. in two dimensions, a simplex represents a triangle. In each iteration, the error of each vertex is computed and geometric operations are performed in order to exclude the worst point of the simplex. These operations are: Reflexion, Expansion, Contraction and Shrink. While measuring in space in this way, the simplex moves in a downhill direction, until an optimum point is found. The convergence properties and the implementation of the Nelder-Mead simplex method are discussed by Lagarias *et al*[83] and Gao *et al*[84].

- Order. Order the $n + 1$ vertices based on the error function: $f(x_1) \leq f(x_2) \leq \dots \leq f(x_{n+1})$
- Reflect. Compute the reflexion point: $x_r = \bar{x} + \rho(\bar{x} - x_{n+1})$. Where $\bar{x} = \sum_{i=1}^n (x_i/n)$ is the centroid of all the vertices except x_{n+1} . If $f(x_1) \leq f(x_r) < f(x_n)$, accept the point x_r and terminate the iteration.
- Expand if $f(x_r) < f(x_1)$. $x_e = \bar{x} + \chi(x_r - \bar{x})$. If $f(x_e) < f(x_r)$, accept x_e and terminate the iteration, otherwise accept x_r .
- if $f(x_r) \geq f(x_n)$, perform a contraction step:
- Contract outside. If $f(x_n) \leq f(x_r) < f(x_{n+1})$. $x_c = \bar{x} + \gamma(x_r - \bar{x})$. If $f(x_c) \leq f(x_r)$, accept the point x_c and terminate the iteration, otherwise perform a shrink step.
- Contract inside. if $f(x_r) \geq f(x_{n+1})$. $x_{cc} = \bar{x} - \gamma(\bar{x} - x_{n+1})$. If $f(x_{cc}) < f_{n+1}$, accept the point x_{cc} and terminate the iteration, otherwise perform a shrink step.
- Shrink. The vertices $v_i = x_1 + \sigma(x_i - x_1)$ $i = 2, \dots, n+1$. The unsorted simplex for the next iteration would be x_1, v_2, \dots, v_{n+1} .

The parameters for Reflexion, Expansion, Contraction and Shrink are: $\rho > 0$, $\chi > 1$ and $\chi > \rho$, $0 < \gamma < 1$, and $0 < \sigma < 1$, respectively. Usually these values are set to $\rho = 1$, $\chi = 2$, $\gamma = 0,5$, and $\sigma = 0,5$.

The advantages of the Nelder-Mead method is that it does not require evaluation of derivates. However, the algorithm can be slow and converge to local minima. Hence, the initial conditions of the algorithm should be set in different segments of the parameters in order to assess global convergence. Also, there is no theoretical way to describe the uncertainty of the fitted parameters.

Maximum-likelihood is a strategy proposed by Fisher *et al* that estimates the likelihood that specific data is observed by the model with a given set of parameters [85]. This method allows for the estimation of uncertainty on the parameters, but it requires a probability model of the data that could be difficult to obtain in some situations. Another issue with this method is that it could be difficult in complex problems to maximize the likelihood [81, 80].

Bayesian methods use a prior distribution of the parameters, this distribution is updated by Bayes' theory based on the likelihood method. In order to update the distribution of the parameters, the Markov chain Monte Carlo method can be applied. The samples provide the exact quantification of the uncertainty levels on the parameters. This method requires a prior information about the parameters' distributions. Multiple prior information of the parameters can be used in order to assess the impact of these initial conditions on the resulting parameters' estimates [81, 80].

Sensitivity analysis

There are basically two different sources of uncertainty in a model's behavior: the uncertainty from the parameters' values and the model structure itself [81]. The set of parameters defined represents an specific behavior of the model. The uncertainty in these parameters proliferate through the model. Hence, in order to quantify the sensitivity of the model's outcome based on the uncertainty of the parameters, a systematic variation in the parameters' values can be performed. From the sensitivity analysis, the set of parameters where more precise values should be assessed to represent an acceptable behavior are identified.

To estimate the uncertainty of the model based on its own structure, the set of assumptions should be verified. The model should be tested with experimental data in order to decide whether it explains qualitatively and quantitatively the different aspects of the data. However, a single dataset cannot accept or reject a model. Hence, the model should be verified with multiple datasets in order to evaluate its performance. These data comparisons is also helpful in establishing the limitations of the model and its applicability [81].

1.4 Document Content

In this chapter, the main problem to be confronted by this thesis was exposed, that is to represent the chikungunya epidemic and optimize a vector control strategy to minimize the impact of an epidemic in Colombia. The concepts related with epidemiological models were mentioned as well as the optimization methods that are implemented to calibrate and optimize the intervention strategy. In chapter 2, the epidemiological data of chikungunya in Colombia is presented as well as an analysis of its relation with previous dengue epidemics. In chapter 3, the construction and validation of a synthetic population of Colombia is presented. This synthetic population constitutes an essential input of the model that allows the representation of the epidemic in a specific geographical place in the country. In chapter 4, the formulation and verification of the model is presented. In chapter 5, the parameters' calibration strategy is illustrated. The chapter includes the selection of parameters and the possible sources of uncertainty. The model is calibrated using one municipality of the country and then its performance was assessed by comparing the results with data from unfitted municipalities. Chapter 6 includes the implementation of vector control strategies in the model based on pupae reduction. The vector control model is calibrated to represent the specific case of the municipality of Santa Marta. Also, the intervention parameters are varied to evaluate their impact on the epidemic curve. The calibrated scenario is extrapolated to the whole country to evaluate geographically focused interventions. Finally, conclusions and future work are presented in chapter 7.

Chapter 2

Dengue and Chikungunya Epidemiology in Colombia

Chikungunya virus is characterized by its rapid spread in susceptible regions. Since its inception, the virus has been considered a major public health concern in Colombia. There is no approved vaccine to protect the population against the virus; therefore, the only alternative available is to reduce the vector population and prevent transmission. However, it requires significant economic effort, and community education to prevent transmission. The amount of economic resources available is limited and the necessary mobilization to educate the community about the risks and prevention of the virus, is expensive. In consequence, costs can be minimized limiting the extension of the intervention programs to chikungunya risk zones. Considering that chikungunya is a new virus in the country, identifying these risk zones could provide useful public health insight.

The transmission of chikungunya occurs under similar conditions as dengue. For instance, both are tropical diseases transmitted by *aedes* mosquitoes with similar pathogenesis. In Colombia, dengue virus is currently circulating heterogeneously across the country. Multiple epidemics have been reported, the largest one was in 2010 with more than 140,000 cases [86, 87]. Taking into account that dengue is endemic in Colombia, it is reasonable to define dengue risk zones as possible chikungunya risk zones. Hence, information from previous epidemics of dengue could help to focus interventions in crucial regions of the country where chikungunya is not yet present. In this chapter, there is a description of dengue epidemiology since 1998 which is focused mainly on the epidemic of 2010. The available chikungunya data is analyzed for different regions and correlated to dengue risk zones. Finally, the correlation between dengue and chikungunya epidemics with multiple covariates is explored, finding that precipitation and temperature are correlated with the occurrence of early epidemics in the country.

2.1 Dengue and Chikungunya Surveillance Reports

In Colombia, the information about the dynamics of transmissible diseases is collected and published by the National System of Surveillance in Public Health (SIVIGILA, Sistema de Vigilancia en Salud Pública). SIVIGILA was created as a sub direction of the Colombian National Institute of Health (INS, Instituto Nacional de Salud www.ins.gov.co) aiming to guide the policies of public health in Colombia.

The dengue and chikungunya weekly incidence surveillance data in Colombia were obtained from the SIVIGILA. The dengue reports include weekly diagnosed cases from 2007 to 2012. The resolution of this dataset differs in 1997 - 2006 compared to 2007 - 2012. In the first period, cases are reported for each of the 33 departments (counties) without any further information. In the last period, the reports include age-structured incidence and department and municipality regional resolution. Chikungunya reports include weekly reports from EW 20/2014 to EW 7/2015 by department and municipality. This dataset includes all the municipalities where more than 200 cumulated cases were reported.

The dengue incidence rate per 10,000 inhabitants by department is presented in figure 2.1. Three different periods can be identified 1997 - 2000, 2000 - 2006, and 2007 - 2012. In the former period,

an outbreak was reported in multiple departments around 1998. It can be seen that these outbreaks were reported mostly in temperate climate departments. Departments with high temperatures did not report an outbreak in this period. Another outbreak was reported around 2002 throughout the same regions. In 2010, an outbreak took place through most of the departments with the number of cases exceeding 140,000.

Most of the departments reported an outbreak in 2010, except from the extreme cold departments and few departments with high annual temperature. In contrast with the last epidemics, there was an evident synchrony between the departments. After this period, some departments show an endemic situation with consistently low cases. It should be noted that in the highest temperature department (Sucre), there were a constant number of cases after 2002, but no particular explosive epidemic, compared to the rest of the departments. In the temperate regions, the impact of the 2010 epidemics suggest a change in the dynamics from wide epidemics to short-lived and sudden.

In figure 2.2, dengue municipality incidence rates are represented weekly from 2007 to 2012. It is

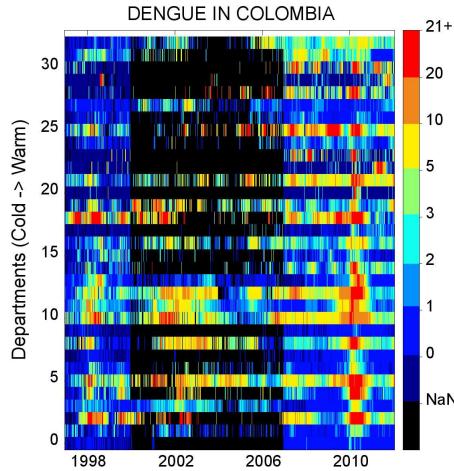


Figure 2.1: Dengue weekly reports by department. Each pixel in X axis represents a week and each pixel in the Y axis represents a department. The number of cases per 10,000 persons per week are represented by a color code. Representing high incidence with red and low incidence with blue. Black represents non-reported weeks.

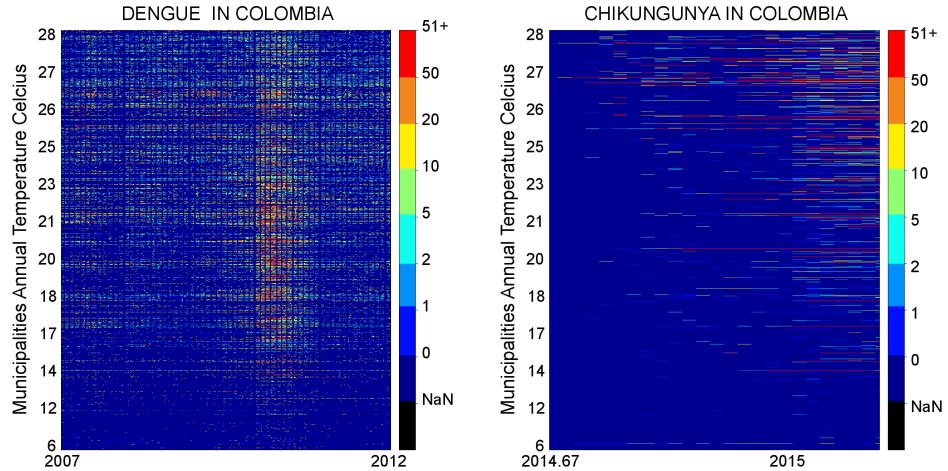


Figure 2.2: Dengue and Chikungunya weekly incidence rates (cases per 10,000 inhabitants) by municipality. Left graphic corresponds to dengue. Right graphic corresponds to chikungunya.

possible to see the synchrony and impact of the 2010 epidemic in most of the warmer municipalities. It is important to point out that municipalities with colder average temperatures are less affected by dengue than warm municipalities. This suggests a correlation between temperature and dengue transmission. Nonetheless, it can also be appreciated that the incidence decreases at a temperature level of around 27 °C.

The incidence of chikungunya by municipality is higher than the incidence for the 2010 dengue epidemic. This could be attributed to the immunity characteristics of the population since dengue has been present for decades in the country. In contrast, chikungunya is a new virus in the region so most of the Colombian population has not been previously exposed. Hence, the susceptible population size is considered larger for chikungunya than for dengue.

Considering the difference in timescales of the chikungunya and dengue data, two different analyses were performed. First, dengue epidemics periodicity and synchrony were analyzed at the department level from 1997 until 2012. Then, the 2010 dengue epidemic was compared to the ongoing chikungunya epidemic in order to understand the similarities within these disease transmission dynamics in Colombia.

2.2 Department Classification Analysis Using Wavelet Analysis

As shown in figure 2.2, dengue is heterogeneously spread across the country, some of the departments show endemic status of the epidemic while others seem to have sporadic or periodic epidemics. Also, for a specific department, the intensity and periodicity of dengue epidemics could change over time. Wavelet analysis is considered a useful tool to understand this type of behavior, as it extracts information on the changing frequency components of a time series as well as its power and noise.

Wavelet analysis has been broadly used in multiple fields such as ecology, economics, geology and epidemiology [88, 89, 90, 91, 92]. In general, wavelet analysis is useful to understand the different frequency components of a signal. For instance, in ecology it has been used to analyze the variability in animal populations and their possible relationships with climate phenomena [93, 94]. In epidemiology, wavelet analysis has been used to investigate synchrony between epidemics in different regions and the periodicity and changes in frequency caused by external interventions such as vaccines [95, 96]. This analysis has also been used in epidemiology to quantify the correlation between outbreaks of dengue and climate fluctuations [97, 98, 99].

Wavelet Analysis

The wavelet analysis is based on the transformation of a time series using wavelets. The wavelet transform can be defined as the convolution integral of a wavelet function and the time series. There are multiple wavelet functions and the selection of a wavelet function depends on the desired application and various other criteria. However, the Morlet wavelet is widely used and is defined as follows [90]:

$$\Psi = \pi^{-1/4} e^{i\omega_0 \eta} e^{\eta^2/2} \quad (2.1)$$

Where ω_0 is the non dimensional frequency that is usually assumed to be 6. In order for Ψ to be a wavelet function it requires to have a zero mean and be localized in both frequency and time space.

The continuous wavelet transform is defined as [92]:

$$\Psi_{a,\tau}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-\tau}{a}\right) \quad (2.2)$$

Where ‘ a ’ is a scale parameter related to the frequency and τ is a translation parameter related to the time. Scaling a wavelet function is an extension or compression, which is a frequency change. This property is shown in figure 2.3. Translating a wavelet function shifts its position in time as displayed in figure 2.3, where the wavelet function is translated by a factor of 2, while maintaining its shape. Using these two properties, a wavelet transform of a signal evaluates the correlation of the signal to different frequencies at different times. This means that a wavelet transform could provide information for the variability in the periodic components of an epidemic curve.

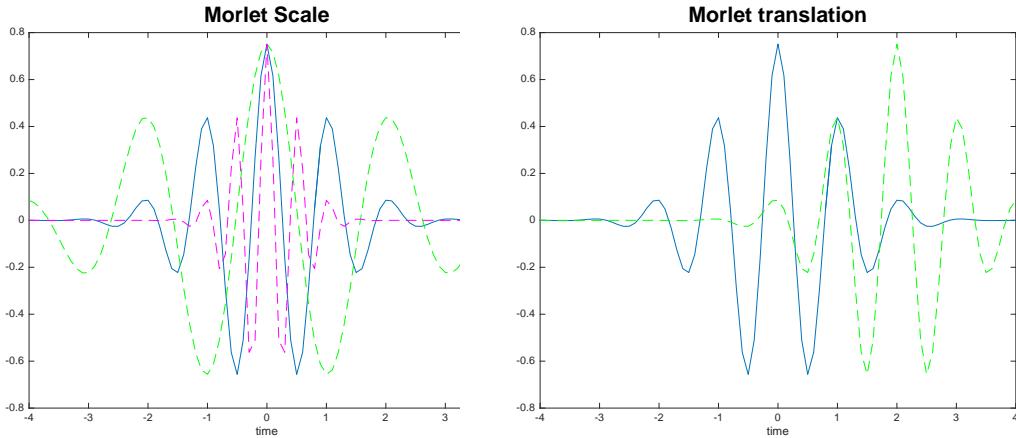


Figure 2.3: Wavelet function scaling and translation properties. Left figure corresponds to scaling the Morlet wavelet by a factor of 0.5 (magenta) and 2 (green). At the right, the wavelet is translated by a factor of 2.

Considering the previously mentioned properties of wavelet transforms, they can be used to represent epidemiological time series in their various frequency components at determined points in time. Solares *et al.* developed a toolbox for wavelet transforms in Matlab® that is suitable to analyze different time series [100]. ASToolbox was used to analyze the epidemiological time series of dengue incidence in Colombia since 1997 to 2012. In order to avoid bias at high frequencies, extreme data points were removed from the analysis. The data was normalized using a logarithm transform and the linear trends were also removed. Additionally, the signals were padded with zeros in order to avoid edge effects.

In figures 2.4, 2.5, 2.6, and 2.7, incidence rates, wavelet transform and average power spectrum are shown for four different departments. These four departments are characterized by different dengue dynamics. The department of Norte de Santander has reported periodic dengue outbreaks since 1997. As shown in figure 2.4, the time series show four different outbreaks, the first two reported within a period of approximately 4 years. There is a difference of 8 years between the last outbreak and the second one. The characteristics in frequency are shown in the wavelet transform power spectrum. It is possible to see that in the first period (1997-2000), the dengue dynamics was characterized by a period of 4 years, but from 2000 - 2008, the time series is distinguished by two multi annual epidemic components: 4 and 3 years. Finally, in the last outbreak the signal has a unique component of four years.

The time series of the department of Norte de Santander shows an extreme case that was not taken into account in this analysis. As shown in figure 2.5, four epidemics are identified. In the first period from 1998 to 2008, the occurrence of these outbreaks was determined by a four year component. Nonetheless, the frequency of these outbreaks seem to increase slowly to every three years. In contrast, Barranquilla's wavelet transform, shown in figure 2.6, suggests that the epidemics in this region are occurring with less frequency, moving from four years to five years periodicity.

Lastly, dengue cases are not expected to transmit in the city of Bogotá because of its altitude (2700 msnm) that translates to cold temperatures that prevent the development of *aedes aegypti* mosquitoes. As shown in figure 2.7, cases in Bogotá were reported for the first time after 2006, and an outbreak was reported in 2010 with a frequency component of approximately 2.5 years. This could mean that cases reported in Bogotá were imported from different regions of the country. Because of the large amount of connections made in Bogotá to other regions, the outbreaks in Bogotá could mirror the incidence of the country.

These four examples show how diverse dengue transmission dynamics are within Colombia. The departments in the country can be grouped based on their frequency components in order to find relationships that explain these differences in transmission.

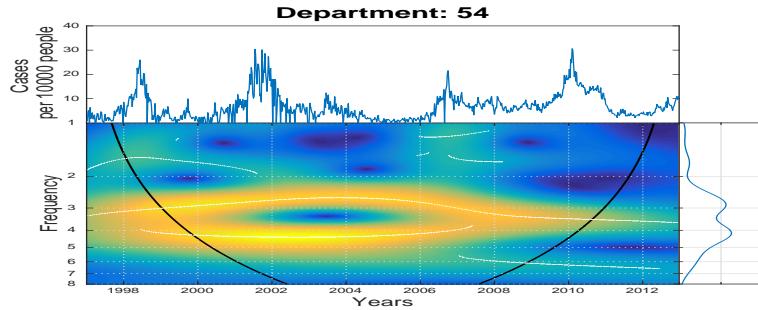


Figure 2.4: Wavelet transform of Norte de Santander's weekly incidence reports. The hot colors represent high power spectrum.

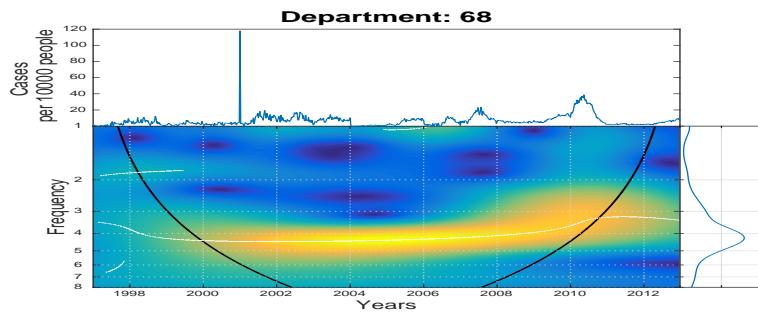


Figure 2.5: Wavelet transform of Santander's weekly incidence reports. The hot colors represent high power spectrum.

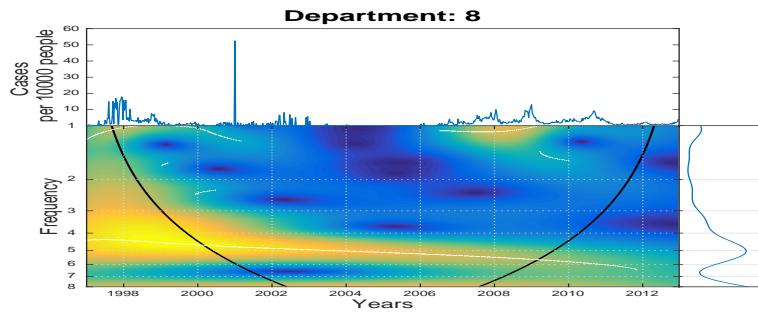


Figure 2.6: Wavelet transform of Barranquilla's weekly incidence reports. The hot colors represent high power spectrum.

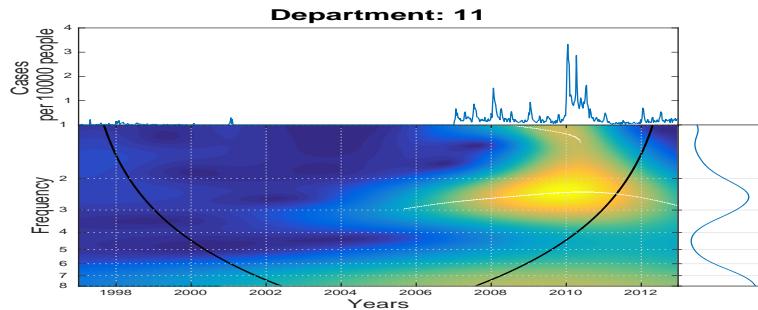


Figure 2.7: Wavelet transform of Bogotá's weekly incidence reports. The hot colors represent high power spectrum.

Transmission clusters using the maximum covariance analysis

The classification analysis of wavelets using maximum covariance analysis (MCA) was first proposed by Rouyer *et al.* [92]. This method is used to compare pairs of wavelets and classify them based on their power spectrum. The MCA uses singular value decomposition of the covariance matrix of a pair of wavelets in order to find common patterns between two wavelets. Then, a distance unit is computed to define the difference between two spectra. With this, a distance matrix can be constructed and used in a cluster tree analysis. This method has been applied to various fields such as biology, economics and electrical distribution [92, 101, 102].

The covariance matrix between two wavelet spectra W_i and W_j is defined as:

$$R_{i,j} = W_i W_j^t \quad (2.3)$$

Where W_j^t represents the transpose of W_j .

This covariance matrix can be expressed using singular value decomposition as:

$$R_{i,j} = \mathbf{U} \boldsymbol{\Gamma} \mathbf{V}^t \quad (2.4)$$

Where \mathbf{U} and \mathbf{V}^t are characterized by orthogonal columns that contain the singular vectors of W_i and W_j , respectively. $\boldsymbol{\Gamma}$ is a diagonal matrix whose elements contain the singular values that are placed in decreasing order. This singular values are proportional to the squared covariance of each axis. Hence, these values can be associated with the common patterns between the pairs of wavelets in order of importance. $\boldsymbol{\Gamma}$ has k non-zero singular values that are less than or equal to the total number of common analyzed frequencies. These values can be expressed as: λ_k . The proportion of the explained covariance associated with each singular vector k can be seen as $\lambda_k^2 / \sum_k \lambda_k^2$.

The number of leading patterns \mathbf{K} used to compare the spectra is selected using a threshold θ , so

$$\frac{\sum_j^K \lambda_k^2}{\sum_k \lambda_k^2} \geq \theta$$

In order to analyze how frequencies evolve over time, the wavelet spectrum can be projected to the leading patterns obtained as follows:

$$\begin{aligned} L_i^k(t) &= \sum_{f=1}^{F} \mathbf{U}^k \times W_i(f, t) \\ L_j^k(t) &= \sum_{f=1}^{F} \mathbf{V}^k \times W_j(f, t) \end{aligned} \quad (2.5)$$

Where F is the maximum common frequency. The difference between the two wavelet spectra is measured based on the leading patterns and the singular vectors using a robust version of the correlation between the derivatives of the leading patterns and the leading vectors. The distance is defined as:

$$d(L_i^k, L_j^k) = \sum_{t=1}^{n-1} \text{atan} [\|(L_i^k(t) - L_j^k(t)) - (L_i^k(t+1) - L_j^k(t+1))\|] \quad (2.6)$$

where n is the length of the vectors. A matrix of distances is computed between each pair of wavelets as follows:

$$D(i, j) = \frac{\sum_{k=1}^K \lambda_k \times (d(L_i^k, L_j^k) + d(\mathbf{U}_i^k, \mathbf{V}_j^k))}{\sum_{k=1}^K \lambda_k} \quad (2.7)$$

Using this method, a distance matrix was constructed for each pair of the 33 wavelets. Then, this matrix was fed into a distance matrix that was appropriate for clustering analysis. Using matlab®, a dendrogram was constructed as illustrated in figure 2.8. This tree representation elucidates four

well-defined clusters. Wavelet spectra is shown in figures 2.9, 2.10, 2.11 and 2.12 in order to analyze each of the clusters and understand their differences.

The first cluster is characterized by 10 departments as presented in figure 2.9. This cluster shows a low power spectra (>0) in most of the analyzed period and a high power spectrum disbursement in most of the cluster departments for the 2010 epidemic. Consequently, this cluster could be called a low incidence with one epidemic. This cluster can be compared to cluster 2 in figure 2.10, where low power spectrum is present in most of the period. However, cluster two includes departments with more than one epidemic in two different time periods. Hence, this cluster could be classified as low incidence with more than one epidemic.

The dynamics in cluster 3 are significantly different than the first two clusters. In this group, the five wavelets have a strong correlation around the 4 years frequency band. This suggests that dengue has been constantly present in these departments with periodic epidemics around 4 years. Three of them show high power spectrum in two different frequency bands, which indicate that the dynamics in these departments are characterized by two components of multi annual epidemics. In consequence, this cluster could be specified as an endemic cluster with periodic epidemics.

In contrast, cluster number 4 is characterized by null power spectrum in the first years. Also, in the latter years one frequency band presents high power spectrum. These departments do not show dengue presence except for the 2010 epidemic. Hence, these regions could be classified as emergent. These clusters are located on the map in order to observe any regional effect in the clusters.

Figure 2.13 displays the geographical location of the color coded departments for each cluster. It can be seen that there is a geographical impact in the distribution of these clusters. For instance, departments in cluster 3, characterized by periodic outbreaks, are connected from the north east to the mid-west of the country. Cluster 2, where there is a constant low-incident presence of dengue, is split into two subgroups in the north and middle sections of the country. The cluster in the north is comprised mostly of the Caribbean region of the country.

The department of Atlántico, from cluster 1, is geographically located inside the cluster 2. This could indicate that geographic characteristics are insufficient to describe dengue transmission. The part of the cluster in the middle region of the country includes departments located on the Andean mountains. The department of Bogotá, from cluster 4, is in the middle of this cluster, which indicates that there is a difference in the transmission in Bogotá compared to the surrounding area. This could

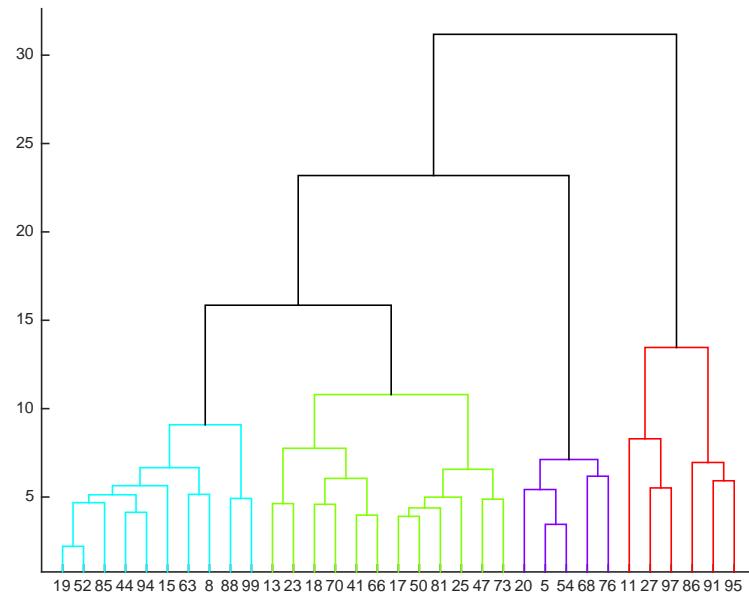


Figure 2.8: Department clusters in a dendrogram. A 'ward' distance measure was used. The threshold was set to 14 in order to obtain four clusters.

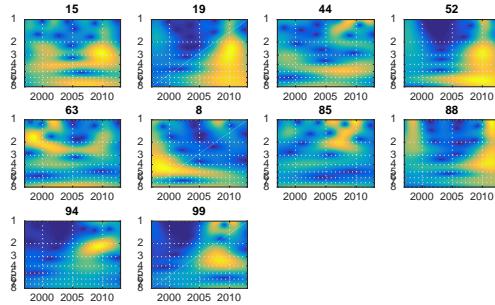


Figure 2.9: Wavelet spectra in cluster 1.

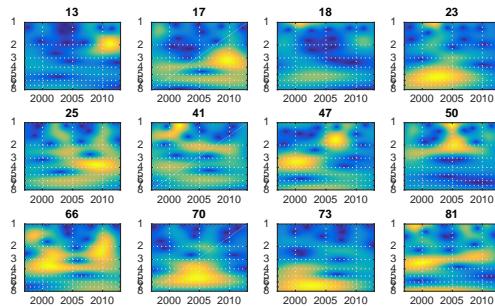


Figure 2.10: Wavelet spectra in cluster 2.

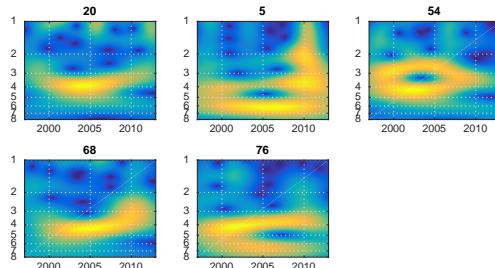


Figure 2.11: Wavelet spectra in cluster 3.

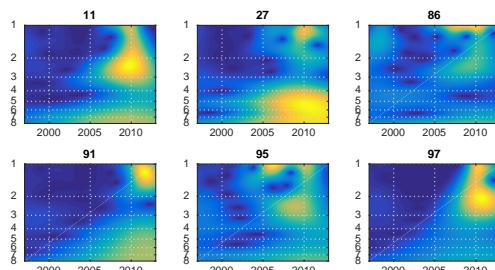


Figure 2.12: Wavelet spectra in cluster 4.

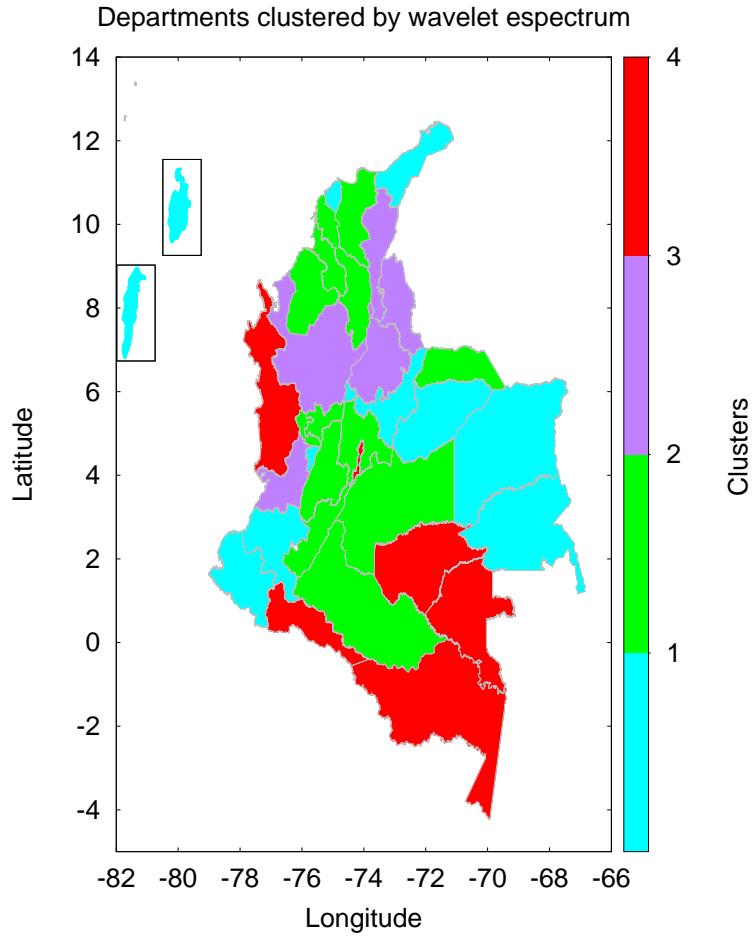


Figure 2.13: Department clusters based on wavelet spectrum. Each color represents a different cluster that agrees with the color code in figure 2.8.

be explained by Bogotá's altitude (2700 meters above the sea).

Finally, there is low transmission in cluster 1 and cluster 2 which include various departments in the Amazon. Also included in the clusters are low density regions, such as San Andrés y Providencia which are comprised of islands that gain access to the rest of the country exclusively by airplane. In consequence, the low transmission in these regions could be explained by the low number of contacts with the rest of the country.

This clustering analysis is helpful to elucidate geographical patterns that affect dengue transmission. However, even inside each department, dengue or chikungunya transmission could present disparate dynamics. Hence the next section deals with an analysis by municipality of the most recent, significant epidemics of dengue and chikungunya.

2.3 Recent dengue and chikungunya data classification analysis

The purpose of this analysis is to classify the municipalities of Colombia based on their recent epidemics of dengue and chikungunya. The period of interest for dengue consists of the 2010 epidemic, since it is the largest epidemic reported in the dataset. For chikungunya, all the reports from the dataset are included in this analysis. Since not all the municipalities reported an outbreak in these periods, a selection process was needed to obtain a final set of municipalities to analyze.

The inclusion criteria is based on various factors of the incidence curves that would constitute an epidemic. These factors are the magnitude, shape and consistency of reports. Hence, a municipality with high incidence reports but multiple zeros between them is excluded from the analysis. In total, six rules were defined for the inclusion criteria, and are listed below:

1. The sum of total cases > threshold (Dengue = 52, chikununya = 10)
2. Start of period of study: Five consecutive non zero samples
3. The Peak of the epidemic is greater than two persons/week
4. Five consecutive zeros: end of the period of study
5. 60 % of the samples within the period of study are > 0

In figure 2.14 two example reports are shown for dengue dynamics. In the left, the municipality of Medellín (5001) has an evident shape of an epidemic curve. The peak-time of this curve is represented by a blue dashed line. On the right, incidence reports are displayed for the municipality of San Andrés (88001). This municipality is rejected and is not included in the study because of the number of zero values and the low number of cases.

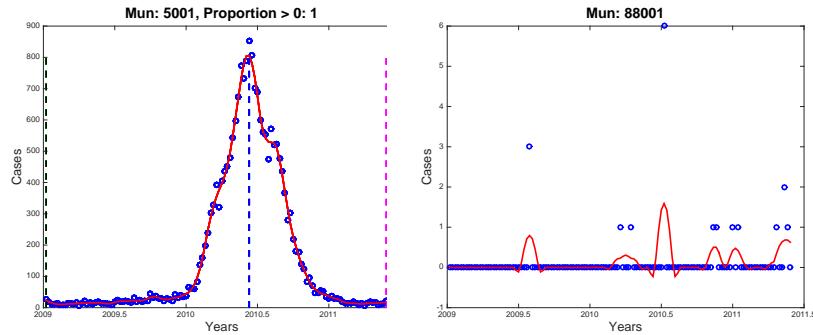


Figure 2.14: Examples of two municipalities when the selection criteria has been applied.

According to this criteria, 105 municipalities have reported an outbreak of chikungunya, and a total of 342 experienced a dengue epidemic in 2010, as summarized in table 2.1. There are 69 municipalities with chikungunya and dengue epidemics, this represents 65 % of the municipalities with a chikungunya epidemic. The total population size of these municipalities is approximately nine million. The percent agreement for this table is 72 % which suggest a high correlation between municipalities that have reported chikungunya epidemics and dengue municipalities. Hence, if dengue can be seen as a predictor of chikungunya cases, a total of 273 municipalities are still at risk.

The time of the incidence peak is extracted from the previously identified municipalities, in order to find clusters of chikungunya and dengue epidemics throughout the country. These peak times are presented in figures 2.15 and 2.16 using light colors to represent early peak times and dark colors for late peak times. Municipalities that enter the epidemic stage later than other areas could be due to interventions against the virus or weak connections with the rest of the country, these variables could delay the epidemic.

In chikungunya, the first cases appear in the north west side of the country and quickly spread to its surroundings. Also, a couple of municipalities in the middle of the territory reported cases early in

	Chik +	Chik -
Den +	69 (9'038.009)	273 (23'030.578)
Den -	36 (1'718.022)	744 (13'875.178)

Table 2.1: Municipalities with epidemics of Dengue and Chikungunya

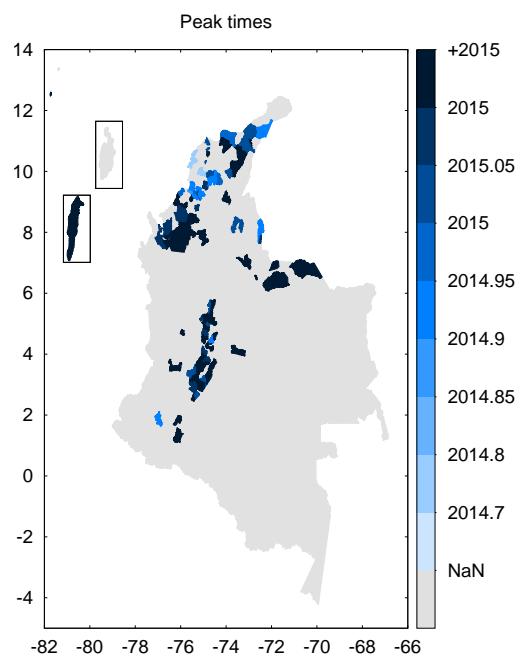


Figure 2.15: Municipalities' peak time of chikungunya epidemic

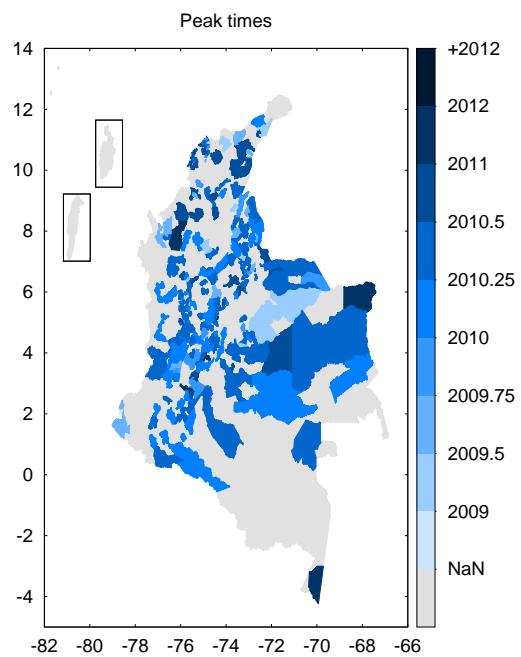


Figure 2.16: Municipalities' peak time of dengue epidemic

the epidemic. Finally, the last municipalities that have reported cases are found south of the starting point. The early cases reported in the middle and southern parts of the country suggest that either these places are well connected to the rest of the country, that they were seeded from outside of the country, or that they are a cause of stochastic effects.

In the case of dengue, early epidemics were reported in the eastern part of the country, spreading through the middle to the south and north. The last municipalities reporting dengue are located in the extremes of the country, such as Leticia that is located in the Amazon.

Municipality clusters

The objective of this analysis is to create groups of municipalities based on how fast they spread the epidemic to other regions. This can be interpreted as the pairwise difference of the epidemic peak times. Also, we want to correlate these groups with different demographic, geographic and climate variables in order to see what kind of variables affect the transmission of the two diseases. Based on those results, a comparison of chikungunya against dengue can be performed to compare their characteristics.

The additional variables that were computed for each municipality are population, temperature, precipitation and mosquito infestation indexes. The population estimates are available at the National Statistics Bureau (www.dane.gov.co). Annual average temperature and precipitation grids have a resolution of 30 arc seconds and are available at the WorldClim Project [103]. Since *aedes aegypti* is considered the main vector for dengue and chikungunya in Colombia [86], infestation indexes of the *aedes aegypti* are included as well from the National Institute of Health of Colombia. The indexes are: household index, reservoir index and brute index. The reproductive number R_0 was calculated for each epidemic using the exponential growth rate method proposed by Wallinga [1].

Based on the epidemic peak time, a matrix of the pairwise peaktime difference was computed. The difference in peak times of one municipality compared to the rest indicates how fast the epidemic moves from the municipality to other regions. Consequently, short peak-time differences suggest fast spreaders, and long peak-time differences suggest slow spreaders. This matrix was used to feed a distance matrix suitable for cluster analysis implemented in Matlab®. As a result, the municipalities are grouped in five clusters for each disease, and ordered by average peak-time difference.

These clusters suggest the possibility of a relationship between climate variables and dengue spread, as displayed in table 2.3. The first cluster, which corresponds to the fast spreaders, has an average peak-time difference of 0.2557 years (3 months). This group ranks second regarding time of epidemic, and is characterized by high precipitation values and mild average temperature (22 °C) compared to the rest of the clusters. Also, the R_0 is the highest value of all the dengue clusters. In contrast, the fifth group is characterized by high temperature, precipitation and mosquito infestation indexes. The average difference in peak-time is 1 year and it has the lowest reproductive number of all the clusters. This group was the first group to report an epidemic of dengue.

High temperature and mosquito infestation values seem to characterize the first municipalities with reported chikungunya epidemics, as shown in table 2.2. The reproductive number for this cluster is the largest with an average of 2.7. The fast spreading cluster is characterized by the lowest temperature values of all the groups. Also, peak time difference for the first cluster is an average 0.066 years (24 days).

In comparison to dengue, larger and homogeneous temperature values characterize chikungunya clusters. All chikungunya clusters have mean temperatures above 25 °C, compared to dengue where the ranges varied from 22 - 26 °C. Also, the impact of the chikungunya epidemic is larger, with four clusters having an average reproductive number above 2.0; whereas for dengue, all the clusters have average $R_0 < 2.0$. It should be taken into consideration the homogenous transmission characteristics of chikungunya in this early stage of the epidemic. Later stages of the epidemic could produce more heterogeneous conditions.

Nonetheless, higher temperatures and infestation indexes indicate the start of the epidemic in both groups. While for dengue the Breteau index is around 28, in chikungunya is 48. Additionally, mild temperatures characterize the fast spreading clusters for both diseases, with 24.5 °C for chikungunya and 22.1 for dengue. This could imply a relationship between connectivity and temperature. While

Cluster	Pop	Temp	Prec.	House	Resrvr.	Breteau	Dif. T.Peak	T.Peak	R0
1	97279	22.1	2082	14.1	6.3	16.9	0.2557	2010.3	1.82
2	115570	24.2	1840	17.1	7.6	25.1	0.44587	2010.7	1.60
3	55907	23.9	1773	21.7	6.4	30.1	0.59535	2009.7	1.75
4	39028	24.7	1942	15.1	4.1	17.1	0.98735	2011.2	1.47
5	67504	25.8	2320	24.1	9.6	27.9	1.0273	2009.2	0.75

Table 2.2: Dengue clusters sorted by peak-time difference. Average climate variables are included as well as mosquito infestation indexes and population size. The R_0 was calculated using the initial growth rate proposed by Wallinga [1].

Cluster	Pop	Temp	Prec.	House	Resrvr.	Breteau	Dif. T.Peak	T.Peak	R0
1	81971	24.5	1845	22.0	9.2	33.7	0.066117	2015.1	2.12
2	69619	25.6	1639	31.3	8.6	34.6	0.069231	2015.1	1.76
3	149870	25.8	1368	17.4	6.4	19.7	0.086241	2015	2.18
4	107610	26.2	1297	33.1	13.6	46.7	0.14018	2014.9	2.68
5	348250	27.2	1168	38.0	18.5	48.0	0.2931	2014.8	2.70

Table 2.3: Chikungunya clusters sorted by peak-time difference. Average climate variables are included as well as mosquito infestation indexes and population size. The R_0 was calculated using the initial growth rate proposed by Wallinga [1].

warm municipalities could be suitable for mosquito development, their connections to other regions could be weak. Whereas, temperate weather regions could be somewhat suitable for *aedes aegypti* development while connecting faster to other regions in the country.

The municipality clusters are represented geographically in figure 2.17. In the case of dengue, the virus spread through most of the territory. Whereas for chikungunya, three different regions can be identified (North coast, NorthEast border and Center). In the center of the country there is a gap where epidemics occurred in dengue and where chikungunya has been absent. These are municipalities located in high altitudes where climate conditions prevent *aedes aegypti* development.

Municipalities from cluster 1 are predominant in both chikungunya and dengue, suggesting that in this region both chikungunya and dengue spread quickly. In the case of chikungunya, cluster five is located in the north coast section of the country. Whereas, for dengue, municipalities of cluster five are located mostly in the eastern part of the country.

The coherence of the clusters between chikungunya and dengue are illustrated with the common municipalities in table 2.4. Most of the municipalities in the various clusters of chikungunya are found in cluster 1 of dengue, few are found in the rest of the clusters, and the remaining municipalities with chikungunya are missing in the dengue clusters.

These results suggests that most of the municipalities of chikungunya are fast spreaders in dengue. Hence, it could be expected from this point, that the disease will spread rapidly through the rest of the municipalities found in cluster 1 of dengue. However, the impact of chikungunya could be expected to be higher than dengue since multiple municipalities without dengue epidemics have already reported a chikungunya epidemic.

	Den 1	Den 2	Den 3	Den 4	Den 5	NO dengue
Chik 1	29	4	2	1	0	16
Chik 2	9	1	1	1	1	3
Chik 3	5	2	2	0	1	6
Chik 4	3	2	1	1	1	10
Chik 5	1	1	0	0	0	1

Table 2.4: Clusters of dengue and chikungunya

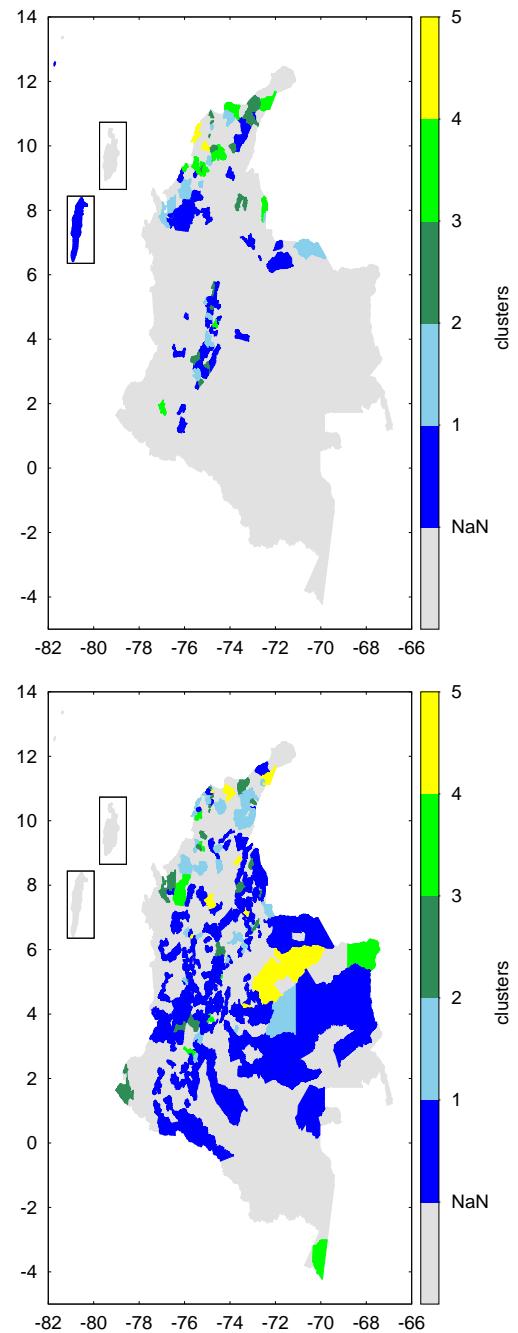


Figure 2.17: Clusters of dengue and chikungunya epidemics by municipality. Top graph shows chikungunya clusters. Bottom graph shows dengue clusters.

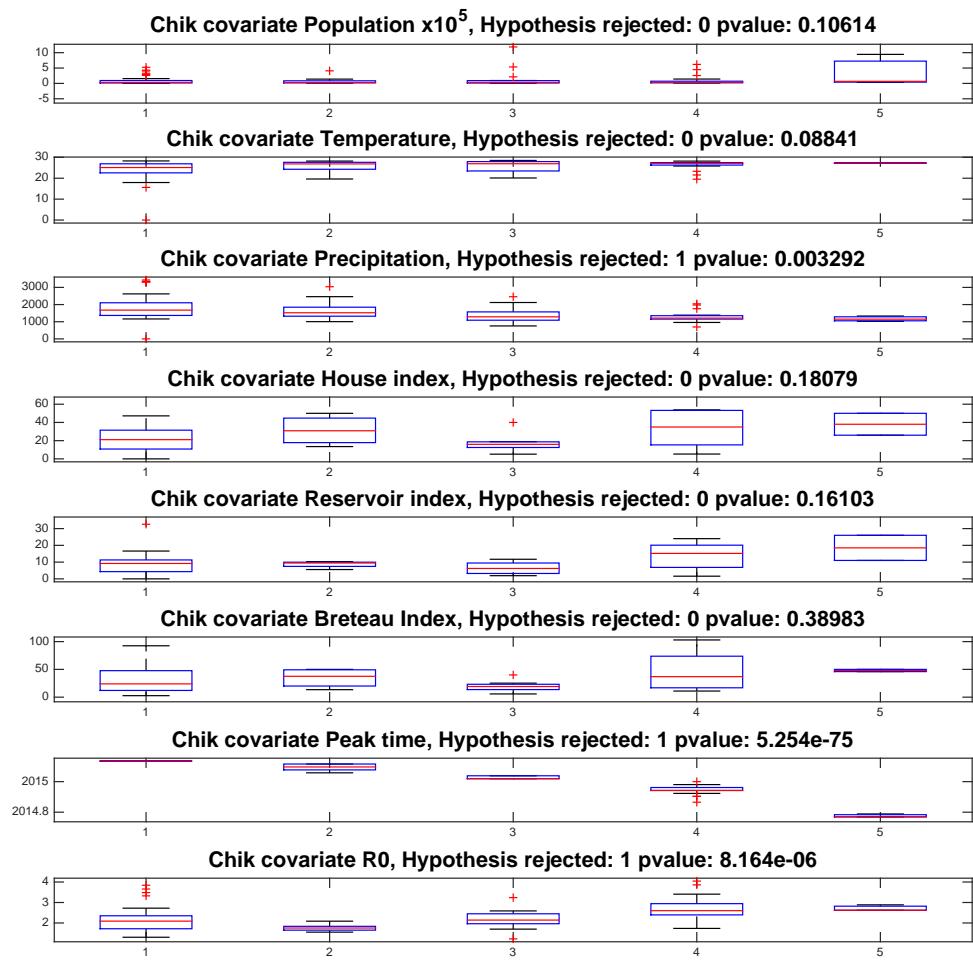


Figure 2.18: Multivariate Analysis of Variance for chikungunya clusters. Precipitation, peak time, and reproductive number variate within the difference clusters.

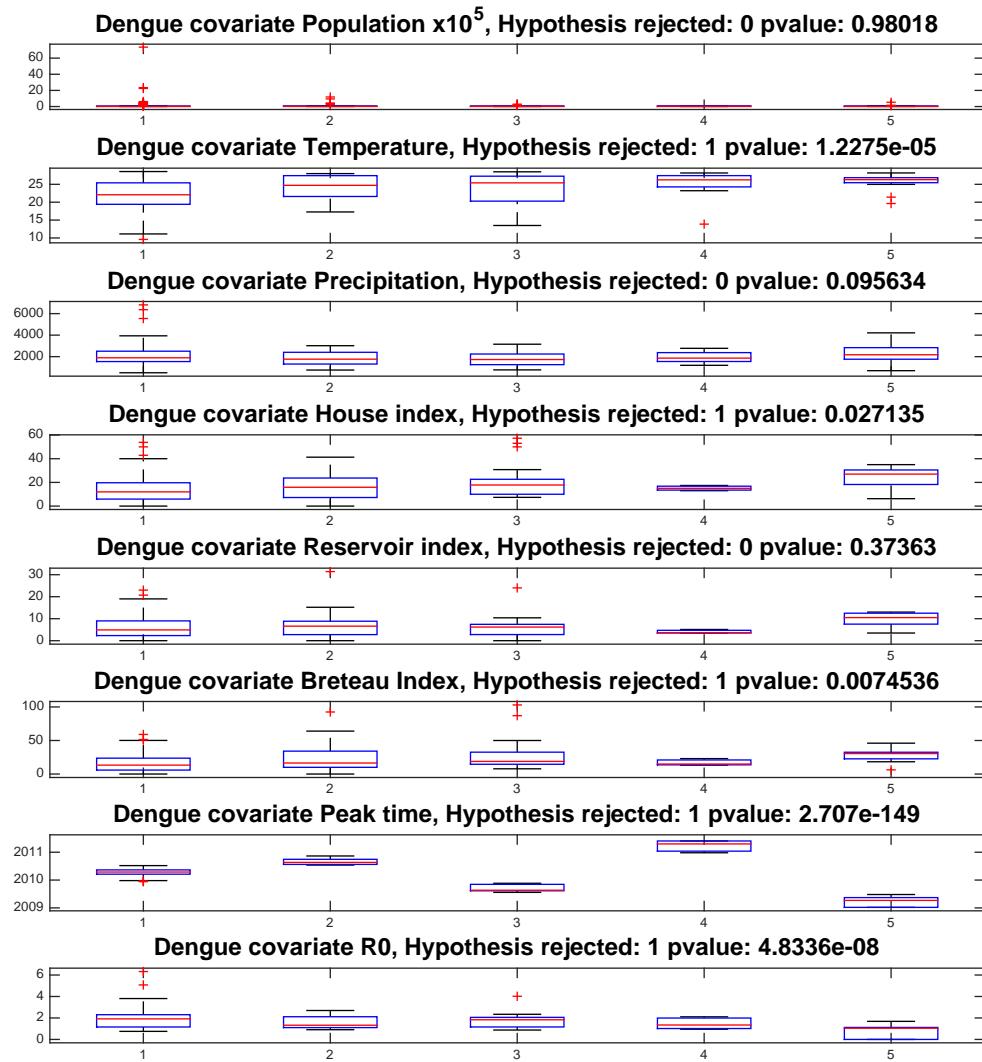


Figure 2.19: Multivariate Analysis of Variance for dengue clusters. Temperature, house index, Breutau index, peak time, and reproductive number variate within the difference clusters.

Multivariate Analysis of Variance (MANOVA) of Clusters' Covariates

The five clusters include information about the average values of covariates that are possibly related to the spread of the virus throughout the territory. However, it is difficult to make conclusions based only on tables 2.3 and 2.2. Hence, a MANOVA analysis was implemented in Matlab® to determine the variation of these 8 covariates within the multiple clusters.

Precipitation and peak-time differences present variability among the five groups of the chikungunya infected municipalities, as shown in figure 2.18. High precipitation values are related to groups with short peak-time differences, when the municipalities reported epidemics late in the analysis period. Whereas, long peak-time differences are characterized by low precipitation values and early epidemic manifestations.

Furthermore, reproduction numbers fluctuate between clusters. Low reproduction number values are found in clusters with short peak-time difference. While high reproduction number values characterize long peak-time difference. This relationship contrasts with the precipitation. Municipalities with high precipitation experience a low impact epidemic that has short transmission time difference. While low precipitation municipalities report a high impact epidemics early in the period of analysis.

Temperature is homogeneous within the chikungunya clusters. Nonetheless, for dengue, the MANOVA analysis suggests that the temperature varies within the groups. Higher temperatures distinguish clusters with long peak-time differences, while clusters with short peak-time differences have mild temperature values. Similar connection between infestation indexes and clusters are perceived.

Reproduction number is also significantly different among the clusters. Low reproduction numbers are associated to long peak-time difference clusters (4, 5). The groups with short peak-time differences have higher reproduction number values. This, differs from the observations in chikungunya, where high R_0 is related to long peak-time difference. Also, the values are significantly lower than those calculated for chikungunya, implying a higher impact in chikungunya epidemic than in dengue.

Peak times are also different among the clusters, the first and last groups to report epidemics are slow spreaders. Whereas fast spreaders are in the middle of the reports. This is contradictory to the case of chikungunya, where fast spreaders are related to the groups with late epidemics. Nonetheless, chikungunya reports include early transmission stages of the virus throughout the country. Consequently, at the end of the epidemic this group will be expected to rank in the middle of the epidemic, as in dengue.

In both cases, the population does not determine the transmission in any cluster. The differences of transmission are attributed to other characteristics such as climate variables like temperature and precipitation. The homogeneous temperature in chikungunya indicates that the epidemic has been focused to high temperature regions. This behavior is expected to change when the epidemic reaches more regions.

2.4 Summary

In this chapter, dengue and chikungunya incidence in Colombia were described using various methods. Wavelet analysis was used to describe the differences among departments in dengue transmission. Epidemics of chikungunya and dengue were analyzed for recent reports in the surveillance system.

From 1997 -2012 a wavelet analysis illustrates the characteristics of transmission of the virus in various regions of the country. Departments were clustered using a distance matrix from the maximum covariance analysis that quantifies the difference of the wavelet spectra of pairs of time series. Four clusters were identified with various characteristics. In general, northern regions are characterized by higher impact of dengue.

Municipalities with dengue can be used as an indicator of chikungunya epidemics. An agreement of 65 % was found between municipalities infected with chikungunya compared to dengue. Chikungunya epidemics were identified for 105 municipalities while 342 municipalities reported dengue in the 2010 epidemic. However, 35 % of the municipalities with chikungunya did not report dengue in 2010. This suggests that the chikungunya epidemic impact can be higher than the 2010 dengue epidemic.

The municipalities with dengue and chikungunya were clustered using the difference between peak times of the epidemic. The majority of the municipalities with chikungunya were found in the cluster

of dengue corresponding to short peak-time differences. Using dengue as a predictor of chikungunya, the epidemic has reached a region of rapid spread. In general, the municipalities with chikungunya showed more homogeneity among multiple covariates related to vector-borne transmission, compared to dengue.

MANOVA analysis show that the clusters have heterogeneous reproductive numbers for chikungunya and dengue. Precipitation was significantly different among chikungunya groups, while temperature was more homogeneous. For dengue, temperature showed more variability than the precipitation. This suggests that the transmission of chikungunya has been constricted to homogeneous climate regions. More heterogeneity in climate conditions can be expected at the end of the epidemic.

Chapter 3

Synthetic Population of Colombia

Agent-Based models represent the interactions of agents in an environment. FRED simulates epidemic dynamics in realistic populations. Hence, this type of model require agents and environment that resemble statistical counts of the real population. This is called a synthetic population and is used as an input in the simulation, as illustrated in figure 3.1. Several inputs are needed to simulate epidemics. This chapter describes the process of creating a synthetic population for Colombia.

The synthetic population consists of a list of individuals' information with demographic characteristics such as race, sex, age, etc. Household characteristics are also included, e.g. household size, head of the household, etc. Additional activities like school attendance and employment status are included in this specific synthetic population. The synthetic population of Colombia includes these types of characteristics that match the demographics of the country according to the latest national census conducted in 2005.

Specific geographical location of houses, workplaces, and schools, provide information useful to represent geographical specific epidemic dynamics. This is important to incorporate in the model since disease transmission can vary from rural to urban areas. Hence, households are located to match population density and land-use high resolution grids. Workplaces and schools are located based on the synthetic population density.

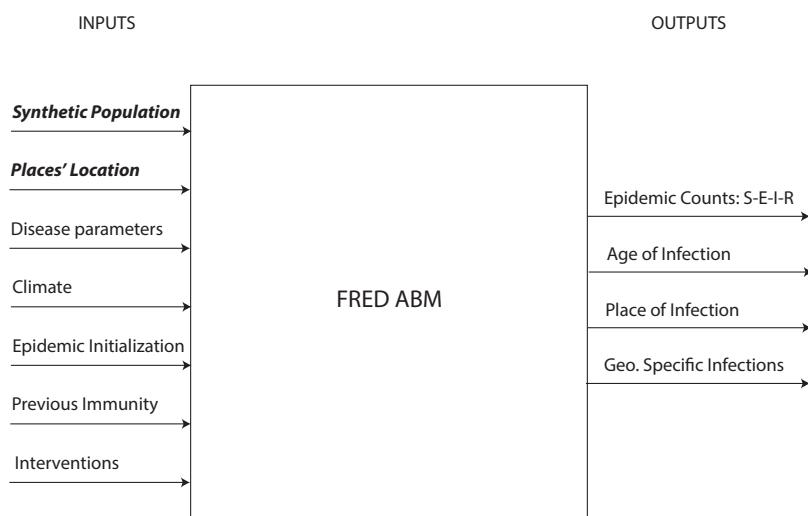


Figure 3.1: Model's inputs and outputs. The synthetic population and places locations are inputs of a facsimile ABM.

3.1 Generation of the Persons and Households Attributes of the Synthetic Population

The Public Use Microdata Series - International data base (IPUMS) database has the largest collection of census micro-data samples around the world, and it provides access to researchers free of charge [104]. In the case of Colombia, there are samples available from five different censuses (1964,1973,1985,1993,2005). For this synthetic population, individual samples were selected from the most recent census dataset (2005). The dataset includes a sample of 4,006,168 persons, which represents 9 % of the population for 2005. There are multiple variables included in the census for both individual and household records. Variables relevant to the study are listed in table 3.1.

Colombia consists of two different geographical levels: departments and municipalities. There are 33 departments in the country and 1122 municipalities with a total population of 45,509,584 inhabitants [105]. The distribution by gender and age is shown in figure 3.2. As shown, there is a high proportion of people under 30 years of age, compared to the other age ranges. Also, the gender proportion of the population is homogeneous for most of the age groups.

The variables of interest to match the census are: population size by municipality, age, school attendance and employment status. Census estimates for these variables are available at the National Bureau of Statistics of Colombia (DANE) [105]. The population estimates and the age distribution chosen to fit are based on 2010 estimates, since it is the most recent dataset with geographical density distributions available.

Multiple algorithms have been proposed to generate synthetic populations from data such as the IPUMS datasets [106, 107, 108]. These algorithms are based on the Iterative Proportional Fitting (IPF), first proposed by Deming and Stephan in 1941 [109]. A variation of the IPF called the General Iterative Proportional Updating (IPU) algorithm proposed by Ye *et al*[106] can match both household and individual characteristics of a population. This algorithm can be implemented with ease and is computationally efficient. Because of its characteristics, the IPU algorithm is used in this study to generate the synthetic population of Colombia.

The algorithm matches the household attributes by adjusting each household weight based on the IPF. Then, for each attributes, the weights of the households that have the attribute are updated to match the objective distribution. This process is performed iteratively, until the error between the synthetic records and the data is considered low enough. The household records are then withdrawn randomly from the fitted database. Specifically, this algorithm was implemented to match each municipality population, urban status, age distribution, and number of students and workers. After each iteration, the goodness-of-fit was calculated using the χ^2 test shown in equation 3.1. If the synthetic population fails to meet the criteria, the households were drawn again from the dataset until the criteria was met or a maximum number of tries was achieved. The procedure is described in algorithm 1.

$$\chi^2 = \sum_j^N \frac{(O_j - E_j)^2}{E_j} \quad (3.1)$$

A linear regression was performed to evaluate the difference of the population by municipality compared to the synthetic population. As shown in figure 3.3, the synthetic population represents adequately the population size of each of the municipalities of the country. The R^2 of the linear regression is approximately 1. Moreover, the χ^2 goodness-of-fit test is computed to assess the similarity between the age distributions of the census and the synthetic population. The rejection region with 5 % of significance level, for the 17 groups of age is $\chi^2 \geq 26,296$. Hence, if the test value is less than

Resolution	Attributes
Households	Persons, Weight, Urban,GeoLevel1,GeoLevel2,House Serial
Persons	House Serial, Weight, Relate, Age, Sex, Race, School Attendance, Employment Status

Table 3.1: IPUMS household records included in the synthetic population.

5 % the distributions are considered statistically similar. In figure 3.4, it can be observed that the majority (97 %) of the municipalities can be considered to have a statistically similar distribution to the census reports.

In figure 3.5 there are opposing examples of the age-structure of two municipalities. One passed the χ^2 test and the other did not. There is no appreciable difference between the census and the synthetic population counts for each age group for the example that met the criteria. In the case of the municipality with a poor χ^2 result, the synthetic population overestimates the size of the first age group and the groups over 45 years of age. However, these differences are low in qualitative terms. The difference on population sizes is appreciable for both populations. In general, the algorithm has better performance for high populated municipalities compared to low populated municipalities.

In addition to the age structure of the population, the number of students and workers were fitted as shown in figure 3.6. The R^2 is approximately 1 for both attributes, which means that the number of students and workers by municipality represents adequately the census estimates. Furthermore, the gender proportion is shown in figure 3.6 as females by municipality, although this characteristic was not fitted by the algorithm, it matches closely to the census.

In general, the characteristics of the population covered in this section are represented by the synthetic population of the country. This synthetic population will be used as an input of the agent-based model to assign the demographic characteristics to each agent. Hence, the model will be able to simulate the spread of the disease in an age-structured population that closely matches the census. Each agent is related to a household from the micro data obtained by IPUMS. The next step is to assign a geographical location to each household to represent appropriately the geographic distribution of the population across the country.

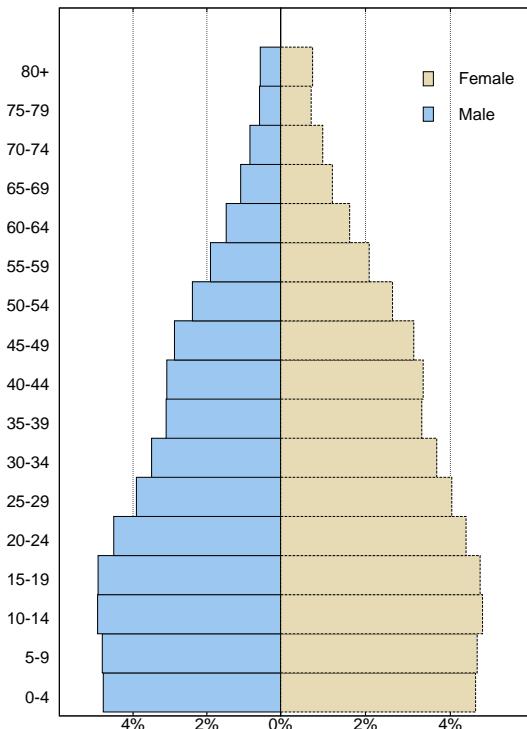


Figure 3.2: Population pyramid of Colombia in 2010. Blue bars represent the males proportion in each age group. Beige bars represent the females proportion in each age group.

```

for  $mun = 1$  to 1122 do
    Estimate the set of constraints for each variable (Person and Household):  $C_j$ ;
    Create a matrix D with dimensions  $N \times m$ , Initialize an error measure  $\delta = \frac{\sum_j \|\sum_i d_{i,j} w_i - c_j\|}{m}$ ;
    where  $N$  is the number of households in the sample and  $m$  is the number of characteristics to fit ;
    Adjust the initial weights based on the standard IPF:  $W_j$ ;
    for  $j = 1$  to  $m$  do
        |  $S_j$ : Column of characteristic  $j$  with entry  $s_{qj}$  where  $q$  are non-zero elements;
    end
    Initialize  $r = 1$ ;
    while ( $\Delta > threshold$ ) and ( $r \leq MaxIter$ ) do
        | Calculate the adjustment factor for the constraint  $j$ ,  $p = \frac{c_j}{\sum_q d_{s_{qj}} \times W_{s_{qj}}}$ ;
        | Update weight for records with the characteristic  $j$ ,  $w_{s_{qj}} = p \times w_{s_{qj}}$ ;
        | Update the error:  $\delta$ ;
        | Calculate improvement in goodness-of-fit:  $\Delta = \|\delta - \delta_{prev}\|$ ;
        |  $r++$ 
    end
    Initialize  $k = 1$ ;
    while ( $\chi^2 > Critical \chi^2$ ) and ( $k < MaxIter2$ ) do
        | Initialize  $upop = 0$ ;
        | Initialize  $rpop = 0$ ;
        | while ( $upop < UrbanPop(mun)$ ) do
            | | Withdraw a house randomly from the urban matrix with updated weights;
            | |  $upop = upop +$  household size;
        | end
        | while ( $rpop < RuralPop(mun)$ ) do
            | | Withdraw a house randomly from the rural matrix with updated weights;
            | |  $rpop = rpop +$  household size;
        | end
        |  $k++$ ;
        | Calculate  $\chi^2$  ;
    end
end

```

Algorithm 1: Implementation of the IPU algorithm used to match synthetic population attributes to census data [106]

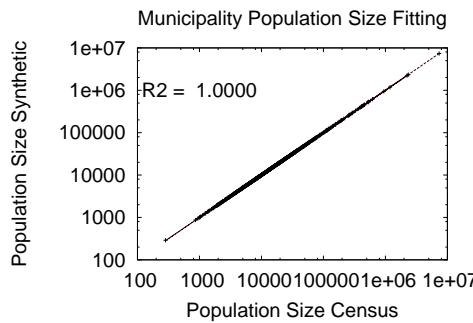


Figure 3.3: Regression analysis of the synthetic population estimates compared to the census data for each municipality.

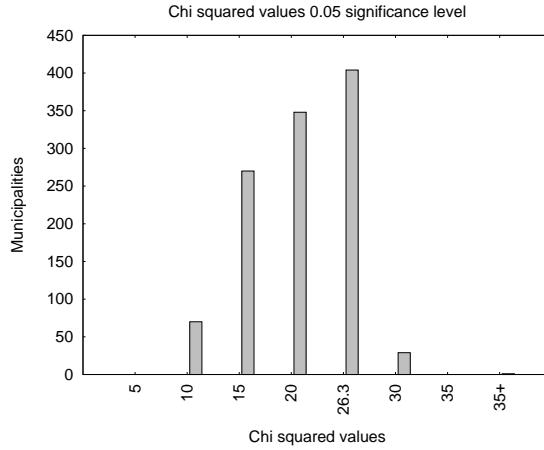


Figure 3.4: Histogram of χ^2 values for the age-distribution of the 1122 municipalities of Colombia

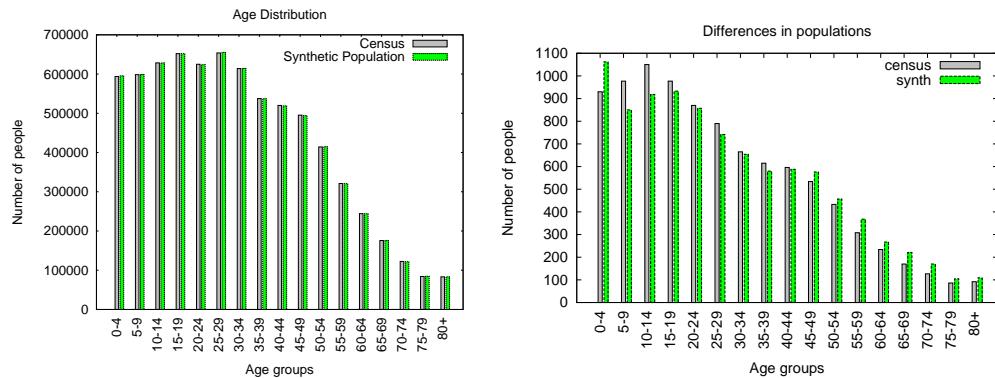


Figure 3.5: Examples of age distributions of the synthetic population compared to the census. The left graph shows a municipality with an acceptable χ^2 . The right graph shows a municipality with a high χ^2 that represents a poor fitting.

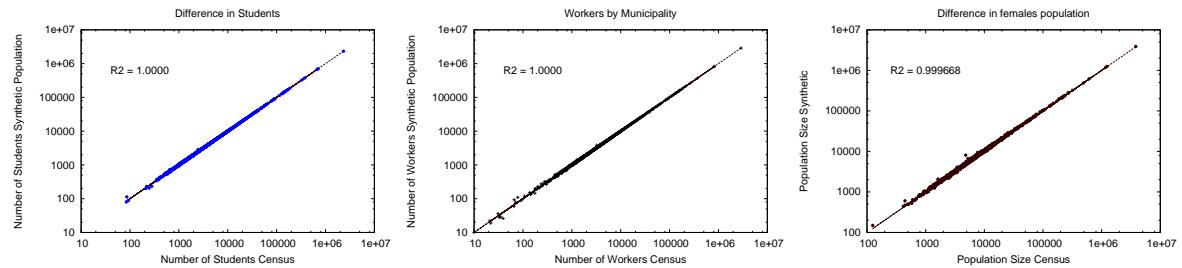


Figure 3.6: Regression analysis of the synthetic population estimates of students, workers, and females compared to the census, by municipality.

3.2 Geographical Location of Households

The IPUMS database includes two levels of geographical resolution for each household record: department and municipality. Also, the records include the urban status of the household, i.e. urban or rural. Population density and land-use grids from open source databases were used in order to assign households' geographical location. A population density grid of 2010 was obtained from the Worldpop project with a resolution of approximately 100 m^2 . The grid contains estimates for people per pixel and is adjusted to the national UN population estimates [110, 111]. Worldpop datasets have been compared to other open access datasets such as the Global Rural-Urban Mapping Project (GRUMP) [112, 113, 114], and has proven to be more up to date and accurate for public health applications [115, 116].

Moreover, the urban-rural status grids from GRUMP with a resolution of 30 arc seconds [112, 114]. These datasets differentiate rural and urban areas based on population counts, nighttime lights, and settlement points. The urban-status grid is used to match the urban status of household records from the census, hence maintaining the urban and rural household attributes. Additionally, buildings, house, and school locations were downloaded as well as roads of Colombia from the OpenStreetMap project (www.openstreetmap.org). The latter is a dataset of free geospatial data built by a community of mappers around the world. The administrative boundaries for departments and municipalities were also downloaded from the Colombian Geographic Information System For National And Regional Comprehensive Land-Use Planning and Management Project [117].

A grid of the entire country was created with a 1 Km^2 resolution, and the detailed information was assigned to each cell, as illustrated in figure 3.7. For each municipality, household records were withdrawn from the synthetic population tables and assigned to a possible set of cells depending on the rural or urban status of the household. Next, each household was assigned randomly to a cell of the subset of cells not yet filled to capacity. The household size was then added to the patch, reducing its capacity. Each household location from the open street database was assigned to a location close to that point. If there was not a building in the cell, then each household was located randomly inside the cell. If all the cells were full before the households were located, then the capacity of the cells were incremented. This procedure continued until all the houses were located.

The population density grid from WorldPop and the synthetic population density grid are shown in figure 3.8. Qualitatively, both grids look similar in that the high and low populated areas are well represented by the synthetic population. However, the synthetic population is less homogeneous than the WorldPop density grid in the medium populated regions.

In order to statistically estimate the accuracy of the location of the households based on the population density grids, we used the spatial correlation (equation 3.2) that has been proposed as a

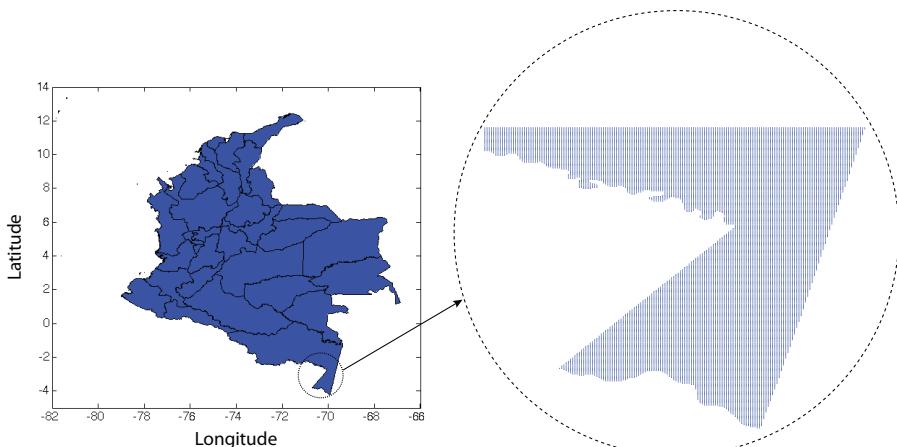


Figure 3.7: The territory of Colombia divided in 1 Km^2 cells. Each cell contains information about the municipality, department, urban status, and population density.

metric to compare grid-based population datasets by Sabesan *et al*[118]. The correlation was calculated for each department as illustrated in figure 3.9. The results show that most of the departments have a high correlation value. However, some small departments show poor performance. Nonetheless, it can be said that the majority of the departments reasonably represent the population density of the country.

$$\rho = \sum_i^N \sum_j^N \frac{(X_{(i,j)} - \bar{X})(Y_{(i,j)} - \bar{Y})}{N^2(\sigma_x \sigma_y)} \quad (3.2)$$

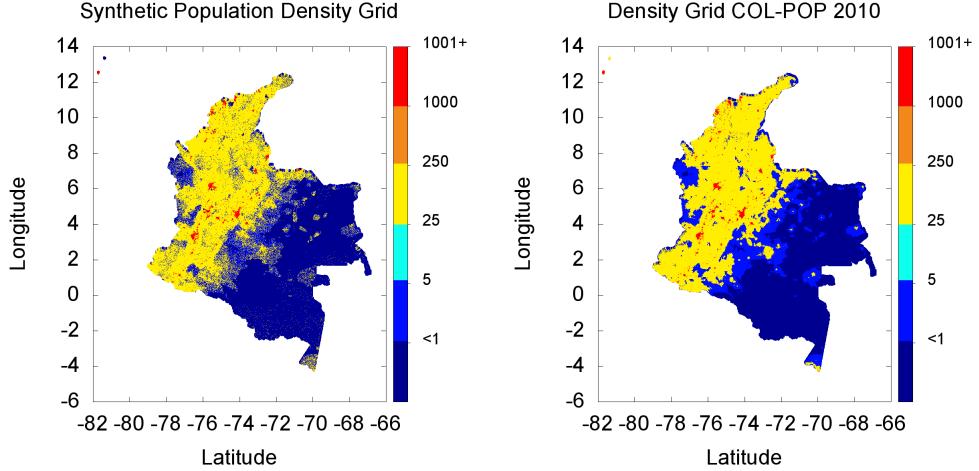


Figure 3.8: Graphical representation of the population density for the synthetic population (Left) and WorldPop (Right).

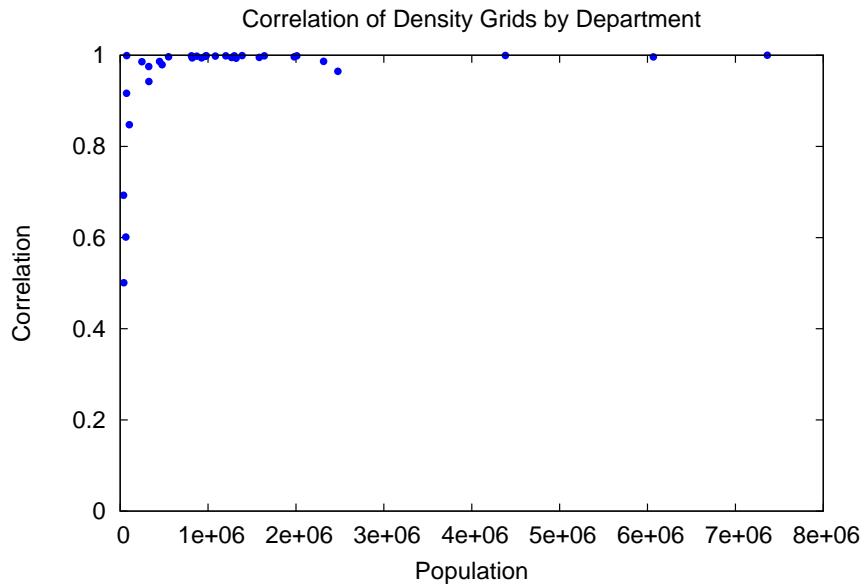


Figure 3.9: Spatial correlation between WorldPop and the synthetic population by department.

3.3 Schools' location and assignment

In addition to households, vector-borne diseases can be transmitted in different places such as schools. If a school is suitable for the *aedes* mosquitoes, then due to the high density of children, it will play an important role in dengue and chikungunya transmission. The transmission in schools could increase the scope of an epidemic to multiple areas within and even outside of a town. Hence, agents were assigned to schools based on the school attendance variable of the synthetic population. The synthetic students were assigned to different types of schools i.e. pre-school, primary, secondary, or a university. Datasets were obtained with individual school information from the ministry of education of Colombia (www.mineducacion.gov.co) with the characteristics shown in table 3.2. Whereas the university dataset includes the capacity of students by institution, the school dataset excluded this information.

In order to represent the geographical position of schools in a realistic way, two different methods were used to locate the schools inside each municipality. The two methods were based on the OpenStreetMap database and student density. When there were enough school locations from the OpenStreetMap dataset by municipality, the schools were located at those points. If the points were insufficient, the rest of the schools were located based on student density by municipality. These densities were computed using the location of households and calculating the density of students by each grade. After the schools were located, an initial capacity was assigned to each school, based on the total number of students and schools by municipality and grade. At last, students were assigned to a school based on their age, distance, and municipality. Enrollment differences in public versus private school was ignored, since the IPUMS and the Ministry of Education datasets omit this information. Hence, all students were assigned to schools using the same algorithm.

Basically, the school assignment algorithm was based on the methods reported by RTI international [119]. As a first step, students were assigned to attend to a school in the same municipality as their home, or outside the municipality in proportion to the available data of the 2005 census of Colombia. Then, a list of schools based on municipality, grade, and capacity was created for each agent. From this list, a school was randomly selected from the five closest schools for each agent. If the school was full and there were still agents to enroll, then the capacity of the school was increased. In the case of college students, the same procedure was applied except that the universities were selected by department and not by municipality, since it can be assumed that college students attend to a university not necessarily close to their residence.

Figure 3.10 illustrates an example of these distributions. The final capacities of the schools are summarized by national level and for the departments of Quindío and Amazonas. Real data about the capacities of the schools was unavailable, hence these capacities could not be compared. The resultant school capacities could be improved if the appropriate data were accessible.

Dataset	Attributes
Schools	Individual school record, Urban status, GeoLevel1, GeoLevel2, Grades
Universities	Individual university record, GeoLevel1, GeoLevel2, Enrolled students by year (2000 - 2013)

Table 3.2: Schools datasets from the Ministry of Education of Colombia

Grade	Age
Kindergarten	< 6
Primary	6 - 10
Secondary	11 - 17
University	≥ 18

Table 3.3: Schools grades equivalent in age

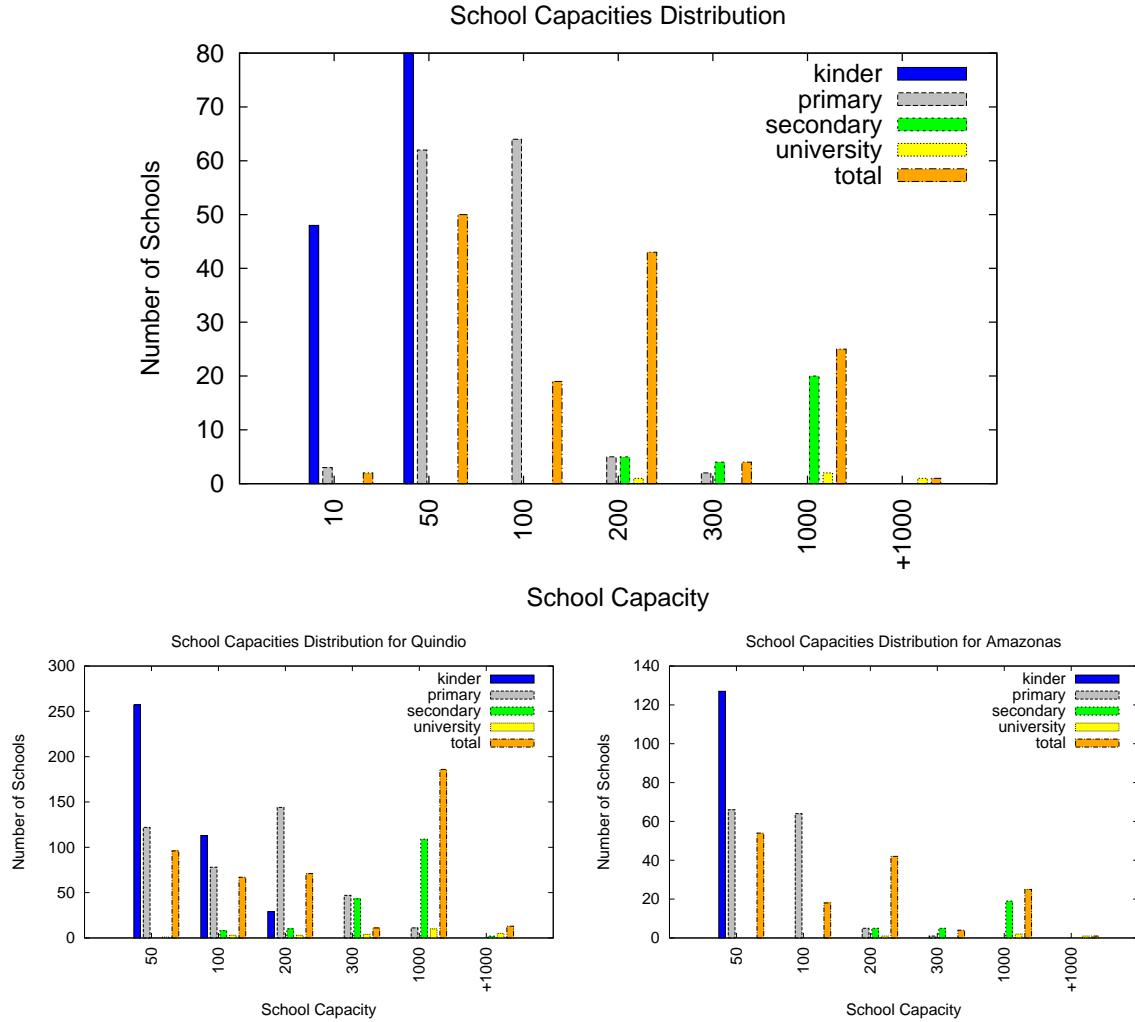


Figure 3.10: Synthetic school capacities by grade and total capacity by school.

3.4 Workplaces location and assignment

Similarly to schools, it is important to include the contacts that agents may have with mosquitoes in the workplace, in order to represent the dynamics of vector-borne disease transmission. Nonetheless, some assumptions for school location and assignment are different for workplaces. For instance, workers do not necessarily live close to their workplace, so the assignment assumption would not be based on the closest workplace. In addition, the Colombian census includes variables for workplaces, such as capacities that are omitted in the school database. These workplace variables that are available on the census by municipality appear in table 3.4.

Variable	Range
Number of workplaces by category	1, 2-5, 6-9, 10, 11-20, 21-50, 51-200, 201-500, 501+
Commuting times distribution (min)	< 10 , 10 - 15, 16 - 30, 31 - 45, 46 - 60, > 60
Proportion of people that work out of the municipality of residence	%In, %Out

Table 3.4: Workplace attributes from the National Census of 2005.

Table 3.4 summarizes a dataset of workplaces that was created based on files from the census. An attempt was made to create a list with the exact number of workplaces in each municipality group. However, the census information was incomplete or incoherent for some municipalities. For example, in small municipalities, the number of people that reported to work inside was larger than the total capacity of the workplaces in the same municipality. For this reason, a list of workplaces was created that retained the same capacity distribution as the census, but with the necessary amount of workplaces. This adjustment, guarantees that for each municipality, there is a sufficient capacity of workplaces for its workers.

In order to accurately locate workplaces inside each municipality, points were located for buildings from OpenStreetMap dataset. School points were also included since they are workplaces as well. Workplaces were designated to these points when possible. When points were unavailable, workplaces were located based on the workers' density of the synthetic population.

There were two different options when assigning people to workplaces, employment within or outside the municipality. For those who were chosen to work in the municipality, the allocation to workplaces was assigned based on commuting times. These commuting times were translated to distance using a baseline speed of 30 Km/h. Hence, people were not assigned to the closest workplace, but to one of the closest workplaces to a predetermined commuting distance as shown in figure 3.11. For each agent, five workplaces were selected within the commuting distance. Then, from those five, one workplace was randomly picked. In contrast, if an agent worked outside the municipality, a random workplace was selected from the list of the other municipalities within each department.

The capacities of the workplaces were differentiated as small (1-50), medium (51-200) or large (>200) in order to make comparisons to the census reports. In figure 3.12, the linear regression of the census contrasted to synthetic workplaces is shown. The graphs illustrate that for the most populated municipalities the results are closer to the census than they are for the small municipalities. This can be attributed to the fact that small municipalities had more discrepancies between workplace capacities and number of workers. Nonetheless, the fit is acceptable for the three categories. The χ^2 goodness-of-fit test was used to assess the similarity between the capacities distribution. Around 80% of the municipalities had an acceptable χ^2 value as shown in figure 3.13.

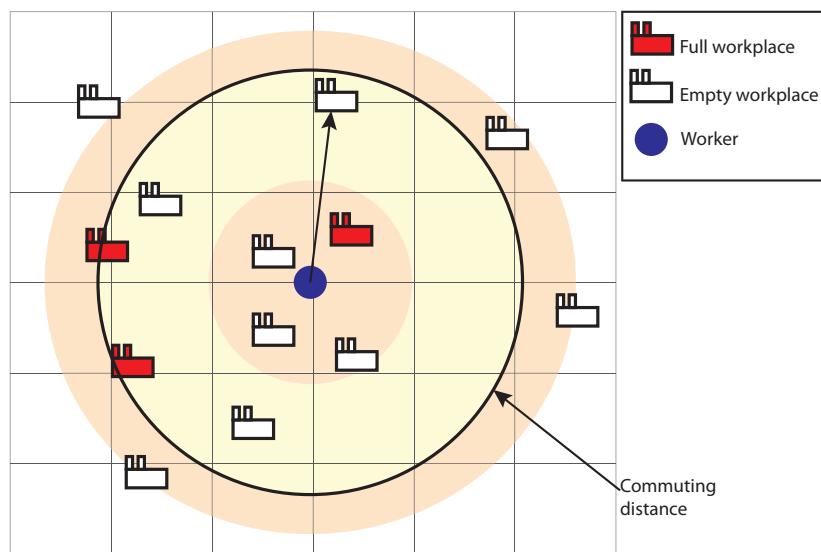


Figure 3.11: Graphical representation of the workplace assignment based on commuting distances.

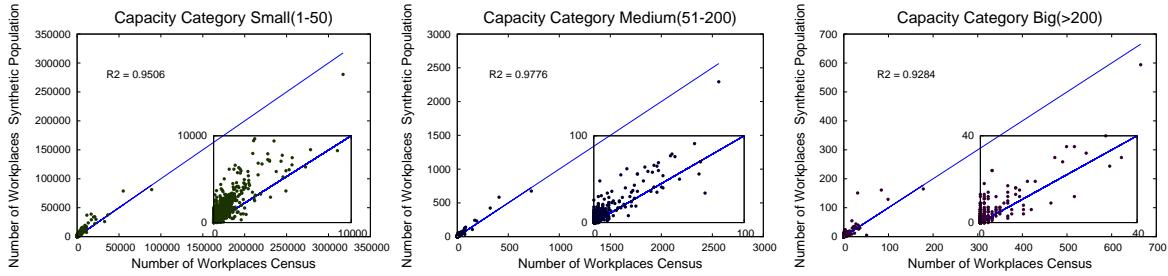


Figure 3.12: Regression analysis of the number of workplaces by each capacity category. From left to right: small, medium, large.

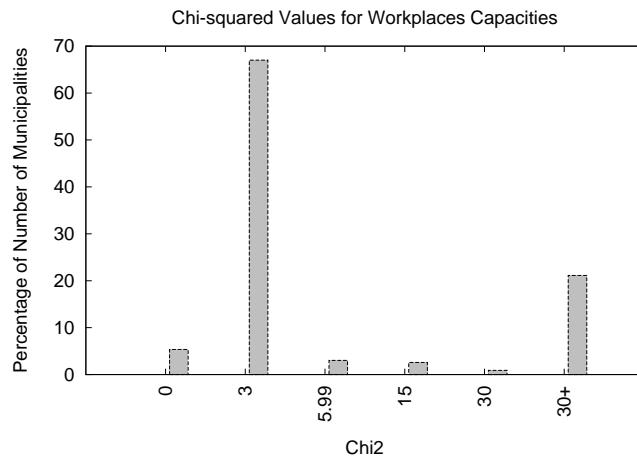


Figure 3.13: Histogram of the number of workplaces by χ^2 goodness-of-fit measure. 70 % of the workplaces are drawn from the same distribution than the census reports. 20 % are highly dissimilar.

Summary

In this chapter the process of generating and validating the synthetic population was described. The synthetic population that was generated accurately matches the population size and the age-structure of the Colombian population of 2010. Also, the number of students and workers matched the estimates of the census. In addition, a number of synthetic places including households, schools and workplaces were created. The households were located to match population density grids and urban status. Schools and workplaces were located based on student and worker density. Furthermore, each agent was assigned to an activity, either study, work, or stay at home. The students were enrolled in schools based on proximity and age. Whereas, workers were assigned to workplaces based on commuting distances. In general, the synthetic population describes an age-structure population of Colombia with the basic activities included. This dataset can be used to reproduce vector-borne disease dynamics or to support other agent-based models that incorporate similar variables.

Chapter 4

Large-Scale Vector-borne Agent-Based Model

The vector-borne disease model was implemented using the FRED platform developed by the University of Pittsburgh [79]. FRED is an open-source model that simulates epidemics based on realistic synthetic populations. The platform allows for the inclusion of multiple diseases and strains. Also, because it is highly modular, it can be adapted to model diverse diseases with different transmission patterns. In general, FRED is divided by agents, places, and layers.

Agents in FRED represent humans from a specific population with predetermined demographic and health characteristics. Each agent performs a set of activities each day and have different behavior characteristics as shown in table 4.1. The environment of FRED is represented by places such as households, schools, workplaces, and neighborhoods (Neighborhoods are represented by $1km^2$ cell). Other types of places such as hospitals, prisons, and college dorms are ignored in this study.

FRED is a discrete time simulator with each time unit equal to one day. Each day, agents perform different activities and have contacts with other agents that share activity locations in the same day. These locations are visited daily by agents with varying health status. Whenever an infectious agent visits a place, that place is marked as infectious and other agents may acquire the disease. Agents can also transport the pathogen to different cities as they travel. In order to represent vector-borne disease transmission, a vector layer was added to the FRED platform. The mosquito population is controlled by the vector layer using additional variables important in vector-borne disease transmission, such as specific temperature by location, initial number of infected mosquitoes, mosquitoes/human ratio, among others.

4.1 Vector Layer

The vector layer is composed of $1km^2$ patches. In the vector layer, the initial setup of the vector model establishes all the variables needed for vector-borne disease transmission, such as the number of mosquitoes per human, average mosquito daily biting rate, mosquito lifespan, etc. The initial

Model	Variables
Demographics	Age, sex, race, household income, date-of-birth
Health	Current infections, immunity, at-risk status, susceptibility, current symptoms levels
Activities	Households, Schools, Classrooms, Workplaces, Offices, Neighborhoods, Travel
Behavior	Stay at home when sick, keep child home when sick, accept vaccines, refuse vaccines, flip behavior, imitate behavior

Table 4.1: Agents' characteristics in FRED

conditions of the immune status of the population is computed for specific regions. Also, the specific temperature by location is specified in the vector layer.

Temperature

The model reads the annual average temperature from input datasets and assigns it to specific patches. Places that are located in those patches are later assigned with the corresponding temperature. For Colombia, the temperature grid was downloaded from the Worldclim project [103]. The dataset comprises the temperature in celsius degrees by each 30 arc seconds ($1Km^2$ near the equator) of land as illustrated in figure 4.1.

Including the temperature in the model is crucial since it influences *aedes aegypti* survival. The *aedes aegypti* is modeled since it is the most common chikungunya and dengue vector in Colombia. The model allows mosquitoes in regions with temperatures above $18^\circ C$, based on observations suggesting that their survival is limited to warm regions [3]. Most of the territory of Colombia has warm annual temperatures, except for the mountains that divide the country. Highly populated towns are located in these mountains, which limits the population that lives in regions where the *aedes aegypti* survives. Although autochthonous transmission is unfeasible in these regions, contagion is possible when traveling to risk zones.

The development of the mosquito also depends on the temperature [120]. In the model, only the female population is considered since they are the ones that bite humans in order to lay eggs, therefore participating in the transmission dynamics. The adult-female population size in terms of pupae and temperature is calculated using equation 4.1, where PD_{temp} is the development time based on the temperature.

$$\text{Female Adults} = \text{Pupae} \cdot \frac{1}{PD_{temp}} \cdot \text{females proportion} \cdot \text{rate of successful emergence} \cdot \text{lifespan} \quad (4.1)$$

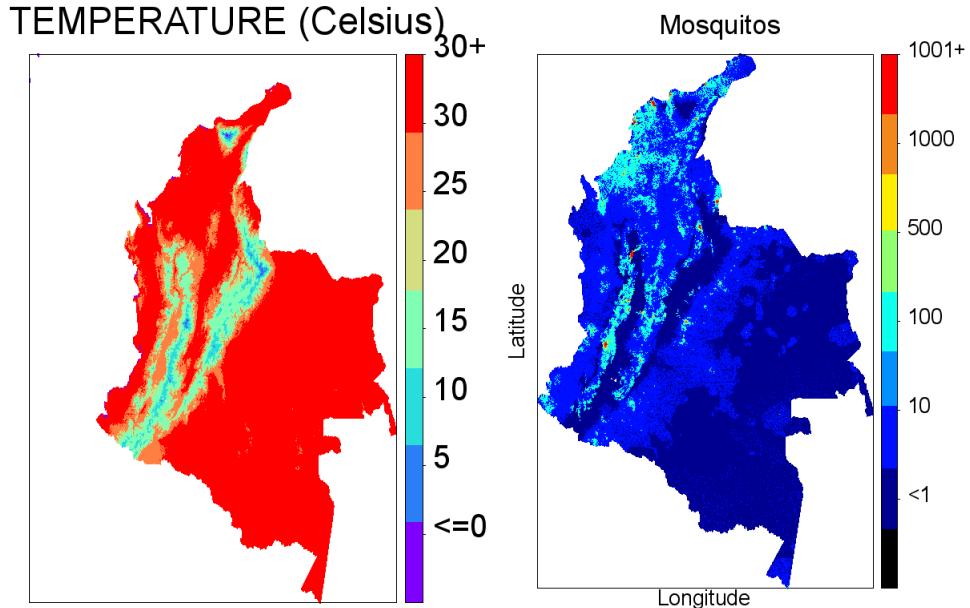


Figure 4.1: Mosquito density based on annual temperature values. Left, annual temperature in Celsius for a grid of 30 arc seconds of resolution, obtained from WorldClim project. Right, mosquito resultant grid of 30 arc seconds of resolution based on temperature and population density, calculated using equation 4.1.

High density areas located in warm regions lean toward high mosquito density in the model. Figure 4.1 illustrates the influence of temperature and population density on the mosquito pupae density. As shown in previous chapter, several high density areas are characterized by low temperatures. Hence, the population size living at risk areas is significantly reduced.

Travel model

Chikungunya and dengue disease spread is often restricted to human mobility because of the residential nature of *aedes aegypti* mosquitoes. In this model, human mobility is represented by daily commuting activities and overnight travel. Daily commuting activities include school attendance, workplace attendance, and neighborhoods interactions. These movements can occur in various levels, in or out the municipality of residence, including surrounding departments. Overnight travel occurs between departments.

Interdepartmental travel impacts the epidemic dynamics because it connects different regions of the country. This is specially important to study national level epidemics, where the virus is introduced in a region and it spreads throughout the country. In Colombia, long-distance travel could affect vector-borne diseases because of the heterogeneous climate conditions. Susceptible people move within regions of low risk to high-risk regions participating in the transmission dynamics and, possibly, introducing the pathogen to new areas.

In Colombia, long-distance travel is predominantly done by airplane, bus or car. Ground transportation data is unavailable from open sources. Whereas, air traffic information is collected by the air traffic department in Colombia (www.aerocivil.gov.co). Hence, in this study air traffic data was used to represent interdepartmental movement. Airports information, such as IATA codes, were obtained from the *Openflights* project (www.openflights.org).

Air transportation is an incomplete representation of the long-distance travel of Colombia since some departments are disconnected by plane routes. However, a transportation model can be implemented to represent the missing connections for those departments. Human movement between cities has been described using mainly gravity models [121], and recently by alternative representations like diffusion models [122].

Gravity models represents the movement of people between two large population areas using the newton law of gravity. Gravity model for movement of persons was first proposed by Zipf in 1946 [121]. Zipf *et al.* compared the model to the flow of people within cities of the United Stated by train, bus, and plane. The results suggested that people transportation between pairs of cities is directly proportional to the product of their population size and inversely proportional to the distance between them, as shown in equation 4.2.

$$Y \sim (P_1 \cdot P_2)/D \quad (4.2)$$

Gravity models have been used to represent transportation within large areas [121, 123, 124]. Jung *et al.* represented highway transportation within cities in Korea adjusting a gravity model to tolls data of 2005 [123]. Additional parameters have been included in the gravity model to incorporate political, economical, and demographic factors. For instance, Karemara *et al.* implemented a modified version of a gravity model to analyze international migration to North America, finding population of country of origin the most important sole factor in migration figures [124].

Flight data in Colombia shows a correlation between people flow compared to population and distance between departments' center points, as illustrated in figure 4.2. Large number of people transported is associated to large populated departments separated by short distances (high Y). While small departments separated by long distances (low Y) have low flow of people between them. In consequence, this data can be represented by a gravity model.

The number of people traveling from each pair of departments is represented by equation 4.3.

$$Y_{i,j} = C_{i,j} \frac{(P_i^\alpha \cdot P_j^\beta)}{f(D_{i,j})} \quad (4.3)$$

where $P_{i,j}$ represents the population of two different cities, $D_{i,j}$ represents the shortest transportation distance between the pair of departments, $f(D_{i,j})$ represents a function of the distance, here assumed to be $f(D_{i,j}) = \gamma \cdot D_{i,j}$, and $C_{i,j}$ is a constant.

The parameters of the travel model were adjusted to represent the flow of people for each pair of departments with flight traffic data available. The difference between the model and the data was minimized using a Global Optimization algorithm from Matlab®. The resultant parameter values are listed in table 4.2. The gravity model outcome shows a linear correlation with the travel data with a positive slope. The regression analysis of the model against the data is shown in figure 4.3. Although, the model fails to reproduce parts of the data with intermediate transportation values, producing a coefficient of determination of $R^2 = 0,46$.

The calibrated model was used to reproduce the missing information about connections between departments without airplane routes. Nevertheless, this travel model can be improved if ground transportation estimates are available.

Parameter	Value
α	0.1363
β	0.9776
γ	2.4800
C	1.7985e-09

Table 4.2: Adjusted parameters of the gravity model.

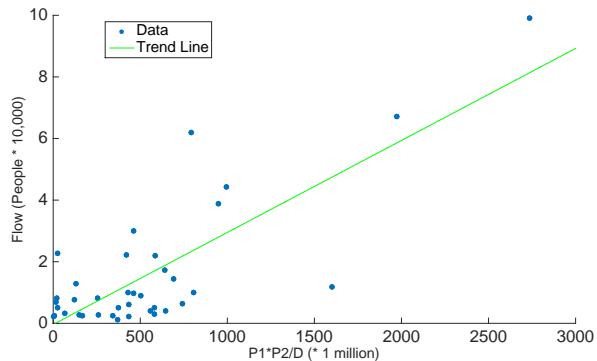


Figure 4.2: Passengers flow between city airports. The green line is shown to represent the linear trend of the data.

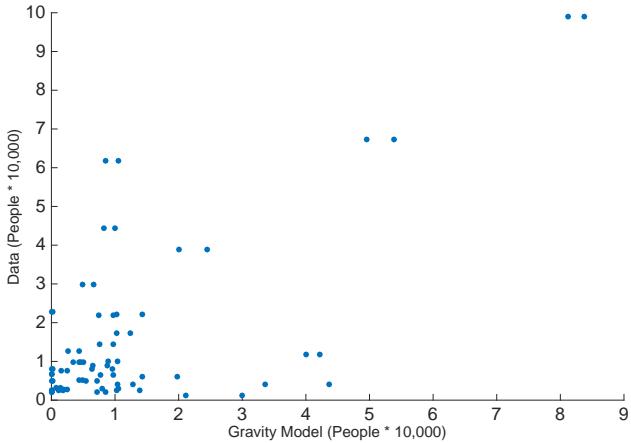


Figure 4.3: Gravity model comparison with data. The coefficient of determination is $R^2 = 0,46$.

4.2 Vector-borne transmission model

In the model, the spread of a disease is attributed to humans with defined daily activities. The transmission of the pathogen occurs inside a structure, identified in the model as places: household, workplace or school. When bitten by a mosquito, a human can become infected. Then within the model, the agent would follow a SEIR structure. Mosquitoes can also become infected when biting an infectious human, then it would follow a SEI transmission. The compartmental representation of the model is shown in figure 4.4.

In the model, each place is assigned with a number of mosquitoes based on the temperature and the number of hosts in the place. Mosquitoes are initially susceptible, however this original condition can be altered with an initial seed of the infection in the mosquitoes. Every day the model updates the activities of the agents, based on their favorite activities. The places visited by each agent and the contact that these agents have with mosquitoes in a specific place are recorded. All the agents are initially susceptible unless specified in the initial conditions.

For each place the number of bites are calculated using a binomial probability that depends on the number of hosts (N_h), the mosquito population size at each place (N_v), and the mosquito's biting rate. When susceptible humans visit a place with infectious mosquitoes, then the number of hosts infected by the virus are computed depending on the number of mosquitoes that are infectious (I_v), the number of susceptible hosts (S_h), the biting rate (b), and a transmission efficiency (β_h), as represented in equation 4.4. Then, these infections are randomly assigned to susceptible hosts, that become exposed to the virus. Exposed individuals (E_h) are not able to transmit the virus to any mosquito for a specific time, denominated as latency period (ϵ). After the latency period, each agent (human) will move to the infectious state. In this stage, the agent could develop symptoms of the disease (I_h) or could be asymptomatic (I_a). Regardless of the presence of symptoms, a human would be able to transmit the virus to susceptible mosquitoes through bites. A human would stay in this stage for a short period ($\gamma \sim 6$ days), then the human would recover from the infection or die. After recovery, a human will gain immunity to the serotype of the virus infected. In the case of chikungunya, a human will be considered completely immune to the virus. Whereas, an infection of dengue would provide immunity only to the infecting serotype. Nonetheless, after two sequential infections of different dengue serotypes, a human would be considered immune to all the four existing serotypes of dengue virus.

$$E_{h_T} = \left(1 - (1 - \beta_h)^{(b \cdot I_{vT})/N_h}\right) \cdot S_h \quad (4.4)$$

The transmission model for the mosquitoes was implemented as a discrete compartmental model with three different compartments: Susceptible (S_v), Exposed (E_v) or Infectious (I_v). There are no

individual mosquitoes considered in the model, but rather a count of the population size of the mosquitoes in the different compartments. A closed vector population was considered which means that mortality and birth rates are equivalent (μ_v). In places visited by infectious agents, susceptible mosquitoes (S_v) can be infected through bites. The number of newly infected mosquitoes depends on the infectious hosts visiting that place (I_h), the number of susceptible mosquitoes (S_v), the total number of hosts (N_h), the average mosquito biting rate (b), and the mosquito infection efficiency (β_v), as illustrated in equation 4.5. These new infections for each serotype would move to the exposed compartment (E_{vi}), where ' i ' indicates the infecting serotype. The length of stay in the compartment is determined by the extrinsic incubation period ($1/\tau$). After that period, the mosquitoes would be able to transmit the virus to a susceptible agent and would move to the Infectious compartment (I_{vi}), where the mosquitoes will stay until they die.

$$\begin{aligned}\dot{S}_v &= \mu_v \cdot N_v - \mu_v \cdot S_v(t) - E_{v_T} \\ \dot{E}_{vi} &= E_{v_T} \cdot \left(\frac{I_{h_i}}{I_{h_T}} \right) - (\mu_v + \tau) \cdot E_{vi}(t) \\ \dot{I}_{vi} &= \tau \cdot E_{vi}(t) - \mu_v \cdot I_{vi}(t)\end{aligned}\quad (4.5)$$

Where:

$$E_{v_T} = (1 - (1 - \beta_v)^{(b \cdot I_{h_T})/N_h}) \cdot S_v$$

τ : 1/Extrinsic incubation period

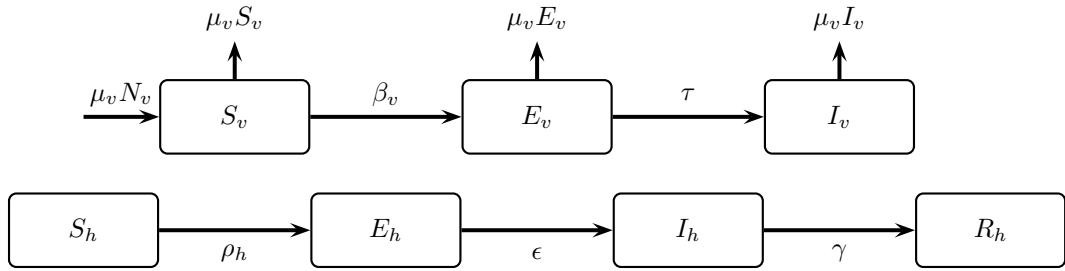
I_{h_i} : Human population infectious with the serotype i

$$I_{h_T} : \sum_{i=1}^4 I_{h_i}$$

Initial Conditions

The initial conditions of a system can change its behavior. For instance, the proportion of the immune population can reduce the amount of people at risk, which consequently decreases the impact of an epidemic. It is assumed that immunity to the four dengue serotypes is gained after two infections. Hence, the previous exposure to one or multiple serotypes in the population, would determine the invasion success of an epidemic caused by a specific serotype.

The level of exposure of a population to a particular disease can be inferred from the age-structure immune status of the population. In general, susceptibility restricted to young ages implies long-term exposure to the pathogen. Similarly, disease reports with high incidence in the youth population can suggest long-term exposure of the population to the pathogen. While non-structure age incidence



$$\begin{aligned}\rho_h &= (1 - (1 - \beta_h)^{(b \cdot I_{v_T})/N_h}) \\ \rho_v &= (1 - (1 - \beta_v)^{(b \cdot I_{h_T})/N_h}) \cdot S_v\end{aligned}$$

Figure 4.4: Model Structure: SEIR for Humans and SEI for Mosquitoes.

suggests low levels of exposure in the population. In order to represent this effect in the model, age-structured immunity levels can be specified for one or multiple serotype by departmental level.

Since dengue has been present in Colombia for decades, previous exposure levels were calculated by department based on historical reports, using the methods proposed by Chao *et al*[28]. Age specific weekly reports were used for the 2007-2012 period from the INS [87]. From 1997 - 2006, weekly reports without age structure were used. For this period, age incidence was assumed to be the same as the average of the age incidence of the period 2007-2012. The annual incidence before 1997 was computed as an average incidence of the 1997-2012 period.

Examples of two different departments' susceptibility status to the four serotypes of dengue are shown in figure 4.5. The department of Norte de Santander has had a high incidence of dengue over the years. Whereas autochthonous dengue or chikungunya transmission is unsustainable in Bogotá because of its high altitude. Consequently, the susceptible proportion in Bogotá is larger for any of the four serotypes than in Norte de Santander. The entire population is considered susceptible for Chikungunya since the virus was recently introduced.

Another initial condition of the model is the number of infectious agents that spark the epidemic. These can be specified using an external text file specifying the number of new infections annually and the serotype of the disease. In addition, these infectious agents can be restrained to a geographical region, specifying the latitude, longitude and a radius. Also, initial infectious mosquitoes can be specified using a text file with similar specifications.

4.3 Model Verification

Basic phenomena expected from the model should be contrasted with the model's outcomes to assess the model's coherency. The effects of temperature, probabilities of contagion, and previous immune status of the population were tested to validate the basic behavior of the model.

In order to see the impact of the temperature in the model, simulations were performed using four different municipalities with different climate conditions as shown in table 4.4. The rest of the parameters are constant through the simulations, although they are not particularly calibrated to any data. The results are shown in figure 4.6. As expected, the epidemic die out in the municipality with average annual temperature bellow 18 °C.

The two municipalities with temperate temperature values (19.50 and 22.80 °C) experience an outbreak with similar peak times and peak cases. This is coherent with the assumption that similar temperature values imply similar mosquito density. The municipality with the highest temperature produce the higher peak value as expected. The epidemic curve is delayed compared to the other ones,

Parameter	Symbol
Mosquito incubation rate	τ
Mean vector daily biting rate	b
Mean mosquito lifespan	$1/\mu_v$
Mean mosquito birth rate	μ_v
Mean mosquito death rate	μ_v
Success development rate	D_r
Vector female ratio	F_r
Days latent distribution	ϵ
Days symptomatic distribution	γ_{symp}
Days asymptomatic distribution	γ_{aymps}
Mosquito infection probability	β_v
Mosquito transmission probability	β_h
Symptomatic Rate	Γ
Pupae per person	PPI

Table 4.3: Parameters of the vector-borne disease transmission model.

probably because of contact patterns in the municipality that vary from the other three. In general, the temperature reflects the expected behavior of higher impact for higher temperatures.

The impact on the model of the probabilities of contagion for mosquitoes and human was assessed varying these values from 0 to 1, while keeping constant the rest of the parameter values. The simulations were performed for the municipality of Belén, where the annual temperature is around 23 °C.

Higher probabilities of contagion produce higher impact epidemics as shown in figure 4.7. The

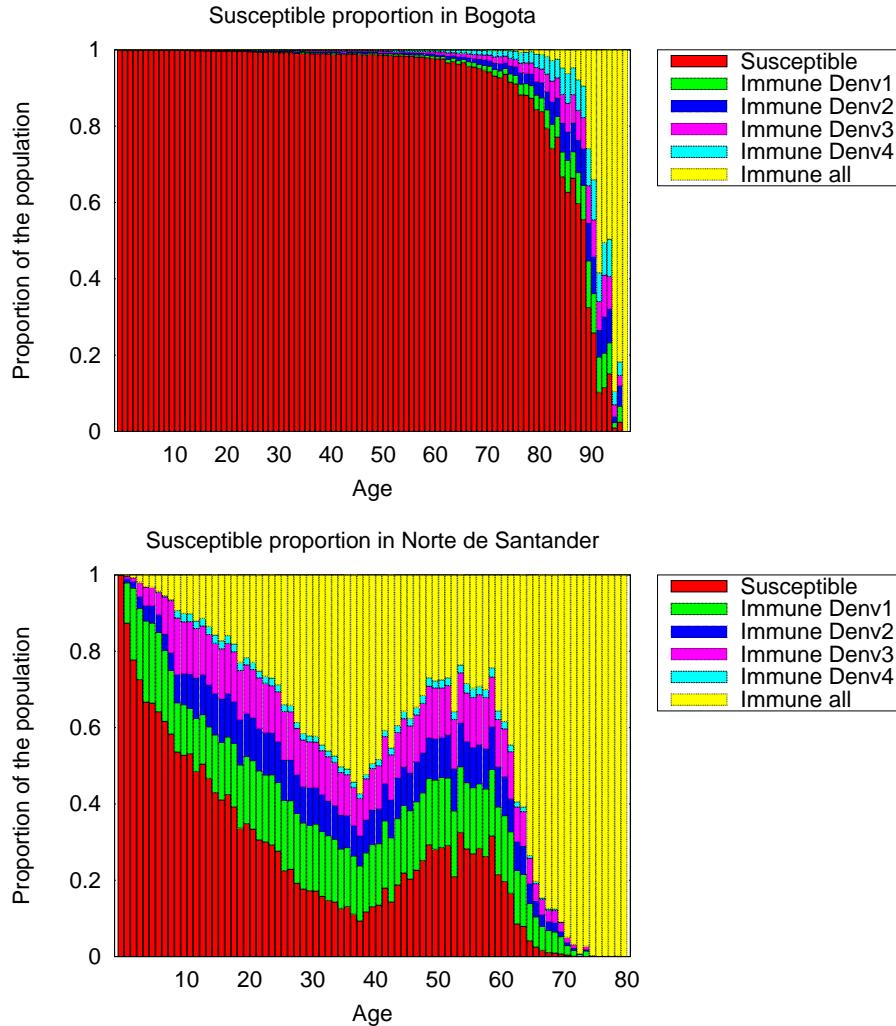


Figure 4.5: Susceptible proportion for each dengue serotype in the departments of Bogotá and Norte de Santander. Computations are made on historic data reports and estimations of underreporting ratios.

Municipality	Code	Mean Temperature (Celcius)
Berbeo (Boyacá)	17015087	10.75
Belén (Boyacá)	17015090	19.50
Soplaviento (Bolívar)	17013760	22.80
Riohacha (Guajira)	17044001	26.33

Table 4.4: List of municipalities to explore the effect of temperature in the disease transmission model.

human's probability of contagion affects the curve in the total number of infections as well as in the peak time. High probabilities produce fast epidemics with large number of cases while low probabilities generate slow epidemics with low number of infections. No epidemic is originated when the probability of contagion is null.

A similar behavior is observed for the mosquito's probability of contagion. However, the epidemic is only sustained when a determined threshold is exceeded. This can be explained by the short lifespan of the mosquito and the representation of mosquitoes by a compartmental model. When not enough mosquitoes are infected, the death rate can be higher than the rate of flow between susceptible, exposed to infectious.

Two scenarios were simulated in Norte de Santander in order to assess the impact of the initial immune status of the population on the model. The first scenario was simulated for a population completely susceptible to the virus, while the second scenario uses the previous immunity proportions illustrated in figure 4.5. An epidemic of DENV 1 was simulated in the two scenarios assuming the rest of the parameter values constant. Although, not calibrated to any particular scenario.

The epidemic has a larger impact in the population when a completely susceptible population is assumed as shown in figure 4.8. The epidemic size is almost five times bigger when the population is susceptible because the number of effective contacts of infectious mosquitoes with susceptible humans is larger. In contrast, when a population has been exposed to a virus for extended periods of time, the number of susceptible humans that a mosquito can bite is restrained.

Previous immunity to the virus also changes the age-structure incidence of the epidemic. As shown in figure 4.8, the age structure of the epidemic incidence is more homogeneous through the age groups when there is no previous exposure to the pathogen, while the incidence of the epidemic on the population previously exposed to the virus is focused on the young age groups.

The simulations performed to verify the behavior of the model are coherent with the expectations from the reality. This could imply that the basic components of the model are working properly. However, the outcome of the model does not represent any realistic scenario. In order to analyze specific outbreaks, it is imperative to adjust the parameters to represent epidemic curves based on historic data.

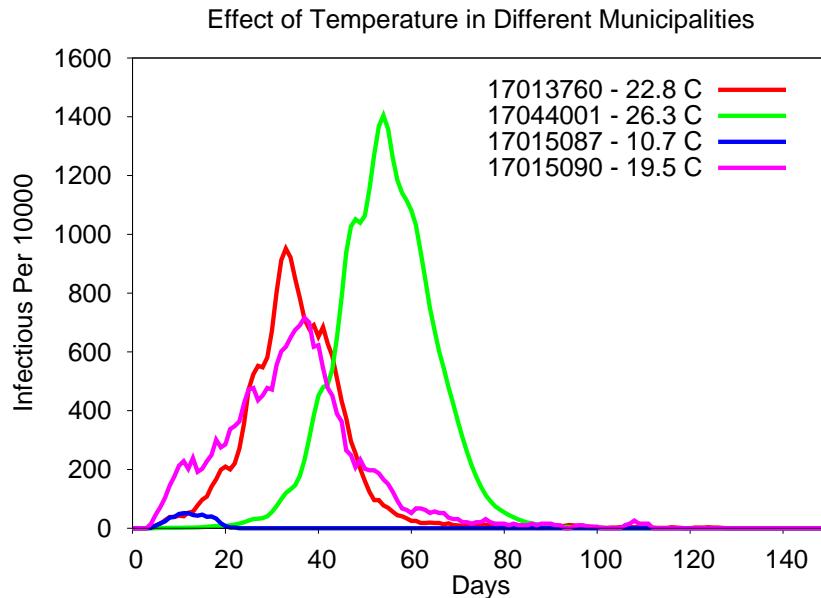


Figure 4.6: Impact of temperature in the model. Four municipalities with different climate conditions were simulated. The temperature influences the model's response as assumed, i.e. high temperatures implies high incidence.

4.4 Model Assumptions and Limitations

Some assumptions were made in order to represent vector-borne diseases in the model. For instance, the model assumes homogeneous pupae density inside a municipality. Also, socio-economical structure of the population is omitted. The seasonality was ignored, so the temperature in the model represents an average annual temperature of each region simulated. Other climate effects such as precipitation and humidity were excluded from the model. Further, the compartmental model of the mosquito ignores specific interactions of mosquitoes in different ages.

Some assumptions were made about the mosquito model. For example, the mosquito population is represented by a homogeneous compartmental model. This model ignores the effects of different stages of the pupae and adult mosquitoes. Furthermore, the interactions among mosquitoes were ignored, attributing the spread of the pathogen exclusively to human movement. The contacts inside specific structures are assumed homogeneous. While this could be dependent on multiple factors, such as age.

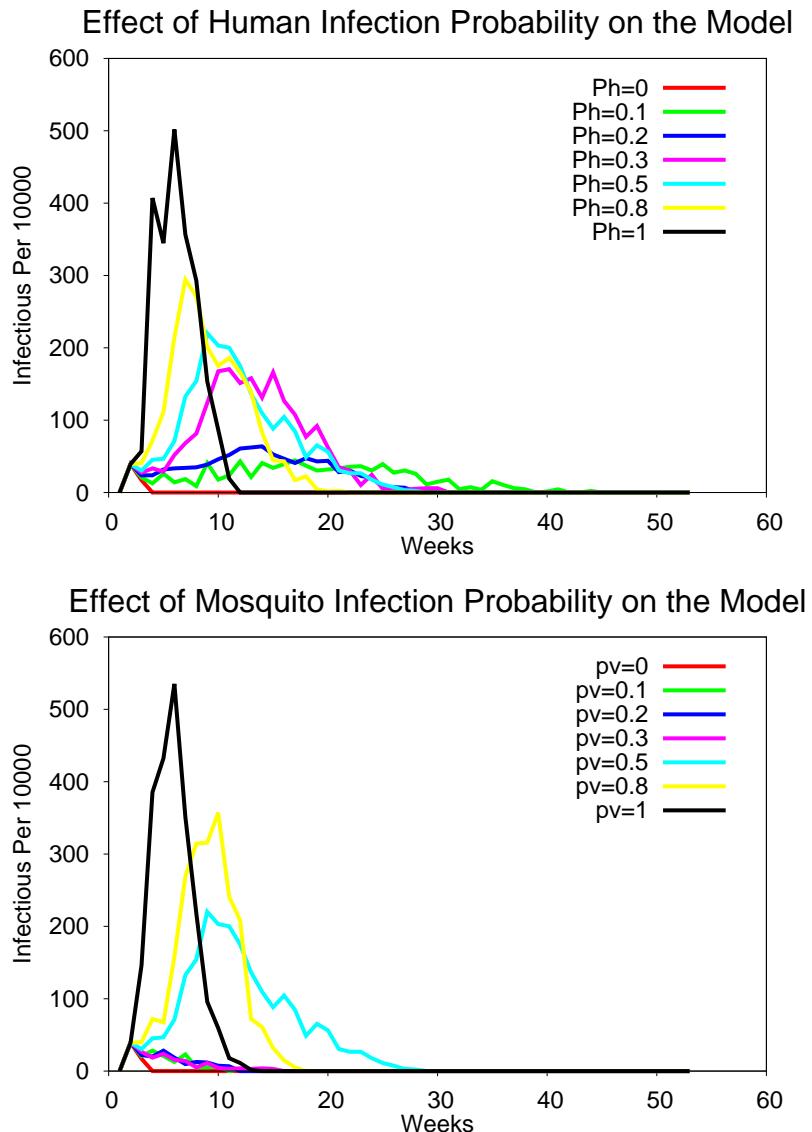


Figure 4.7: Impact of probabilities of contagion in the model's response. Small values of probability of contagion produce unsustainable epidemics.

Human contacts with mosquitoes are assumed to be homogeneous among places. While this could change from place to place. For instance, household contacts with mosquitoes are believed to be larger than in other locations. However, there is a lack of data about this phenomenon. Finally, the long-distance-travel model is based on a gravity model that assumes that only distance and population size affects the travel among cities.

4.5 Summary

In this chapter, the formulation of the model was presented. The model was implemented in the FRED platform including a new vector layer. The effect of temperature was included in the model to represent the differences in transmission among different climates. Mosquito density was calculated to be higher in warm dense regions, while in the model, as believed in reality, cold regions have null presence of the mosquito *aedes aegypti*. The human mobility was represented using two different approaches: daily commuting and long-distance travel. Agents' daily commuting patterns were previously calculated in the synthetic population. Long-distance travel was represented through a calibrated gravity model using air traffic data. The validation of the model was performed by observing expected patterns while

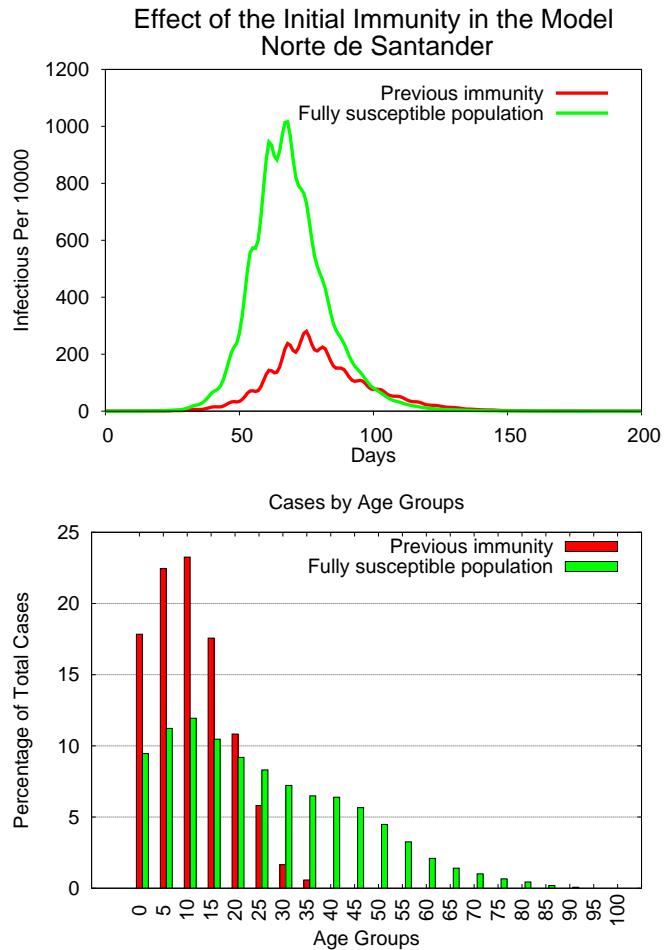


Figure 4.8: Impact of previous exposure to the virus in the model's response. Previous immunity to the virus reduce the susceptible population, hence the impact of the epidemic is reduced as well. The age-structure incidence of the disease moves toward the young population when the population has been previously exposed to the virus.

varying the parameters corresponding to the temperature, contagion probabilities and initial immune status. Finally, a set of assumptions and limitations of the model has been given.

Chapter 5

Model Calibration to Represent Chikungunya Epidemics in Colombia

Epidemiological models can be used as a tool for public health decision makers. These models can be used to make predictions for the spread of a specific disease. Also, models can be used to evaluate possible scenarios for different intervention strategies that may be implemented. Some deductions can be made from the model without representing a real scenario. For example, using the model to analyze the impact of a new pathogen's proliferation under different attack ration scenarios. On the other hand, if the purpose of the model is to make predictions or to analyze the impact of control interventions in an ongoing outbreak, then the model needs to represent reality in a reasonable manner.

5.1 The Data

Historic reports of the disease are used as a representation of the reality. The weekly incidence reports of Chikungunya in Colombia, obtained from the National Institute of Health of Colombia, are discussed in chapter 3. We selected the municipality of Riohacha to calibrate the model, since it was one of the first municipalities to have an epidemic and the impact of the outbreak in the municipality was large enough to calibrate the model. Riohacha is the capital of the department of Guajira and it's located in the north coast of the country ($11^{\circ}32'39''N$, $72^{\circ}54'25''W$) at the Caribbean Sea. The average annual temperature is $26^{\circ}C$ and its current (2015) population is 259.492 inhabitants. These characteristics are suitable for *aedes aegypti* survival, hence for chikungunya and dengue transmission.

The chikungunya weekly reports in Riohacha are shown in figure 5.1. The first case was introduced in October, 2014. After a period of a couple of weeks, the epidemic was established and peaked at almost 1200 cases reported in January, 2015. Data is available up to February, 2015, when the epidemic appears to have died out. The curve shows considerable noise in the reports. This could be attributed to mistakes in the data input. In total, 6517 cases were reported in this period. However, the real impact of chikungunya on this municipality could be underestimated because of asymptomatic cases and underreporting. The total extent of chikungunya in this municipality can be estimated using the reported data and the calibrated parameters of the model.

Riohacha surveillance reports were filtered to reduce the noise, as presented in figure 5.2. The total number of cases were conserved while approximating the reports to a normal distribution. It can be seen in figure 5.2, that the shape of the filtered curve is more consistent than the original data, removing extreme values. Hence, the filtered curve is used in the following sections to calibrate the model.

5.2 Calibration Strategy

In order to fit the model to the weekly incidence reports of Riohacha, a subset of the model's parameters needs to be selected for adjustment. The validity of the selected parameters can be assessed using a

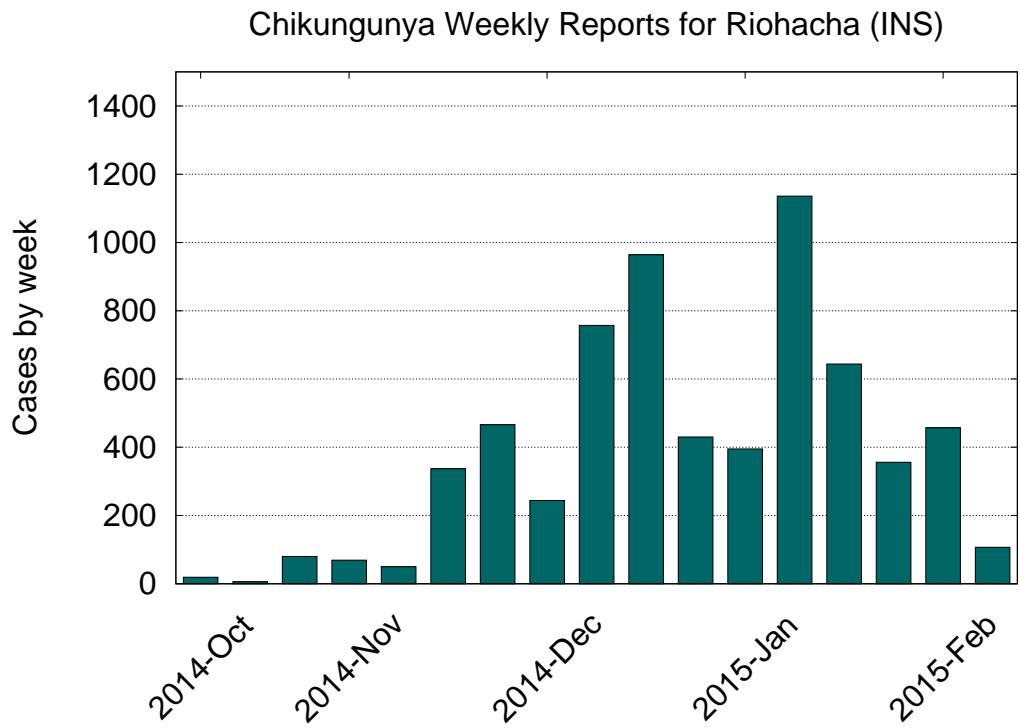


Figure 5.1: Weekly incidence chikungunya reports in Riohacha, Guajira. Obtained from the National Institute of Health.

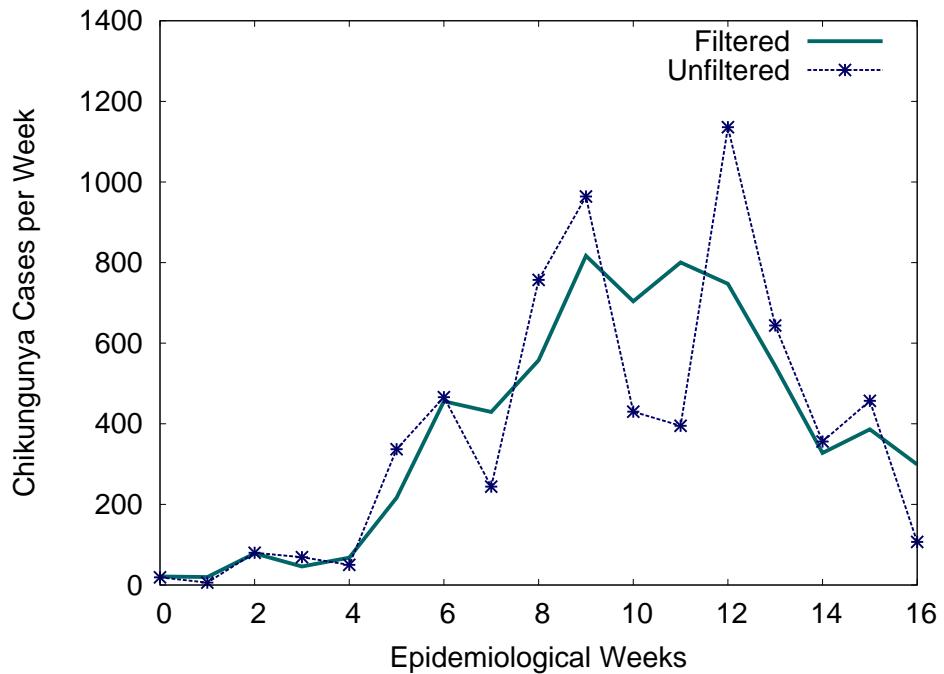


Figure 5.2: Weekly incidence chikungunya reports filtered to remove extreme points while maintaining the cumulative number of cases.

search strategy that is defined in such a way that it finds a global optimal value for the parameters. The selected goodness-of-fit measure can affect the results of the calibration process.

Selection of Parameters to Estimate

The model's parameters can be split into three categories: geographic, demographic and disease transmission parameters. Geographic parameters include population density, schools and workplaces location, climate, etc. Demographic parameters are related to the population size, age-structure, gender, activities, etc. As described in chapter 3, these parameters have already been calibrated in order to represent a population that statistically represents the country in the demographic and geographic scope. Since these parameters have already been adjusted, they are not included in this calibration step.

Fourteen parameters are related to vector-borne disease transmission in the model, as summarized in table 4.3. Most of the parameters were adjusted to represent values from dengue and chikungunya literature, as listed in table 5.1. A total of three parameters were selected to calibrate the model: mosquito infection probability (β_v), mosquito transmission probability (β_h) and the symptomatic rate (Γ). Because of the complexity of mosquito and human probability of infection, these parameters are difficult to obtain from population studies. Also, these parameters could change from model to model, since they include other effects misrepresented by the model. The symptomatic rate could also be seen as the diagnosis rate. There are some estimates of the symptomatically for chikungunya, however the underreport can vary in different populations based on diverse factors such as the capabilities of the health institutions.

Goodness-of-Fit Measures

Different error measurements could lead to varying calibration results. Three different goodness of fit measures were considered: lsq, weighted lsq and χ^2 . These measures give more priority to different aspects of the curves. In order to demonstrate this point, in figure 5.3 there are two simulations with different parameters. Simulation 1 has a curve with a slow increase, a large width and a peak close to the data. On the other hand, simulation 2 has a curve with a steep growth, a short width and a large peak. Qualitatively, simulation 2 fits the data more accurately except at the peak, whereas simulation 1 fits the peak of the data, but misses the information for the initial growth rate.

Depending on the error measure used, different results are obtained. For both, LSQ and Weighted LSQ, simulation 1 has a smaller error than simulation 2. But, for χ^2 , simulation 2 aligns closer to the data. Because of these differences, the calibration step will be performed using both *LSQ* and χ^2 .

Parameter	Value	Reference
Mosquito incubation rate	1/11	[125]
Mean vector daily biting rate	0.76	[126, 28]
Mean mosquito lifespan	18	[28]
Mean mosquito birth rate	1/18	[28]
Mean mosquito death rate	1/18	[28]
Success development rate	0.83	[120]
Vector female ratio	0.5	[120]
Days latent distribution	mean 6, sd 1.4	[125]
Days symptomatic distribution	mean 4.83, sd 1.2	[125]
Days asymptomatic distribution	mean 4.83, sd 1.2	[125]
Pupae per person	1.02	[34]

Table 5.1: Model's parameter values adjusted to reported values in the literature. A closed population was assumed for the *aedes aegypti*, hence the mortality and birth rate are assigned to the same value. Latency and infectious period were obtained from dengue distributions.

- $LSQ = \sum_t (f(t; \theta) - y(t))^2$
- $WLSQ = \sum_t y(t) \cdot (f(t; \theta) - y(t))^2$
- $\chi^2 = \sum_t \frac{(f(t; \theta) - y(t))^2}{y(t)}$

Parameter Search Strategy

The *Nelder-mead simplex* algorithm was used in order to minimize the difference between the data and the model's output. This method was selected because it requires limited information about the system and usually converges quickly to optimal solutions. Agent-Based Models usually demand large amounts of computational resources, compared to equation-based models. For instance, a single simulation of the municipality of Riohacha takes about 2 minutes in FRED. Therefore, the speed of the calibration algorithm is crucial. The *Nelder-mead simplex* algorithm was implemented in *Perl*, following the methods of Lagarias *et al*[83] and Gao *et al*[84].

In order to evaluate the impact of the inclusion of the parameter Γ (Underreporting) in the calibration process, two different scenarios were utilized. The first scenario was to assume a diagnosis rate value of $\Gamma = 0,1$ to represent a reporting of 1 out of 10 infectious cases. The other scenario include this value to adjust in the optimization process. Also, in order to examine the effect of the goodness of fit measure, two more scenarios were performed: χ^2 and LSQ.

1. GOF χ^2 with 2 parameters: β_v , β_h and $\Gamma = 0,1$
2. GOF LSQ with 2 parameters: β_v , β_h and $\Gamma = 0,1$
3. GOF χ^2 with 3 parameters: β_v , β_h and Γ
4. GOF LSQ with 3 parameters: β_v , β_h and Γ

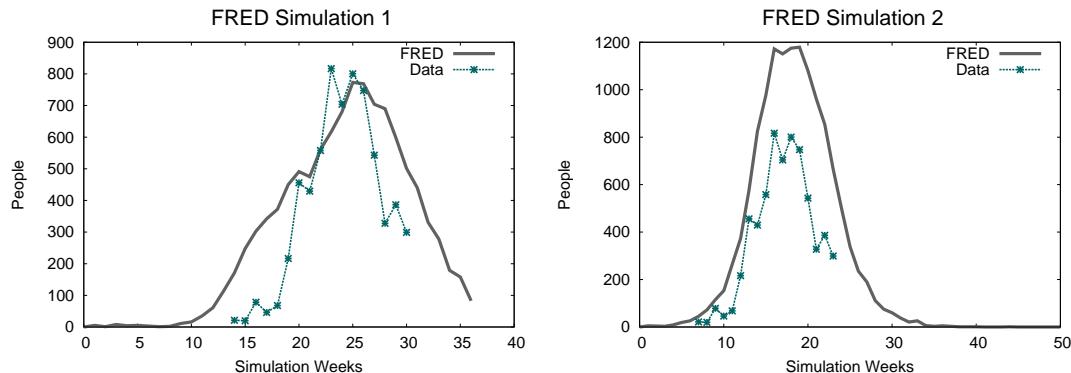


Figure 5.3: Two simulations of the model contrasted with the weekly incidence reports of chikungunya in Riohacha. Left, the model's response is characterized by a wide curve that fits the peak of the data but misrepresents the initial growth rate. Right, the model's response is characterized by a narrow curve that fits the initial growth rate of the data but overestimates the peak.

Simulation	Parameters	LSQ	WLSQ	χ^2
1	$\Gamma = 0.1$ $\beta_v = 0.23$ $\beta_h = 0.45$	647,790	100,480,066	5,784
2	$\Gamma = 0.1$ $\beta_v = 0.29$ $\beta_h = 0.3$	2,120,867	1,088,511,998	5,455

Table 5.2: Errors' estimations using different goodness-of-fit measures for the model's output presented in figure ??.

The magnitude of the outbreak is reproduced by the model calibrated using the first two strategies, while the growth rate is not captured. Graphs of the calibration results are shown in figure 5.4. The resultant curves are significantly closer to the data compared to the initial point. However, for these strategies, the model's curve show a initial growth rate slower than the surveillance reports.

Using χ^2 measure, the model fits the initial growth closer than the *LSQ*, while the peak is overestimated. This is achieved assigning a larger value for the parameter β_h , which produces a larger reproductive number, as illustrated in table 5.3. It is difficult to determine which strategy produces the model to fit the data closer, since both yield to discrepancies in different aspects of the data. This is a result of the initial assumptions in the diagnosis rate value. Hence, strategies 3 and 4 varies this parameter to increase the flexibility in the model's outcome.

The model fits the data more accurately adjusting the underreporting parameter value as well as the transmission probabilities, as presented in figure 5.5. In table 5.3, it is possible to see that using both error measures, the algorithm converged to the same value for the three parameters. The differences between the resultant curves in figure 5.5 are probably caused by random effects of each simulation. Even though, the model fits considerably closer to the data, it should be pointed out that this algorithm converges to local solutions. Hence, a global approach is needed in order to obtain a global optimal solution for the model's parameters.

Global Calibration

The Nelder-mead simplex converges to a local minima. Hence, these results can depend on the initial projection of the parameter values. In order to obtain a global minima, this method can be applied multiple times with random initial points. In this way, the optimal solution depends less on the initial value of the parameters.

A global calibration strategy was implemented for the same two and three parameters used in the local calibration step. The χ^2 GOF was used in the global strategy, since it fits focuses on the shape of the curve. The underreporting value was set to 0.05 for the first strategy, based on the previous estimations obtained from the local calibration process. The two different global calibration strategies are summarized bellow.

Paramter	Min	Max	Initial Guess	Strategy 1	Strategy 2	Strategy 3	Strategy 4
β_v	0.1	1.0	0.4	0.198	0.209	0.296	0.296
β_h	0.1	1.0	0.4	0.600	0.402	0.517	0.517
Γ	0.05	1.0	0.1	0.100	0.100	0.052	0.052

Table 5.3: Parameter values fitted using the local optimization algorithm.

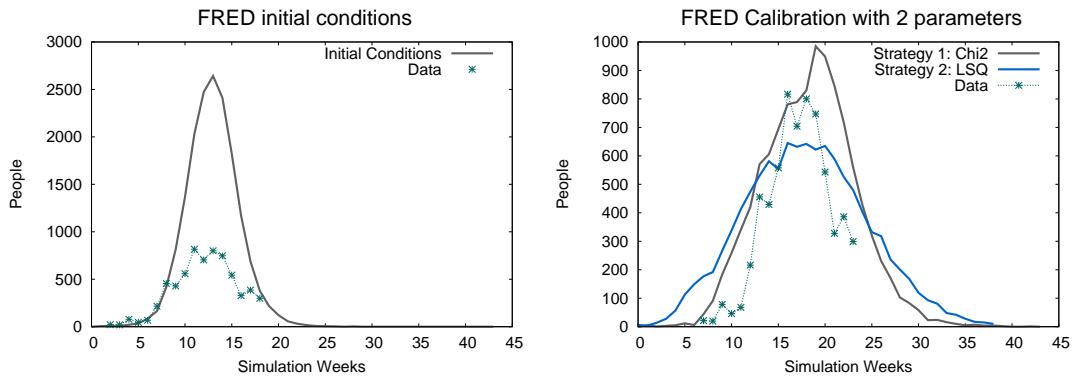


Figure 5.4: Model calibrated to Riohacha reports using two parameters and two different goodness of fit measures. In both cases the Nelder-mead optimization algorithm was used with two parameters and a maximum number of iterations of 100.

Paramter	Min	Max	Strategy 1	Strategy 2
β_v	0.01	0.9	0.295	0.297
β_h	0.01	0.9	0.473	0.447
Γ	0.05	1	0.05	0.053

Table 5.4: Parameter values fitted using global optimization algorithm.

1. Two parameters: β_v and β_h , $\Gamma = 0,05$
2. Three parameters: β_v , β_h and Γ

The results of the first global calibration strategy are presented in figure 5.6. The graph shows a correlation between the parameters β_h and β_v . Large values of β_h are compensated by low values of β_v . The algorithm converged to multiple points of β_h above 0.3, producing fitting errors of $\chi^2 < 400$. In contrast, the algorithm converged to values of β_v above 0.2 and bellow 0.4. One point of the algorithm converged to a solution with a χ^2 larger than 2000. The parameters with minimum error are considered the global optimal solution, and are denoted by a circle. They are also summarized in table 5.4.

These multiple local minima that were derived using the global calibration algorithm show that the data can be represented by multiple values for the two parameters following the correlation in figure 5.6. Hence, there is not a unique solution for the optimization problem. These parameter values could be constrained if previous knowledge of their values were known. However, collecting data for these parameters is a difficult process that could involve ethical issues, and it's out of the scope of this study. Also, the value of this parameter depends on each model's assumptions and limitations.

Strategy 2 attempts to find an optimal value for the three parameters of interest in the model. In table 5.4, the calibrated values of these three parameters are shown. It can be determined that the values are similar to the calibrated values of strategy one where only two parameters were selected. The diagnosis ratio is 0.053 which represents a reporting of 1 out of 18 infectious cases. As shown in figure 5.7, assigning a larger value to this parameter results in the model not converging to an optimal point. While assigning a lower value produces multiple local minima points. For this reason, the fitted values of global strategy 2 were selected to represent the chikungunya epidemic dynamics in Riohacha.

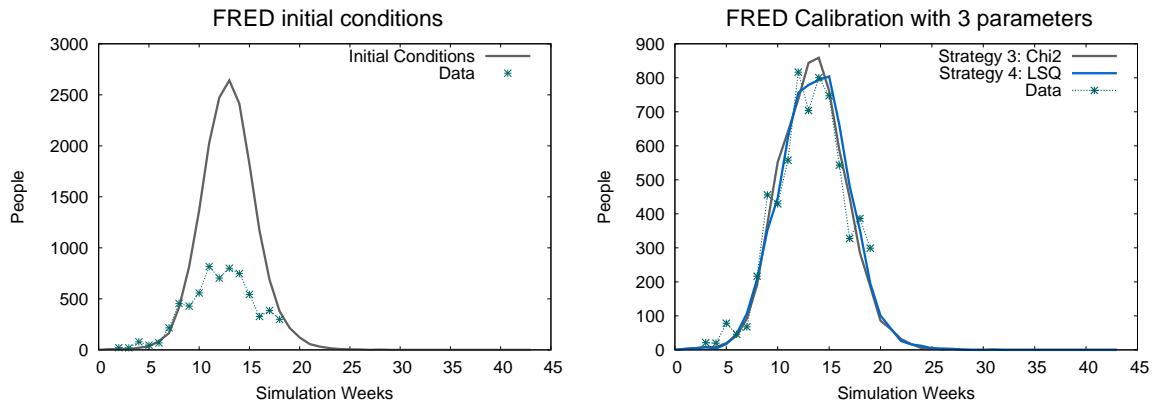


Figure 5.5: Model calibrated to Riohacha reports using three parameters and two different goodness of fit measures. In both cases the Nelder-mead optimization algorithm was used with three parameters and a maximum number of iterations of 100.

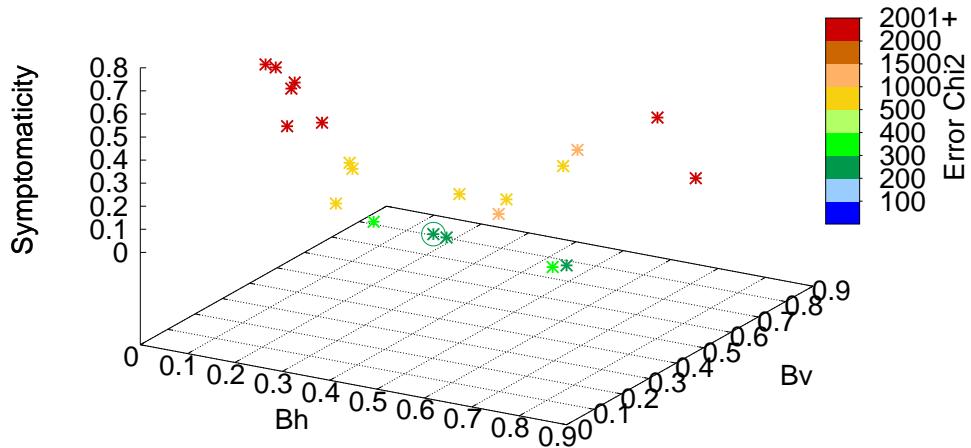


Figure 5.7: Calibration convergence points from the global strategy 2. Random initial values to calibrate three parameters. The circle denotes the minimum point found by the optimization algorithm. The color represents the error between the model and the reports. 40 initial points were tested with maximum 100 iterations for each calibration step.

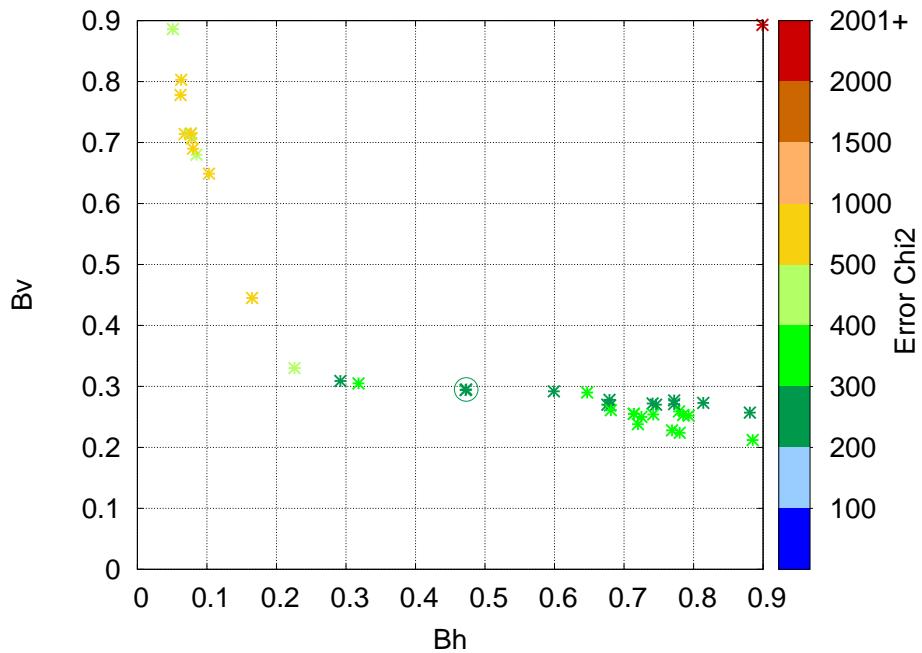


Figure 5.6: Calibration convergence points from the global strategy 1. Random initial values to calibrate two parameters. The circle denotes the minimum point found by the optimization algorithm. 40 initial points were tested with maximum 100 iterations of each calibration step.

5.3 Calibration Results

The calibrated model's goodness-of-fit compared to the data is $\chi^2 = 254$. In order to calculate the uncertainty of the model caused by inherent randomness, one hundred runs were performed using these calibrated parameter values. The mean and standard deviation of the diagnosed cases were calculated over time and compared to the weekly reports.

Figure 5.8 includes the comparison of the model to the unfiltered and filtered signal of the incidence in Riohacha. Qualitatively, the model follows the data closely. The standard deviation for each week is represented as an error bar in the model. The cumulative cases reported in the surveillance system are 6517. The model estimates 6637 ± 675 diagnosed cases as shown in table 5.5. Also, the model estimates a total of 7014 ± 748 diagnosed cases in the simulation period. It should be noted, that according to the model, the total of infectious cases is 18 times larger than those diagnosed.

The attack rate represents the portion of the population that has been infected by the virus, which is the opposite of the susceptible curve. According to the model's outcome shown in figure 5.9, the attack rate grew exponentially in the initial phase stabilizing around 60% of the population. This suggests that the epidemic had a significant impact in the population of Riohacha, infecting more than half of the population.

The infectious and exposed populations are shown in figure 5.10. The exposed population reached a peak of 12 thousand individuals a week, which is above the infectious peak of approximately 8 thousand. The epidemic grew exponentially until the epidemiological week 53 (14 in the simulation). At this point, the susceptible population dramatically decreased, such that the effective contacts per infectious individual was below 1, i.e. when there is herd immunity in the population.

In general, the model represents the diagnosed cases and suggests that the impact of the epidemic is larger than what is represented in the surveillance reports. These results are valid for the municipality of Riohacha, where the model has been calibrated. The model represents the climate and population specifics of all the municipalities in Colombia. However, some parameters like the diagnosis rate and the pupae per person rate can be specific for each municipality. Hence, the model can be simulated in other municipalities, in order to observe its performance with unfitted data.

Performance to unfitted data

A group of 25 municipalities with an epidemic curve capable of comparison with the model was selected in order to verify the behavior in other municipalities with chikungunya where the model was not adjusted. The same parameters were utilized in the model to represent these outbreaks. The results vary from high performance, acceptable performance, to poor performance.

In table 5.6, the municipalities with high performance using unfitted data are listed. The error was calculated using the χ^2 measure and unfiltered data. The table shows the total diagnosed cases for the data and the model in both the data period and the total simulation period. Also, two columns are included to show the standard deviation of the diagnosed cases that were predicted by the model in both periods.

From 25 municipalities, 60% have high or acceptable fit performance. Eleven municipalities are represented by the model with high performance, which is based on the error, the qualitative curve, and the difference between the reported total cases and model predictions. Four more municipalities have acceptable performance considering that the data has extreme points that affect the error measures. The plots that correspond to these municipalities are shown in figures 5.11, 5.12. The interpretation

Variable	Cumulative)	Standard deviation
Reported Cases Model	6637	675
Reported Cases Data	6517	-
Reported Cases Model (Simulation period)	7014	784

Table 5.5: Cumulative cases of chikungunya reported by the surveillance system and estimated by the model.

of the error measurements should be done with caution, since the data is noisy and could lead to incorrect calculations.

A group of seven municipalities (28 %) showed poor performance as listed in table 5.7. Four of them had an epidemic with more cases than predicted from the model as presented in figure 5.13. The magnitude of the reported cases is around 5 - 6 times larger than the model. Three more municipalities had poor performance, but in this case the magnitude of the predictions overstate the reports from the surveillance system as shown in figure 5.14. There could be multiple reasons to explain these differences, for example a difference in the diagnosis rate, a difference in the pupae per person index, control interventions, among others. However, there is no factual data to support these assumptions.

Finally, a set of three municipalities (12 %) showed initial good fit with the model until a sudden decrease in the cases occurred as summarized in table 5.8. The model predicts more cases than were reported as shown in figure 5.15. A reason to explain this behavior could be the implementation of vector control programs in the middle of the epidemic. These three municipalities are: Córdoba (23001), Santa Marta (47001) and Mariquita (73443).

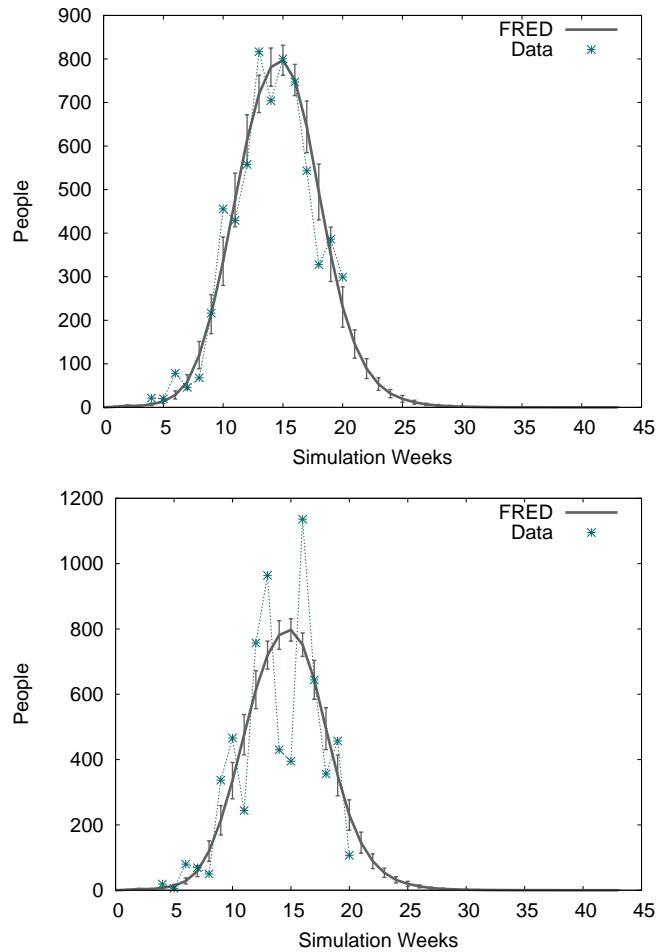


Figure 5.8: Model fitted to chikungunya weekly reports in Riohacha (Guajira). The top figure shows the mean response of 100 simulations of the model compared to the filtered data. The bottom figure shows the model compared to the original data from the National Institute of Health in Colombia. In both graphs, the bars represent the standard deviation.

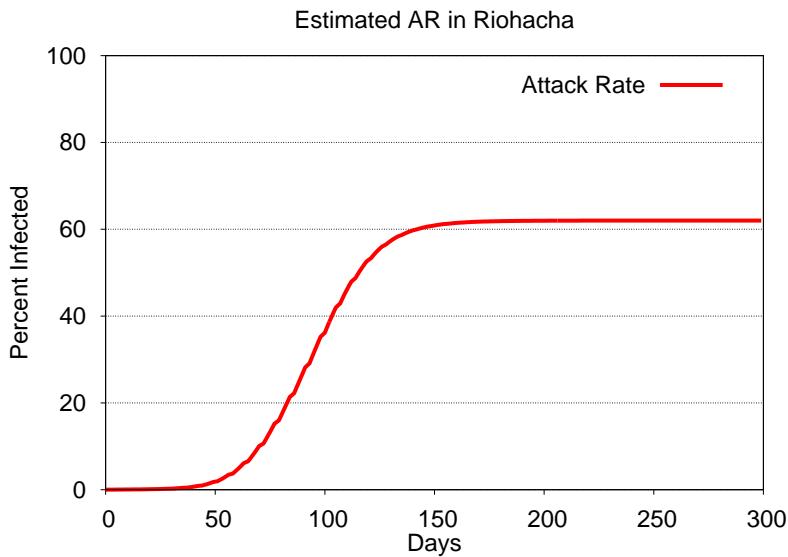


Figure 5.9: Mean attack rate of 100 simulations of the model fitted to chikungunya data. 60 % of the population is infected at some point in the epidemic.

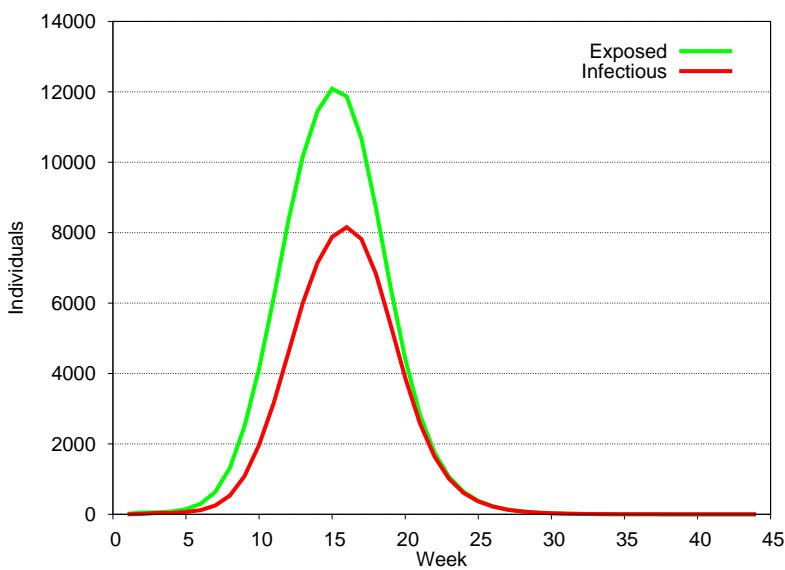


Figure 5.10: Exposed and Infectious variables of 100 simulations of the model fitted to chikungunya data. The green curve represents exposed individuals per week, whereas the red curve represents infectious individuals per week. There is a peak of 120000 exposed individuals and 8000 infectious individuals.

CODE	ERROR	DATA	MODEL	MODEL TOTAL	MODEL SD	MODEL SD T
44001	1532	6517	6623	6957	512	575
73449	116	797	778	1286	73	170
73671	486	419	227	367	65	137
70820	607	1437	1030	1089	188	218
47460	1458	297	462	601	178	295
70670	1893	681	1121	1232	421	545
47798	2298	310	327	392	76	106
47555	9106	1211	1472	1788	474	625
54874	9301	936	1524	2312	510	996
70215	14515	1366	1971	2098	353	412
70001	24021	10373	9742	10858	3463	4618
54001	24806	23951	24238	25153	3431	3693
13836	4618	5929	1895	2339	314	425
41001	8613	13677	8928	13561	1578	2799
25307	4340	6387	2485	3438	510	849
44430	106076	2660	3831	3978	276	334

Table 5.6: 16 Municipalities with good performance. Gray background color indicates Riohacha (44001) which is the municipality used to calibrate the model. 15 municipalities' data are closely represented by the model without fitting. Light gray background indicates the four municipalities with good fit except for extreme values of the data.

CODE	ERROR	DATA	MODEL	MODEL TOTAL	MODEL SD	MODEL SD T
73055	1038	1328	184	316	56	110
23675	2281	2358	788	1113	266	475
41132	2309	2550	823	980	197	295
13657	3498	3525	611	1009	111	201
8001	531978	4680	45200	45727	4382	4514
8758	312525	1494	10747	13504	2124	3430
20001	480871	292	9214	14656	984	1713

Table 5.7: Seven Municipalities with poor fit performance. Light Gray background indicates that the model predictions are larger than the incidence reports.

CODE	ERROR	DATA	MODEL	MODEL TOTAL	MODEL SD	MODEL SD T
23001	18119	7252	10665	17274	3082	5977
47001	29639	5214	14381	14841	1151	1259
73443	1542	206	596	1051	86	213

Table 5.8: Three municipalities with initial good fit but sudden decay of cases in the data.

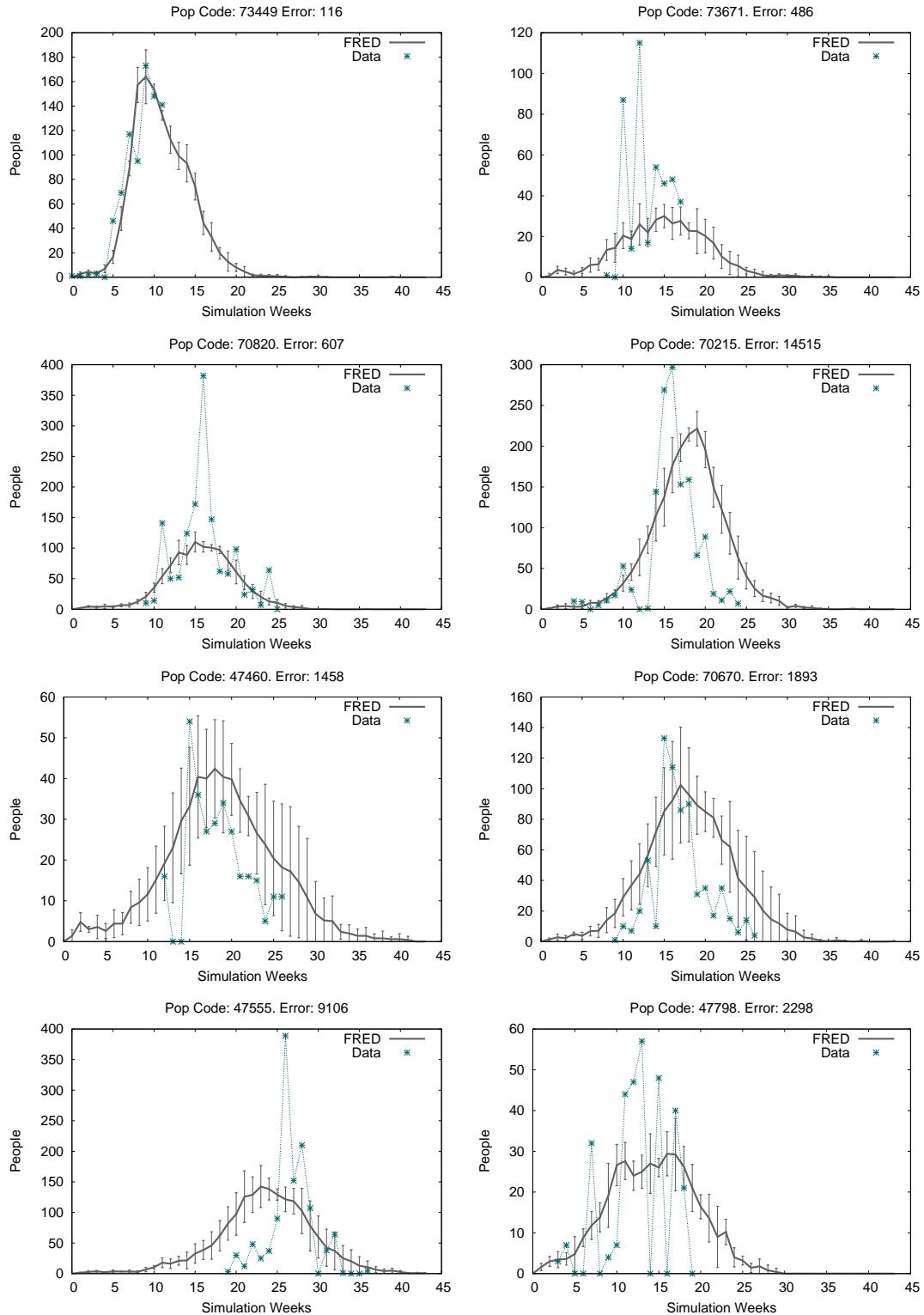


Figure 5.11: Municipalities that showed good fit performance when simulated with the calibrated parameter values to the Riohacha epidemic curve.

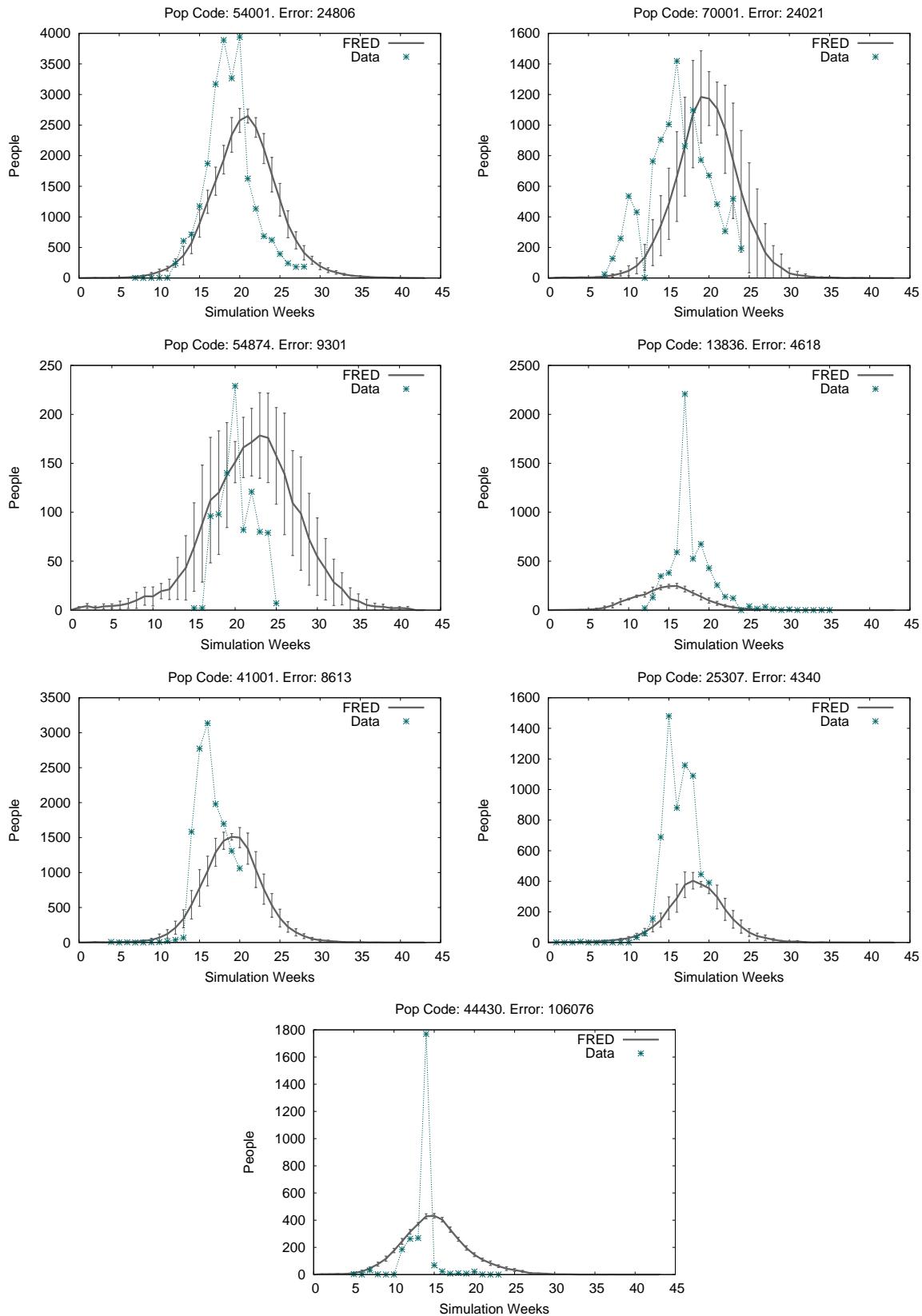


Figure 5.12: Municipalities that showed good fit performance when simulated with the calibrated parameter values to the Riohacha epidemic curve.

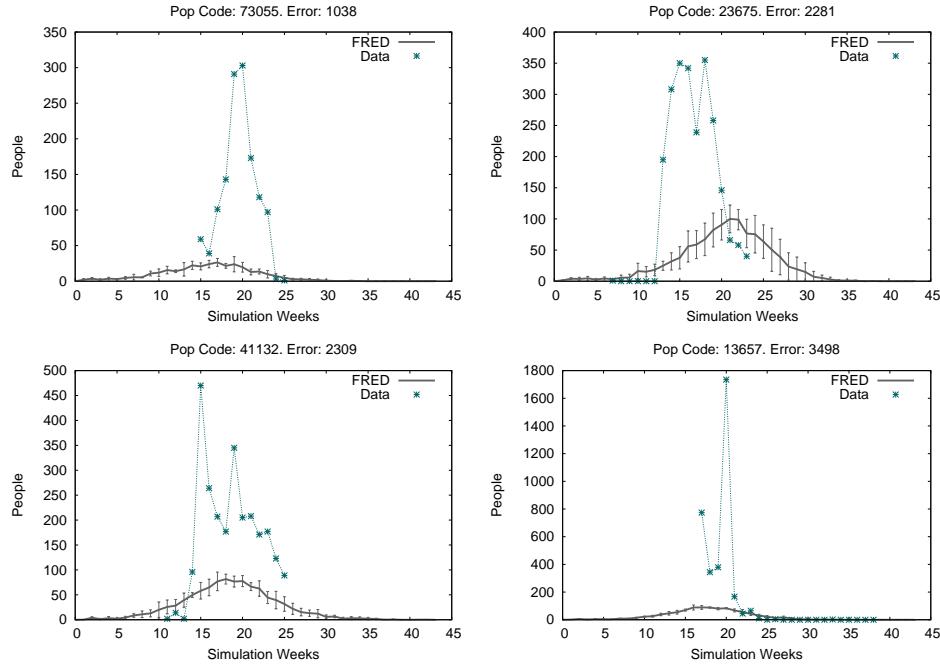


Figure 5.13: Municipalities that showed good fit performance when simulated with the calibrated parameter values to the Riohacha epidemic curve. The reports present larger impact than the model's predictions.

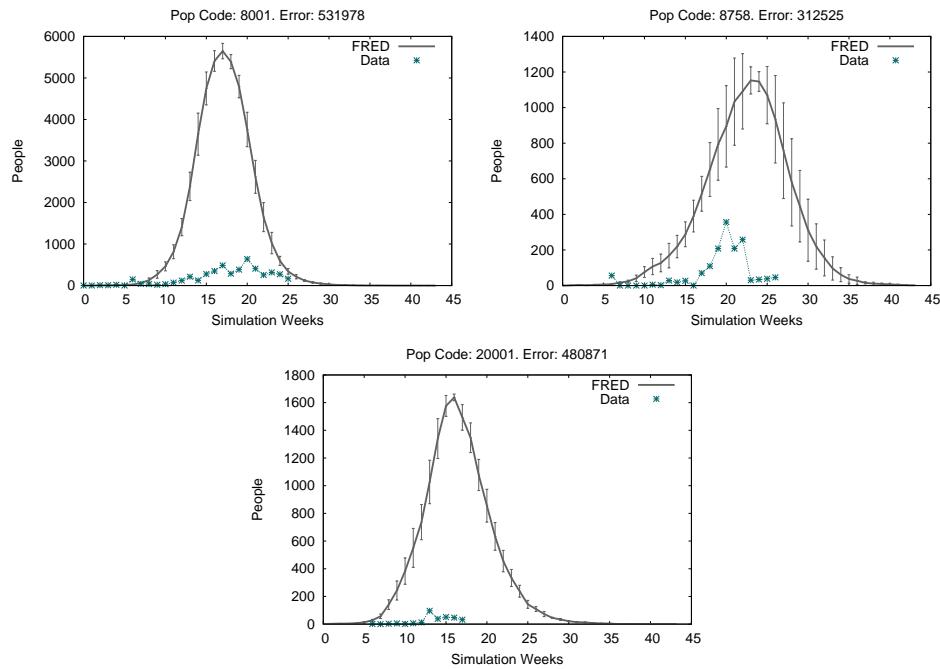


Figure 5.14: Municipalities that showed good fit performance when simulated with the calibrated parameter values to the Riohacha epidemic curve. The model's predictions present larger impact than the reports.

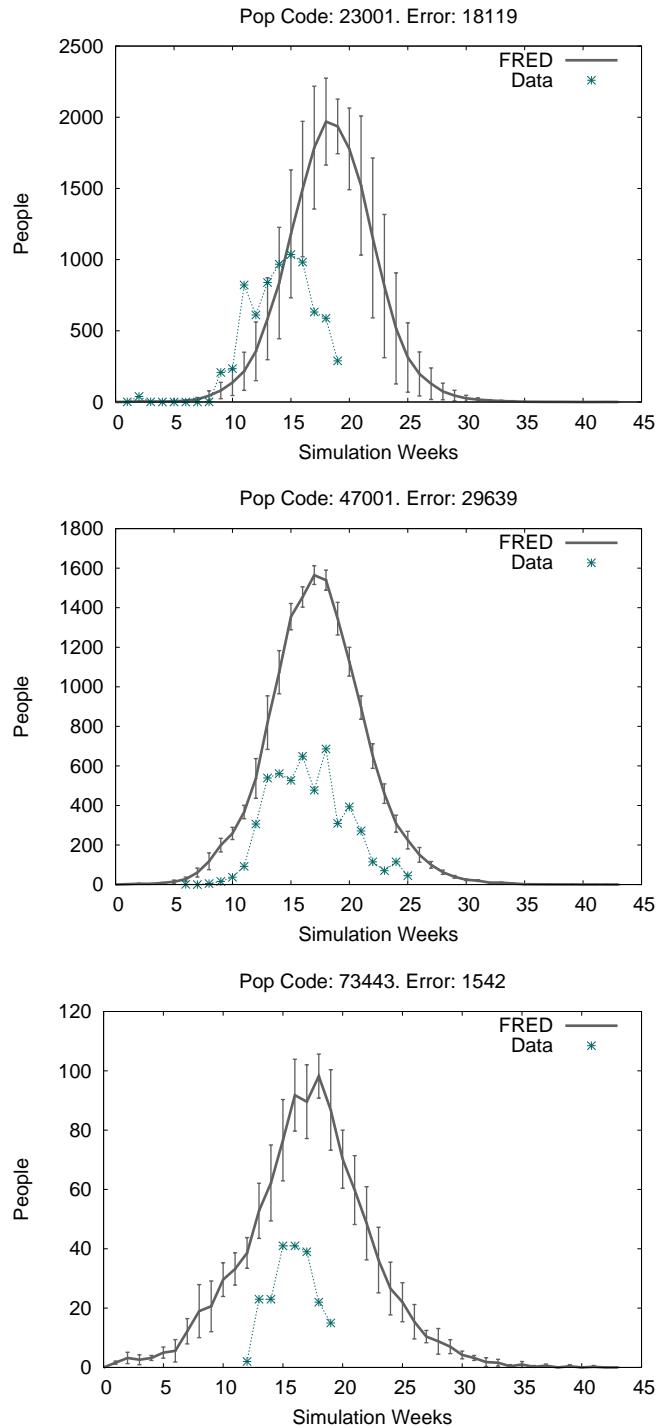


Figure 5.15: Three municipalities that showed initial good fit performance and sudden decay of cases in the middle of the epidemic, compared to simulations of the calibrated parameter values to the Riohacha epidemic curve.

5.4 Comparison at National Scale Epidemic

As mentioned previously, the characterization of the parameters for the entire country requires a more extensive database that includes estimates of the underreporting for each municipality or department. The inclusion of these factors exceeds the extent of this thesis.

In addition, the course of an epidemic in the country depends on the connectivity of the different regions, so the gravity model has been implemented within the study. However the mobility of people within regions includes more complex behavior not included in this model. Additional factors such as the socioeconomic status of households in each municipality can affect the outcome of the epidemic. Alas, the microdata used omitted the income level of the occupancies.

Even if a more complex model that includes travel, mosquito density, and socioeconomic status were implemented, it would not account for random events that can shape the epidemic in many different ways. This means that the observed data from the country is one realization of all the possible paths that the epidemic could have followed. This makes the calibration at the national level even more complicated. Nonetheless, a national-scale epidemic was simulated in order to compare the results and identify the areas where the model needs improved parameterization in order to represent accurately the epidemics in distinct geographical regions of Colombia.

Initialization of the Epidemic in the Model

The epidemic was seeded in the same region where, according to the data, the first cases of chikungunya developed, aiming to represent the same initial conditions of the outbreak in Colombia. The municipality with the first case reported was Turbaco, Bolívar, located in the north west of Colombia ($10^{\circ}34' N$, $75^{\circ}37' W$) as illustrated in figure 5.16. The population size of Turbaco is approximately 72000 people for 2015 and the annual temperature is around $27^{\circ}C$. A number of infectious mosquitoes was seeded in the houses of the municipality within a radius of 20 Km of the center point.

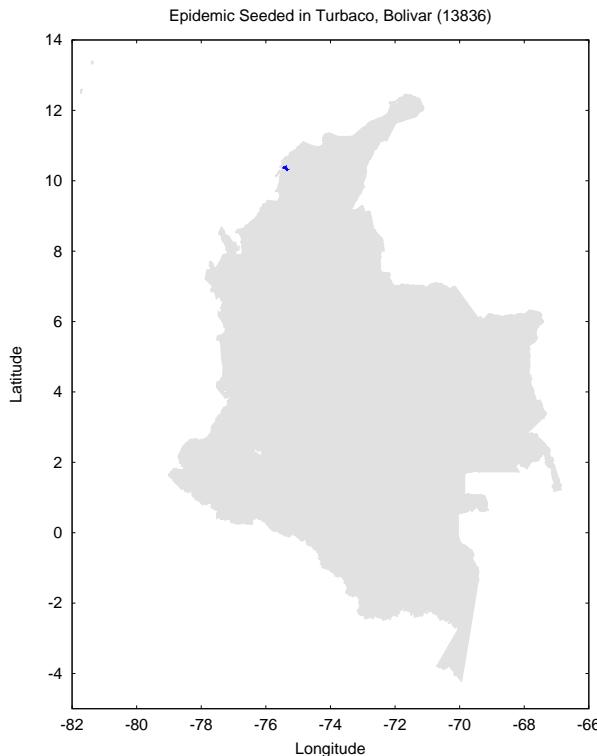


Figure 5.16: Municipality of Turbaco, Bolívar. The blue color represents the location of the municipality. The epidemic of chikungunya initiated in this municipality.

Weekly incidence comparisons

The weekly incidence for the entire country is shown in figure 5.17. It can be seen that the data follows the model up to week 20 (Week 0 being the beginning of the epidemic). Between week 20 and week 24, the model overestimates the incidence of chikungunya in the country.

There are two aspects that should be taken into consideration: the magnitude of the epidemics and their synchrony between regions. In the case of the magnitude, for the simulated period, the model predicts 250,277 diagnosed cases, whereas the reports showed 162,816 cases, as shown in table 5.9. There could be multiple explanations for the discrepancy in the number of cases, such as the commission of additional parameters that affect the chikungunya transmission, or the lack of vector control in the model.

In order to propose explanations for the differences in the model, multiple approaches are made. First, at the level of departments the outbreaks are compared to the data in order to locate with the majority of discrepancies. The discrepancies are represented geographically, aiming to highlight possible regional clusters in the prediction errors. The errors are also organized by socioeconomic indexes and temperature. In order to verify the synchrony of the events, the departments are also sorted by epidemic and geographical location.

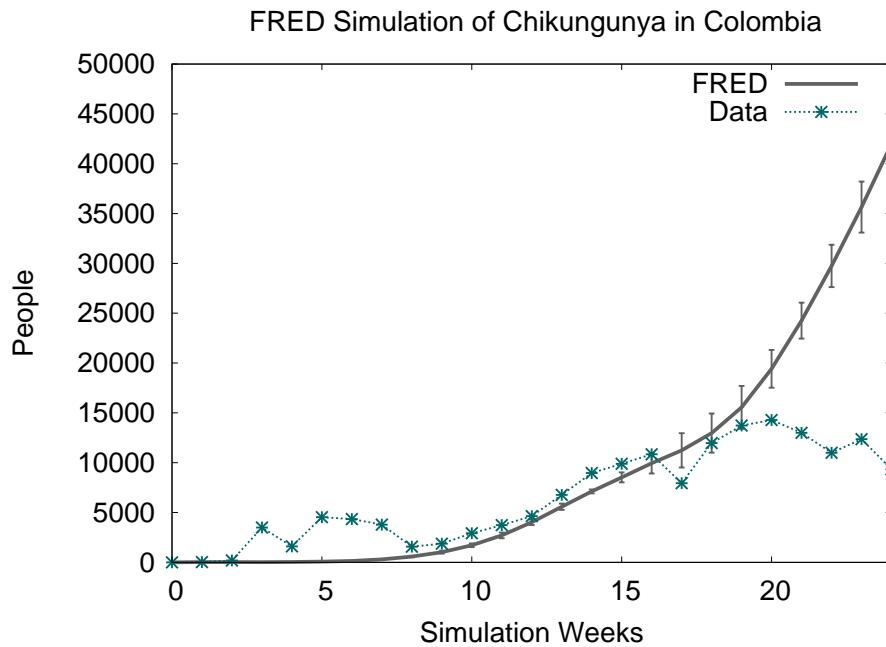


Figure 5.17: Weekly incidence of Chikungunya cases in the model and surveillance reports.

Cumulative Reports	Cumulative Model	Standard Deviation Model
162816	254277	23346

Table 5.9: Cumulative diagnosed cases in the model and in the surveillance reports.

Epidemic DATA	Epidemic FRED	Common	DATA not FRED	FRED not DATA
18	23	17	1	6

Table 5.10: Departments with cumulative incidence > 10 people per 10,000 inhabitants

Incidence by Department

As shown in table 5.10, the model and the surveillance reports agree that the same 17 departments reported an epidemic. One department reported an epidemic in the surveillance system and not in the model and 6 more reported an epidemic in the model and not yet in the surveillance system. Aiming to find the source of the errors between the predictions and the incidence reports, the incidence by departments are sorted using different covariates.

Colombia can be divided into 6 different regions characterized by different environmental conditions. They are: 1. Andina, 2. Amazonía, 3. Caribe, 4. Insular, 5. Orinoquía, 6. Pacífico. These regions are shown in figure 5.18. In order to find the region where the discrepancies are higher, the errors of the data and the model by department are sorted by region in figure 5.19.

According to the errors summarized in figure 5.19, the Caribbean region of the country shows the most significant discrepancy of the cumulative incidence by departments, specifically Atlántico

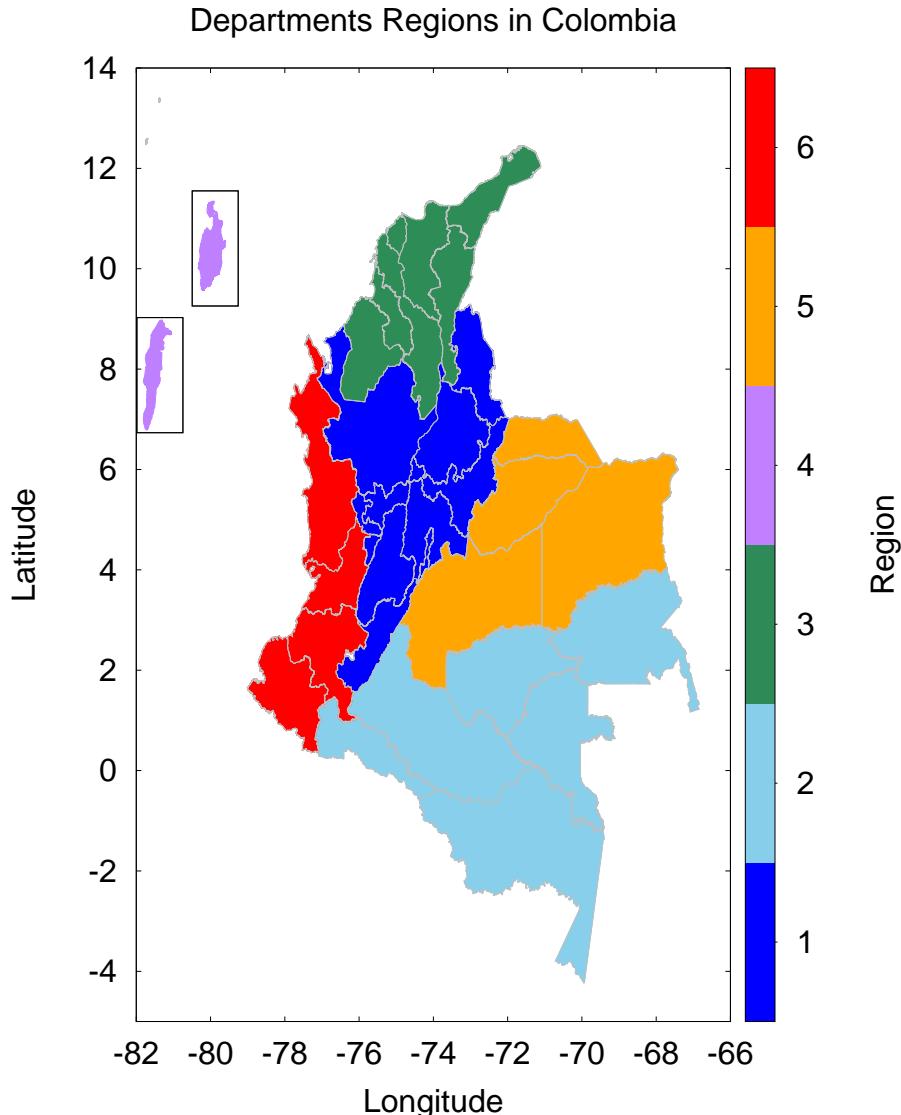


Figure 5.18: Regions of Colombia. 1. Andina, 2. Amazonía, 3. Caribe, 4. Insular, 5. Orinoquía, 6. Pacífico.

(17008) and Bolívar (17013). In this region, the model predicts significantly more cases than those reported in the national health care system, with the exception of the department of Sucre (17070). Sucre has reports that are higher than the predictions of the model, although the incidence is close to the prediction.

The incidence curves by region are presented in figure 5.20. The Andean region shows a discrepancy in the curves, however their magnitudes are comparable. The data curve has a steeper initial growth that decreases after week 20. Whereas in the model, the epidemics takes longer to develop followed by an exponential increase.

The Amazonian region has a low number of cases in both the model and the data, reaching less than 1 case per 10,000 inhabitants. This agreement shows that in both the model and reality, the amazonian region has few contacts with the rest of the country because its difficult access.

The caribbean region, that was previously identified as the region with the larger error, shows an initial similarity between both curves. After week 10, the cases reported in the surveillance system have a slower incremental rate, in contrast with the model, where the number of cases continue to increase up to the total simulation period.

The Insular region is composed by one department (San Andrés) composed of small islands. The connectivity with the country is strictly restricted to plane travel. Due to its distance from the rest of the country, migration cannot be represented only by the contacts within the country. The region has complex chikungunya dynamics, the first cases are reported before week 9, however there is not a clear pattern in the curve. The reports contain multiple weeks with null cases within the epidemic period. Hence, these cases could be attributed to migration of cases instead of autochthonous transmission. In the model, the epidemic starts in week 20 with a slow growth rate. This region should be simulated separately from the country in order to analyze the possible impact of a chikungunya epidemic.

Finally, the Orinoquía and Pacific regions show the more accurate fit of the six regions, considering the magnitude, shape and timing of the curves. However, the available data for these regions show that the epidemic is under the early stages, so future conditions of the epidemic in these regions could shape differently the incidence curve.

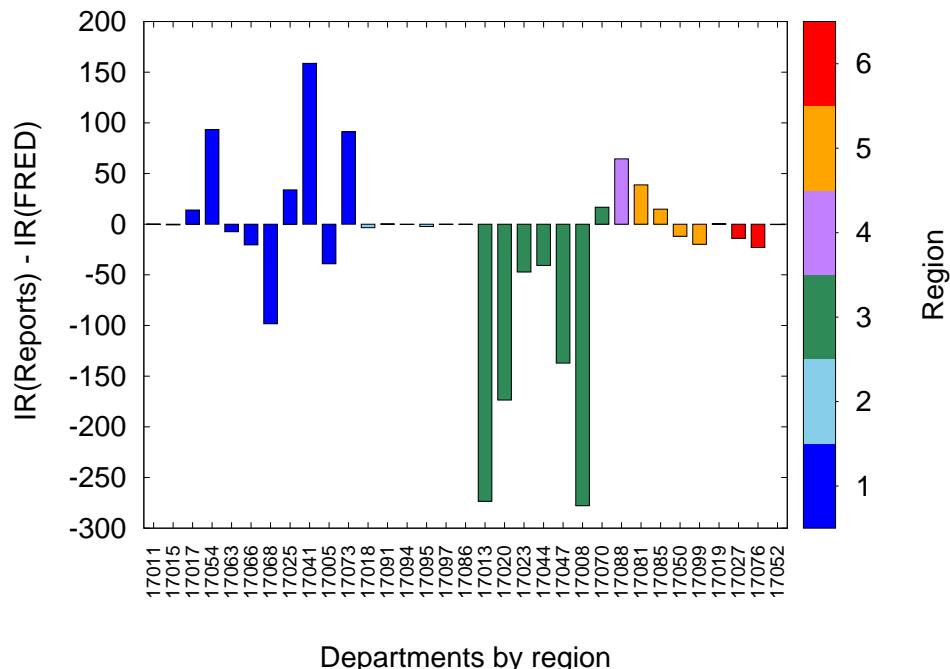


Figure 5.19: Error of reports (Individuals per 10000 inhabitants) - Model predictions (Individuals per 10000 inhabitants) by region.

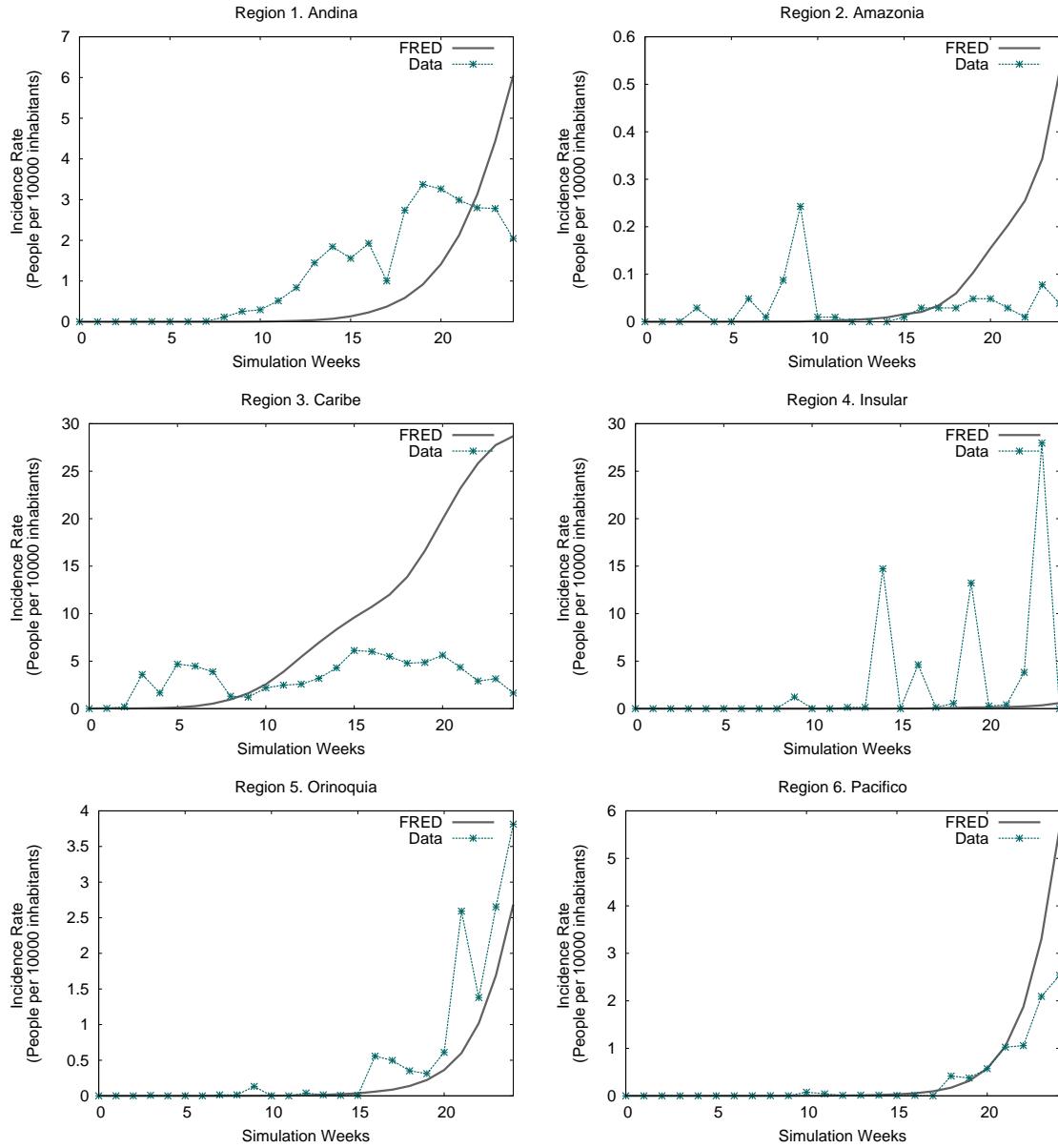


Figure 5.20: Model and reports incidence curves by region in Colombia.

Considering the above results, there are numerous explanations for the discrepancies. First, the model was calibrated using only one municipality of the country. The variation between the model and the data can also be attributed to heterogeneity in the mosquito density parameters across the different regions, departments, and municipalities. In addition, the model assumes that no vector control interventions take place in any of the departments. Whereas in reality, some departments might have implemented multiple strategies to reduce the impact of the epidemic.

There are vector control actions that can be implemented by each department that could lead to reduction of the pupae density before the epidemic started. One hypothesis is that these interventions can be linked to the income level of the department, therefore low income departments would be linked to higher pupae density areas. The temperature parameter, which is included in the model, can also affect the mosquito density. In the next chapter, the control actions will be discussed further. However, in this chapter the baseline scenario is described.

In figures 5.21 and 5.22, the difference in the cumulative incidence by department are sorted by temperature and an unsatisfied basic needs index, respectively. The largest errors are found in the high temperature regions, where the model overestimated the total number of cases. Although, there are other departments with large errors in low temperature regions.

The model overestimates the number of cases predominantly in regions with low and high income. In contrast, the model underestimates the number of cases in regions with medium income level. This effect could suggest that there are socioeconomic factors that the model is ignoring and that affect chikungunya dynamics. In order to formally understand the sources of error in the model, a regression analysis go the errors was performed using additional covariates listed in table 5.11.

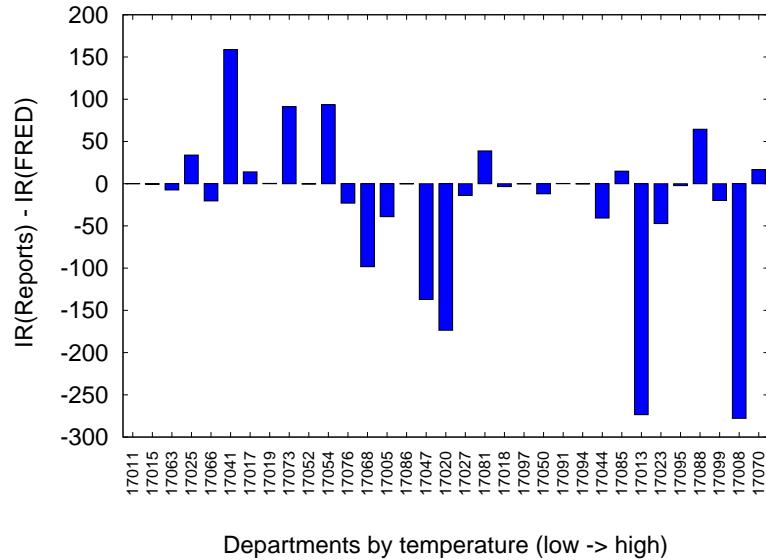


Figure 5.21: Error of reports (Individuals per 10000 inhabitants) - Model predictions (Individuals per 10000 inhabitants) by region.

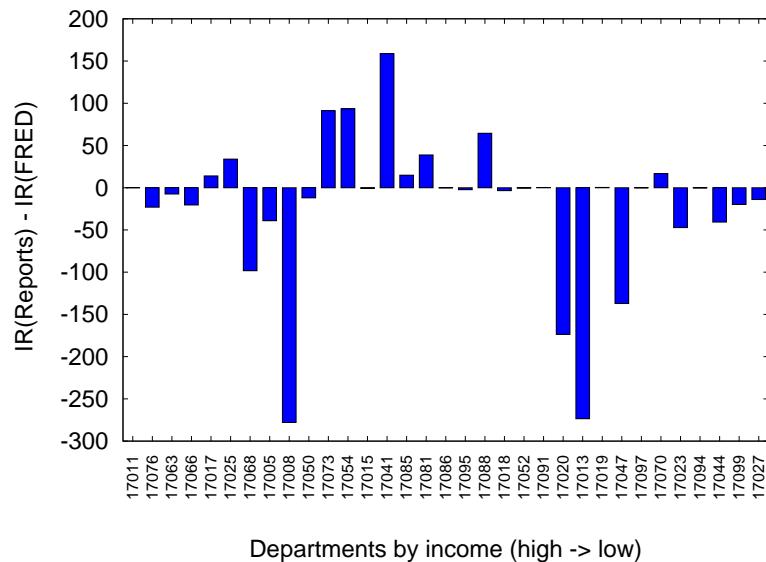


Figure 5.22: Error of reports (Individuals per 10000 inhabitants) - Model predictions (Individuals per 10000 inhabitants) by region.

The regression analysis was performed in Matlab using the function *regress*. The difference between the data and the model was set as the observations vector and the covariates listed in table 5.11 were used as the input matrix. The region attribute was assigned as six boolean values. The resulting coefficients are: $\beta_{population} = -4,76$, $\beta_{temperature} = 4,85$, $\beta_{area} = -25,39$, $\beta_{BNU} = 57,81$, $\beta_{precipitation} = -36,63$, $\beta_{region1} = 0$, $\beta_{region2} = -29,17$, $\beta_{region3} = -273,34$, $\beta_{region4} = -45,28$, $\beta_{region5} = -24,71$, $\beta_{region6} = -59,67$.

These coefficients show that the error between the data and the model are highly affected by the socioeconomic factors (represented as the unsatisfied basic needs) and the precipitation. Low-income regions could have higher pupae density because of the lack of education or the cost of pupae reduction activities. Also, precipitation has shown to increase the density of pupae per household, according to the WHO [127]. Hence, including the variation of pupae due to precipitation could improve the model predictions. Because of the complexity of the model, these variables have not been included. Additionally, the Caribbean region (3) shows the highest uncertainty from this analysis, figure 5.23. The residuals from the regression analysis are presented by region, displaying the increased error between the model and the surveillance incidence reports in the Caribbean region.

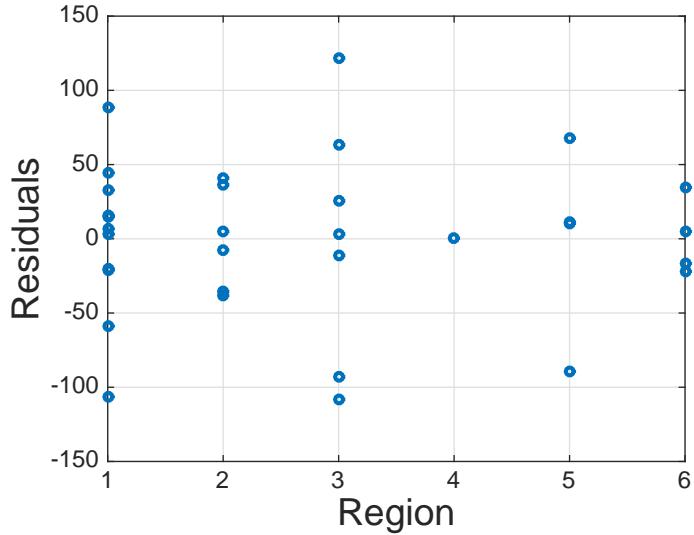


Figure 5.23: Residuals of regression analysis of the model's prediction errors.

Incidence by Municipality

The incidence rates per week by municipality are presented in figure 5.24. The left figure shows the reports from the surveillance system in Colombia. A few important characteristics can be derived from the graphs. The first cases were reported in warm areas that have similar annual temperatures. The epidemic was limited to the warm regions of the country up to 2015, when the epidemic started to spread to some municipalities with colder temperatures. In comparison, the model shares these attributes, with the epidemic starting in the same region and spreading to cold areas after 2015. Nonetheless, the model over predicts the spread of the disease in warm areas. As discussed in the previous section, this discrepancy can be attributed to limitations in the model or to the implementation of vector control programs in the municipalities that were at risk.

The errors of cumulative incidence rates were also computed by municipality as shown in figure 5.25. It is possible to see that the error is focused on the Caribbean region (3) of the country, which is in agreement with the department level analysis. Also, the errors caused by the existence of higher case counts in the data compared to what was used in the model are focused in the Andean region (1).

Code	Error	Reports	Model	Population	Region	Temperature	Area (Km^2)	NBI
17011	0.08	0	0	7363782	1	9.47	1876	9.20
17066	-20.44	0	20	925117	1	18.41	4203	17.47
17017	14.02	18	4	978342	1	19.26	8705	17.76
17025	33.90	37	3	2477036	1	18.09	26324	21.30
17054	93.47	198	105	1297951	1	20.64	25790	30.43
17063	-7.42	0	7	549662	1	17.06	2272	16.20
17068	-98.29	6	104	2010393	1	21.89	35995	21.93
17005	-39.03	7	46	6066003	1	22.92	74484	22.96
17073	91.18	108	17	1387621	1	19.85	28208	29.85
17015	-0.68	0	0	1267652	1	15.53	27268	30.77
17041	158.73	163	5	1083189	1	19.23	21355	32.62
17097	-0.10	0	0	41534	2	25.52	62256	54.77
17018	-3.47	0	3	447767	2	25.34	105351	41.72
17086	-0.12	1	1	326093	2	23.62	30353	36.01
17091	0.25	0	0	72017	2	25.84	127439	44.41
17094	-0.21	0	0	38328	2	25.97	82857	60.62
17095	-2.35	0	2	103307	2	26.83	64948	39.89
17013	-273.49	109	382	1980012	3	26.57	31742	46.60
17044	-40.67	120	161	818740	3	26.01	24584	65.23
17020	-173.56	13	187	966450	3	25.06	26926	44.73
17070	16.77	219	202	810664	3	27.40	12693	54.86
17008	-277.84	28	306	2314460	3	27.21	3977	24.74
17047	-137.07	65	202	1201501	3	24.53	27262	47.68
17023	-47.20	81	129	1582784	3	26.62	29436	59.09
17088	64.47	67	2	73320	4	27.00	2466	40.84
17085	14.87	22	7	325621	5	26.29	51952	35.55
17050	-11.99	1	13	870921	5	25.56	100198	25.03
17099	-19.81	0	19	63670	5	27.14	116597	66.95
17081	38.78	45	6	247541	5	25.14	27946	35.91
17027	-13.92	0	13	476149	6	25.14	56398	79.19
17052	-0.43	0	0	1639560	6	20.37	36034	43.79
17076	-23.08	14	37	4383277	6	21.14	24134	15.68
17019	0.36	1	1	1319120	6	19.31	36112	46.62

Table 5.11: Errors of prediction of cumulative incidence rates (per 10000 inhabitants)and the list of covariates by department.

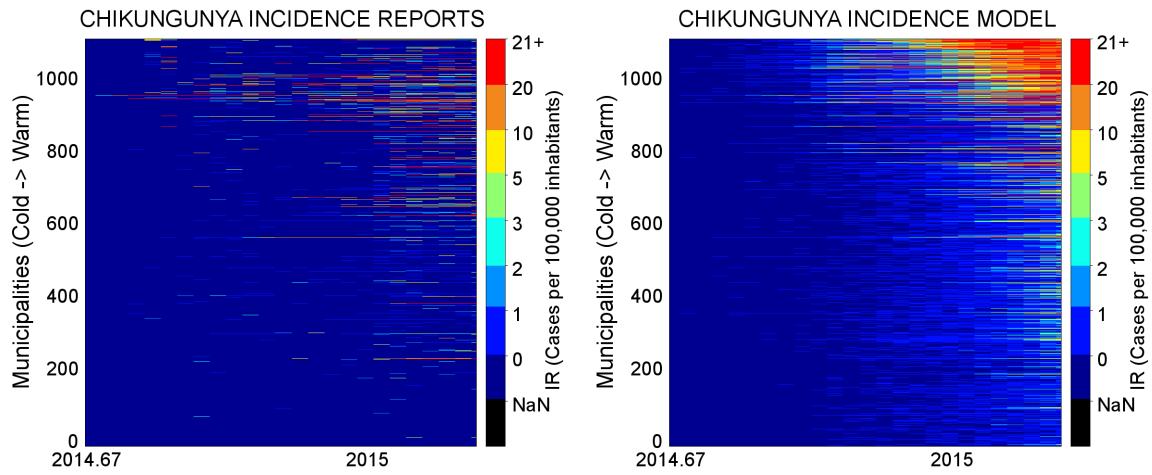


Figure 5.24: Incidence cases by municipality reported by the surveillance system (Left) and predicted by the model (Right).

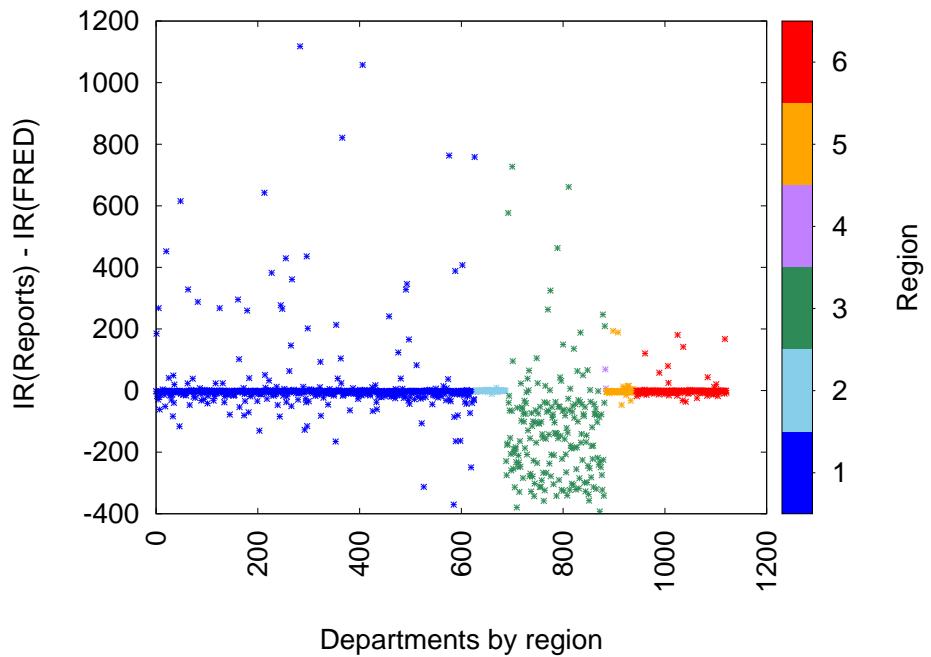


Figure 5.25: Error of reports (Individuals per 10000 inhabitants)- Model predictions (Individuals per 10000 inhabitants) by municipality.

5.5 Summary

In this chapter, the data used to calibrate the model was presented, as well as different strategies used to calibrate it. A global optimization strategy was used to calibrate the model to the epidemic of one municipality of the country that experienced an outbreak with a significant impact (Riohacha, Guajira). These parameters were calibrated: the mosquito probability of infection, the human probability of infection, and the diagnosis rate. The model fits closely to the surveillance reports of this Riohacha.

The performance to unfitted data was evaluated using 25 municipalities that experienced an outbreak of chikungunya at various levels. The model presents acceptable fit performance for around 60 % of the unfitted data. However, some municipalities' epidemic are overestimated by the model.

The model was also setup to reproduce a similar epidemic in the entire country of Colombia. Factors, such as the synchrony of the epidemic in different regions were captured by the model, mostly for the Orinoquía and Pacific regions. Nonetheless, the model overestimates the cases in the Caribbean region. These errors were analyzed using a regression model.

The coefficients of the regression model suggest that the socioeconomic factors and the precipitation should be included in the model to improve its performance. Possibly, vector control programs were implemented in reality, hence, pupae density was decreased in various regions. Including vector control in the model can help to understand these differences.

Finally, the behavior of the model compared to the data raises the question of what vector programs were implemented in the country. With this information, the model could be calibrated to represent the actual distribution of the virus across the country. Unfortunately, information about specific vector control strategies implemented in each region is currently unavailable. Nonetheless, it is possible to analyze the impact of vector control strategies that were implemented in some municipalities like Santa Marta (47001), where the initial growth rate agrees with the model's output, but then suddenly decreases.

Chapter 6

Vector Control Strategies

Vector control is the only intervention currently available to control the spread of dengue and chikungunya virus. In the case of dengue, multiple vaccine candidates are currently being developed and a dengue vaccine is expected to be introduced in endemic regions in the short term. In regards to chikungunya, there are no expectations for a vaccine in the near future. Consequently, the effective implementation of vector control programs is crucial to decrease the impact of the ongoing chikungunya outbreak in the region. As discussed in the previous chapter, multiple factors can affect the transmission of chikungunya within a population. In this chapter, the previously exposed chikungunya calibrated model will be used as a baseline for the comparison and evaluation of a vector control strategy based on pupae reduction.

6.1 Santa Marta Case

The Santa Marta case is selected as an example of vector control interventions within an epidemic. As shown in figure 6.1, the epidemic curve in Santa Marta is characterized by a sharp initial growth

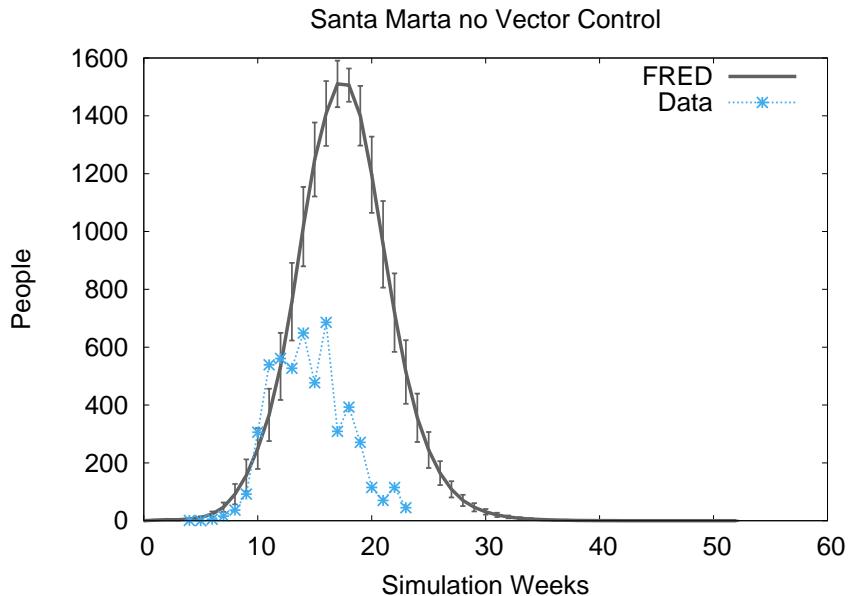


Figure 6.1: 100 simulations of the model for Santa Marta without implementing vector control programs contrasted to the surveillance reports. The model fits the initial growth rate but the data shows a sudden decay that the model misrepresents.

rate, that is consistent with the model predictions using the calibration from Riohacha. However, after week 10 of the simulation, the shape of the curves differ.

Experimental and theoretical curves agree in the initial phase of the epidemic, but the model estimates a larger epidemic than the cases reported. The model's peak is approximately twice as large as the reports. Also, the model's curve is considerably wider than the observed cases. The surveillance reports show a duration of approximately 7 weeks when the slope of the epidemic curve tends to zero. At week 17, the number of cases decrease drastically until low incidence is achieved at week 20. In contrast, the FRED model curve continues growing to a peak of around 1600 cases reported a week; followed by an exponential decrease until week 30.

Hence, there are multiple questions that arise from this scenario. First, could the model reproduce the data when implementing vector control? If so, could the interventions have been improved to obtain a significant reduction of reported cases? What is the impact of the different parameters in the vector control strategies for the reduction of symptomatic cases? Also, what was the impact in terms of reduction of symptomatic cases of the applied vector control strategies? In order to address these questions, a vector control strategy was implemented into the model.

6.2 Vector Control Strategies in the Model

Contingency guidelines of the chikungunya virus advise to implement community based vector control strategies in order to reduce the impact of the epidemic [44, 45]. Also, in Colombia the ministry of health promotes the social mobilization with different campaigns and social media. The country's response focuses on the reduction of the *Aedes* population in risk areas, which are defined by previous dengue epidemics [128].

Within the model, a community based strategy is implemented to reduce the immature states of the pupae. The reduction is represented through the pupae density, based on the efficacy of the intervention. Hence, a steady reduction in the adult mosquito population would be achieved as well.

The strategies begin after the number of weekly diagnosed cases exceeds a defined threshold. Neighborhoods are enrolled at a constant rate in the vector control programs based on a determined parameter called “recruitment rate”. The enrollment prioritize at-risk regions defined by occurrence of infectious cases. Neighborhoods in the model are defined as $1Km^2$ regions. A proportion of establishments participate in vector control programs by each enrolled area. This proportion is defined as the “participation rate”. Finally, the programs would be carried out until a stopping time is reached. The parameters of the vector control strategies are listed below:

- **Threshold** [Cases per 10000 inhabitants]: After the threshold is exceeded in a municipality, the interventions begin . If the threshold is exceeded by the capital of the department, the remaining municipalities enroll immediately in the vector control programs.
- **Recruitment Rate** [%area / day]: When the simulation starts, areas delimited by $1Km^2$ are classified as infectious or non-infectious areas. When the threshold of vector control is exceeded in a municipality, a percentage of the area is covered starting daily with the infectious areas and continuing to the non-infectious areas.
- **Participation Rate** [%places / $1Km^2$]: Inside each $1Km^2$ area, a proportion of places (households, schools, or workplaces) enroll in the vector control programs. The remaining places not to apply vector control. The places are selected randomly.
- **Efficacy** [%pupae reduction/place]: The steady state efficacy of the vector control strategy that is to be used. This parameter may combine the efficacy of multiple interventions, such as behavioral, biological or chemical.
- **Stop Time** [days]: The day when the interventions stop. After this, the pupae density inside each location returns to the initial state.

Parameter	Min	Max	Value	Reference
Threshold	0	10000	18.979	Calibrated
Recruitment Rate	0	1	0.078	Calibrated
Participation Rate	0	1	0.8	
Efficacy	0	1	0.8	
Stop Time	0	-	∞	-

Table 6.1: Parameter values of the vector control programs adjusted from the literature and using the *Simplex Nelder-Mead* optimization algorithm.

6.3 Calibration of Vector Control to Surveillance Data

The vector control parameter values are adjusted to represent the Santa Marta observations, using values from the literature and a calibration algorithm. The participation rates and the efficacy of reduction were set to 80 % based on values obtained from the literature [129, 130, 131, 132, 133, 134, 135].

The threshold and *recruitment rate* parameters were calibrated using the same calibration algorithm described in chapter 5, since data about these parameters were inconclusive. However, the Colombian guidelines recommends applying control in a municipality immediately after a case is reported in the health system, and covering the area as quickly as possible [128]. Finally, the stop time was set to be longer than the simulation period, since there is no reason to believe that the vector control strategies stopped within the period of the epidemic. The values of these parameters are listed in table 6.1.

The model captures qualitatively the peak and shape of the data as shown in figure 6.2. However, with this vector control strategy, the model is unable to reproduce precisely the growth and decay rates of the incidence curves. Nonetheless, this approximate response can be used to explore the effects of the different parameter's values in the control strategy.

The parameter values estimated suggest that the control was implemented early in the epidemic

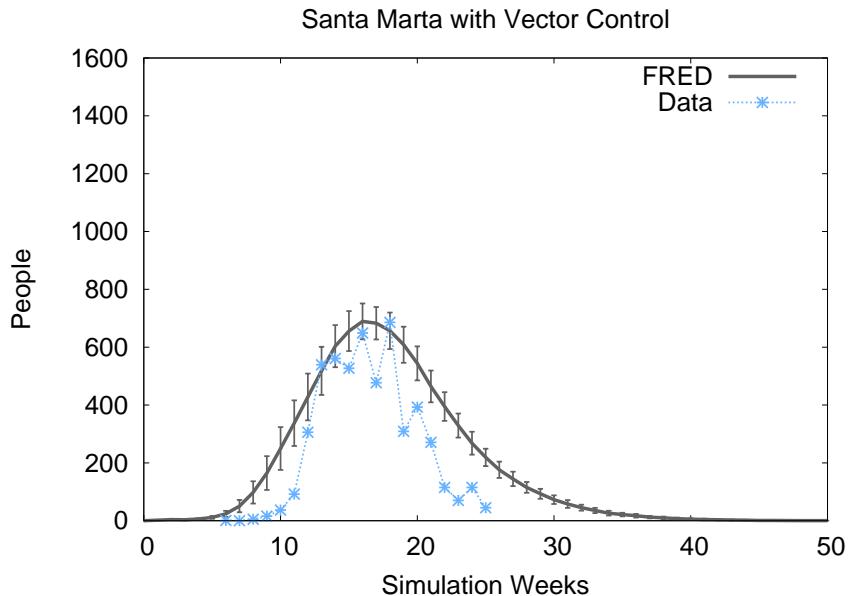


Figure 6.2: 100 simulations of the model for Santa Marta implementing vector control programs contrasted with the surveillance reports. The model fits the data better compared to the simulations without vector control.

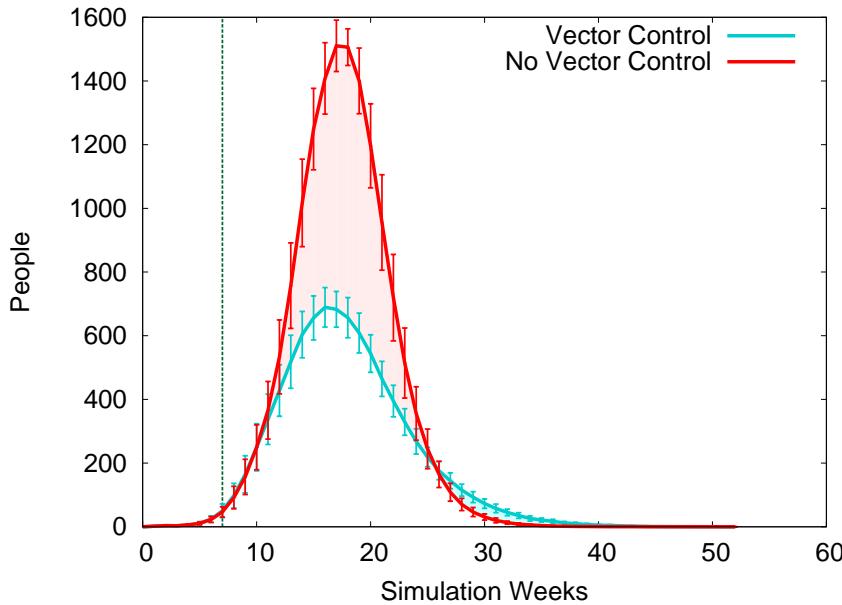


Figure 6.3: Model simulation of Santa Marta with and without vector control strategies. The red shade represents the cases overestimated by the model without including vector control. The blue shade corresponds to the additional case counts estimated by the model with vector control.

with a fast requirement rate, causing a significant reduction in the reported cases as presented in figure 6.3. From the graph, it can be appreciated that the peak of the cases when implemented interventions was reduced to half the estimated without vector control. Also, the initial growth rate of both curves is similar, even a few weeks after the interventions are implemented. This shows that the interventions effects are delayed even if they implemented early in the epidemic.

The interventions could have had a greater impact if implemented before the epidemic, because the reduction of the mosquito population is delayed. Figure 6.4 shows how the mosquito population was reduced after the interventions were implemented. A stable condition of the population is reached after approximately 10 weeks of the beginning of the interventions. This suggests that the time of reaction is very short in order to prevent the spread of the disease, specifically in Santa Marta. However, even with this delay, the campaigns showed a positive impact on the population.

Approximately eight thousand diagnosed cases were prevented using vector control in Santa Marta, according to the model comparison with the data. As shown in table 6.2, the model estimated 13,717 cases whereas 5214 cases were reported to the health system. In the total simulation period, the model estimates a reduction of roughly six thousand reported cases.

Moreover, from the model results, almost 40 % of the population in Santa Marta would be infected by the chikungunya virus, even though a small fraction of these cases were reported. Although this is a high estimate, without the vector control interventions the attack rate was estimated to be above 60 %. Hence, the model estimates that around 20 % of the population was protected by the vector control programs. It should be noted that the underreporting proportion is an estimation from the calibration process, it has not been measured in the city, hence these results could vary.

6.4 Impact of Parameters' Values in the Vector Control Strategies

The five parameters of vector control are varied in order to observe their effect on the epidemic curve. This can help to determine critical parameters to optimize future vector control programs.

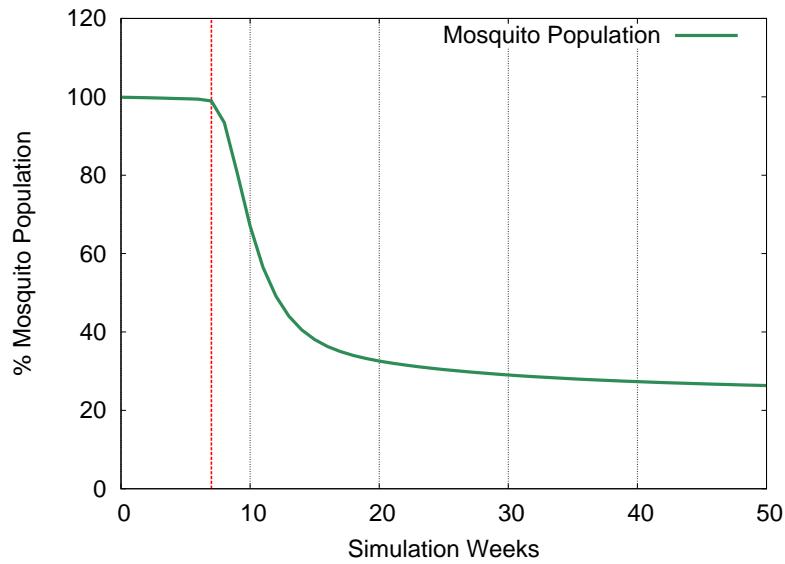


Figure 6.4: Mosquito population when applying vector control. The red dashed line represents the introduction time of the interventions.

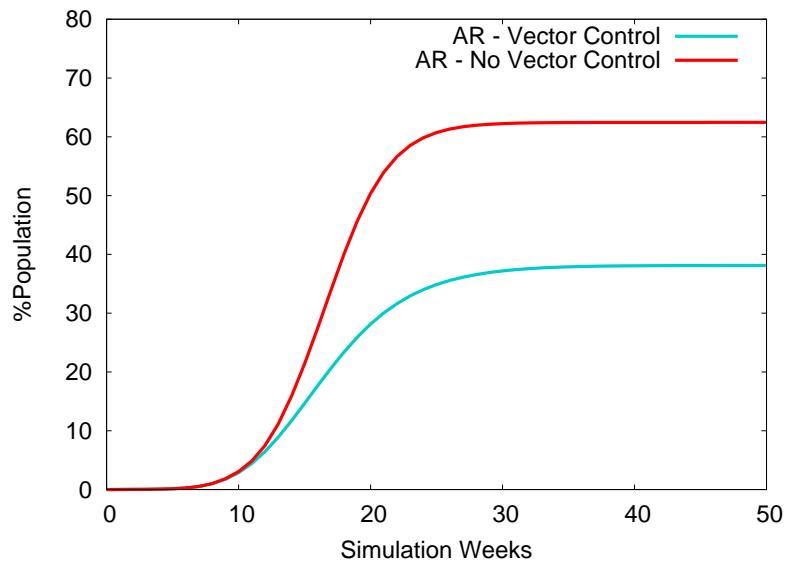


Figure 6.5: Estimated Attack Rate of the chikungunya epidemic in Santa Marta with and without vector control.

Threshold

The threshold of the introduction of interventions was varied from 1 to 500 cases per 10000 inhabitants. For each value of this parameter, the percentage of reduction was calculated as:

$$\%Reduction_{th} = \frac{\sum_t C_{baseline}(t) - \sum_t C_{sth}(t)}{\sum_t C_{sth}(t)} \times 100$$

Where th : threshold $\in \{1, 2, \dots, 500\}$, Cs : Diagnosed cases.

As seen in figure 6.6, the percentage of reduction decreases exponentially when the threshold increases. The model estimates a 70 % in reduction of cases when the threshold is set to 1 (cases/10,000 people). Also, this analysis shows that the impact of the interventions is reduced to 20 % when the threshold is set to 90 (cases/10,000 people). After 200 (cases/10,000) the scenario with vector control produces similar results than without vector control.

This analysis also shows how the shape of the curve varies with different values of the threshold parameter. When the threshold is low, the peak is low as well, but the shape of the curve is wider. While, narrow curves are obtained when the threshold is set to high values. This could be beneficial for the health care system, because a curve that is more spread means that the cases are distributed within the period, hence fewer number of cases are reported weekly, allowing the health system to provide more access to new infection cases.

In general, the faster the interventions are implemented the higher the percentage of reduction. Also, the benefit from this type of vector control is null after the epidemic has reached values around 200 cases/ 10,000 per week.

Scenario	Total Symptomatic Cases
Reports	5214
Model without control (Data period)	13717 ± 1681
Model without control (Total simulation)	14805 ± 1974
Model with control (Total simulation)	8866 ± 1306

Table 6.2: Total count of cases reported by the surveillance system in Santa Marta and predicted by the model.

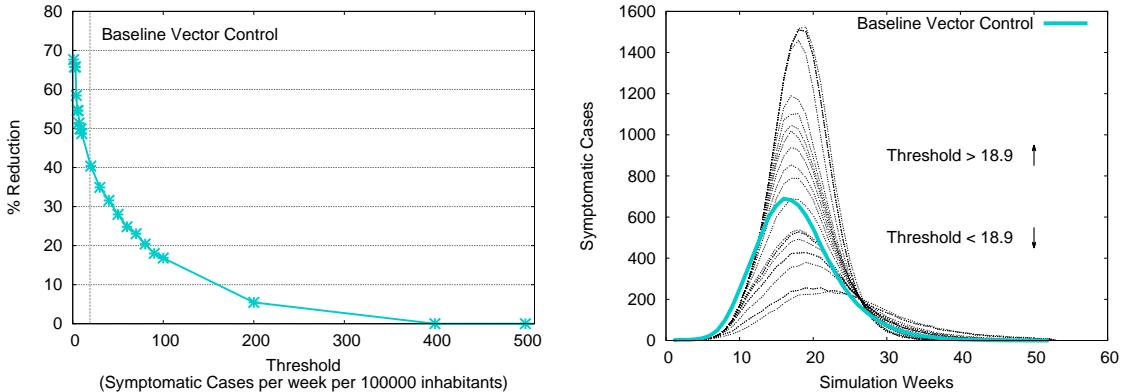


Figure 6.6: Variation of the threshold of cases to start the intervention programs. The figure at the left represents the percentage of reduction in symptomatic cases when varying the threshold to initiate vector control interventions. The figure at the right shows the different curves produced with the multiple threshold values.

Recruitment Rate

The benefit of increasing the recruitment rate shows a saturation effect around the estimated value, as presented in figure 6.7. The recruitment rate was modified from 0.0001 (0.01 % of total area covered per day) to 0.2 (20 % of total area covered per day). The percentage of reduction was calculated in a similar way than for the threshold variations. It should be considered that this vector control strategy gives priority to infectious neighborhoods. Hence, these results suggests that the recruitment rate is important to cover infected areas, but the incremental benefit is minimal when protecting noninfectious areas.

The interpretation of this parameter is also linked to the value of the threshold of cases that initiates the interventions. Due to the early implementations of the interventions, high speed recruitment is not relevant. However, when larger areas have been infected, the recruitment rate would need to be higher in order to control the epidemic. This also indicates that interventions implemented early can be effective at a small recruitment rate. For instance, preventive measures can benefit the population when they focus on previously identified risk areas.

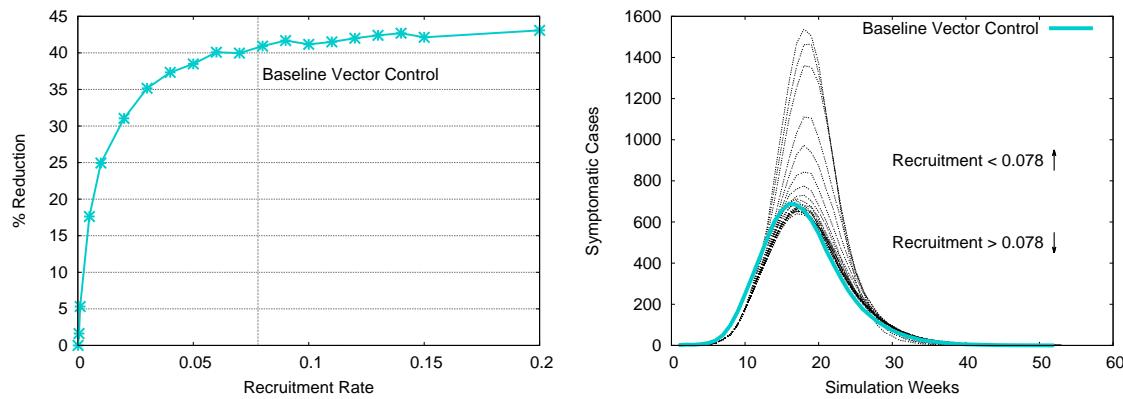


Figure 6.7: Variation of the recruitment rate of the intervention. The figure at the left represents the percentage of reduction in symptomatic cases when varying the recruitment rate. The figure at the right shows the various curves produced with the multiple recruitment rate values.

Participation Rate

As presented in figure 6.8, 12 % more cases are reduced the participation rate is set to 100 %. This parameter was modified from 10 % to 100 % suggesting a linear correlation between the percentage of reduction and the participation rate with a positive slope of around 0.5 reduction per increase in participation rate.

Efficacy

Figure 6.9 shows the correlation between the reduction of cases and the values of efficacy, varied from 10 % to 100 %. The results suggest that it is necessary to have efficacy values above a threshold of 50 % to obtain a significant reduction in reported cases. Above 60 % of efficacy, the reduction of cases follows a rectilinear shape with a positive slope of 1.75 [%reduction / %efficacy].

These results portray the importance of implementing interventions that effectively reduce the pupae population, since the reduction of cases is null for values under 50 % of efficacy. The model's results suggest that control programs should verify the efficacy of the strategies at different times of the intervention in order to ensure significant reduction.

However, these results are limited to the baseline scenario, in which the interventions were implemented once the epidemic initiated. Further estimations are necessary to evaluate the required efficacy in other scenarios, such as preventive control. Also, this analysis elucidates the sensitivity of the mo-

del's response to variations in the efficacy parameter. The model would benefit from the ability to obtain measurements of this parameter in the vector control programs implemented in Santa Marta.

Stop Time

The model suggests that the reduction of cases is insignificant when the control programs are ceased in the early stages of the epidemic, figure 6.10. The model was simulated using various lengths of the interventions, from 60 to 700 days. A constant migration of infectious agents was included to simulate possible re-invasions of the virus.

Secondary epidemics are produced when the interventions are stopped after the peak of the first outbreak, i.e. around 20 simulation weeks. The peak of these epidemics are smaller than the scenario without vector control. This suggests that the programs should be implemented even after the epidemic has passed in order to prevent new infectious cases.

From the above parameters, the threshold parameter showed the highest impact in reduction of cases. Initiating the interventions early in the epidemic showed an improvement 30 % in reduction of cases, compared to the baseline scenario. Further, the benefit in reduction of cases is minimal when the interventions are implemented after a threshold of 200 [cases per 10,000] is reached.

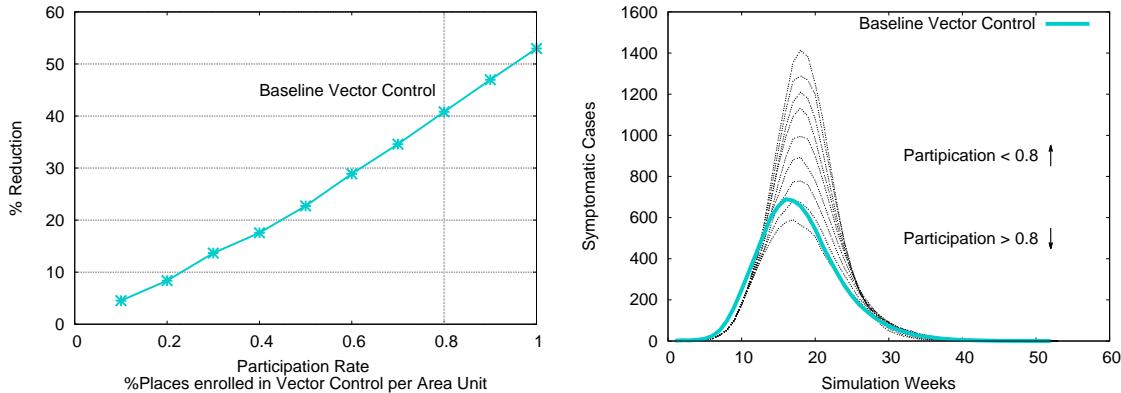


Figure 6.8: Variation of the recruitment rate of the intervention. The figure at the left represents the percentage of reduction in symptomatic cases when varying the participation rate. The figure at the right shows the different curves produced with the multiple participation rate values.

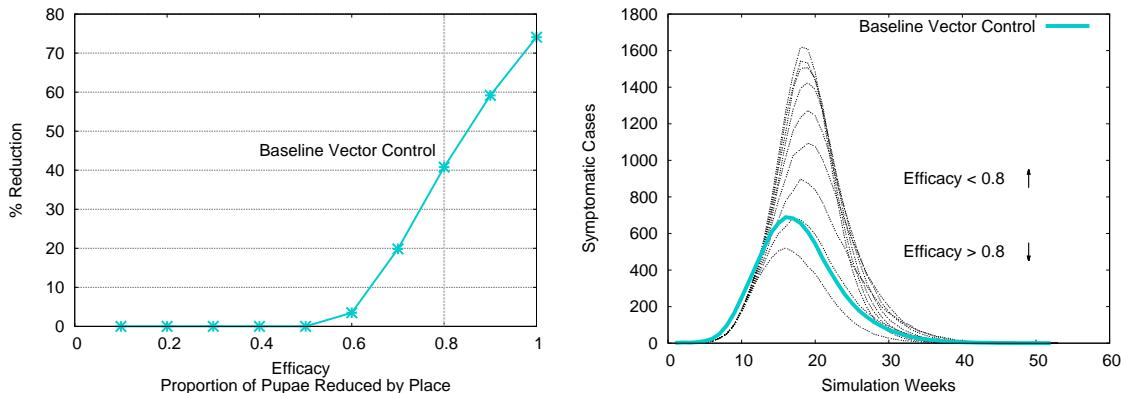


Figure 6.9: Variation of the recruitment rate of the intervention. The figure at the left represents the percentage of reduction in symptomatic cases when varying the efficacy of the interventions. The figure at the right shows the different curves produced with the multiple efficacy values.

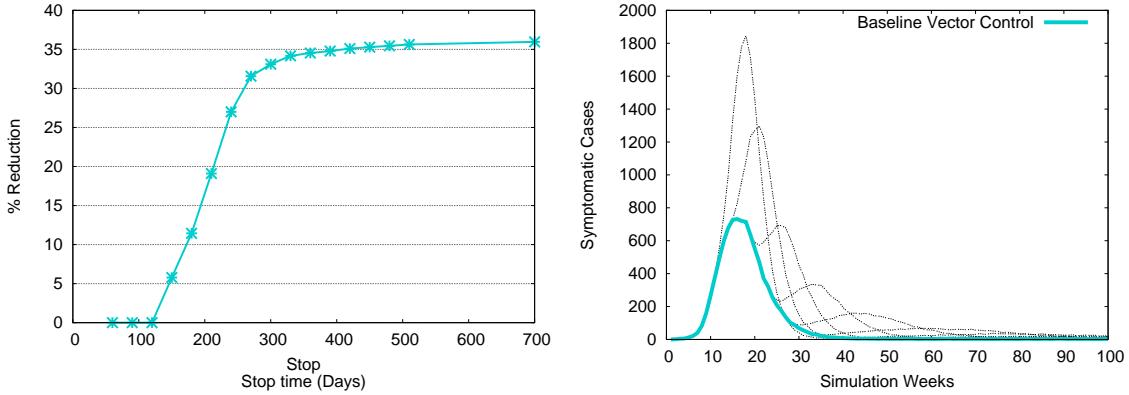


Figure 6.10: Variation of the stop time of the intervention programs. The figure at the left represents the percentage of reduction in symptomatic cases when varying the stop time. The figure at the right shows the different curves produced with the multiple parameter values.

After analyzing the impact of the parameters related to vector control strategies in the model, it is possible to remark that the threshold showed the highest impact in terms of reduction of symptomatic cases. Improving the start of the programs could benefit significantly the population at risk. Also, starting the vector control programs late in the epidemic does not show significant benefit. Improving the control programs by implementing them early in the epidemic would also mean that the recruitment rate could be slower, consequently, less economic effort would be needed.

The participation rate and efficacy values showed a linear correlation with the reduction of cases around the calibrated point. However, significant reduction of cases was obtained for efficacy values above 60%. Control programs should measure this parameter to assess the effectiveness of the interventions. Also, values of the pupae reduction efficacy in Santa Marta control programs would benefit the model's calibration and estimations.

The goal is to analyze the effect of implementing vector control strategies in the entire country. The calibrated vector control program for Santa Marta was implemented in different municipalities in order to evaluate which municipalities should be the focus when assigning resources for vector control.

6.5 Vector Control Strategies at National Level

From the 1121 municipalities of Colombia, 770 register annual temperatures above 18°C. These municipalities allow the development of the *aedes aegypti*, and therefore the transmission of dengue or chikungunya. As mentioned previously in chapter 2, a subset of 342 municipalities reported a dengue outbreak in the 2010 epidemic, compared to 105 municipalities with chikungunya. The focus of this section is to evaluate the benefit of focusing a vector control intervention in diverse sets of municipalities, based on dengue and chikungunya risk.

Scenario	Symp. Cases (Data Period)	Symp. Cases (Total)	Error (%)
No vector control	239620 ± 26896	1138129 ± 81337	47
All municipalities in vc	122120 ± 15967	481886 ± 38218	-24
Municipalities with dengue epidemics	158319 ± 18847	650662 ± 42945	-2.76
Municipalities with chik epidemics	162704 ± 15993	945735 ± 61269	-0.06
hline 342 Random municipalities	211012 ± 26467	942552 ± 74532	29
105 Random municipalities	227609 ± 19840	1088670 ± 59435	39
Capital Cities	179386 ± 24187	899492 ± 65436	10

Table 6.3: Multiple scenarios of vector control. Error = (Model - Data)/Data * 100.

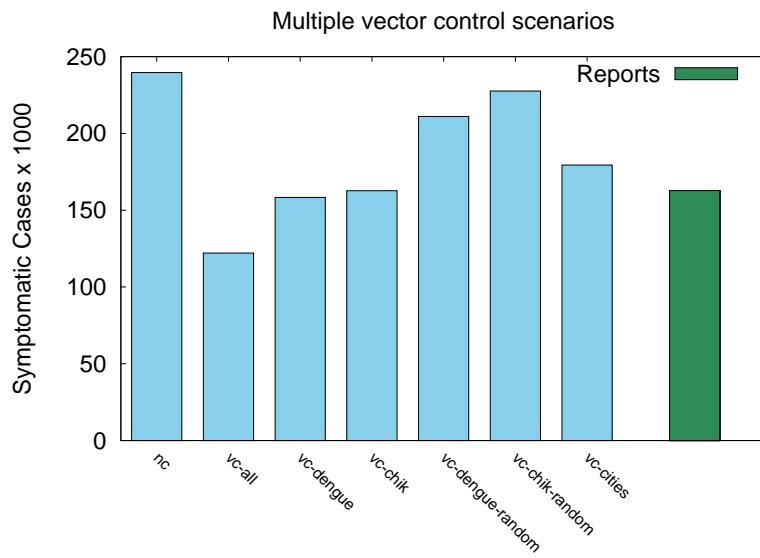


Figure 6.11: Cumulative symptomatic incidence in Colombia.

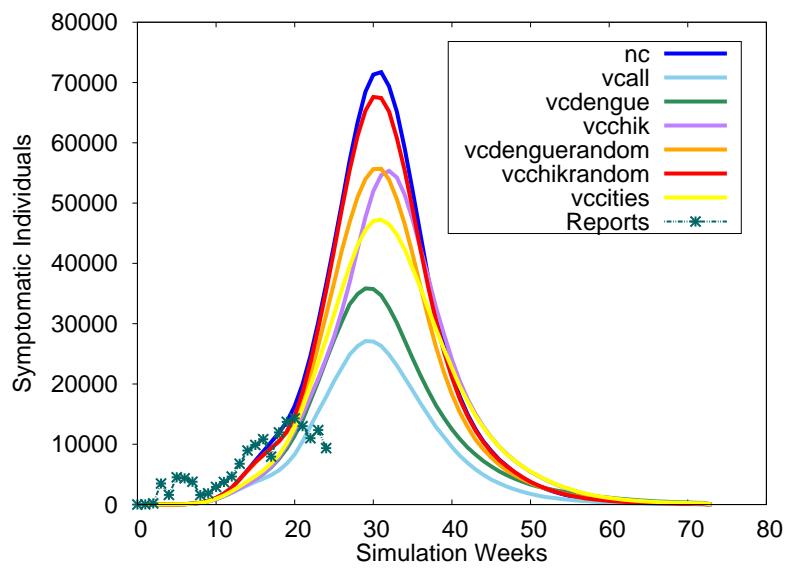


Figure 6.12: Symptomatic Incidence curves of different control scenarios in Colombia.

A total of seven different scenarios were simulated as summarized in table 6.3: no vector control, vector control implemented in all the municipalities with temperatures above 18°C, vector control in the 342 municipalities with previous dengue epidemics, vector control in the 105 municipalities where chikununya outbreaks have been reported, vector control in 342 random municipalities, vector control in 105 random municipalities, vector control only in the capital cities of the 33 departments in Colombia. These simulations were undertaken for 500 days in order to cover the full extent of the epidemic.

A reduction of around 40 % in the total reported cases is achieved when focusing the interventions in the set of dengue municipalities. More than 600 thousand cases are reported in total using this intervention, compared to approximately 1.1 million cases reported without vector control. In contrast, when selecting 342 municipalities at random, the total reduction is around 17 %. In a similar manner, a reduction of 17 % is achieved when the interventions are implemented in the 105 chikungunya identified municipalities. While less than 5 % of cases are prevented, when distributing the vector control program efforts in 105 random selected municipalities. In addition, focusing the efforts in only capital cities reduce the incidence in symptomatic cases in around 22 %.

Implementing the control programs in municipalities with previously detected dengue epidemics shows the largest benefit of the focused interventions. In comparison, 57 % of cases are prevented when all the municipalities above 18°C are selected for vector control. This suggest that if there are enough resources, the control programs should cover the entire territory that has ideal climate conditions for *aedes aegypti* development. In contrast, if the resources are limited, the control programs can be focused in dengue identified areas.

In addition, the model suggests that any structured program benefits the population more than an unstructured programs. The unstructured programs showed the smaller reduction of cases of the seven simulated scenarios. Even focalizing control interventions in capital cities provide larger benefit that any of the scenarios with random selected municipalities.

According to the model's results, the data shows that vector control strategies have been implemented in the country. As presented in figure 6.11, the cumulative cases are closer represented by scenarios with focused interventions. However, comparing the time series of the data and the model, the data adjusts closer to the curve of the vector control in all the municipalities, as shown in figure 6.12. It should be noted that, lacking from real data, the interventions simulated are a hypothetical scenario where all the municipalities implemented the same actions implemented in Santa Marta. This introduces unquantified uncertainty that can affect the results.

When comparing the dengue and chikungunya focused vector programs presented in figure 6.12, the incidence curves have the same growth rate up to week 25. After this, the virus spreads to areas unprotected by the chikungunya focused control strategies, whereas dengue focused programs significantly prevents the spread of the epidemic to this regions. The impact of the program selecting 342 municipalities randomly to implement vector control is compared to the impact of the chikungunya focused programs with only 105 municipalities. This highlights the significance of focusing vector control strategies when the resources are limited.

The model follows closely the data for the first two weeks, as discussed in chapter 5. However, after week 20 the data drops down. It is qualitatively more similar to the model when implementing vector control, but their trend disagree. In order to understand this behavior, the data and the model are contrasted by different regions in the following section. However, the analysis presented in this thesis is constricted to the unavailability of real data about the control programs implemented in Colombia.

Incidence by Regions

In order to examine the synchrony of the epidemic through the diverse geographical regions of the country, the incidence by week over the six different regions is presented below.

Andean region

The time series from the surveillance reports and the model's results for the seven scenarios in the Andean region are compared in figure 6.13. The figures show a delay of around 10 weeks of the

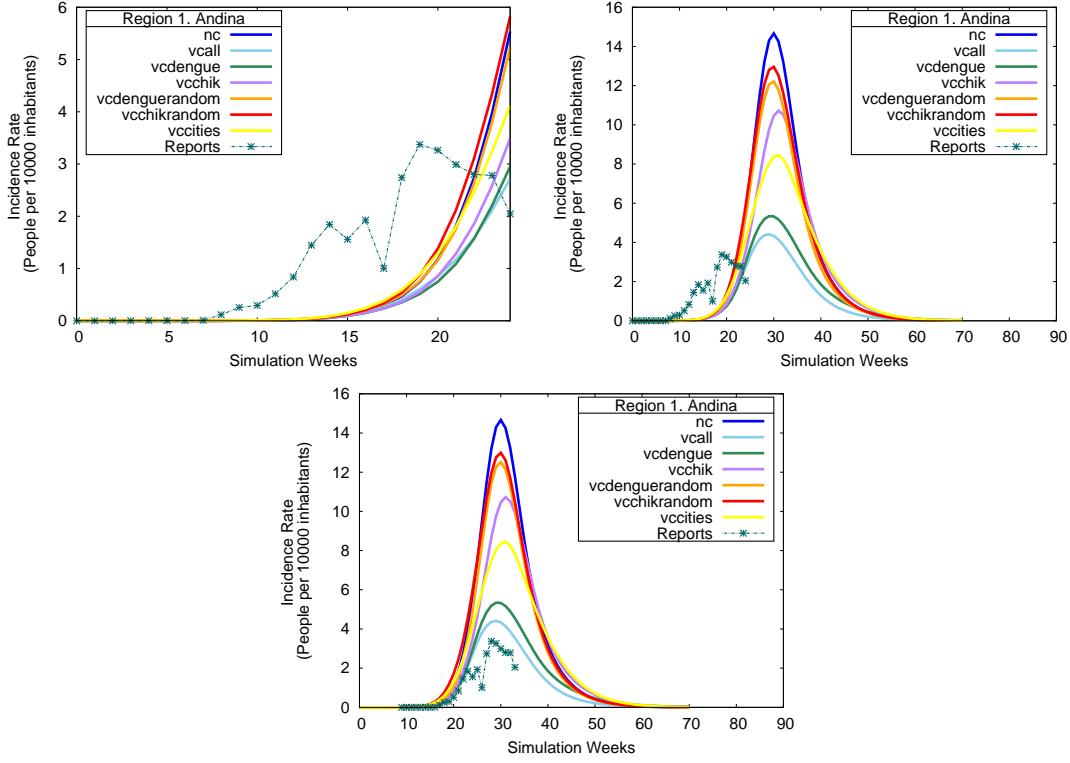


Figure 6.13: Model and reports incidence curves for the Andean region in Colombia. Top left figure compares the model and the data in the data period. Top right figure displays the data and the model for the whole simulation period. Bottom figure presents the model compared to the data delayed 9 weeks.

model's results compared to the data. This disparity can be attributed to the fact that the model assumes a single introduction of infectious cases on the Caribbean region. The Andean region can have experienced other introductions of the virus from outside of the country that are not represented by the model.

As presented in figure 6.13, when the data is delayed 9 weeks it is represented closely by the scenarios of vector control in dengue municipalities and in all the municipalities. Although, the model overestimates the magnitude of the curve in both scenarios. The remaining control programs show insignificant benefit in the reduction of cases in the Andean region. The improvement of fit to the data using vector control suggests that vector control programs were implemented in this region.

Amazon region

The incidence reports and model's response for the Amazon region are illustrated in figure 6.14. The incidence of chikungunya is low in both, the model and the reports. Less than one case per 10,000 inhabitants per week have been reported in this region. The model's results for the study period also show low incidence values. This coherence suggest that in the model and reality the number of contacts of this region with the others are low. Hence, the disease is delayed considerably to this region.

The model's response shows how the impact of chikungunya is affected only by those campaigns where dengue all the municipalities participate in vector control. Vector control programs focused on the municipalities with chikungunya affects the region with a similar magnitude than the scenario without vector control. While vector control programs focused on dengue municipalities reduces at least three times the impact of the epidemic in the population. Nonetheless, the number of data points are insufficient to assess the model's predictive value in this region.

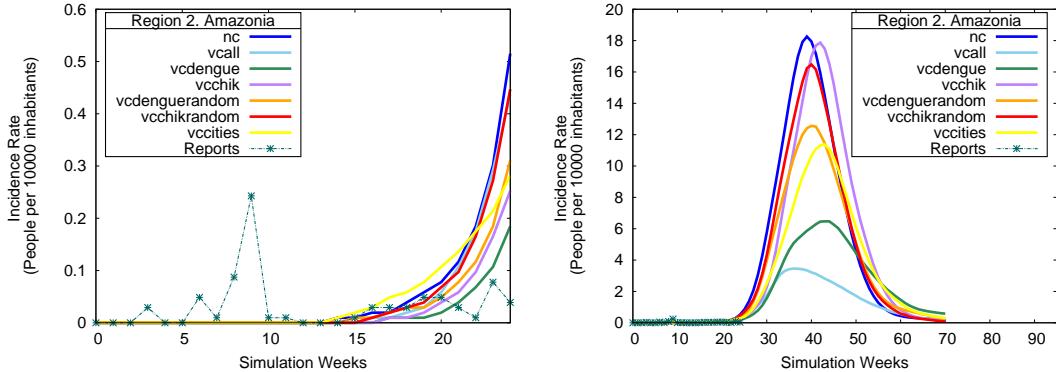


Figure 6.14: Model and reports incidence curves for the Amazon region in Colombia. Top left figure compares the model and the data in the data period. Top right figure displays the data and the model for the whole simulation period. Bottom figure presents the model compared to the data delayed 9 weeks.

Caribbean region

Figure 6.15 illustrates the comparison between the data and the model's simulations for the Caribbean region. None of the vector control strategies in the model reproduce the reports of this region. Although the scenario of all the municipalities participating in vector control follows closely the data up to week 15. After week 15, the data follows a continuous decay, while the number of cases estimated in the model increases up to week 25, where a constant decrease characterizes the curve.

As discussed in chapter 5, the Caribbean region is described by heterogeneous dynamics that the model is unable to reproduce without additional information. For instance, the model without vector control was calibrated using data from Riohacha, a municipality from this region. Also, the vector control programs were fitted using data from Santa Marta, another municipality from this region. Even including vector control, the model overestimates the impact in other municipalities such as Barranquilla. This suggest the implementation of heterogeneous vector control programs over this region, where some municipalities applied vector control before the epidemic, some others applied vector control in the epidemic and others did not apply vector control programs.

Another feature of the results shown in figure 6.15 is that the chikungunya and dengue focalized interventions have similar impact in the reduction of cases. Although, both are considerably higher than the scenario with all the municipalities enrolled in vector control. This, reinforces the idea that regions at risk of dengue are a reasonable predictor of chikungunya epidemics, and that when resources are limited their distribution in dengue areas maximize the reduction of cases. Also, randomized selection of the municipalities for chikungunya and dengue have minimum impact in the transmission dynamics of this region.

Insular region

The Insular region is composed by islands that are mainly disconnected from the country. Because of this, it is difficult to account for all the connections with the rest of the country. Also, it is a small region with a total population of 70 thousand people, compared to the Andean region that has around 24 million inhabitants. As seen in figure 6.16, the cases reported in the surveillance system include multiple weeks with zero values. Since there is not a clear trend of transmission, it is difficult to formulate a comparison within the model. Also, the reports omit the information of autochthonous cases and imported cases.

The chikunguya transmission in this region is unaffected by dengue focused interventions. Because of the separated transmission dynamics from the rest of the territory, the epidemic is influenced by interventions that include all the municipalities, capital cities, or previously identified chikungunya areas, because these strategies involve municipalities from the region.

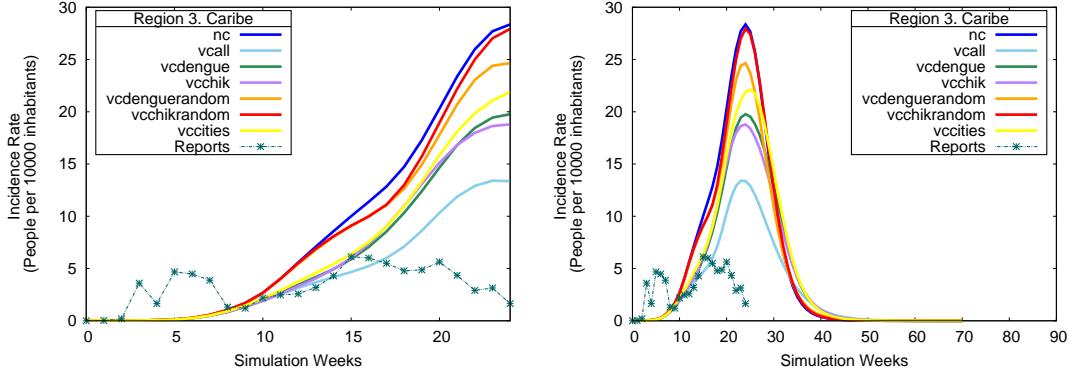


Figure 6.15: Model and reports incidence curves for the Caribbean region in Colombia. Top left figure compares the model and the data in the data period. Top right figure displays the data and the model for the whole simulation period. Bottom figure presents the model compared to the data delayed 9 weeks.

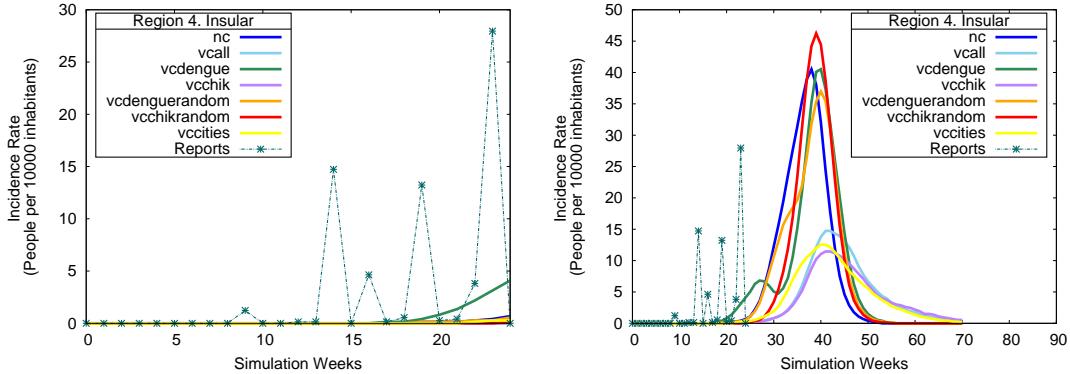


Figure 6.16: Model and reports incidence curves for the Insular region in Colombia. Top left figure compares the model and the data in the data period. Top right figure displays the data and the model for the whole simulation period.

Orinoquía region

Time series of incidence rates in Orinoquía are presented in figure 6.17. The model appropriately describes the introduction of the virus in this region when compared to the data. This synchrony suggest that the travel model captures appropriately the connections between for this region. Also, the magnitude of the cases are reconstructed by the model for various scenarios. However, it is difficult to assess the validity of the model since the data points represent early stages in the epidemic.

The data follows closer those scenarios of the model without vector control programs implemented in the region. This suggests that no vector control programs have been implemented in this region. For this scenario, the model estimates a total of 56,545 diagnosed cases in the region. This figure is reduced in the model when vector control programs are simulated. From the seven scenarios, all the municipalities participating in vector control and dengue focused programs showed the largest reduction in cases.

In fact, the scenario with dengue focused interventions reduces the impact of chikungunya in this region by approximately 50 %. In total 24,342 cases are obtained with this strategy while 20,386 cases are obtained when all the municipalities participate in vector control. Future data availability will help to assess the goodness-of-fit of these predictions, and to understand the real interventions implemented in the region.

Pacific region

Figure 6.18 illustrates the time series comparison between the surveillance reports and model's estimations for the Pacific region. The start of the epidemic is captured closely by the seven scenarios simulated, suggesting that the contacts of this region are reproduced correctly by the model. The magnitude of the epidemic is also represented by the model for the 24 weeks of reports.

Compared to the data, the model's output agrees with the data for the scenarios with vector control focused on dengue municipalities and all the municipalities of the territory. This result suggests that this region has implemented similar vector control programs than those simulated. Also, the response of the model with vector programs focused on dengue municipalities is qualitatively the same as when all the municipalities are included in the interventions.

The model estimates that focusing the interventions in dengue regions, reduce the impact of the chikungunya in more than 55 %. A total of 214 thousand cases are estimated without vector control actions. While 96 thousand cases are predicted when interventions based on dengue municipalities are implemented in the region, compared to 76 thousand cases produced when all the municipalities participate in vector control.

In general, the model estimates the initial growth of the incidence rates reported in the six regions of the country. Particularly for the Orinoquía and the Pacific regions. While the Andean region is represented with a delay by the model of vector control in all the municipalities. The Caribbean region presents complex heterogeneous intervention programs that are not captured by the model, hence the predictions of the model after the initial growth rate differ from the observations.

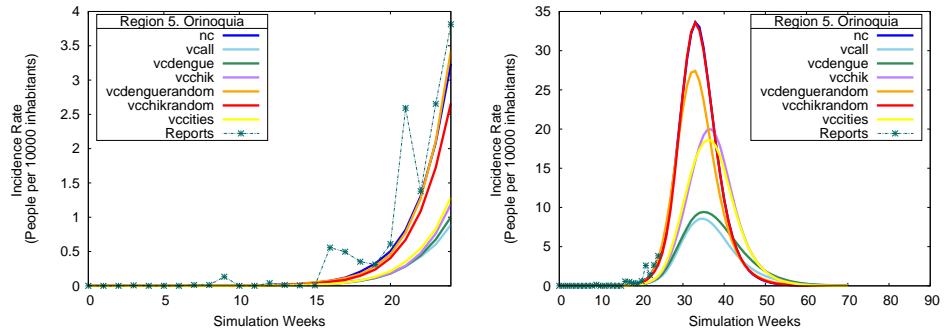


Figure 6.17: Model and reports incidence curves for the Orinoquía region in Colombia. Top left figure compares the model and the data in the data period. Top right figure displays the data and the model for the whole simulation period.

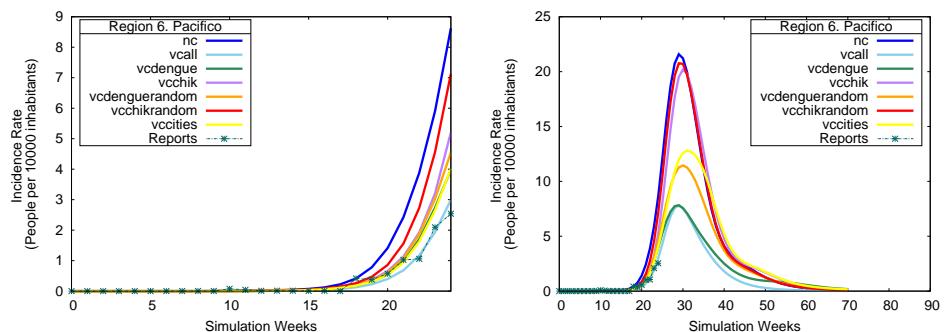


Figure 6.18: Model and reports incidence curves for the Pacific region in Colombia. Top left figure compares the model and the data in the data period. Top right figure displays the data and the model for the whole simulation period.

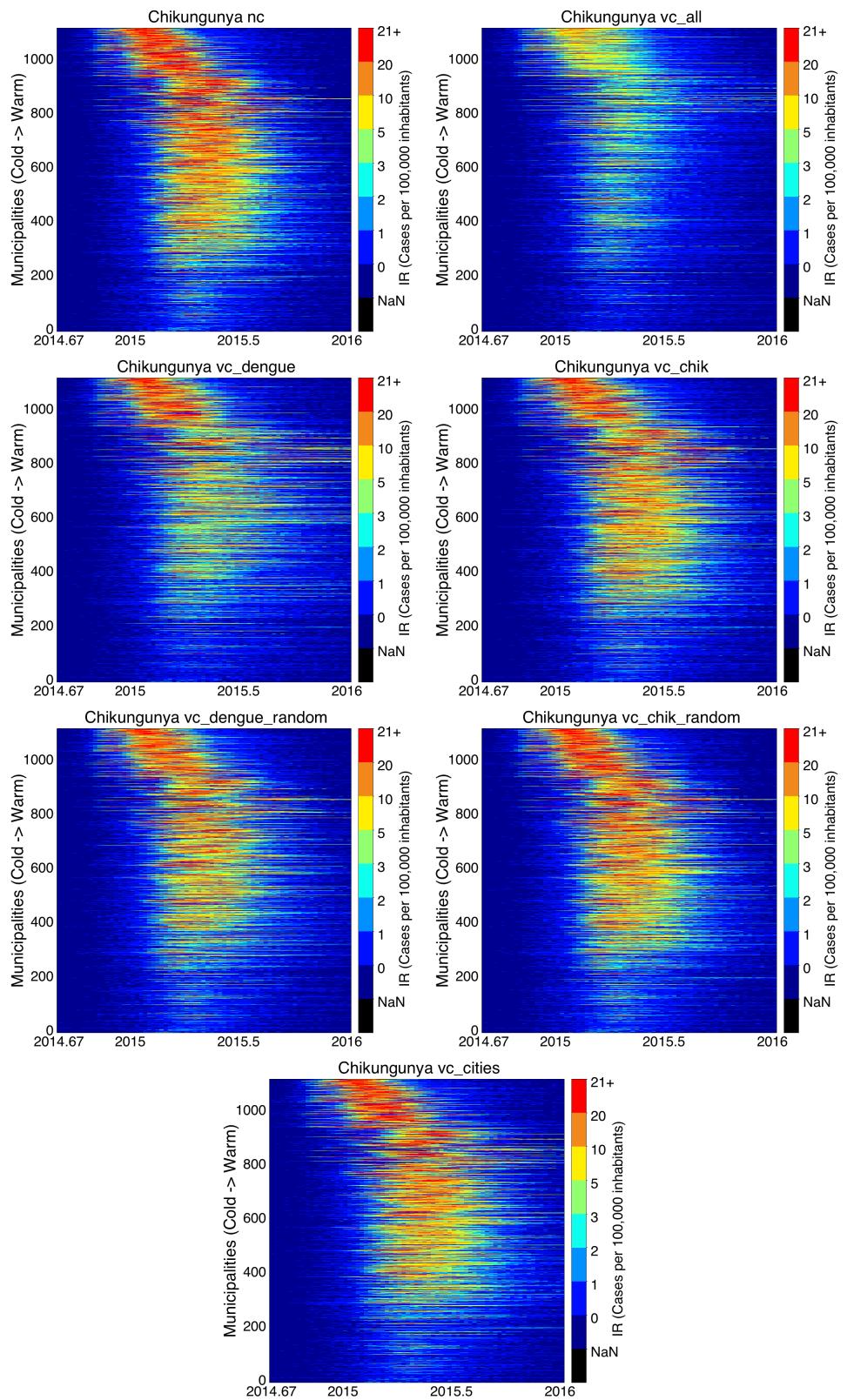


Figure 6.19: Heatmaps of incidence rates by municipalities for seven different scenarios.

Incidence by Municipalities

In figure 6.19, seven heat maps describe the distribution of the virus across the country. For each column, the color of each pixel represents the incidence rate for a municipality (rows) reported in a specific week. The municipalities are organized from cold to warm areas using the average annual temperature.

In all the scenarios simulated the virus spreads through the warm regions before spreading to the cold regions. Also, it can be seen in all the simulations, that the temperature limits the transmission of the chikungunya virus, since cold areas in the bottom of the figure are considerably less affected than warm areas.

Vector control strategies applied in all the municipalities at risk are the most effective in the temperate climate municipalities. Also, the transmission in the warmer municipalities is partially reduced. In comparison, the interventions focused on dengue municipalities reduce transmission in the temperate municipalities, but the warmer municipalities are insignificantly affected.

A characteristic noticed in the dengue strategy is that the length of the epidemics in the temperate regions are the same than without vector control. In contrast, the length of the epidemics when applying vector control in all the municipalities, are characterized by a shorter duration.

It is also notable that the rest of the strategies do not reduce significantly the impact of the epidemic on any region. Supporting the point that limited resources should be distributed to the 342 identified municipalities with previous dengue epidemics.

6.6 Summary

In this chapter, the vector control strategies focused on pupae reduction were implemented in the model and calibrated to the specific case of Santa Marta.

Five parameters affect the dynamics of these vector interventions. The importance of these parameters was assessed performing a parameter sweep from the baseline scenario calibrated to Santa Marta. The number of cases reported per week to initiate the vector control strategies, named as 'threshold', affects the model significantly. This effect suggests that starting the vector control programs early, or before the epidemic manifests itself, provides a significant reduction on the total cases reported in the health care system.

More information about the specific efficacy and participation rate of the vector control programs can provide the model with better estimates of the probable reduction of cases when applying vector control. Also, the model's results show that these interventions should be implemented for as long as possible in order to prevent re-invasion of the virus in the population.

National epidemics were simulated applying seven different scenarios of vector control. The results for focused interventions on risk areas are comparable in magnitude to the number of cases reported in the surveillance system. This could be an indication that there are vector control strategies implemented in the country. However, in order to assess the impact of the vector control strategies implemented in the country, more data is needed with the specifics of the different parameters affecting the strategies.

In general, the model captures properly the synchrony and magnitude of two out of the six regions of the country (Orinoquía and Pacífico). The Amazon region did not report enough cases in the data period to compare with the model. The reports for the Insular region are noisy and difficult to compare. Also, the travel model implemented for the country was insufficient to represent the importation of cases in the Insular region. The model represents the magnitude with reasonable error for the Andean region, however the model showed to be out of synchrony. This error can be attributed to the transportation model, or simultaneous importation of cases into this region from international territories.

The Caribbean region could not be represented by the model by any of the interventions simulated. This suggests a complex distribution of the vector control programs that are not captured for the model. For instance, high income municipalities have the ability to implement more effective vector control than low income municipalities. More data would be needed to capture this phenomenon in the model.

The simulations' results elucidate that dengue municipalities are a reasonable predictor of chikungunya epidemics. This agrees with the Colombian recommendations for vector control. Hence, these regions should be prioritized when distributing resources for vector control. Nonetheless, the vector control strategies simulated in the country were implemented after the epidemic was sparked. Prevention of these epidemics could be achieved if vector control strategies were implemented with consistent effort before the virus is introduced in a region.

Chapter 7

Conclusions and Future Work

This thesis presents a model designed to reproduce vector-borne disease dynamics, and to evaluate and optimize vector control interventions, particularly for dengue and chikungunya. A large-scale agent-based model was designed using the FRED platform developed by the University of Pittsburgh. The model represents transmission between mosquitoes and humans through realistic populations. Specifically, the model was implemented to represent chikungunya transmission in Colombia. Since chikungunya is a new virus in Colombia, a comparison was performed to analyze dengue epidemics using historical reports of both diseases. Municipalities and departments were classified based on their exposure to these diseases. From this, a synthetic population of Colombia, based on census data, was developed. Both demographic and geographic characteristics of the country are represented by the synthetic population. The synthetic population also includes basic activities of the population, such as going to work and school. Using the data, the model was calibrated by applying parameterization algorithms. The model was tested and accurately reproduces the chikungunya dynamics in multiple municipalities. Also, vector control strategies were implemented in the model using five extra parameters. These strategies were calibrated according to the historical data of chikungunya, where the data suggests that vector control actions have the greatest effect. Variations in the parameters were proposed to optimize the control programs. The duration of the intervention was proven to be a crucial parameter, resulting in an increase in the program impact. Finally, vector control programs were presented that focus on regions at risk.

Historical surveillance reports for dengue and chikungunya were analyzed to understand the transmission characteristics and similarities between the two diseases. Each department was classified using the wavelet spectrum of dengue incidence for the period between 1997 and 2010. After four groups of departments were identified and analyzed, a geographic correlation between dengue and chikungunya was observed. Areas at risk of dengue also predicted chikungunya epidemics; in fact 65 % of the municipalities with chikungunya experienced dengue in the 2010 epidemic. Municipalities were also classified based on their time of exposure to dengue and chikungunya using data from the largest, most recent, epidemic. Most municipalities at-risk for chikungunya were found as clusters with a short transmission time of dengue. Also, chikungunya epidemics are expected to have a greater impact on the country than dengue, since most of the R_0 computed for the chikungunya epidemics were higher than dengue. Furthermore, a multivariable analysis of variance showed that temperature defines dengue transmission, while chikungunya has been more affected by precipitation.

A synthetic population was generated to represent the demographic and geographic characteristics of Colombia. A sample of the census data, available from the IPUMS database, was used to represent the population of each municipality in the country. The General Iterative Proportional Updating algorithm was used to match the individual and household attributes of the population, while also maintaining the population distribution estimated by the census. The age-structure of the population was assessed using the χ^2 goodness-of-fit measure. Most of the municipalities had an acceptable fit according to this criterium. The number of students and workers were also fitted to represent census estimates. Households were located using population density grids from the WorldPop project, land-use grids from the GRUMP project, and geographical point locations obtained from the OpenStreetMap

project. The correlation of the synthetic population and the WorldPop density grids was computed showing acceptable values for most of the territory. Additionally, geographic locations were assigned to schools and workplaces based on the density of students and workers, respectively. Information from OpenStreetMap was also used to locate these structures. Students were assigned to schools based on proximity and availability for a specific grade, while workers were assigned to workplaces based on commuting times that were translated to a commuting distance using an average speed of commuting. Capacities of workplaces were obtained from the census; whereas, capacities for schools were artificially created for each municipality. This synthetic population constitutes an input for the disease model, in order to reproduce epidemics as realistic populations.

In the model, the transmission of the virus is accredited to contacts between humans and mosquitoes. Humans are represented by agents with characteristics assigned from the synthetic population. Hence, agents are assigned to households, neighborhoods, workplaces or schools. Agents perform daily activities in these places where mosquitoes can be present, depending on climate conditions. The mosquito density in each location depends on the annual temperature and the number of hosts. Mosquitoes were represented by a population model with three different health status: Susceptible, Exposed or Infectious. Transmission of the virus occurs in places with infectious agents and susceptible mosquitoes, or with susceptible agents and infectious mosquitoes. Human's evolution of the disease was represented by four compartments: Susceptible, Exposed, Infectious, and Recovered. The combination of detailed activities, climate conditions, and virus transmission allows the model to reproduce epidemics with high resolution geographic and demographic characteristics. The variation of these climate conditions was explored as well as the change in the initial conditions of the model. Additionally, a gravity model was implemented to represent long-distance travel that contributes to the spread of the virus through the country.

The model's parameters were adjusted to represent the chikungunya epidemic reported by the city of Riohacha, Guajira in 2014-2015. Eleven parameters of the model were adjusted to values found in the literature. Disease specific parameters were obtained from dengue studies, due to the lack of research on chikungunya. Three more parameters specifically related with disease transmission were calibrated using a global approach of the *simplex Nelder-Mead* algorithm. The reporting rate was estimated to be 1 out of 18 infectious cases, this is a low reporting rate considering that the symptomatic rate for chikungunya is estimated to be around 75 %. Further research in this parameter should be conducted to contrast the model results. The calibrated model's results contrasted with surveillance reports of the recent chikungunya epidemic for 25 municipalities. From these, 15 municipalities showed an acceptable fit. This comparisons suggested that interventions were implemented in some of the municipalities where the model overestimated the incidence of chikungunya. Moreover, national-scale epidemics were simulated in the model attempting to reproduce the surveillance reports of chikungunya. The model adjusted the data for the first stages of the epidemic, but the data showed a drastic decline that the model was unable to reproduce. A conclusion is that the discrepancies between the model and the data are mostly affected by socioeconomic factors, specially in the Caribbean region. These socioeconomic factors could imply constant vector control in wealthy areas that could yield to heterogeneous density of the mosquitoes in different municipalities.

The final chapter of this document presented a methodology to identify and optimize the impact of parameters involved in vector control strategies. A vector control strategy was implemented in the model based on pupae reduction by location. Five parameters were defined to describe this strategy: threshold of introduction, participation rate, recruitment rate, efficacy, and stop time. These parameters were adjusted to represent the epidemic reported in the city of Santa Marta, Magdalena, using values found in the literature and the *simplex Nelder-Mead* algorithm. The number of cases prevented by vector control were calculated for Santa Marta. Also, the parameters of the vector control strategies were modified, and the benefit of improvement in each of these was calculated in terms of cases prevented. The threshold to start the interventions showed to affect produce the largest number of cases prevented when set to low values. Lastly, vector control strategies were implemented in multiple scenarios for national-level epidemics. The prevented cases were quantified for each of these scenarios. Strategies that prioritized dengue municipalities showed a similar reduction than when the strategies were applied in the entire country. The time series of the data adjusted better for the scenarios with vector control, suggesting that control actions have been implemented in the country. More knowledge

about these strategies can improve the model's estimates.

The main conclusions of this dissertation are drawn as follows. In the second chapter, the dengue and chikungunya epidemics were compared. Important results are:

- Geographical patterns were found for dengue transmission using the Maximum Covariance Analysis to classify signals based on their frequency spectrum generated by wavelet analysis.
- The municipalities that have experienced dengue in the past epidemics can be considered at risk of chikungunya. 65 % of the municipalities with chikungunya epidemics experienced dengue in the 2010 epidemic.
- Dengue and Chikungunya municipalities were classified based on their epidemic peak time. A multivariable analysis of variance showed that temperature influences dengue transmission. For the analysis period, chikungunya clusters showed that chikungunya first epidemics have been limited to regions with homogeneous temperature values.
- The reproductive numbers of the chikungunya epidemics were calculated using the initial growth method. The chikungunya epidemics are characterized by higher reproductive numbers than those reported in the 2010 dengue epidemic. Consequently, the impact of the chikungunya is expected to be higher than the largest dengue epidemic reported in last two decades.

In chapter three, a synthetic population to represent the Colombian population was proposed.

- A methodology to represent synthetic populations using open-access databases was presented. The synthetic population matches individual, household and population characteristics based on micro-data samples and census estimates.
- The synthetic population proposed matches demographic and geographic characteristics of the population. The χ^2 test was used to assess the age structure of the population by municipality. Spatial correlation metrics were used to verify the correlation between the synthetic and the WorldPop population density.
- Synthetic workers were assigned to locations based on commuting times instead of proximity. This reproduce more realistically the population mobility.
- A conclusion from this work is that better quality data are needed to generate more reliable synthetic populations. Various inconsistencies were found in the census databases. Important information such as schools capacities are omitted in the open-access databases.

The synthetic population was used as an input for the agent-based model. Main conclusions from this step are:

- A large-scale agent-based model was proposed using compartmental representation of mosquitoes and individual humans. The model represents transmission between mosquitoes and humans including temperature effects.
- Intercity spread of the virus was represented using a gravity model calibrated to flight data. The model was used to represent missing connections from the air travel database.
- The model represents the effect of the initial immunity status in the population. These immunity levels determine the impact of the epidemic in a population and the incidence's age-structure.
- Some of the parameters in the model have been measured for dengue disease. However, to the authors' knowledge, these parameters are not available for chikungunya in the same extent.

The model was calibrated to represent real epidemics' dynamics of chikungunya in Riohacha, Guajira. Some conclusions drawn from chapter 5 are:

- The parameters of the model were adjusted to represent the dynamics of the disease in one municipality. None of the calibration attempts produced reasonable adjustments with high reporting rates. Hence, the model can be misrepresenting some characteristics of transmission, or the quality of the data underrepresented the impact of chikungunya in the population. Serological studies can be implemented to conclude a value about this parameter.
- The model can reproduce epidemics in various municipalities. And it can be adjusted to represent specific cases of interest.
- However, some municipalities' dynamics were reproduced poorly by the model. This suggests that temperature itself cannot reproduce the heterogeneity of transmission in the entire country.
- Errors of prediction were analyzed by additional covariates such as temperature, socioeconomic status, precipitation, population, etc. Using a regression model in the error, the most significant sources of error seem to be the precipitation and the socioeconomic status. Most of this error was present in the Caribbean region of the country.
- The analysis of the cases by municipality suggest that the model overestimates the incidence of chikungunya in warm municipalities while it approximates to the incidence in temperate and cold municipalities.
- These differences suggest that the model should include socioeconomic structure in the population to represent the difference in the transmission over low income and high income areas. This could be related with the pupa density since low-income areas would allow different artificial containers of water that can favor the reproduction of the mosquitoes.

The last chapter of this document presented the vector controls strategies included the model. Some conclusions from this chapter are:

- The sensitivity analysis of the intervention's parameters showed the importance of the time of the introduction of the interventions. Interventions applied late in the epidemic generate null benefit in the population. This supports the idea that structured preventive programs are important to reduce transmission, whereas the impact of responsive strategies to sudden epidemics are limited.
- The efficacy of the interventions presented a minimum threshold to prevent significant amount of cases. Measures of the present programs will improve model's predictions.
- Focused control programs in municipalities at risk of dengue showed a larger reduction of cases compared to randomly distributed strategies.
- The model quantified the number of possible cases reported without vector control programs. The comparisons with the data suggest that vector control have heterogeneously been implemented across the country.

7.1 Original Contributions

To the authors' knowledge some of the results obtained from this thesis are new contributions to the field:

- A synthetic population was designed and validated to properly represents the demography, geography, and individual activities of the population in Colombia.
- A Large-scale agent-based model was proposed and calibrated to reproduce chikungunya dynamics in Colombia. This is the first study that represents vector-borne diseases in large populations and that attempts to calibrate it to represent national-scale epidemics.

- The burden of chikungunya disease was estimated in this thesis. This is the first study that evaluate the number of cases prevented in municipalities that implemented vector control.
- An evaluation of the parameters that describes the vector control strategy implemented in the model was performed. This is the first study that quantify the effect of different types of a vector control strategy on the chikungunya spread.

7.2 Future Work

Dengue and Chikungunya epidemics were compared using surveillance reports of the 2010 dengue epidemic and the current chikungunya epidemic. This analysis was performed for the initial stages of the chikungunya epidemic. This cluster analysis can be applied to further available data for both dengue and chikungunya diseases. Also, alternative approaches may enhance the knowledge of the transmission dynamics of these diseases. Furthermore, dynamic analysis of this data can provide useful information about the transmission specifics of these diseases in Colombia. For instance, time series of dengue and chikungunya reported cases can be compared to climate conditions variable in time.

A synthetic population that matches demographic and geographic characteristics was proposed and validated. Socioeconomic variables were not included due to the lack of data. The inclusion of these variables would allow more sophisticated analysis in the study of epidemics, and other possible implementations of the synthetic population. Also, the composition of places was restricted to households, neighborhoods, schools, and workplaces. Future versions of the synthetic population might include Hospitals, Military bases, Parks, Churches, Malls, among other locations that congregate large groups of people.

In addition, individuals were enrolled to schools and workplaces based on distance measures. However, this algorithm ignores other types of geographic and socioeconomic variables. For instance, school enrollment is usually influenced by income-level. Other future geographic variables may include presence of waterbodies or mountains that discourage people to attend to places apparently close.

The model represents specific characteristics about the Colombian population using age-structure population, temperature, and human mobility. However, comparisons with the surveillance reports showed that additional variables as the socioeconomic factors should be included to represent the heterogeneity in chikungunya transmission across the country. Additionally, the inclusion of geographic factors to improve the travel model would improve the model's representation of the disease transmission dynamics.

In the model, the mosquito model is presented using a compartmental model for each location that includes pupae and, susceptible, exposed, and infectious mosquitoes. The pupae development to adult mosquito is simplified using development rates values for different temperatures. Development of more complex models to represent the mosquito development stages would allow to understand better dengue and chikungunya transmission and, to investigate various control strategies.

Additional information about chikungunya parameters would benefit the model's calibration. Further research may include studies to quantify the underreporting ratio in specific municipalities. Also, alternative approaches may be attempted to calibrate the transmission in the entire country rather than one municipality at the time.

Lastly, in the model, vector control strategies are represented by a constant rate of pupae reduction. More sophisticated models can be implemented to compare different control strategies based on reduction of pupae or adult mosquitoes. Also, the model's performance and value for public health stakeholders may be improved by collecting data and recalibrating the model to specifics vector control strategies implemented to prevent the spread of chikungunya in Colombia.

References

- [1] Jacco Wallinga and Marc Lipsitch. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1609):599–604, 2007.
- [2] Jennifer L. Kyle and Eva Harris. Global spread and persistence of dengue. *Annual Review of Microbiology*, 62:71–92, 2008.
- [3] Samir Bhatt, Peter W Gething, Oliver J Brady, Jane P Messina, Andrew W Farlow, Catherine L Moyes, John M Drake, John S Brownstein, Anne G Hoen, Osman Sankoh, et al. The global distribution and burden of dengue. *Nature*, 2013.
- [4] World Health Organization et al. *Global strategy for dengue prevention and control 2012-2020*. World Health Organization, 2012.
- [5] World Health Organization. *Dengue guidelines for diagnosis treatment, prevention and control*, 2009.
- [6] Maria G Guzmán and Gustavo Kouri. Dengue: an update. *The Lancet infectious diseases*, 2(1):33–42, 2002.
- [7] Pan American Health Organization. Number of reported cases of chikungunya in the americas, by country or territory. <http://www.paho.org>, October 2014.
- [8] Simon-Djamel Thiberville, Nanikaly Moyen, Laurence Dupuis-Maguiraga, Antoine Nougairede, Ernest A Gould, Pierre Roques, and Xavier de Lamballerie. Chikungunya fever: epidemiology, clinical syndrome, pathogenesis and therapy. *Antiviral research*, 99(3):345–370, 2013.
- [9] Olivia Brathwaite Dick, José L San Martín, Romeo H Montoya, Jorge del Diego, Betzana Zambrano, and Gustavo H Dayan. The history of dengue outbreaks in the americas. *The American journal of tropical medicine and hygiene*, 87(4):584, 2012.
- [10] Bruno Guy, Beatrice Barrere, Claire Malinowski, Melanie Saville, Remy Teyssou, and Jean Lang. From research to phase iii: preclinical, industrial and clinical development of the sanofi pasteur tetravalent dengue vaccine. *Vaccine*, 29(42):7229–7241, 2011.
- [11] M Derouich, A Boutayeb, and EH Twizell. A model of dengue fever. *BioMedical Engineering OnLine*, 2(1):4, 2003.
- [12] Lourdes Esteva and Cristobal Vargas. A model for dengue disease with variable human population. *journal of mathematical biology*, 38:220–240, 1999.
- [13] Lourdes Esteva and Cristobal Vargas. Analysis of a dengue disease transmission model. *Mathematical Biosciences*, 150:131–151, 1998.
- [14] Helen J. Wearing and Pejman Rohani. Ecological and immunological determinants of dengue epidemics. *PNAS*, 103(31):11802–11807, 2006.

- [15] Isao Kawaguchi, Akira Sasaki, and Michael Boots. Why are dengue virus serotypes so distantly related? enhancement and limiting serotype similarity between dengue virus strains. *Proceedings of the Royal Society*, 270:2241 – 2247, 2003.
- [16] Lourdes Esteva and Cristobal Vargas. Coexistence of different serotypes of dengue virus. *Mathematical Biology*, 46:31–47, 2003.
- [17] Mario Recker, Konstantin B. Blyuss, Cameron P. Simmons, Tran Tinh Hien, Bridget Wills, Jeremy Farrar, and Sunetra Gupta. Immunological serotype interactions and their effect on the epidemiological pattern of dengue. *Proceedings of the Royal Society*, 276:2541–2548, 2009.
- [18] Lora Billings, Ira B. Schwartz, Leah B. Shaw, Marie McCrary, Donald S. Burke, and Derek A.T. Cummings. Instabilities in multiseroype disease models with antibody-dependent enhancement. *Journal of Theoretical Biology*, 246:18–27, 2007.
- [19] Paul S. Wikramaratna, Cameron P. Simmons, Sunetra Gupta, and Mario Recker. The effects of tertiary and quaternary infections on the epidemiology of dengue. *PLOS ONE*, 5(8):12347–12354, 2010.
- [20] Annelise Tran and Marcel Raffy. On the dynamics of dengue epidemics from large-scale information. *Theoretical Population Biology*, 69:3–12, 2006.
- [21] Yoshiro Nagao and Katia Koelle. Decreases in dengue transmission may act to increase the incidence of dengue hemorrhagic fever. *PNAS*, 105(6):2238–2243, 2008.
- [22] Laith Yakob and Archie CA Clements. A mathematical model of chikungunya dynamics and control: the major epidemic on reunion island. *PloS one*, 8(3):e57448, 2013.
- [23] Yves Dumont, Frédéric Chiroleu, and Caroline Domerg. On a temporal model for the chikungunya disease: modeling, theory and numerics. *Mathematical biosciences*, 213(1):80–91, 2008.
- [24] Djamila Moulay, MA Aziz-Alaoui, and Hee-Dae Kwon. Optimal control of chikungunya disease: larvae reduction, treatment and prevention. *Mathematical Biosciences and Engineering*, 9(2):369–392, 2012.
- [25] Djamila Moulay and Yoann Pigné. A metapopulation model for chikungunya including populations mobility on a large-scale network. *Journal of theoretical biology*, 318:129–139, 2013.
- [26] Diego Ruiz-Moreno, Irma Sanchez Vargas, Ken E Olson, and Laura C Harrington. Modeling dynamic introduction of chikungunya virus in the united states. *PLoS neglected tropical diseases*, 6(11):e1918, 2012.
- [27] Thomas Seyler, Francesco Grandesso, Yann Le Strat, Arnaud Tarantola, and Evelyn Depoortere. Assessing the risk of importing dengue and chikungunya viruses to the european union. *Epidemics*, 1(3):175–184, 2009.
- [28] Dennis L Chao, Scott B Halstead, M Elizabeth Halloran, and Ira M Longini Jr. Controlling dengue with vaccines in thailand. *PLoS neglected tropical diseases*, 6(10):e1876, 2012.
- [29] Carlos J Dommar, Rachel Lowe, Marguerite Robinson, and Xavier Rodó. An agent-based model driven by tropical rainfall to understand the spatio-temporal heterogeneity of a chikungunya outbreak. *Acta tropica*, 129:61–73, 2014.
- [30] Carlos Castañeda-Orjuela, Hernando Díaz, Nelson Alvis-Guzman, Andres Olarte, Heidy Rodriguez, Guido Camargo, and Fernando De la Hoz-Restrepo. Burden of disease and economic impact of dengue and severe dengue in colombia, 2011. *Value in Health Regional Issues*, 1(2):123–128, 2012.

- [31] Claudia Torres, Samier Barguil, Miguel Melgarejo, and Andrés Olarte. Fuzzy model identification of dengue epidemic in colombia based on multiresolution analysis. *Artificial intelligence in medicine*, 2013.
- [32] Santiago Velez, Claudia Nuñez, and Daniel Ruiz. Hacia la construcción de un modelo de simulación de la transmisión del dengue en colombia. *Revista EIA*, (5):23–43, 2006.
- [33] Dana A Focks, Eric Daniels, Dan G Haile, James E Keesling, et al. A simulation model of the epidemiology of urban dengue fever: literature analysis, model development, preliminary validation, and samples of simulation results. *American Journal of Tropical Medicine and Hygiene*, 53(5):489–506, 1995.
- [34] Harish Padmanabha, David Durham, Fabio Correa, Maria Diuk-Wasser, and Alison Galvani. The interactive roles of aedes aegypti super-production and human density in dengue transmission. *PLoS neglected tropical diseases*, 6(8):e1799, 2012.
- [35] Annelies Wilder-Smith and Duane J Gubler. Geographic expansion of dengue: the impact of international travel. *Medical Clinics of North America*, 92(6):1377–1390, 2008.
- [36] Organización Panamericana de la Salud. *Dengue y Dengue Hemorrágico en las Américas: guías para su prevención y control*, 1997.
- [37] Pan American Health Organization (PAHO). *A timeline for Dengue in the americas to december 31, 2000 and noted first occurrences*, 2001.
- [38] Jorge Boshell, Hernando Groot, Manuel Gacharna, Gladys Márquez, Martha González, Martha O Gaitán, Christine Berlie, and Mancel Martínez. Dengue en colombia. *Biomédica*, 6(3-4):101–106, 1986.
- [39] Cassie C. Jansen and Nigel W. Beebe. The dengue vector aedes aegypti: what comes next. *Microbes and infection*, 12:272–279, 2010.
- [40] SIVIGILA. Vigilancia rutinaria. Instituto Nacional de Salud. <http://www.ins.gov.co/lineas-de-accion/Subdireccion-Vigilancia/sivigila/Paginas/vigilancia-rutinaria.aspx>, September. Accessed: 2014-09-25.
- [41] Babatunde Olowokure, Lorraine Francis, Karen Polson-Edwards, Roger Nasci, Philippe Quénel, Sylvain Aldighieri, Dominique Rousset, Cristina Gutierrez, Pilar Ramon-Pardo, Thais dos Santos, et al. The caribbean response to chikungunya. *The Lancet Infectious Diseases*, 14(11):1039–1040, 2014.
- [42] Pan American Health Organization. Number of reported cases of chikungunya in the americas, by country or territory. <http://www.paho.org>, October 2014.
- [43] Olaf Horstick, Silvia Runge-Ranzinger, Michael B. Nathan, and Axel Kroeger. Dengue vector-control services: hoy do they work? a systematic literature review and country case studies. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 104:379–386, 2010.
- [44] World Health Organization et al. *Guidelines for prevention and control of chikungunya fever*. World Health Organization, Regional Office for South-East Asia, 2009.
- [45] CDC OPS and OMS Preparación. respuesta ante la eventual introducción del virus chikungunya en las américa. *Washington, DC: OPS*, page 159, 2011.
- [46] Willem G Van Panhuis, John Grefenstette, Su Yon Jung, Nian Shong Chok, Anne Cross, Heather Eng, Bruce Y Lee, Vladimir Zadorozhny, Shawn Brown, Derek Cummings, et al. Contagious diseases in the united states from 1888 to the present. *N Engl J Med*, 369(22):2152–8, 2013.
- [47] Beth-Ann G Coller and David E Clements. Dengue vaccines: progress and challenges. *Current opinion in immunology*, 23(3):391–398, 2011.

- [48] Julia Schmitz, John Roehrig, Alan Barrett, and Joachim Hombach. Next generation dengue vaccines: a review of candidates in preclinical development. *Vaccine*, 29(42):7276–7284, 2011.
- [49] Annelies Wilder-Smith and Paul Macary. Dengue: challenges for policy makers and vaccine developers. *Current infectious disease reports*, 16(5):1–8, 2014.
- [50] Maria Rosario Capeding, Ngoc Huu Tran, Sri Rezeki S Hadinegoro, Hussain Imam HJ Ismail, Tawee Chotpitayasunondh, Mary Noreen Chua, Chan Quang Luong, Kusnandi Rusmil, De-wa Nyoman Wirawan, Revathy Nallusamy, et al. Clinical efficacy and safety of a novel tetravalent dengue vaccine in healthy children in asia: a phase 3, randomised, observer-masked, placebo-controlled trial. *The Lancet*, 2014.
- [51] A.K. Supriatna, E. Soewono, and S.A. van Gils. A two-age-classes dengue transmission model. *Mathematical Biosciences*, 216:114–121, 2008.
- [52] Michael A. Johansson, Joachim Hombach, and Derek A.T. Cummings. Models of the impact of dengue vaccines: A review of current research and potential approaches. *Vaccine*, 29:5860–5868, 2011.
- [53] Mathieu Andraud, Niel Hens, Christiaan Marais, and Philippe Beutels. Dynamic epidemiological models for dengue transmission: a systematic review of structural approaches. *PloS one*, 7(11):e49085, 2012.
- [54] Nicholas G Reich, Sourya Shrestha, Aaron A King, Pejman Rohani, Justin Lessler, Siripen Kalayanarooj, In-Kyu Yoon, Robert V Gibbons, Donald S Burke, and Derek AT Cummings. Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity. *Journal of The Royal Society Interface*, 10(86):20130414, 2013.
- [55] Eriko Chikaki and HIrofumi Ishikawa. A dengue transmission model in thailand considering secuential infections with all four serotypes. *J Infect Dv Ctries*, 3:711–722, August 2009.
- [56] Sz-Chieh Chen and Meng-Huan Hsieh. Modeling the transmission dynamics of dengue fever: implications of temperature effects. *Science of the Total Environment*, 431:385–391, 2012.
- [57] DO Fuller, A Troyo, and JC Beier. El nino southern oscillation and vegetation dynamics as predictors of dengue fever cases in costa rica. *Environmental Research Letters*, 4(1):014011, 2009.
- [58] Pei-Chih Wu, How-Ran Guo, Shih-Chun Lung, Chuan-Yao Lin, and Huey-Jen Su. Weather as an effective predictor for occurrence of dengue fever in taiwan. *Acta tropica*, 103(1):50–57, 2007.
- [59] Robert C Reiner, T Alex Perkins, Christopher M Barker, Tianchan Niu, Luis Fernando Chaves, Alicia M Ellis, Dylan B George, Arnaud Le Menach, Juliet RC Pulliam, Donal Bisanzio, et al. A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010. *Journal of The Royal Society Interface*, 10(81):20120921, 2013.
- [60] Nigel Gilbert. *Agent-based models*. Sage, 2008.
- [61] Volker Grimm, Uta Berger, Finn Bastiansen, Sigrunn Eliassen, Vincent Ginot, Jarl Giske, John Goss-Custard, Tamara Grand, Simone Heinz, Geir Huse, Andreas Huth, Jane Jepsen, Christian Jorgensen, Wolf Mooij, Birgit Muller, Guy Pe'er, Cyril Piou, Steven Railsback, Andrew Robbins, Martha Robbins, Eva Rossmanith, Nadja Ruger, Espen Strand, Sami Souissi, Richard Stillman, Rune Vabo, Ute Visser, and Donald Deangelis. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198(1-2):115–126, 2006.
- [62] Catherine Linard, Nicolas Ponçon, Didier Fontenille, and Eric F Lambin. A multi-agent simulation to assess the risk of malaria re-emergence in southern france. *Ecological Modelling*, 220(2):160–174, 2009.

- [63] J.M. Lloyd. An agent-based approach to epidemiological modeling of coupled systems. Master's thesis, University of Waterloo, 2009.
- [64] Weidong Gu and Robert J Novak. Agent-based modelling of mosquito foraging behaviour for malaria control. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 103(11):1105–1112, 2009.
- [65] Guillaume Muller, Pascal Grébaut, and Jean-Paul Gouteux. An agent-based model of sleeping sickness: simulation trials of a forest focus in southern cameroon. *Comptes rendus biologies*, 327(1):1–11, 2004.
- [66] Dana A. Focks, Eric Daniels, Dan G. Haile, and James E. Keesling. A simulation model of the epidemiology of urban dengue fever: a literature analysis, model development, preliminary validation and samples of simulation results. *The American society of tropical medicine and hygiene*, pages 489 – 506, 1995.
- [67] Alicia M. Ellis, Andres J. Garcia, Dana A. Focks, Amy C. Morrison, and Thomas W. Scott. Parameterization and sensitivity analysis of a complex simulation model for mosquito population dynamics, dengue transmission, and their control. *American Journal of Tropical Medicine and Hygiene*, 85:257 – 264, 2011.
- [68] Marcelo Otero, Nicolás Schweigmann, and Hernán G. Solari. A stochastic spatial dynamical model for aedes aegypti. *Bulletin of mathematical biology*, 2008.
- [69] Marcelo Otero, Daniel H. Barmak, Claudio O. Dorso, Hernán G. Solari, and Mario A. Natiello. Modeling dengue outbreaks. *Mathematical Biosciences*, 232:87–95, 2011.
- [70] Carlos Isidoro, Nuno Fachada, Fábio Barata, and Agostinho Rosa. Agent-based model of dengue disease transmission by aedes aegypti populations. In *Advances in Artificial Life. Darwin Meets von Neumann*, pages 345–352. Springer, 2011.
- [71] Paula Mendes Luz, Cláudia Torres Codeço, Eduardo Massad, and Claudio José Struchiner. Uncertainties regarding dengue modeling in rio de janeiro, brazil. *Memórias do Instituto Oswaldo Cruz*, 98(7):871–878, 2003.
- [72] Eduardo Massad, Marcelo Burattini, Francisco Bezerra, and Luiz Lopez. Dengue and the risk of urban yellow fever reintroduction in são paulo state, brazil. *Rev saúde Pública*, 37(4):477–484, 2003.
- [73] Norberto Aníbal Maidana and Hyun Mo Yang. Describing the geographic spread of dengue disease by traveling waves. *Mathematical biosciences*, 215(1):64–77, 2008.
- [74] M. Otero and H.G. Solari. Stochastic eco-epidemiological model of dengue disease transmission by aedes aegypti mosquito. *Mathematical Biosciences*, 223:32–46, 2010.
- [75] Amy C Morrison, Sharon L Minnick, Claudio Rocha, Brett M Forshey, Steven T Stoddard, Arthur Getis, Dana A Focks, Kevin L Russell, James G Olson, Patrick J Blair, et al. Epidemiology of dengue virus in iquitos, peru 1999 to 2005: interepidemic and epidemic patterns of transmission. *PLoS neglected tropical diseases*, 4(5):e670, 2010.
- [76] Kelly A Liebman, Steven T Stoddard, Amy C Morrison, Claudio Rocha, Sharon Minnick, Moises Sihuinchá, Kevin L Russell, James G Olson, Patrick J Blair, Douglas M Watts, et al. Spatial dimensions of dengue virus transmission across interepidemic and epidemic periods in iquitos, peru (1999–2003). *PLoS neglected tropical diseases*, 6(2):e1472, 2012.
- [77] Steven F Railsback, Steven L Lytinen, and Stephen K Jackson. Agent-based simulation platforms: Review and development recommendations. *Simulation*, 82(9):609–623, 2006.

- [78] S Arunachalam, R Zalila-Wenkstern, and R Steiner. Environment mediated multi agent simulation tools—a comparison. In *Proc. of IEEE Workshop on Environment-Mediated Coordination in Self-Organizing and Self-Adaptive Systems, Venice, Italy*, 2008.
- [79] John J Grefenstette, Shawn T Brown, Roni Rosenfeld, Jay DePasse, Nathan TB Stone, Phillip C Cooley, William D Wheaton, Alona Fyshe, David D Galloway, Anuroop Sriram, et al. Fred (a framework for reconstructing epidemic dynamics): an open-source software system for modeling infectious diseases and control strategies using census-based populations. *BMC public health*, 13(1):940, 2013.
- [80] Tazio Vanni, Jonathan Karnon, Jason Madan, Richard G White, W John Edmunds, Anna M Foss, and Rosa Legood. Calibrating models in economic evaluation. *Pharmacoconomics*, 29(1):35–49, 2011.
- [81] Ben S Cooper. Confronting models with data. *Journal of Hospital Infection*, 65:88–92, 2007.
- [82] John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.
- [83] Jeffrey C Lagarias, James A Reeds, Margaret H Wright, and Paul E Wright. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on optimization*, 9(1):112–147, 1998.
- [84] Fuchang Gao and Lixing Han. Implementing the nelder-mead simplex algorithm with adaptive parameters. *Computational Optimization and Applications*, 51(1):259–277, 2012.
- [85] Ronald Aylmer Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, pages 309–368, 1922.
- [86] Julio Padilla, Diana Rojas, and Roberto Saenz-Gomez. *Dengue en Colombia de la reemergencia a la hiperendemia*. Los autores, 1 edition, 2012.
- [87] Instituto Nacional de Salud de Colombia. www.ins.gov.co.
- [88] Aslak Grinsted, John C Moore, and Svetlana Jevrejeva. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear processes in geophysics*, 11(5/6):561–566, 2004.
- [89] Luís Aguiar Conraria and Maria J Soares. The continuous wavelet transform: a primer. *NIPE Working Paper*, 16:1–43, 2011.
- [90] Christopher Torrence and Gilbert P Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*, 79(1):61–78, 1998.
- [91] Bernard Cazelles, Mario Chavez, Guillaume Constantin de Magny, Jean-Francois Guégan, and Simon Hales. Time-dependent spectral analysis of epidemiological time-series with wavelets. *Journal of the Royal Society Interface*, 4(15):625–636, 2007.
- [92] Tristan Rouyer, Jean-Marc Fromentin, Nils Christian Stenseth, and Bernard Cazelles. Analysing multiple time series and extending significance testing in wavelet analysis. 2008.
- [93] Bernard Cazelles, Mario Chavez, Dominique Berteaux, Frédéric Ménard, Jon Olav Vik, Stéphanie Jenouvrier, and Nils C Stenseth. Wavelet analysis of ecological time series. *Oecologia*, 156(2):287–304, 2008.
- [94] Tristan Rouyer, J-M Fromentin, Frédéric Ménard, B Cazelles, K Briand, R Pianet, Benjamin Planque, and N Chr Stenseth. Complex interplays among population dynamics, environmental forcing, and exploitation in fisheries. *Proceedings of the National Academy of Sciences*, 105(14):5420–5425, 2008.

- [95] Derek AT Cummings, Rafael A Irizarry, Norden E Huang, Timothy P Endy, Ananda Nisalak, Kumnuan Ungchusak, and Donald S Burke. Travelling waves in the occurrence of dengue haemorrhagic fever in thailand. *Nature*, 427(6972):344–347, 2004.
- [96] BT Grenfell, ON Bjørnstad, and J Kappey. Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414(6865):716–723, 2001.
- [97] Michael A Johansson, Derek AT Cummings, Gregory E Glass, et al. Multiyear climate variability and dengue—el nino southern oscillation, weather, and dengue incidence in puerto rico, mexico, and thailand: a longitudinal data analysis. *PLoS medicine*, 6(11):1262, 2009.
- [98] Hoang Quoc Cuong, Nguyen Tran Hien, Tran Nhu Duong, Tran Vu Phong, Nguyen Nhat Cam, Jeremy Farrar, Vu Sinh Nam, Khoa TD Thai, and Peter Horby. Quantifying the emergence of dengue in hanoi, vietnam: 1998–2009. 2011.
- [99] KT Thai, Bernard Cazelles, Nam Van Nguyen, Long Thi Vo, Maciej F Boni, Jeremy Farrar, Cameron P Simmons, H Rogier van Doorn, and Peter J de Vries. Dengue dynamics in binh thuan province, southern vietnam: periodicity, synchronicity and climate variability. *PLoS Negl Trop Dis*, 4(7):e747, 2010.
- [100] Joana Solares and L. Aguilar-Conraria. The astoolbox. wavelets, economics and politics. <https://sites.google.com/site/aguiarconraria/joanasoares-wavelets/the-astoolbox>, 2014.
- [101] Luís Aguiar-Conraria and Maria Joana Soares. Business cycle synchronization across the euro area: A wavelet analysis. *NIPE, Universidade De Minho*, 2009.
- [102] Anestis Antoniadis, Xavier Brossat, Jairo Cugliari, and Jean-Michel Poggi. Clustering functional data using wavelets. *International Journal of Wavelets, Multiresolution and Information Processing*, 11(01):1350003, 2013.
- [103] Robert J Hijmans, Susan E Cameron, Juan L Parra, Peter G Jones, Andy Jarvis, et al. Very high resolution interpolated climate surfaces for global land areas. *International journal of climatology*, 25(15):1965–1978, 2005.
- [104] Integrated Public Use Microdata Series, International: Version 6.3 [Machine-readable database]. *Minneapolis: University of Minnesota*. Minnesota Population Center, 2014.
- [105] Departamento Administrativo Nacional de Estadística (DANE), www.dane.gov.co. *Estimación de población Colombiana*.
- [106] Xin Ye, Karthik Konduri, Ram M Pendyala, Bhargava Sana, and Paul Waddell. A methodology to match distributions of both household and person attributes in the generation of synthetic populations. In *88th Annual Meeting of the Transportation Research Board, Washington, DC*, 2009.
- [107] Kirill Müller, Kay W Axhausen, Kay W Axhausen, and Kay W Axhausen. *Hierarchical IPF: Generating a synthetic population for Switzerland*. Eidgenössische Technische Hochschule Zürich, IVT, 2011.
- [108] Stephen E Fienberg. An iterative procedure for estimation in contingency tables. *The Annals of Mathematical Statistics*, pages 907–917, 1970.
- [109] W Edwards Deming and Frederick F Stephan. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, 11(4):427–444, 1940.
- [110] Forrest R Stevens, Andrea E Gaughan, Catherine Linard, and Andrew J Tatem. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PloS one*, 10(2):e0107042, 2015.

- [111] Worldpop Project, www.worldpop.org.uk. *WorldPop Americas dataset*, 2015.
- [112] Center for International Earth Science Information Network CIESIN Columbia University, International Food Policy Research Institute IFPRI, The World Bank, and Centro Internacional de Agricultura Tropical CIAT. Global rural-urban mapping project, version 1 (grumpv1): Urban extents grid, 2011.
- [113] D Balk, F Pozzi, G Yetman, U Deichmann, and A Nelson. The distribution of people and the dimension of place. *Proc. Urban Remote Sensing, Int. Soc. Photogrammetry and Remote Sensing, Tempe, Arizona, March 2005*, 2005.
- [114] DL Balk, U Deichmann, G Yetman, F Pozzi, SI Hay, and A Nelson. Determining global population distribution: methods, applications and data. *Advances in parasitology*, 62:119–156, 2006.
- [115] SI Hay, AM Noor, A Nelson, and AJ Tatem. The accuracy of human population maps for public health application. *Tropical Medicine & International Health*, 10(10):1073–1086, 2005.
- [116] Catherine Linard and Andrew J Tatem. Large-scale spatial population databases in infectious disease research. *Int J Health Geogr*, 11(7), 2012.
- [117] Instituto Geográfico Agustín Codazzi. Geographic information system for national and regional comprehensive land-use planning and management. <http://sigotn.igac.gov.co/sigotn/>, 2005.
- [118] Aarthy Sabesan, Kathleen Abercrombie, Auroop R Ganguly, Budhendra Bhaduri, Eddie A Bright, and Phillip R Coleman. Metrics for the comparative analysis of geospatial datasets with applications to high-resolution grid-based population data. *GeoJournal*, 69(1-2):81–91, 2007.
- [119] James C Cajka, Philip C Cooley, and William D Wheaton. Attribute assignment to a synthetic population in support of agent-based disease modeling. *Methods report (RTI Press)*, 19(1009):1, 2010.
- [120] Dana A Focks, Richard J Brenner, Jack Hayes, and Eric Daniels. Transmission thresholds for dengue in terms of aedes aegypti pupae per person with discussion of their utility in source reduction efforts. *The American journal of tropical medicine and hygiene*, 62(1):11–18, 2000.
- [121] George Kingsley Zipf. The p1 p2/d hypothesis: On the intercity movement of persons. *American sociological review*, pages 677–686, 1946.
- [122] Filippo Simini, Marta C González, Amos Maritan, and Albert-László Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100, 2012.
- [123] Woo-Sung Jung, Fengzhong Wang, and H Eugene Stanley. Gravity model in the korean highway. *EPL (Europhysics Letters)*, 81(4):48005, 2008.
- [124] David Karemra, Victor Iwuagwu Oguledo, and Bobby Davis. A gravity model analysis of international migration to north america. *Applied Economics*, 32(13):1745–1755, 2000.
- [125] Hiroshi Nishiura and Scott B Halstead. Natural history of dengue virus (denv)—1 and denv—4 infections: Reanalysis of classic studies. *Journal of Infectious Diseases*, 195(7):1007–1013, 2007.
- [126] M Yasuno and Robert J Tonn. A study of biting habits of aedes aegypti in bangkok, thailand. *Bulletin of the World Health Organization*, 43(2):319, 1970.
- [127] D Focks, N Alexander, and E Villegas. Multicountry study of aedes aegypti pupal productivity survey methodology: findings and recommendations. *Geneva: WHO*, 2006.
- [128] Ministerio de salud y protección social. *Plan Nacional de Respuesta Frente a la introducción del virus chikungunya en Colombia*. Ministerio de salud y protección social, 2014.

- [129] Mauricio Fuentes-Vallejo, Diana Rocío Higuera-Mendieta, Tatiana García-Betancourt, Lucas Andrés Alcalá-Espinosa, Diana García-Sánchez, David Alejandro Munévar-Cagigas, Helena Luisa Brochero, Catalina González-Uribe, and Juliana Quintero. Territorial analysis of aedes aegypti distribution in two colombian cities: a chorematic and ecosystem approach. *Cadernos de Saúde Pública*, 31(3):517–530, 2015.
- [130] Juliana Quintero, Tatiana García-Betancourt, Sebastian Cortés, Diana García, Lucas Alcalá, Catalina González-Uribe, Helena Brochero, and Gabriel Carrasquilla. Effectiveness and feasibility of long-lasting insecticide-treated curtains and water container covers for dengue vector control in colombia: a cluster randomised trial. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 109(2):116–125, 2015.
- [131] Clara B Ocampo, Camila González, Carlos A Morales, Mauricio Pérez, Dawn Wesson, and Charles S Apperson. Evaluation of community-based strategies for aedes aegypti control inside houses. *Biomedica*, 29(2):282–297, 2009.
- [132] Clara B. Ocampo, Myriam J. Salazar-Terreros, Neila J. Mina, Janet McAllister, and William Brogdon. Insecticide resistance status of aedes aegypti in 10 localities in colombia. *Acta Tropica*, 118:37 – 44, 2011.
- [133] TE Erlanger, J Keiser, and J Utzinger. Effect of dengue vector control interventions on entomological parameters in developing countries: a systematic review and meta-analysis. *Medical and veterinary entomology*, 22(3):203–221, 2008.
- [134] Tatiana García-Betancourt, Catalina González-Uribe, Juliana Quintero, and Gabriel Carrasquilla. Ecobiosocial community intervention for improved aedes aegypti control using water container covers to prevent dengue: lessons learned from girardot colombia. *EcoHealth*, 11(3):434–438, 2014.
- [135] SHPP Karunaratne, TC Weeraratne, MDB Perera, and SN Surendran. Insecticide resistance and, efficacy of space spraying and larviciding in the control of dengue vectors aedes aegypti and aedes albopictus in sri lanka. *Pesticide biochemistry and physiology*, 107(1):98–105, 2013.