Mathematical Biosciences and Engineering

https://www.aimspress.com/journal/mbe

MBE, 23(2): 266–290.
DOI: 10.3934/mbe.2026011
Received: 26 August 2025
Revised: 09 November 2025
Accepted: 25 November 2025
Published: 30 December 2025

Research article

Stochastic modeling, analysis, and simulation of Dengue in Valle del Cauca: A case study

Diego Alejandro Becerra-Becerra[1], **Jhonier Rangel**[1,2] **and Viswanathan Arunachalam**[1,*]

[1] Department of Statistics, Universidad Nacional de Colombia, Bogotá, Colombia
[2] Program of Statistics, Universidad ECCI, Bogotá, Colombian

* **Correspondence:** Email: varunachalam@unal.edu.co.

Abstract: Dengue remains a major public health challenge in Colombia, with Valle del Cauca experiencing recurrent outbreaks characterized by seasonal fluctuations and long-term variability. Understanding the transmission dynamics of Dengue across age groups is critical for targeted interventions. In this study, we developed an age-structured stochastic host–vector model, incorporating a compartmental SIR–SI framework within a stochastic differential equation (SDE) approach. The population is stratified into youths (0–17 years) and adults (18 years and older), enabling analysis of age-specific infection and recovery patterns. Simulations and forecasts were performed using the Euler–Maruyama method, informed by fixed parameters from the literature, estimated disease-specific parameters, and epidemiological data from Colombia's Public Health Surveillance System (SIVIGILA) spanning 2013–2023. Additionally, a Seasonal Autoregressive Integrated Moving Average (SARIMA) model was employed as a complementary approach to capture and forecast monthly Dengue incidence. Our results highlighted distinct epidemic patterns across age groups, the higher infection burden among adults, and the complementary roles of mechanistic SDE modeling and SARIMA forecasting for surveillance and control planning.

Keywords: Dengue; stochastic differential equations; Euler–Maruyama method; SARIMA; forecasting

1. Introduction

Dengue outbreaks represent a major public health challenge across tropical and subtropical regions worldwide. In Colombia, the disease remains a persistent concern, particularly in the department of Valle del Cauca, where 18,112 cases were reported in the epidemiological bulletin for week 13 of 2024. The transmission dynamics of Dengue are strongly associated with the presence of the *Aedes aegypti* mosquito, the principal vector of the virus. The limited effectiveness of vector control

strategies has facilitated the continued spread of Dengue, underscoring the urgent need for more efficient preventive measures.

Dengue is a viral infection caused by a member of the *Flaviviridae* family, comprising four distinct but closely related serotypes: DENV-1, DENV-2, DENV-3, and DENV-4. Infection with any of these serotypes can result in two clinical manifestations: Classic (mild) Dengue and severe (hemorrhagic) Dengue, as reported in [1]. Successive infections involving different serotypes substantially increase the risk of developing the severe form of the disease, which can be fatal, particularly among individuals with a history of infections.

Mathematical modeling provides a fundamental framework for understanding the dynamics of epidemic outbreaks such as Dengue. Deterministic models, as discussed in [2, 3], provide consistent predictions through systems of differential equations; however, they often fail to capture external environmental factors, such as temperature, rainfall, and humidity, which play a critical role in the transmission process. To address these limitations, stochastic models incorporate random perturbations into the transmission parameters, thereby offering a more realistic representation of the inherent variability observed in epidemic dynamics (see [4–7]).

Within this framework, we introduce a compartmental SIR-SI host-vector structured model, formulated as a system of stochastic differential equations (SDEs). The model incorporates population stratification into two age groups: youth (0–17 years) and adults (18 years and older). This stratification is essential for accurately characterizing Dengue transmission in Valle del Cauca, as infection risk and exposure conditions vary by age. For younger individuals, transmission risk is primarily associated with school environments and public transportation, whereas adults are more likely to be exposed in workplaces and social settings. This demographic segmentation contributes to a more nuanced understanding of the long-term transmission dynamics of Dengue within the region.

Model parameters and initial conditions are derived from other studies and Dengue incidence data reported by the Public Health Surveillance System (SIVIGILA) [8] for Valle del Cauca. The parameters governing transmission dynamics and stochastic fluctuations were estimated by solving a nonlinear least-squares problem using the Trust-Region Reflective (TRF) optimization algorithm. Using these parameter estimates, simulations were performed covering the historical observation period and a one-year forecast horizon. The simulation results were subsequently compared with empirical data to assess the model's goodness-of-fit and predictive performance.

Complementing the SDE-based modeling framework, a Seasonal Autoregressive Integrated Moving Average (SARIMA) model was implemented to forecast the temporal evolution of monthly Dengue cases in Valle del Cauca. This time-series model effectively captures trend and seasonal components of the data, characterized by periodic variations in the mean, as discussed in [9]. Seasonal fluctuations in Dengue incidence largely reflect climatic conditions, such as temperature and humidity, that directly influence mosquito population dynamics. Incorporating these seasonal effects enhances the robustness and accuracy of the predictive analysis.

The remainder of this paper is organized as follows: In Section 2, we present the formulation of the proposed stochastic model for Dengue transmission and its numerical discretization via the Euler–Maruyama scheme. In Section 3, we describe the Dengue case data for Valle del Cauca, the parameter calibration procedure, and the simulation results, including the short-term forecasts. In Section 4, we detail the implementation of the SARIMA model and evaluate its predictive accuracy. Finally, In Section 5, we summarize the major conclusions and discuss the implications of the findings for public

health surveillance and epidemic control.

2. A stochastic model: Equations and methods

To analyze the dynamics of Dengue transmission in Valle del Cauca, we propose a stochastic compartmental *SIR–SI* model that extends the deterministic frameworks developed in [10, 11]. To incorporate the uncertainty associated with disease transmission, a stochastic component is introduced to represent environmental variability and other sources of random fluctuations, following the methodology of [4, 5]. In this context, each equation is modeled as a stochastic process defined on a complete probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$, where the filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfies the usual conditions of completeness and right-continuity.

Suppose now that a stochastic environmental factor simultaneously affects all individuals in the population. In this case, the transmission rate β becomes a random variable, denoted by $\tilde{\beta}$. More precisely, the number of potentially infectious contacts made by each infected individual in the small time interval $[t, t + dt)$ is expressed as

$$\tilde{\beta}\, dt = \beta\, dt + \sigma\, dB(t), \tag{2.1}$$

where β and σ are positive constants, and $\{B(t)\}_{t\geq 0}$ denotes a standard Brownian motion. Parameter σ quantifies the intensity of the environmental noise.

We assume that the number of infectious contacts a single infected individual makes with another individual during the infinitesimal interval $[t, t+dt)$ is normally distributed with mean $\beta\, dt$ and variance $\sigma^2\, dt$. As $dt \to 0$, the variance tends to zero, ensuring that the stochastic model remains consistent with the deterministic case in the limit. This formulation provides a biologically plausible representation of random fluctuations in the transmission process, enabling the explicit incorporation of environmental noise into population dynamics models.

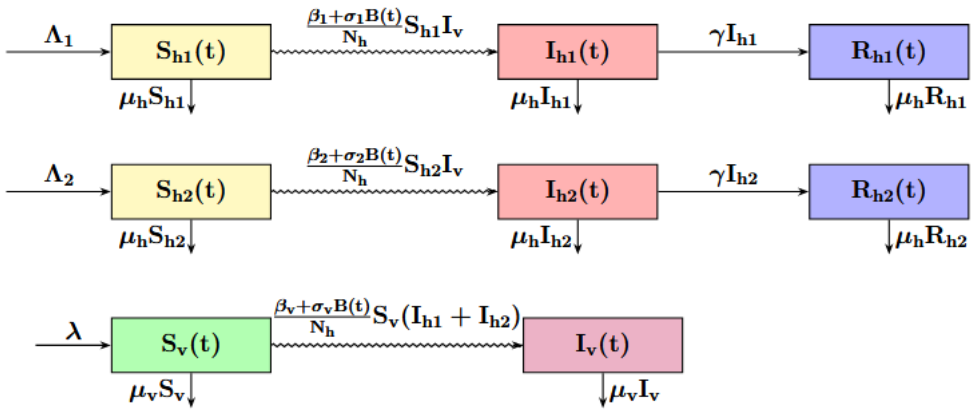

Figure 1. Flow diagram of the dynamics of the proposed stochastic SIR-SI model.

The flow diagram in Figure 1 illustrates the dynamics of the proposed stochastic model. Individuals transition between compartments at specific rates and are influenced by stochastic perturbations, thereby simulating the inherent randomness observed in epidemic processes. The total human population is assumed to be homogeneous and constant, denoted by N_h.

In this study, the model captures the transmission dynamics of the Dengue virus between humans and vectors (mosquitoes) through a compartmental structure of type *SIR–SI*. The human population N_h is stratified into two age groups: h_1 (young individuals) and h_2 (adults). The vector population, denoted by N_v, consists entirely of mosquitoes. For the human population, the compartments are susceptible individuals $S_{h1}(t)$ and $S_{h2}(t)$; infected individuals $I_{h1}(t)$ and $I_{h2}(t)$; and recovered individuals $R_{h1}(t)$ and $R_{h2}(t)$, who are assumed to have acquired immunity to the infecting serotype. The mosquito population is divided into susceptible $S_v(t)$ and infected $I_v(t)$ compartments. All state variables and parameters are assumed to take non-negative real values.

Table 1 summarizes the model parameters used in this work. Parameter values were obtained from demographic data and experimental studies conducted in Valle del Cauca. The recruitment rates Λ_1, Λ_2, and λ denote the inflow of new individuals into the young, adult, and mosquito populations, respectively. The natural mortality rates μ_h and μ_v account for deaths unrelated to Dengue in humans and mosquitoes. The recovery rate γ characterizes the per capita rate at which infected humans recover and acquire immunity. The transmission rates β_1, β_2, and β_v quantify the per-contact probabilities of transmission from infected mosquitoes to young and adult humans, and from infected humans to susceptible mosquitoes, respectively. Finally, $\sigma_1, \sigma_2,$ and σ_v denote the intensities of stochastic perturbations affecting the corresponding transmission channels.

Table 1. Model parameters and descriptions.

Parameter	Description
Λ_1, Λ_2	Recruitment rates of susceptible young and adult individuals, respectively.
μ_h	Natural mortality rate for human individuals (both age groups).
γ	Recovery rate of infected humans (both age groups).
λ	Recruitment rate of susceptible mosquitoes.
μ_v	Natural mortality rate of mosquitoes.
β_1	Transmission rate from infected mosquitoes to susceptible young individuals.
β_2	Transmission rate from infected mosquitoes to susceptible adult individuals.
β_v	Transmission rate from infected humans to susceptible mosquitoes.
σ_1, σ_2, σ_v	Intensities of stochastic perturbations affecting the corresponding transmission pathways.

Susceptible individuals $S_{h1}(t)$ and $S_{h2}(t)$ may become infected through contact with the same pool of infected mosquitoes with transmission rates $\tilde{\beta}_1$ and $\tilde{\beta}_2$, respectively. We represent the stochastic infection rate as

$$\tilde{\beta}_i dt = \beta_i\, dt + \sigma_i\, dB_t^i, \quad i = 1, 2, v,$$

where B_t^i denotes independent Brownian motions associated with the youth, adult, and vector populations, respectively. Susceptible humans are also subject to natural mortality at a rate μ_h. Infected individuals $I_{h1}(t)$ and $I_{h2}(t)$ increase via new infections and decrease through recovery at rate γ or death at rate μ_h. Recovered individuals $R_{h1}(t)$ and $R_{h2}(t)$ arise from recovery and decline only due to natural mortality. Susceptible mosquitoes enter the vector population at rate λ and may become infected by biting infected humans $I_{h1}(t)$ and $I_{h2}(t)$ at transmission rate β_v. All mosquitoes, whether susceptible or infected, experience natural mortality at a rate μ_v. We assume that the total human population N_h remains constant over time. Under this constant-population assumption, the

recruitment rates must satisfy $\Lambda_1 + \Lambda_2 = \mu_h N_h$, ensuring that births (or inflow) balance natural deaths in the human population at demographic equilibrium.

Next, we present the system of stochastic differential equations for the modified SIR–SI model with random perturbations (see [5]). Let $\{S_{h1}(t), I_{h1}(t), R_{h1}(t), S_{h2}(t), I_{h2}(t), R_{h2}(t), S_v(t), I_v(t)\}, t \geq 0$, be governed by the following Itô system:

$$\begin{aligned}
dS_{h1}(t) &= \Lambda_1\,dt - \frac{\beta_1 S_{h1}(t) I_v(t)}{N_h}\,dt - \mu_h S_{h1}(t)\,dt - \frac{\sigma_1 S_{h1}(t) I_v(t)}{N_h}\,dB_t^1,\\
dI_{h1}(t) &= \frac{\beta_1 S_{h1}(t) I_v(t)}{N_h}\,dt - (\mu_h + \gamma) I_{h1}(t)\,dt + \frac{\sigma_1 S_{h1}(t) I_v(t)}{N_h}\,dB_t^1,\\
dR_{h1}(t) &= \gamma I_{h1}(t)\,dt - \mu_h R_{h1}(t)\,dt,\\
dS_{h2}(t) &= \Lambda_2\,dt - \frac{\beta_2 S_{h2}(t) I_v(t)}{N_h}\,dt - \mu_h S_{h2}(t)\,dt - \frac{\sigma_2 S_{h2}(t) I_v(t)}{N_h}\,dB_t^2,\\
dI_{h2}(t) &= \frac{\beta_2 S_{h2}(t) I_v(t)}{N_h}\,dt - (\mu_h + \gamma) I_{h2}(t)\,dt + \frac{\sigma_2 S_{h2}(t) I_v(t)}{N_h}\,dB_t^2,\\
dR_{h2}(t) &= \gamma I_{h2}(t)\,dt - \mu_h R_{h2}(t)\,dt,\\
dS_v(t) &= \lambda\,dt - \frac{\beta_v S_v(t)\big(I_{h1}(t) + I_{h2}(t)\big)}{N_h}\,dt - \mu_v S_v(t)\,dt - \frac{\sigma_v S_v(t)\big(I_{h1}(t) + I_{h2}(t)\big)}{N_h}\,dB_t^v,\\
dI_v(t) &= \frac{\beta_v S_v(t)\big(I_{h1}(t) + I_{h2}(t)\big)}{N_h}\,dt - \mu_v I_v(t)\,dt + \frac{\sigma_v S_v(t)\big(I_{h1}(t) + I_{h2}(t)\big)}{N_h}\,dB_t^v.
\end{aligned} \tag{2.2}$$

Here, $\{B_i(t)\}_{t\geq 0}$ for $i = 1, 2, v$ denote independent standard one-dimensional Brownian motions, and $\sigma_1, \sigma_2, \sigma_v \geq 0$ represent the intensities of the corresponding stochastic perturbations.

Under standard local Lipschitz continuity and linear growth conditions on the drift and diffusion coefficients, and for nonnegative initial data, system Equation 2.2 admits a unique strong solution that remains nonnegative for all $t \geq 0$. This follows from classical results on existence, uniqueness, and positivity for stochastic differential equations (see [12–14]). In particular, for epidemiological SDE models with multiplicative noise, positivity of the state variables is preserved almost surely. We therefore state the following lemma without proof:

Lemma 2.1 (Existence, Uniqueness, and Non-Explosion)**.** Let $X_t \in \mathbb{R}^d$ satisfy the stochastic differential equation

$$dX_t = F(X_t)\,dt + \Gamma(X_t)\,dW_t, \qquad X_0 = x_0 \in (0,\infty)^d,$$

where W_t is an m-dimensional standard Brownian motion, and $F : \mathbb{R}^d \to \mathbb{R}^d$, $\Gamma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are measurable functions. Assume that:

1) **(Local Lipschitz condition)** For each compact set $K \subset \mathbb{R}^d$, there exists a constant $L_K > 0$ such that

$$\|F(x) - F(y)\| + \|\Gamma(x) - \Gamma(y)\| \leq L_K \|x - y\|, \qquad \forall\, x, y \in K.$$

2) **(Linear growth condition)** There exists a constant $C > 0$ such that

$$\|F(x)\|^2 + \|\Gamma(x)\|_F^2 \leq C(1 + \|x\|^2), \qquad \forall\, x \in \mathbb{R}^d,$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

Then there exists a unique strong solution X_t to the SDE for all $t \geq 0$, and the solution is non-explosive. Furthermore, if $X_0 \in (0, \infty)^d$ and both F and Γ preserve positivity, then

$$X_t \in (0, \infty)^d \quad \text{for all } t \geq 0$$

almost surely.

2.1. Numerical discretization

The Euler—Maruyama method, described in [13] and introduced by the Japanese mathematician G. Maruyama [15] as an extension of the Euler method, is a numerical integration technique used to obtain approximate solutions to a system of stochastic differential equations given an initial value $X_0 = x_0$. Moreover, a time partition $0 = t_0 < t_1 < \cdots < t_k = T$ with step size $\Delta t = t_{i+1} - t_i$, and increments $\Delta B_i = B(t_{i+1}) - B(t_i) \sim \mathcal{N}(0, \Delta t)$. For each trajectory of the stochastic process $\{X_t, t \geq 0\}$, the value $X_{t_{i+1}}$ is approximated using the value of the previous step X_{t_i}. The approximate solution of the stochastic differential equation is given by

$$X_{t_{i+1}} = X_{t_i} + f(X_{t_i})\Delta t + g(X_{t_i})\Delta B_i, \tag{2.3}$$

for all $i = 0, 1 \cdots, k-1$, and where f and g are the drift and diffusion coefficients of the SDE, respectively.

Applied to our model, the corresponding discretization in Eq (2.2) must be carried out, which is given by:

$$\begin{aligned}
S_{h1}(t_{j+1}) &= S_{h1}(t_j) + \left[\Lambda_1 - \frac{\beta_1 S_{h1}(t_j) I_v(t_j)}{N_h} - \mu_h S_{h1}(t_j)\right]\Delta t - \frac{\sigma_1 S_{h1}(t_j) I_v(t_j)}{N_h}\Delta B^1_{t_j},\\
I_{h1}(t_{j+1}) &= I_{h1}(t_j) + \left[\frac{\beta_1 S_{h1}(t_j) I_v(t_j)}{N_h} - (\mu_h + \gamma) I_{h1}(t_j)\right]\Delta t + \frac{\sigma_1 S_{h1}(t_j) I_v(t_j)}{N_h}\Delta B^1_{t_j},\\
R_{h1}(t_{j+1}) &= R_{h1}(t_j) + \left[\gamma I_{h1}(t_j) - \mu_{h1} R_{h1}(t_j)\right]\Delta t,\\
S_{h2}(t_{j+1}) &= S_{h2}(t_j) + \left[\Lambda_2 - \frac{\beta_2 S_{h2}(t_j) I_v(t_j)}{N_h} - \mu_h S_{h2}(t_j)\right]\Delta t - \frac{\sigma_2 S_{h2}(t_j) I_v(t_j)}{N_h}\Delta B^2_{t_j},\\
I_{h2}(t_{j+1}) &= I_{h2}(t_j) + \left[\frac{\beta_2 S_{h2}(t_j) I_v(t_j)}{N_h} - (\mu_h + \gamma) I_{h2}(t_j)\right]\Delta t + \frac{\sigma_2 S_{h2}(t_j) I_v(t_j)}{N_h}\Delta B^2_{t_j},\\
R_{h2}(t_{j+1}) &= R_{h2}(t_j) + \left[\gamma I_{h2}(t_j) - \mu_h R_{h2}(t_j)\right]\Delta t,\\
S_v(t_{j+1}) &= S_v(t_j) + \left[\lambda - \frac{\beta_v S_v(t_j)(I_{h1}(t_j) + I_{h2}(t_j))}{N_h} - \mu_v S_v(t_j)\right]\Delta t - \frac{\sigma_v S_v(t_j)(I_{h1}(t_j) + I_{h2}(t_j))}{N_h}\Delta B^v_{t_j},\\
I_v(t_{j+1}) &= I_v(t_j) + \left[\frac{\beta_v S_v(t_j)(I_{h1}(t_j) + I_{h2}(t_j))}{N_h} - \mu_v I_v(t_j)\right]\Delta t + \frac{\sigma_v S_v(t_j)(I_{h1}(t_j) + I_{h2}(t_j))}{N_h}\Delta B^v_{t_j}.
\end{aligned} \tag{2.4}$$

3. Computational results from statistical data: A case study of the department of Valle del Cauca

In this section, we analyze the stochastic dynamics of Dengue transmission under environmental noise using real-world data from the Department of Valle del Cauca. The data employed for

simulations, parameter calibration, and analysis were obtained from the National Public Health Surveillance System database [8]. This database includes both quantitative and qualitative variables, such as event codes, notification dates, age, hospitalization and mortality records, and geographical information corresponding to the department of residence. All computational implementations were carried out using `Python 3.11.7`.

For the proposed stochastic model described in Eq (2.2), Dengue case data from January 2013 to December 2023 were used, together with a forward projection for 2024 (January to December). The model was calibrated using age-stratified infection data, distinguishing young and adult populations in Valle del Cauca. The simulated infected populations, I_{h1} and I_{h2}, corresponding to these two demographic groups, were directly compared with the reported Dengue cases. It is important to note that surveillance data represent newly reported cases (incidence) rather than the instantaneous prevalence generated by the model; thus, the simulated results were interpreted accordingly. All parameters were defined on a daily scale, while the simulations were evaluated at monthly intervals (every 30 days), recording the system state at the end of each month. Consequently, the initial conditions were established considering the total population at the beginning of the first month.

The simulations were executed using the discretization scheme defined in Eq (2.4). The parameters considered in the simulations include:

$$(\Lambda_1, \Lambda_2, \beta_1, \beta_2, \beta_v, \gamma, \mu_h, \mu_v, \lambda, \sigma_1, \sigma_2, \sigma_v),$$

while the initial conditions correspond to:

$$N_h(0),\ S_{h1}(0),\ S_{h2}(0),\ I_{h1}(0),\ I_{h2}(0),\ R_{h1}(0),\ R_{h2}(0),\ S_v(0),\ I_v(0).$$

Figure 2 depicts the monthly time series of reported Dengue cases in Valle del Cauca from 2013 to 2023, categorized into two age groups: young and adult individuals. This figure is based on data obtained from [8] for the period between January 2013 and mid-2019. Analysis of the data reveals that the number of infections among young individuals was generally lower than among adults. However, starting in 2019, this difference became less evident: In some periods, infections among minors exceeded those among adults, whereas in others, the opposite occurred.

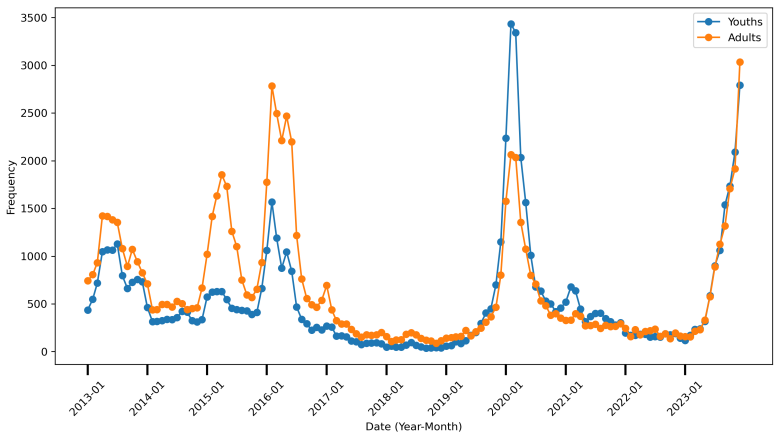

Figure 2. Monthly Dengue cases by age group in Valle del Cauca [8].

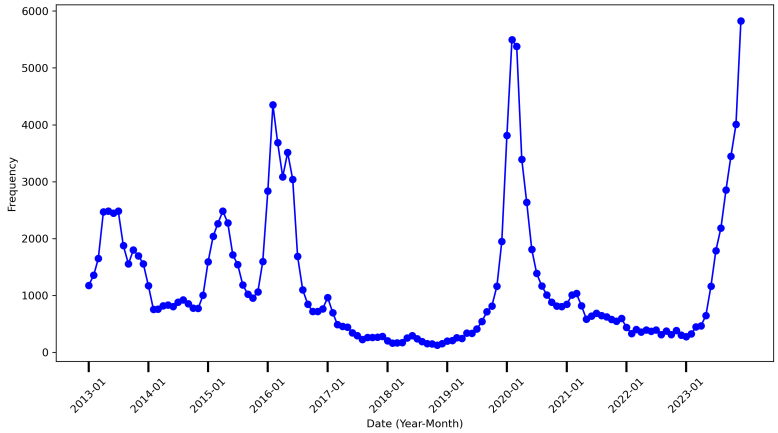

Figure 3. Total monthly Dengue cases in Valle del Cauca [8].

Further analysis shows that Dengue case peaks tend to occur more frequently during the early months of the year. This seasonal pattern may be associated with environmental factors such as higher temperatures and humidity, which enhance the proliferation of mosquito vectors. Additionally, human mobility patterns play a significant role in transmission dynamics: Among younger populations, exposure often occurs in schools and public transportation settings, whereas for adults, transmission risk is typically associated with occupational and social environments.

Figure 3 presents the overall monthly time series of reported Dengue cases from January 2013 to December 2023. The data, obtained from [8], display temporal trends consistent with those observed in Figure 2. Notably, significant peaks occurred in 2016, 2020, and 2023, suggesting a potential cyclical pattern.

This recurrent behavior provides the basis for forecasting future outbreaks. Given the peak observed in 2023 and the emerging trend, an increase in Dengue incidence is anticipated for 2024. Accordingly, these historical data were also used to construct forecasts using a Seasonal Autoregressive Integrated Moving Average (SARIMA) model, enabling an assessment of the risk and potential magnitude of future outbreaks.

3.1. Deterministic and stochastic simulations of Dengue dynamics

Before making predictions with the stochastic model, the deterministic behavior of Dengue in Valle del Cauca was simulated using parameters from previous studies and official data. For the deterministic SIR–SI model simulations, eight main parameters are considered: the recruitment rate of new individuals into the human population (Λ_1, Λ_2), the transmission rates between humans and mosquitoes (β_1, β_2, and β_v), the recovery rate of infected humans (γ), the natural mortality rates of humans and mosquitoes (μ_h and μ_v), the recruitment rate of new mosquitoes (λ), and the stochastic noise intensity (σ).

These parameters were established based on epidemiological and demographic information, complemented by previous studies that have estimated realistic ranges for Dengue transmission dynamics. The parameter values of the model are presented in Table 2, while the initial conditions used in the simulations are shown in Table 3. The procedure for determining these values is described below.

Table 2. Fixed parameters of the stochastic SIR–SI model.

Parameter	Value	Units	Reference
Λ_1	105	humans · day^{-1}	[16]
Λ_2	105	humans · day^{-1}	[16]
β_1	0.1	day^{-1}	[17]
β_2	0.1	day^{-1}	[17]
β_v	0.5	day^{-1}	[17]
γ	0.07142	day^{-1}	[18]
μ_h	0.0000391	day^{-1}	[18]
μ_v	0.008	day^{-1}	[19]
λ	2500	mosquitoes · day^{-1}	[17]
σ_1	0.1	—	Assumed
σ_2	0.2	—	Assumed
σ_v	0.3	—	Assumed

Table 3. Initial conditions of the stochastic SIR–SI model.

Initial conditions	Value	Units	Reference
$S_{h1}(0)$	1,681,310	individuals	[20]
$I_{h1}(0)$	435	individuals	[8]
$R_{h1}(0)$	31	individuals	[18]
$S_{h2}(0)$	2,107,303	individuals	[20]
$I_{h2}(0)$	742	individuals	[8]
$R_{h2}(0)$	52	individuals	[18]
$S_v(0)$	312,500	mosquitoes	[18]
$I_v(0)$	100	mosquitoes	[19]
$N_h(0)$	3,789,874	individuals	[20]

Dengue transmission in Valle del Cauca was first simulated deterministically using the SIR–SI model with parameters and initial conditions summarized in Table 2. Deterministic trajectories (Figure 4) for compartments S_{h1}, S_{h2}, I_{h1}, I_{h2}, R_{h1}, R_{h2}, S_v, and I_v show smooth trends over January 2013 to December 2023. Susceptible humans decline, infections rise and fall, and recovered compartments increase steadily. While deterministic simulations capture general epidemic patterns, they do not reflect stochastic fluctuations or abrupt outbreak peaks.

To account for variability, stochastic simulations were performed using the Euler–Maruyama method applied to the discretized system (Eq (2.4)) based on Eq (2.2), with the same parameters. Figure 5 shows four stochastic realizations per compartment. Susceptible humans decrease, while infections exhibit seasonal fluctuations and occasional intense outbreaks, capturing peaks absent in deterministic trajectories. Recovered compartments increase steadily, and mosquito dynamics reflect ongoing transmission. Stochastic simulations thus provide a more realistic representation of epidemic variability.

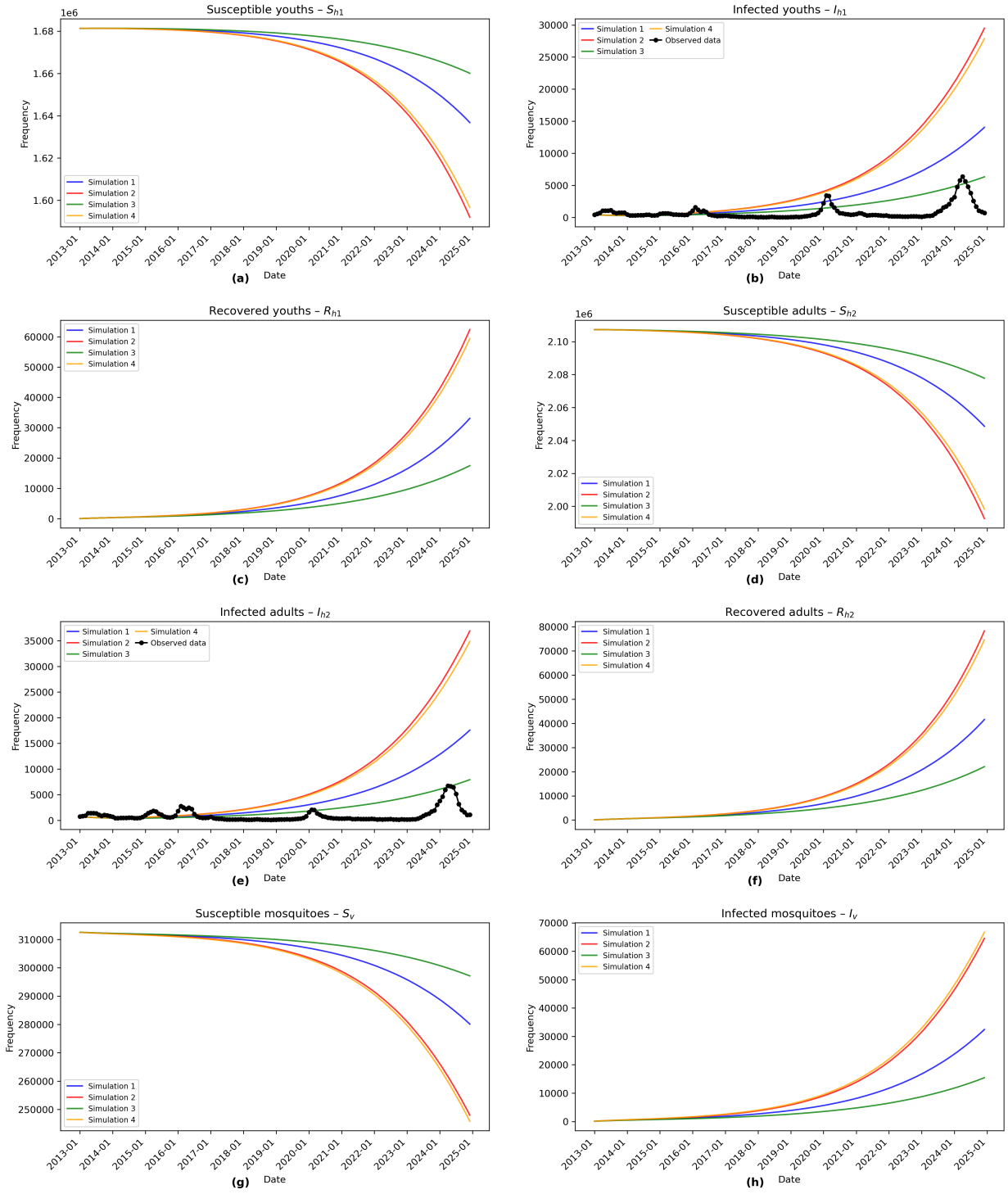

Figure 4. Deterministic simulations of Dengue dynamics. Four trajectories per compartment are shown; observed infections are superimposed.

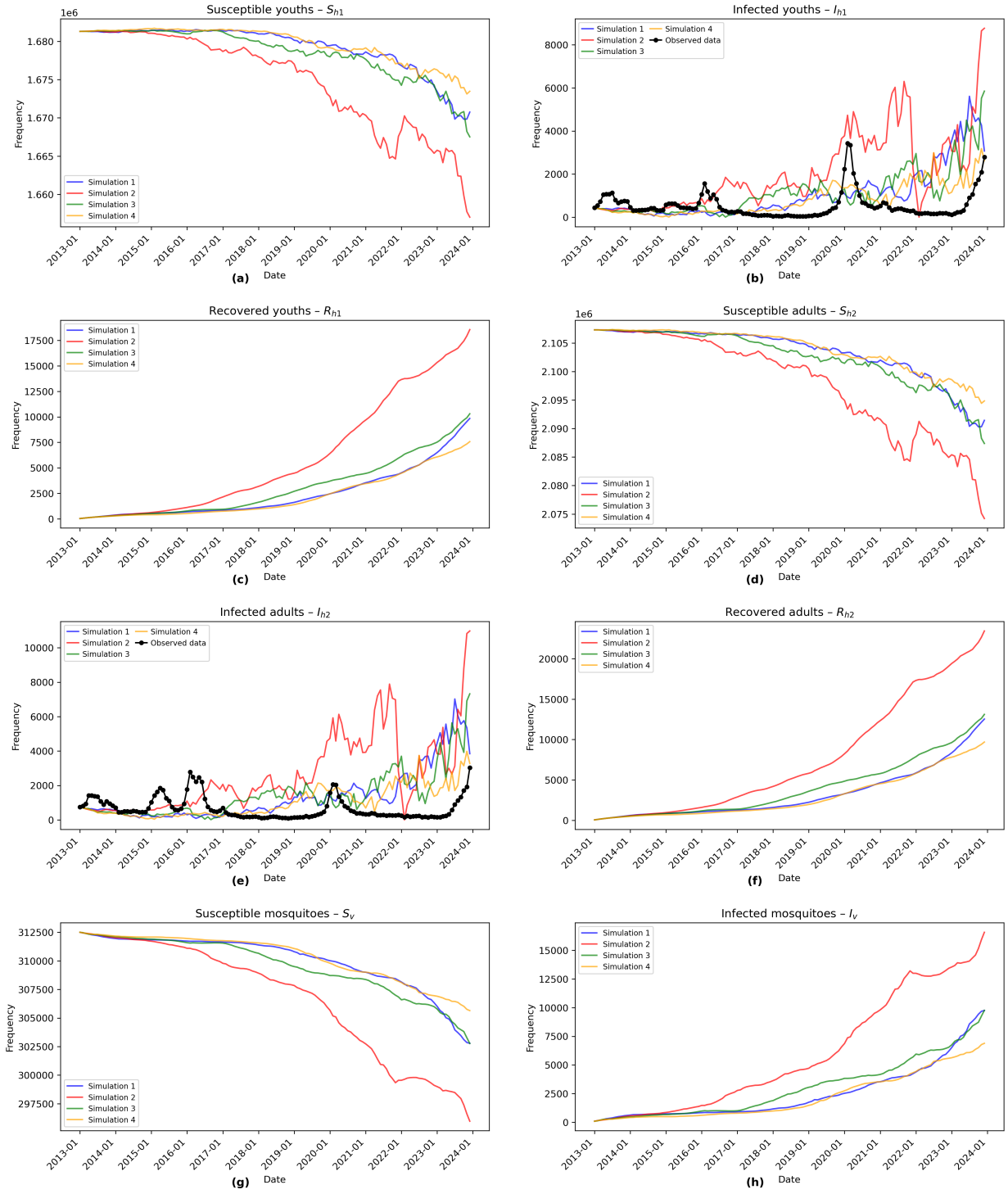

Figure 5. Stochastic simulations of Dengue dynamics. Four trajectories per compartment are shown; observed infections are included for comparison.

Model performance was evaluated using the *Mean Squared Error* (MSE) between simulated and observed infections. Table 4 shows that stochastic simulations generally yield slightly lower MSEs,

particularly for young individuals (I_{h1}), indicating improved fit by capturing temporal variability and epidemic peaks.

Table 4. Comparison of MSE values between deterministic and stochastic models.

Simulation	Deterministic		Stochastic	
	MSE (I_{h1})	MSE (I_{h2})	MSE (I_{h1})	MSE (I_{h2})
1	24,628,150	41,152,402	24,627,909	41,152,091
2	25,661,649	42,547,716	25,661,758	42,548,217
3	23,393,697	39,572,211	23,393,069	39,571,244
4	25,658,699	42,524,967	25,658,880	42,525,718

3.2. Calibration of the model

Model calibration was performed to estimate the optimal parameters (β, β_v, σ) of the Dengue model, ensuring the best fit between model predictions and observed infection data in Valle del Cauca. The procedure involved three major steps: Formulating a constrained optimization problem, implementing the Trust Region Reflective (TRF) algorithm to solve it, and selecting the parameter set that minimizes model residuals in Eq (2.2), while ensuring feasibility and robustness.

Practical identifiability of (β, β_v, σ) was assessed using *profile likelihoods* and *Fisher information*, with confidence intervals quantifying parameter uncertainty. Confidence bands for the state variables illustrate how parameter uncertainty propagates through the stochastic model. Robustness was ensured via *multi-start TRF optimization* from dispersed initial guesses, examining the dispersion of resulting optima.

Parameters were estimated by bounded nonlinear least-squares,

$$\min_{\ell \le x \le u} \frac{1}{2} \|r(x)\|_2^2,$$

using the *Trust-Region Reflective (TRF)* algorithm. TRF handles bounds efficiently via trust-region steps, reflection for active constraints, and Levenberg–Marquardt regularization as needed. Under standard assumptions, TRF converges globally, with local convergence linear or superlinear for well-conditioned problems. Confidence intervals were derived from the Jacobian at the solution, adjusted for active bounds, providing practical evidence of identifiability.

Problem formulation: The calibration targets three key parameters: β (mosquito-to-human transmission rate), β_v (human-to-mosquito transmission rate), and σ (stochastic noise intensity capturing random environmental variations). These parameters govern the disease dynamics and its stochastic fluctuations, whereas the remaining parameters are fixed based on empirical data. Accurate estimation enables the model to effectively describe Dengue dynamics across young and adult populations. Parameter estimation was performed by solving a nonlinear least-squares problem using the TRF algorithm.

Implementation of the TRF algorithm: The Trust Region Reflective (TRF) algorithm [21], implemented via SciPy [22] in Python, combines trust-region strategies with reflective

transformations to handle parameter bounds. For a least-squares problem with n parameters and m residuals ($m \geq n$), the objective is:

$$f(\mathbf{x}) = \frac{1}{2}\|\mathbf{r}(\mathbf{x})\|^2 = \frac{1}{2}\sum_{i=1}^{m} r_i(\mathbf{x})^2, \quad \mathbf{l} \leq \mathbf{x} \leq \mathbf{u},$$

where $\mathbf{x} \in \mathbb{R}^n$ is the parameter vector, $\mathbf{r} : \mathbb{R}^n \to \mathbb{R}^m$ is the residual function, and $\mathbf{l}, \mathbf{u} \in \mathbb{R}^n$ define bounds.

The algorithm uses adaptive scaling:

$$v_i = \begin{cases} u_i - x_i & \text{if } (\nabla f(\mathbf{x}))_i < 0 \text{ and } u_i < \infty, \\ x_i - l_i & \text{if } (\nabla f(\mathbf{x}))_i > 0 \text{ and } l_i > -\infty, \\ 1 & \text{otherwise,} \end{cases} \quad \mathbf{D} = \text{diag}(\mathbf{v}^{1/2}),$$

which rescales the search space to respect bounds. In the transformed space $\hat{\mathbf{x}} = \mathbf{D}^{-1}\mathbf{x}$, first-order optimality is expressed as

$$\mathbf{D}^2 \nabla f(\mathbf{x}) = \mathbf{0},$$

covering interior and boundary conditions. At each iteration k, the trust-region subproblem is

$$\min_{\hat{\mathbf{p}}} \left(\frac{1}{2}\hat{\mathbf{p}}^T \mathbf{B}_k \hat{\mathbf{p}} + \hat{\mathbf{g}}_k^T \hat{\mathbf{p}} \right) \quad \text{s.t.} \quad \|\hat{\mathbf{p}}\| \leq \Delta_k, \tag{3.1}$$

with $\hat{\mathbf{g}}_k = \mathbf{D}\nabla f(\mathbf{x}_k)$ and $\mathbf{B}_k = \mathbf{D}\nabla^2 f(\mathbf{x}_k)\mathbf{D} + \mathbf{C}_k$, where $\mathbf{C}_k$ ensures positive definiteness. Constraint violations are addressed via reflection:

$$x_i^{\text{ref}} = \begin{cases} 2l_i - x_i & x_i < l_i, \\ 2u_i - x_i & x_i > u_i, \\ x_i & \text{otherwise.} \end{cases}$$

The subproblem is solved using SVD on the scaled Jacobian $\mathbf{JD}$, yielding

$$\hat{\mathbf{p}} = -\mathbf{V}(\mathbf{\Sigma}^T\mathbf{\Sigma} + \lambda\mathbf{I})^{-1}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{r}(\mathbf{x}),$$

with λ chosen to satisfy the trust-region constraint $\|\hat{\mathbf{p}}\| = \Delta_k$.

Selection of the best parameter set: The final parameter vector $\mathbf{x}^*$ is selected based on:

1) *Optimality:* $\|\mathbf{D}^2\nabla f(\mathbf{x}_k)\| < \epsilon_1$.

2) *Objective minimization:* $\mathbf{x}^* = \arg\min f(\mathbf{x})$ among evaluated points.

3) *Feasibility:* $\mathbf{l} \leq \mathbf{x}^* \leq \mathbf{u}$.

The TRF algorithm guarantees convergence from any feasible starting point, rapid convergence near the optimum, efficiency for large sparse problems, automatic handling of bounds, and numerical stability. Using Dengue case records from Valle del Cauca and SIVIGILA data [8], the model parameters were fitted via nonlinear least squares with TRF. Optimized parameters with 95% confidence intervals are shown in Table 5.

Table 5. Optimized parameters using the TRF algorithm.

Parameter	Value	Units	Calibration bounds
β_1	0.6306	day^{-1}	[0.3826, 1.0393]
β_2	0.5875	day^{-1}	[0.3560, 0.9695]
σ_1	0.0347	day^{-1}	[0, 84.0347]
σ_2	1.3844	day^{-1}	[0.1369, 14.0011]
β_v	0.0177	day^{-1}	[0.0076, 0.0416]
σ_v	0.0373	day^{-1}	[0.0017, 0.8099]

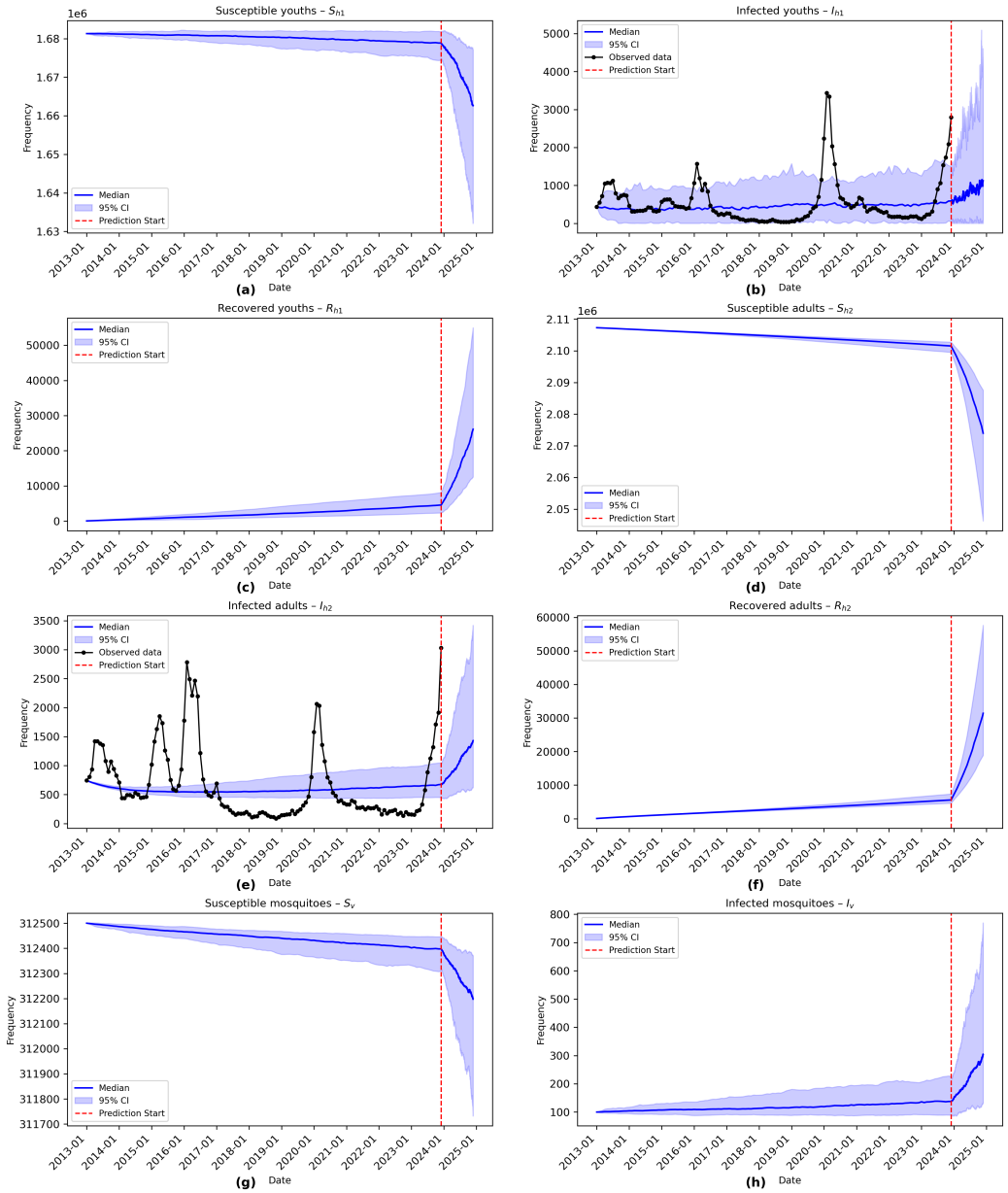

Figure 6. Stochastic model simulations with estimated parameters. Median trajectories and 95% prediction intervals are shown.

Figure 6 illustrates the temporal evolution of human and mosquito populations, presenting the median trajectories along with 95% prediction intervals obtained from 100 stochastic simulations using the Euler–Maruyama method (Eq (2.4)). The vertical red line marks the transition between the interpolation phase, during which the model is fitted to observed data, and the extrapolation phase, which corresponds to model-based forecasts. Simulations employing the calibrated parameters $(\beta_1, \beta_2, \sigma_1, \sigma_2, \beta_v, \sigma_v)$ closely reproduce the observed dynamics, revealing a pronounced decline in susceptible human compartments S_{h1} and S_{h2}, with a steeper decrease among adults, concomitant with elevated infection levels I_{h2} in the adult population exhibiting greater variability, whereas infections among younger individuals remain comparatively stable. Recovered compartments R_{h1} and R_{h2} increase steadily over time, particularly among adults. In the mosquito population, the number of susceptible vectors, S_v, declines while the number of infected vectors, I_v, increases, consistent with sustained transmission. These results indicate that the model effectively captures the temporal patterns and the stochastic variability of Dengue dynamics, providing a robust framework for short- and medium-term epidemic forecasting.

4. SARIMA

In this section, the SARIMA model was employed as an alternative forecasting approach to the proposed SDE-based model, using Dengue case data for the general population, as illustrated in Figure 3. This choice is motivated by the observation that monthly Dengue incidence in Valle del Cauca exhibits marked seasonality: Warm and humid conditions promote the reproduction of the vector, *Aedes aegypti*, whereas colder periods restrict its proliferation. Such environmental variability gives rise to trend and seasonal components that are well captured by time series models.

The SARIMA model incorporates lagged values, representing past observations that influence the current state of the series, thereby enabling the assessment of temporal dependence within the data. This feature makes SARIMA a suitable tool for short- and medium-term forecasting. In this context, two key concepts relevant to the methodology of this study are introduced: *White noise*, which is essential for validating model assumptions, and the SARIMA framework, which is used to fit and forecast the Dengue time series.

Definition 4.1. A stochastic process $\{Z_t\}$ is called **white noise** if it has zero mean, constant variance, and its variables are uncorrelated at all lags. Formally, the following conditions are satisfied [9]:

1) $\mathbb{E}[Z_t] = 0, \quad t = 1, 2, \ldots$

2) $\mathrm{Var}(Z_t) = \sigma^2, \quad t = 1, 2, \ldots$

3) $\mathrm{Cov}(Z_t, Z_{t-k}) = 0, \quad k = \pm 1, \pm 2, \ldots$

Definition 4.2 (SARIMA Process $(p, d, q) \times (P, D, Q)_s$)**.** If d and D are non-negative integers, then $\{X_t\}$ is called a **seasonal ARIMA process** $(p, d, q) \times (P, D, Q)_s$ with period s, if the differenced series $Y_t := (1 - B)^d (1 - B^s)^D X_t$ is a causal ARMA process defined by [23]

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2), \tag{4.1}$$

where:
$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$,

$\Phi(z) = 1 - \Phi_1 z - \cdots - \Phi_P z^P$,
$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$,
$\Theta(z) = 1 + \Theta_1 z + \cdots + \Theta_Q z^Q$.

For the SARIMA model, the following should be considered:

- The parameters d and D represent the orders of ordinary and seasonal differencing, respectively.
- $\phi(B)$ and $\theta(B)$ are the ordinary (non-seasonal) autoregressive and moving average polynomials.
- $\Phi(B^s)$ and $\Theta(B^s)$ are the seasonal autoregressive and moving average polynomials associated with a periodicity s.

The process $\{Y_t\}$ is causal if and only if $\phi(z) \neq 0$ and $\Phi(z) \neq 0$ for $|z| \leq 1$. In practical applications, D is rarely greater than one, and the values of P and Q are usually less than three [24].

Equation (4.1), satisfied by the differenced process $\{Y_t\}$, can be rewritten in the equivalent form:

$$\phi^*(B)Y_t = \theta^*(B)Z_t, \tag{4.2}$$

where $\phi^*(\cdot)$ and $\theta^*(\cdot)$ are polynomials of degree $p + sP$ and $q + sQ$, respectively, whose coefficients can be expressed in terms of $\phi_1, \ldots, \phi_p$, $\Phi_1, \ldots, \Phi_P$, $\theta_1, \ldots, \theta_q$ and $\Theta_1, \ldots, \Theta_Q$.

Given that $p < s$ and $q < s$, the constraints on the coefficients of $\phi^*(\cdot)$ and $\theta^*(\cdot)$ can be expressed as multiplicative relationships:

$\phi^*_{is+j} = \phi^*_i \Phi^*_j$, for $i = 1, 2, \ldots$ and $j = 1, \ldots, s-1$; similarly, $\theta^*_{is+j} = \theta^*_i \Theta^*_j$, for $i = 1, 2, \ldots$ and $j = 1, \ldots, s-1$.

As a result, the modeling of the time series for monthly Dengue cases in Valle del Cauca from January 2013 to December 2023 was carried out following the steps outlined below. First, the variance of the time series was stabilized. Second, a differentiation diagnosis was performed. Subsequently, the model orders were identified. Next, SARIMA models were fitted by varying the values of p, d, q and P, D, Q. The optimal model was selected based on the *Akaike Information Criterion* (AIC). Finally, the statistical assumptions of the selected model were validated, and a forecast of future values was generated to accurately estimate monthly Dengue cases.

4.1. Variance stabilization

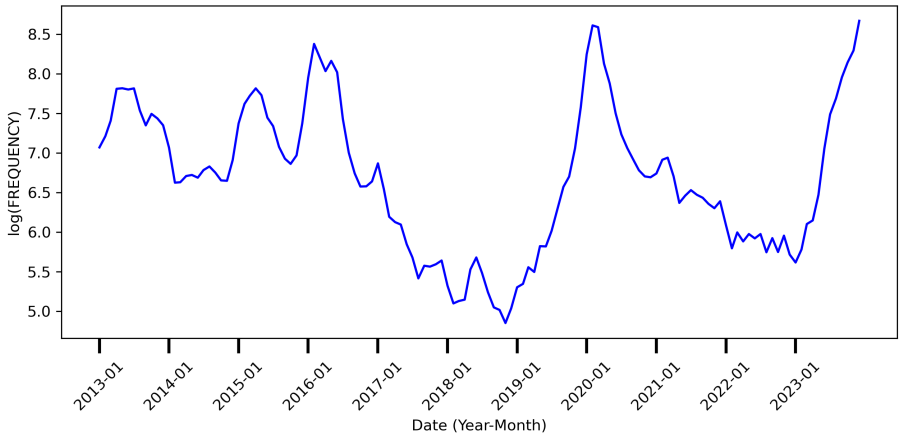

Figure 7. Box-Cox transformation.

The first step involved evaluating the need to apply a Box-Cox transformation to stabilize the variance. This transformation was applied to the time series because, as shown in Figure 3, the variance appeared unstable. The transformation is detailed in the book *Introduction to Time Series and Forecasting* [24]. As a result, in Figure 7, it is evident that the variance has been stabilized following the transformation.

4.2. Differencing diagnosis

The values of d and D were determined to ensure the stationarity of the Box-Cox transformed series, which, in this, case corresponds to the logarithmic transformation. To assess ordinary stationarity, the Augmented Dickey-Fuller (ADF) test was applied, as described in *Introduction to Time Series and Forecasting* [24], to determine whether regular and/or seasonal differencing is required to achieve stationarity. This procedure is essential for the formulation of the SARIMA model, as established in Definition 4.2. The following hypothesis test was performed with the corresponding ADF test results for the monthly Dengue cases in Valle del Cauca, as reported in Table 6:

$$\begin{cases} H_0 : \phi = 1, & \text{the series is non-stationary (unit root present)} \\ H_1 : \phi < 1, & \text{the series is stationary (no unit root).} \end{cases} \tag{4.3}$$

Table 6. Results of the augmented Dickey-Fuller (ADF) test.

Parameter	Value
Test Statistic	0.047
P-value	0.700

Since the p-value is greater than 0.05, we fail to reject H_0, indicating the presence of a unit root and, therefore, non-stationarity in the series. Consequently, the series is differenced to achieve stationarity. In Figure 8, the differenced series is shown, which presents signs of having reached stationarity.

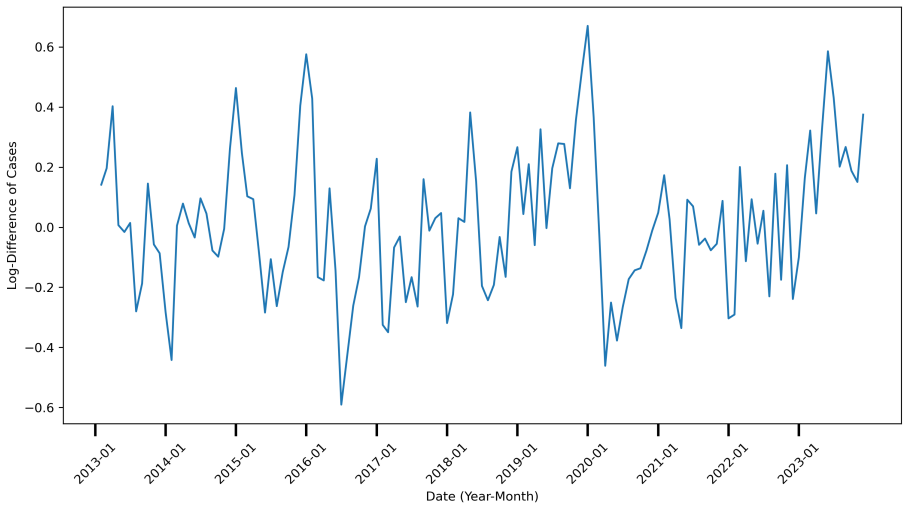

Figure 8. Ordinarily differenced series.

To assess whether a second differencing was necessary, the ADF test was applied to the differenced series again. The resulting p-value of 0.0336 is below the 5% significance level ($p < 0.05$), leading to rejection of the null hypothesis of non-stationarity; thus, no additional differencing is required. The seasonal structure of the series was then examined to determine whether seasonal differencing with period $s = 12$ (annual periodicity) was warranted. For this purpose, exploratory tools such as the monthly boxplot, seasonal subseries plot, and periodogram were used, as shown in Figure 9. Among these, the periodogram most clearly indicates seasonality, displaying a peak at frequency $f = 0.083$, corresponding to a period of $s = 1/0.083 \approx 12$. Harmonics of this dominant frequency also appear at $2f = 0.166$ and $3f = 0.249$.

This seasonal pattern can be attributed to the nature of the time series, which represents Dengue infections, a phenomenon influenced by climate-related factors such as temperature and humidity that vary across months. To confirm this seasonality, the OCSB (Osborn-Chui-Smith-Birchenhall) test was applied using the `pmdarima` library and the command `nsdiffs(series, m=12, test=''ocsb'')`, with a seasonal period $m = 12$ for monthly data. Although the test does not suggest the need for seasonal differencing, graphical analysis reveals clear seasonal behavior in the series. As a result, seasonal differencing is applied with a 12-month lag. This decision is supported visually in Figure 10, where the seasonally differenced series no longer exhibits a clear seasonal structure with period $S = 12$.

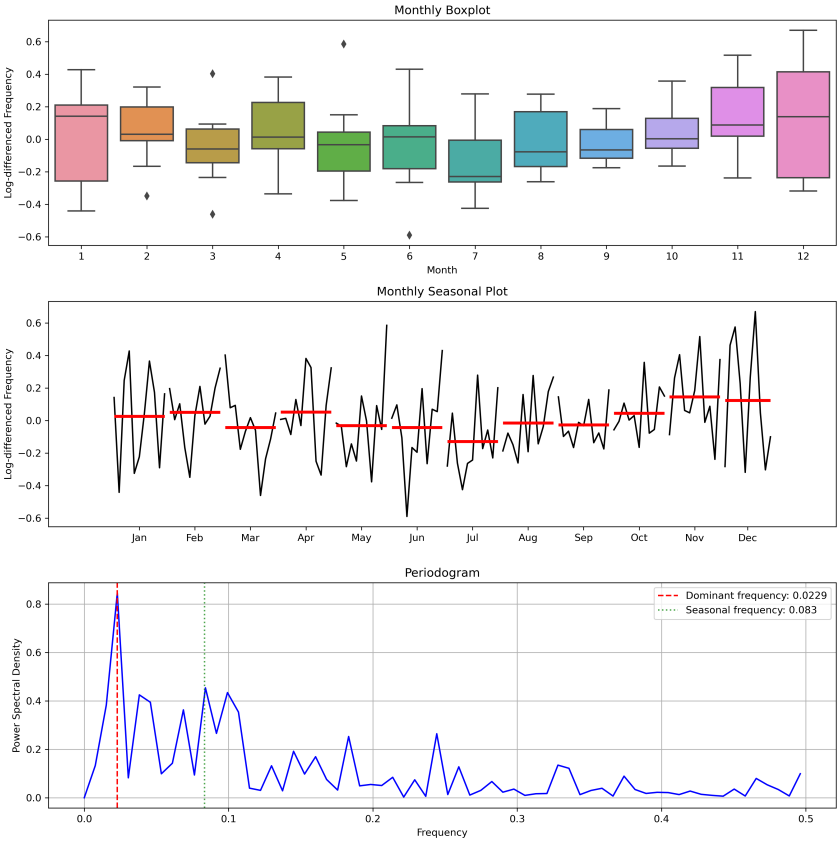

Figure 9. Monthly box-plot, time plot, and periodogram based on the ordinary differenced series.

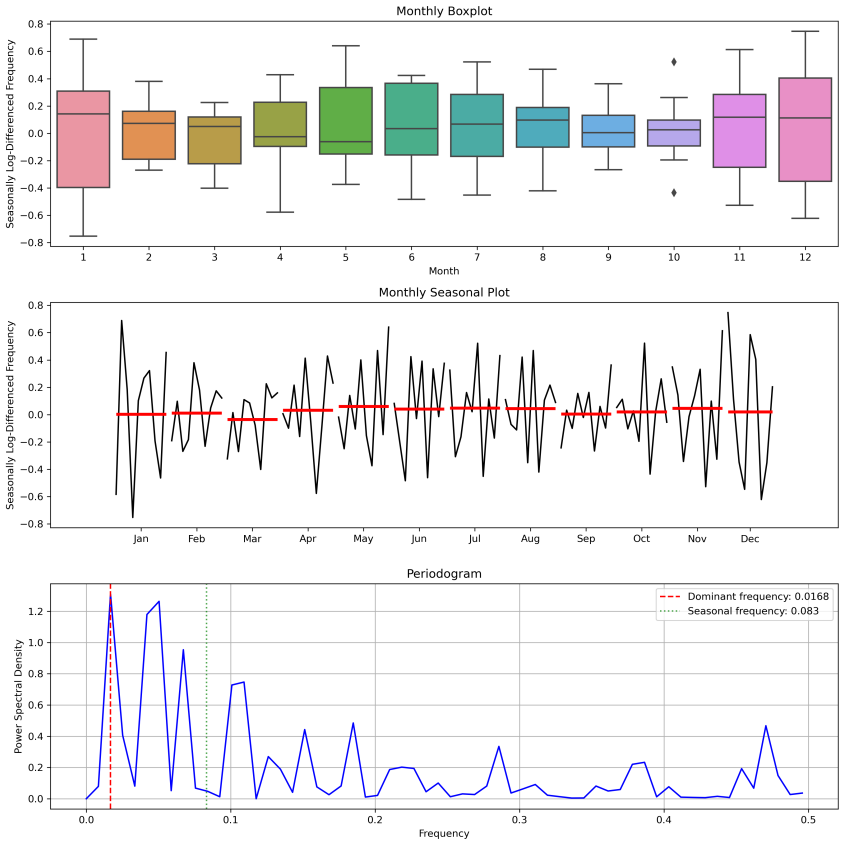

Figure 10. Monthly box-plot, time plot, and periodogram based on the seasonally differenced series.

4.3. Model order identification

The autocorrelation function (ACF) and partial autocorrelation function (PACF) were used to identify potential orders for the SARIMA$(p, d, q) \times (P, D, Q)_s$ model, following the theory presented in *Introduction to Time Series and Forecasting* [24]. In this context, the seasonal autoregressive components P were associated with lags $\hat{\rho}(ks)$, while the non-seasonal components p correspond to $\hat{\rho}(s)$. For the moving average part of the model, the seasonal components Q were identified from $\hat{\alpha}(ks)$, and the non-seasonal components q from $\hat{\alpha}(k)$.

Based on the series that was differenced ordinarily and seasonally, Figure 11 presents the corresponding autocorrelation and partial autocorrelation functions. From the ACF and PACF plots, the following possible model orders are suggested:

- The value of q may be 0, 1, or 2.
- The value of Q may be 0 or 1.
- The value of p may be 0 or 1.
- The value of P may be 0 or 1.
- The seasonal period is $S = 12$.

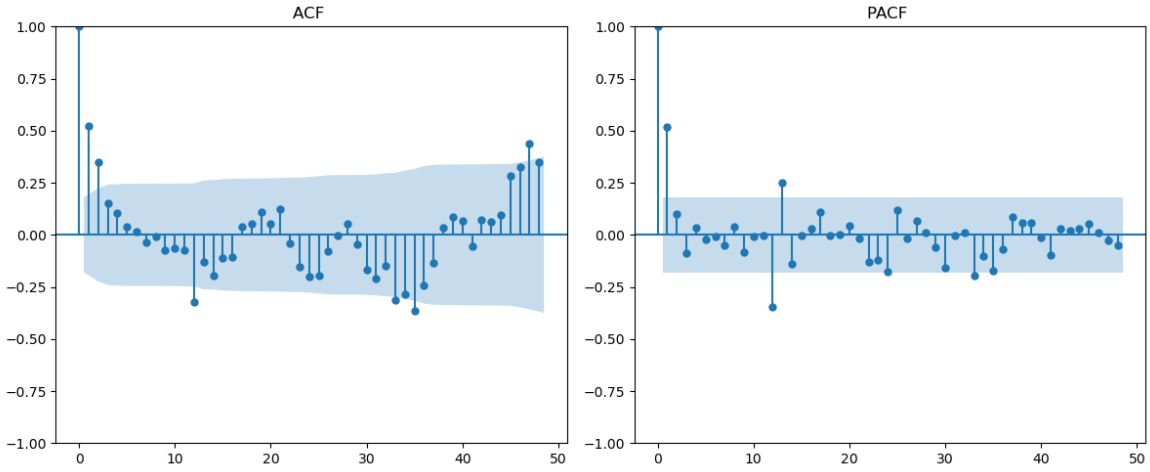

Figure 11. ACF and PACF of the series after ordinary and seasonal differencing.

4.4. Model fitting and selection

The determined values of p, d, q, P, D, and Q were used to estimate the parameters of the SARIMA model: ϕ, θ, Φ, Θ, and σ^2. Parameter estimation and model fitting were performed on the Box-Cox transformed series using the *Python* programming language, specifically with the `statsmodels.api` library, imported as `sm`.

The parameters were estimated using the maximum likelihood method, assuming that the error term follows a Gaussian white noise process $Z_t \sim \mathcal{N}(0, \sigma^2)$. A total of 24 model order combinations were evaluated, and the optimal model was selected based on the lowest value of the *Akaike Information Criterion* (AIC), according to the methodology presented in [24]. The selected model, summarized in Tables 7 and 8, corresponds to a SARIMA$(1, 1, 0)(0, 1, 1)_{12}$ model.

Table 7. Summary statistics of the selected SARIMA model.

Dependent variable	Box–Cox transformed series
Model	SARIMA$(1, 1, 0)(0, 1, 1)_{12}$
Number of observations	132
Log-likelihood	10.878
AIC	−15.756

Table 8. Parameter estimates of the selected SARIMA model.

Parameter	Coefficient	Std. error	z-statistic	p-value	2.5%	97.5%
AR(1)	0.5681	0.081	6.984	< 0.001	0.409	0.727
Seasonal MA$(1)_{12}$	−0.7948	0.123	−6.482	< 0.001	−1.035	−0.555
σ^2	0.0440	0.006	7.540	< 0.001	0.033	0.055

The results indicate that the first-order autoregressive component is positive and highly significant, reflecting temporal dependence in current Dengue cases. The seasonal moving average component

with a 12-month lag is negative and statistically significant, indicating an annual cyclical structure. Furthermore, the low estimated error variance ($\sigma^2 = 0.044$) suggests a good model fit and a limited contribution of the random component. Overall, the results support a seasonal dynamic consistent with the epidemiological behavior of Dengue in Valle del Cauca.

4.5. Validation of statistical assumptions

Once the optimal model was selected, the statistical assumptions about the residuals are evaluated. These residuals were assumed to follow a Gaussian white noise process. To verify this, statistical tests were applied, including the Jarque-Bera test for normality, the Breusch-Pagan test for heteroscedasticity, and the Durbin-Watson statistic to detect the presence of autocorrelation.

The corresponding hypothesis tests were conducted to assess the classical assumptions of the model: normal distribution of errors, homoscedasticity (constant variance over time), and absence of autocorrelation among residuals. These procedures and results are detailed in Table 9.

Table 9. Results of normality, heteroscedasticity, and autocorrelation tests.

Test	Statistic	P-value
Normality (Jarque-Bera)	–	0.667
Heteroscedasticity (Break Test)	–	0.379
Autocorrelation (Durbin-Watson)	1.1889	–

Based on the results presented, the following interpretations can be made regarding the model assumptions:

- **Normality:** The Jarque-Bera test yields a p-value of 0.667, which is greater than the common significance threshold (e.g., 0.05). Thus, the null hypothesis of normality is not rejected, indicating that the model residuals are normally distributed.
- **Homoscedasticity:** The Break Test result for heteroscedasticity shows a p-value of 0.379, also above the typical significance level. This suggests that there is insufficient evidence to claim the presence of heteroscedasticity, indicating that the residuals have constant variance.
- **Autocorrelation:** The Durbin-Watson statistic is 1.1889. This value is noticeably below the ideal value of 2, which may suggest the presence of positive autocorrelation in the residuals. This finding should be interpreted with caution, as it may affect the efficiency of the estimators.

In summary, the results suggest that the assumptions of normality and homoscedasticity are adequately satisfied for the model residuals. However, the potential presence of autocorrelation, as evidenced by the Durbin-Watson statistic, indicates the need to review the model structure or consider additional adjustments to address this temporal dependence.

4.6. Forecasting

A 12-month forecast of Dengue cases for the year 2024 is presented using the fitted SARIMA model. The predicted values are displayed on their original scale, having been transformed back from the logarithmic scale using the exponential function, thus reversing the initial Box-Cox transformation. Figure 12 illustrates the forecast trajectory for each month, alongside the original observed series and

the corresponding 95% confidence intervals, enabling a clear assessment of the uncertainty associated with the predictions.

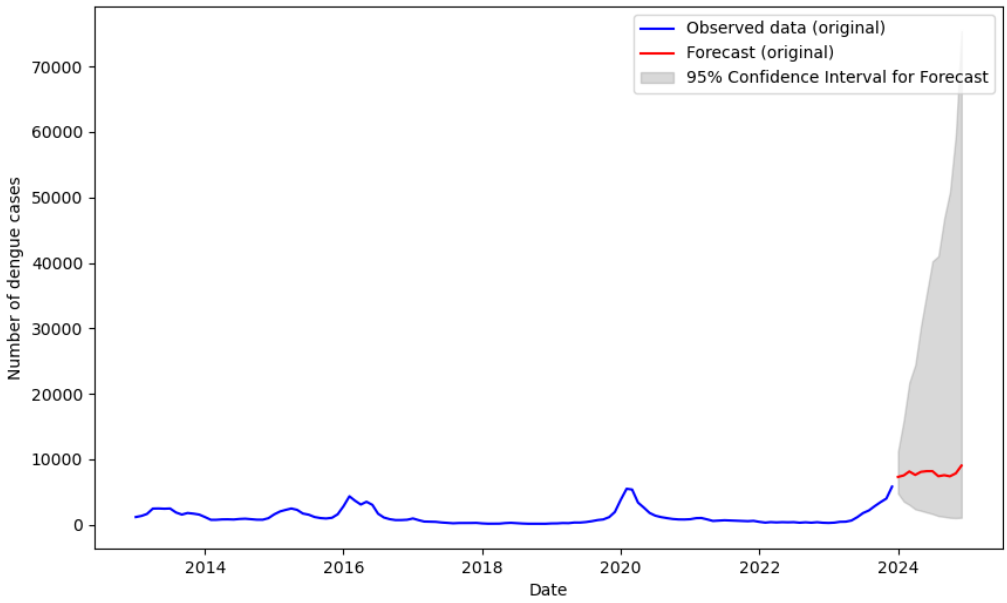

Figure 12. 12-month forecast generated by the SARIMA model.

From Figure 12, the following results can be interpreted:

- **Expected behavior:** The forecast begins in January 2024 with approximately 7303 estimated cases and shows an increasing trend throughout the year. By December 2024, the model predicts around 9039 cases, representing the highest projected value in the forecast horizon.
- **Confidence intervals (95%):** The confidence intervals start off moderate but widen significantly over time:
  - In January, the number of cases is expected to fall between 4783 and 11,151.
  - In December, the range broadens substantially, from 1084 to 75,402 cases, indicating a high level of uncertainty associated with long-term forecasts.

The SARIMA$(0, 1, 1)(0, 1, 1)_{[12]}$ model successfully captures the trend and seasonality in the series, demonstrating good performance for short-term (12-month) forecasting. The projections suggest a possible significant increase in Dengue cases during 2024, serving as an early warning for public health authorities. These estimates can be valuable tools for planning and designing preventive and control strategies.

Specifically, considering that by epidemiological week 13 of 2024, a total of 18,112 cases had been reported, and the model projects a 95% confidence interval ranging from 2375 to 24,412 cases for the month of April. This reflects a reasonable alignment with the observed data and the inherent uncertainty in the forecasting process. However, caution is advised when interpreting estimates for the final months of the forecast horizon, as uncertainty naturally increases with time.

5. Conclusions

In this study, we developed an age-structured stochastic model to elucidate the transmission dynamics of the Dengue virus between human and mosquito populations. Simulation results revealed distinct epidemic patterns, characterized by declining susceptible compartments (S_{h1} and S_{h2}) and pronounced infection peaks (I_{h1} and I_{h2}) across both age classes. Despite inherent stochastic variability, consistent temporal trends were observed, demonstrating the robustness of the model structure. The age-stratified analysis indicated that adult individuals experienced a higher infection burden and greater variability in epidemic trajectories, likely due to increased exposure associated with occupational, social, and healthcare-related activities. These findings highlight the importance of designing targeted intervention strategies that prioritize adult populations, including strengthened vector control measures, enhanced workplace prevention programs, and sustained public health campaigns promoting protective behaviors.

For predictive applications, the seasonal autoregressive integrated moving average (SARIMA) and stochastic differential equation (SDE) models demonstrated satisfactory performance. The SARIMA framework effectively captured the observed temporal regularities, particularly seasonal and cyclical patterns, making it well-suited for short-term forecasting and public health preparedness. In contrast, the SDE model provided a mechanistic interpretation of Dengue transmission dynamics across age classes and mosquito populations, offering deeper insights into system variability and the role of random perturbations. Together, these complementary approaches strengthen the capacity for Dengue surveillance, forecasting, and control planning by integrating statistical prediction with biologically grounded stochastic modeling.

This study has several limitations. Model calibration was based on aggregated surveillance data, and the lack of age-specific entomological information required simplifying assumptions. Human mobility and multiple Dengue serotypes were not explicitly represented, and stochastic perturbations were modeled as Gaussian noise, which may approximate only real environmental variability. Despite these limitations, the proposed stochastic age-structured SIR–SI framework provides a robust basis for incorporating additional heterogeneities and data sources in future work. In future research, we will extend this framework to examine how random perturbations affect disease fade-out probabilities and the emergence of quasi-stationary behavior in endemic regimes. In addition, using a comprehensive sensitivity analysis, we will assess the influence of key epidemiological parameters on temporal dynamics and the effective reproduction number. These developments will place our findings within the broader context of stochastic threshold theory and explicitly link the framework with the basic reproduction number formulation proposed by [5]. This integrated deterministic–stochastic perspective may contribute to early warning systems and adaptive Dengue control strategies, particularly in data-limited endemic settings.

Availability of data and material

Not applicable. All data generated or analyzed during this study are included in this manuscript. The data source can be found in the Dengue case records from the Public Health Surveillance System (SIVIGILA) [8].

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

V.A. was partially supported by VRI-Universidad Nacional de Colombia, HERMES projects No. 63574.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. OMS, *Dengue y Dengue grave*, *OMS*, 2024. Available from: https://www.who.int/es/news-room/fact-sheets/detail/Dengue-and-severe-Dengue
2. O. A. Montesinos-López, C. M. Hernández-Suárez, Modelos matemáticos para enfermedades infecciosas, *Salud Publica Mex.*, **49** (2007), 218–226.
3. Y. Guo, T. Li, Modeling the competitive transmission of the Omicron strain and Delta strain of COVID-19, *J. Math. Anal. Appl.*, **526** (2023), 127283. https://doi.org/10.1016/j.jmaa.2023.127283
4. A. Gray, D. Greenhalgh, L. Hu, X. Mao, J. Pan, A stochastic differential equation SIS epidemic model, *SIAM J. Appl. Math.*, **71** (2011), 876–902. https://doi.org/10.1137/10081856X
5. A. Ríos-Gutiérrez, S. Torres, V. Arunachalam, Studies on the basic reproduction number in stochastic epidemic models with random perturbations, *Adv. Differ. Equ.*, **2021** (2021), 288. https://doi.org/10.1186/s13662-021-03445-2
6. Y. Sabbar, A. Din, D. Kiouach, Influence of fractal–fractional differentiation and independent quadratic Lévy jumps on the dynamics of a general epidemic model with vaccination strategy, *Chaos Solitons Fractals*, **171** (2023), 113434. https://doi.org/10.1016/j.chaos.2023.113434
7. Y. Sabbar, Refining extinction criteria in a complex multi-stage epidemic system with non-Gaussian Lévy noise, *Commun. Nonlinear Sci. Numer. Simul.*, **149** (2025), 108911. https://doi.org/10.1016/j.cnsns.2025.108911
8. I. N. de Salud, *Sistema de Vigilancia en Salud Pública (SIVIGILA)*, 2025. Available from: https://portalsivigila.ins.gov.co/.
9. D. Peña, Análisis de series temporales, Alianza, 2005.
10. A. Din, T. Khan, Y. Li, H. Tahir, A. Khan, W. A. Khan, Mathematical analysis of Dengue stochastic epidemic model, *Results Phys.*, **20** (2021), 103719. https://doi.org/10.1016/j.rinp.2020.103719
11. M. Z. Ndii, Y. A. Adi, B. S. Djahi, Deterministic and stochastic Dengue epidemic model: Exploring the probability of extinction, *Barekeng: J. Ilmu Mat. Ter.*, **16** (2022), 583–596. https://doi.org/10.30598/barekengvol16iss2pp583-596

12. I. Karatzas, S. E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd edition, Springer, 1991. https://doi.org/10.1007/978-1-4612-0949-2
13. P. E. Kloeden, E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, 1992. https://doi.org/10.1007/978-3-662-12616-5
14. X. Mao, C. Yuan, J. Zou, Stochastic differential delay equations of population dynamics, *J. Math. Anal. Appl.*, **304** (2005), 296–320. https://doi.org/10.1016/j.jmaa.2004.09.027
15. G. Maruyama, Continuous Markov processes and stochastic equations, *Rend. Circ. Mat. Palermo*, **4** (1955), 48–90. https://doi.org/10.1007/BF02846028
16. DANE, *Estadísticas vitales: Nacimientos*, 2023. Available from: https://www.dane.gov.co/index.php/estadisticas-por-tema/salud
17. M. Andraud, N. Hens, C. Marais, P. Beutels, Dynamic epidemiological models for Dengue transmission: A systematic review of structural approaches, *PLoS ONE*, **7** (2012), e49085. https://doi.org/10.1371/journal.pone.0049085
18. M. K. Enduri, S. Jolad, Dynamics of Dengue disease with human and vector mobility, *Spatial Spatio-temporal Epidemiol.*, **25** (2018), 57–66. https://doi.org/10.1016/j.sste.2018.03.001
19. D. P. Lizarralde-Bejarano, S. Arboleda-Sánchez, M. E. Puerta-Yepes, Understanding epidemics from mathematical models: Details of the 2010 Dengue epidemic in Bello (Antioquia, Colombia), *Appl. Math. Modell.*, **43** (2017), 566–578. https://doi.org/10.1016/j.apm.2016.11.022
20. DANE, *Identidad gráfica del Censo Nacional de Población y Vivienda 2018*, 2019. Available from: https://sitios.dane.gov.co/cnpv/
21. M. A. Branch, T. F. Coleman, Y. Li, A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems, *SIAM J. Sci. Comput.*, **21** (1999), 1–23. https://doi.org/10.1137/S1064827595289108
22. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, et al., SciPy 1.0: Fundamental algorithms for scientific computing in Python, *Nat. Methods*, **17** (2020), 261–272. https://doi.org/10.1038/s41592-019-0686-2
23. P. J. Brockwell, R. A. Davis, *Time Series: Theory and Methods*, Springer, 1991. https://doi.org/10.1007/978-1-4419-0320-4
24. P. J. Brockwell, R. A. Davis, *Introduction to Time Series and Forecasting*, Springer, 2016. https://doi.org/10.1007/978-3-319-29854-2

AIMS Press