

Stochastic Models and Estimation of Undetected Infections in the Transmission of Zika Virus

Lillian Achola Oluoch · Florent Ouabo Kamkumo · Ralf Wunderlich

Version of November 5, 2025

Abstract Zika fever, a mosquito-borne viral disease with potential severe neurological complications and birth defects, remains a significant public health concern. The epidemiological models often oversimplify the dynamics of Zika transmission by assuming immediate detection of all infected cases. This study provides an enhanced SEIR (Susceptible-Exposed-Infectious-Recovered) model to incorporate partial information by distinguishing between detected and undetected Zika infections (also known as “dark figures”). By distinguishing the compartments, the model captures the complexities of disease spread by accounting for uncertainties about transmission and the number of undetected infections. This model implements the Kalman filter technique to estimate the hidden states from the observed states. Numerical simulations were performed to understand the dynamics of Zika transmission and real-world data was utilized for parameterization and validation of the model. The study aims to provide information on the impact of undetected Zika infections on disease spread within the population, which will contribute to evidence-based decision making in public health policy and practice.

Keywords Stochastic epidemic model · Diffusion approximation · Partial information, · Extended Kalman filter · Zika · Estimation of dark figures

Mathematics Subject Classification (2010) 92D30 · 92-10 · 60J60 · 60G35 · 62M20

Lillian Achola Oluoch

The Technical University of Kenya; Haile Selassie Avenue, P.O. Box 52428 - 00200, Nairobi, Kenya. E-mail: lillianoluoch@tukenya.ac.ke

Florent.OuaboKamkumo / Ralf Wunderlich

Brandenburg University of Technology Cottbus-Senftenberg, Institute of Mathematics, P.O. Box 101344, 03013 Cottbus, Germany; E-mail: Florent.OuaboKamkumo / ralf.wunderlich@b-tu.de

Contents

1	Introduction	3
2	Stochastic Epidemic Modeling Framework	7
3	Zika Models	10
3.1	Introduction	10
3.2	Compartmental Modeling of Zika Transmission	10
3.3	Simplified Zika Model	11
3.4	Base Zika Model	15
3.5	Extended Zika Model	18
4	Estimation of Unobservable States	20
4.1	Kalman Filtering for Conditionally Gaussian State-Space Models	21
4.2	Extended Kalman Filter	23
4.3	Approximation by Conditional Gaussian Sequences	24
5	Numerical Results	26
5.1	Model parameters	26
5.2	Initialization of the Extended Zika Model	27
5.3	Initialization of the Filter Estimate and Covariance Matrix	29
5.4	Impact of Initial Estimates	32
5.5	Impact of Cascade States	32
	Appendix	38
A	Notation	38
B	Coefficients of Discrete-time Recursions of the State Process	39
B.1	Simplified Zika Model	39
B.2	The Base Zika Model	40
B.3	The Extended Zika Model	43
C	Proof of Proposition 5.2	46
	References	47

1 Introduction

Understanding the dynamics of transmission of infectious diseases is fundamental to designing effective public health strategies. Stochastic models, which incorporate random variables to account for uncertainties, have proven essential for predicting the course of disease outbreaks. The Zika virus disease (ZIKV) is a vector-borne disease transmitted by *Aedes* mosquitoes that gained worldwide attention during the major outbreak in the Americas between 2015 and 2016. Although often asymptomatic or mildly symptomatic in many individuals, ZIKV has been linked to severe consequences, including congenital microcephaly and neurological disorders, making its control a priority in global health efforts [2, 4].

The complexity of modeling ZIKV dynamics increases due to the presence of undetected cases: individuals who remain unreported because they are asymptomatic or because their symptoms are mild and do not lead them to seek medical attention [5, 7]. These undetected cases are critical, as they introduce substantial uncertainty into the understanding of disease prevalence and transmission. This unreported component of an epidemic, often referred to as “dark figures”, can cause a significant underestimation of the true scope of an outbreak and lead to misinformed public health interventions.

Furthermore, the challenge of tracking and managing ZIKV transmission is amplified by the partial information available to researchers and health authorities. While some epidemiological data, such as the number of symptomatic reported cases or vector population changes, are readily observable, other key aspects of the epidemic remain hidden. This includes the number of asymptomatic individuals or the extent of underreporting among those with mild symptoms [11]. As a result, modeling ZIKV dynamics with only partial information becomes inherently uncertain, complicating the accurate estimation of vital epidemic parameters like the effective reproduction number and the total number of infections [12, 18]).

To address these uncertainties, it is crucial to adopt stochastic epidemic models that incorporate both the observable and unobservable elements of transmission. Models that integrate the effects of undetected cases allow for more reliable predictions and can inform more effective control strategies. This is particularly relevant for vector-borne diseases like ZIKV, where the transmission process is influenced not only by human hosts but also by environmental conditions and vector biology [19]. Accounting for hidden infections, whether asymptomatic or underreported symptomatic cases, can lead to better estimates of disease spread and improve response strategies. Such stochastic models can support decision making in epidemic control, such as evaluating and adjusting intervention strategies, such as vector control, based on updated data.

In this study, we develop a stochastic epidemic model for ZIKV that explicitly accounts for the partial information regarding undetected cases. These include both asymptomatic individuals and those with mild symptoms who do not report their illness. We explore how these unobserved cases affect the overall transmission dynamics and propose methods for improving the estimation of the true burden of the epidemic. By integrating stochastic control approaches, our model will allow for more effective public health interventions in the context of ZIKV and similar infectious diseases as in [39].

Literature Review on Zika Disease. Zika virus which causes Zika fever is a *Flaviviridae* virus predominantly transmitted by female mosquito bites from *Aedes aegypti* and

Aedes albopictus [44]. It poses a serious public health hazard due to its potential to cause severe birth abnormalities (such as microcephaly) and neurological disorders. Other arboviruses transmitted by mosquitoes include Chikungunya and Dengue. While these diseases are primarily transmitted by mosquitoes, research indicates that Zika, unlike other insect-transmitted infections, can also be transmitted through sexual contact, primarily from men to women [37].

In human-to-human transmission, Zika can spread before, during, and after symptoms appear. There is also a high probability of a mother transmitting the disease to her fetus during pregnancy, delivery, or by blood transfusion and breastfeeding [4, 26]. Asymptomatic carriers may also transmit the virus [11]. Studies reveal that the Zika virus can persist in semen for extended periods, possibly up to six months [38], meaning that even months after recovery, the disease might still be transmissible. The incubation period of the disease is approximately 3-14 days. Infected individuals typically experience no symptoms or moderate symptoms such as muscle and joint soreness, fever, rash, and headaches [53]. Due to these reasons, many infected persons may go undetected, posing a significant risk in transmission of the virus.

Zika virus continues to spread globally, posing a significant hazard to public health as there are currently no identified vaccinations or therapies for treatment and prevention [50]. Consequently, Zika fever is one of the priority diseases in the WHO's Blueprint for Research and Development [54], due to its likely persistence as a significant threat in the future. Therefore, understanding the mechanisms of Zika transmission and the effectiveness of intervention efforts is crucial in mitigating its impact. In many scenarios, a portion of infected individuals may go undetected through testing or surveillance systems, leading to undetected infections. This study thus proposes a classic epidemiological model that can be used to understand the spread of infectious diseases within a population, particularly considering undetected cases. Understanding the dynamics of undetected infections is crucial for effective disease control and public health interventions.

Numerous mathematical models have been developed to predict the spread of Zika virus disease [43, 5]. A stage-structured model was developed to examine the impact of sexual transmission [49]. Another compartmental model of Zika propagation, considering vector-borne and sexual transmission, was proposed by [24], which used an SEIR-type model for humans and an SEI-chain for vectors, distinguishing between asymptomatic and symptomatic infected individuals. Additionally, a compartmental model accounting for vector-borne and sexual transmission, differentiating between sexes and migration, was developed by [2]. Transmission of Zika fever using two vector management strategies, lowering mosquito biting rates and population size, was proposed by [51]. In [20], we developed a non-autonomous model that incorporated many important features of Zika transmission, including weather seasonality, sexual and vector-borne transmission, the prolonged period of infectiousness after recovery, and the role of asymptomatically infected humans. However, in the study, we did not address the fact that within the susceptible and exposed population, there may be those who were infected but undetected.

No Zika model has addressed the fact that not all state variables can be directly observed and others go undetected, even though the inclusion of non-detected compartments has been proposed [32]. In the study, they used the stochastic epidemic model to explore state variables that could not be directly observed and termed them as the "dark figures" problem. The application of the stochastic model in epidemics is not novel. Stochastic

epidemic modeling, with a focus on compartmental models, has been used to determine the distribution of ultimate epidemic size, the impact of various infectiousness patterns, and the quantification of stochastically maintained oscillations [25]. Hence, in the present work, based on our model in [20], we propose a stochastic epidemic model by incorporating partial information on Zika infections, distinguishing between detected and undetected cases.

Literature Review on General Filtering Methods. In the context of stochastic epidemic models with partial observability, filtering techniques serve as indispensable tools for inferring unobservable states and estimating unknown parameters. These models typically classify system components into *observable states*, which can be measured or reported, and *latent states*, which remain hidden from direct observation. A variety of filtering methods have been employed in the literature to recover these hidden dynamics and calibrate model parameters using noisy, incomplete, or indirect observations.

A significant challenge in epidemic modeling is the scarcity or complete absence of data regarding the true infection process. This problem is addressed using a Bayesian framework [41], by implementing Markov chain Monte Carlo (MCMC) techniques to jointly estimate missing data and infer model parameters. Their approach facilitates inference even when key epidemiological events, such as infection times, are unobserved.

Extending this methodology, [6] introduced a model in which the population structure is described by a random graph, resembling an SIR-type process. Their work integrates MCMC methods for simultaneous inference of disease transmission rates and characteristics of the social contact network, allowing a coherent estimation of both epidemic and structural parameters.

Further developments include state-space models incorporating modified SEIR dynamics. [9] proposed a Bayesian particle filtering algorithm for estimating the temporal evolution of both hidden states and unknown parameters from noisy daily infection data related to the COVID-19 pandemic. Their method enables robust estimation even in the presence of measurement noise and under-reporting.

Similarly, [34] employed an ensemble Kalman filter (EnKF) within a SIRD modeling framework to describe COVID-19 progression. The EnKF is particularly well-suited for high-dimensional or nonlinear systems, offering an efficient recursive estimation of both model parameters and unobserved epidemiological compartments.

A discrete-time stochastic SIR model with latent infectious states was considered by [17], where the transmission rate and the actual number of infectious individuals were assumed to be unobservable. Their approach utilizes a hidden Markov model (HMM) formulation, with nested particle filtering techniques employed to recover the hidden states and estimate the reproduction number and other model parameters.

To address the presence of non-Gaussian features such as skewness and outliers in epidemic data, [1] introduced a skew Kalman filter (SKF). Traditional Kalman filters often assume Gaussianity and can be sensitive to extreme values. The SKF generalizes the Kalman framework to accommodate asymmetric distributions, leading to more robust Bayesian inference, particularly for state estimation in the presence of irregular or noisy data.

In general, Bayesian filtering frameworks, including MCMC and sequential Monte Carlo (SMC) methods like particle filters, offer powerful probabilistic tools for estimation under uncertainty [10, 21, 48]. These methods approximate posterior distributions of hidden variables through sampling, making them flexible and suitable for a wide range of

nonlinear and non-Gaussian models. However, they often come with a computational cost: convergence can be slow and the algorithms may be computationally intensive, especially when applied to complex models or large-scale datasets.

Literature Review on Kalman Filtering in Epidemic Modeling. Increasing research has used Kalman filtering techniques to improve state space estimation in compartmental epidemic models. A stochastic SEIR (R) DSD model has been introduced by [59]. An Extended Kalman Filter (EKF) was employed to infer both model parameters and unobserved states, enhancing short-term forecasting accuracy. [30] follow a comparable route, employing an EKF within a classical SIRD model to dynamically track disease progression, enabling refined temporal parameter estimation.

On the other hand, [57] implemented a switching Kalman Filter based on linear Gaussian assumptions. However, such flexibility comes at the cost of increased computational demand and challenges in calibrating transition probabilities among regimes. In an alternative direction, [39] apply EKF within a stochastic optimal-control framework. Under partial information, their methodology addresses decision-making problems faced by policymakers aiming to contain an epidemic efficiently while balancing public health and economic costs. In a related study, [32] estimated the unobservable component from the observations using the extended Kalman filter approach to account for nonlinearity of the state dynamics.

Further expanding the filtering toolbox, [15] explore conditional nonlinear Gaussian systems (CNGS). These systems support fast joint estimation of both latent states and model parameters in nonlinear settings under partial observability. A major advantage lies in the availability of closed-form expressions for conditional distributions, enabling efficient data assimilation and uncertainty quantification. Building on this, [16] highlighted the suitability of conditional Gaussian frameworks for multiscale stochastic systems. Their work emphasizes the capacity of such models to capture non-Gaussian, non-linear dynamics characteristic of epidemics and other natural phenomena.

Several studies have successfully applied ensemble Kalman filters to dengue transmission models, which is another mosquito-based disease such as the Zika virus. For example, [58] utilized an SIR-type compartmental model coupled with the Ensemble Adjustment Kalman Filter (EAKF) to forecast the weekly incidence of dengue in Guangzhou, China, over the 2011-2017 period. Their model incorporated mosquito density and meteorological variables (such as temperature and rainfall) as covariates. The EAKF approach allowed for dynamic estimation of effective transmission rate, population susceptibility, and peak timing, generating accurate short-term outbreak predictions.

In a related effort, [56] developed a metapopulation network model, also driven by EAKF, to predict dengue spread in cities in Guangdong province. By assimilating case reports and climatic data across multiple nodes, the model improved forecasts of spatial and temporal disease dynamics, outperforming single-city models in epidemic peak prediction up to 10 weeks in advance. Furthermore, a comparative analysis by [31] evaluated both EnKF and EKF methods within a SIRS model framework. They reported high forecast accuracy, with the EKF slightly outperforming the EnKF in terms of computational efficiency and convergence speed.

Our Contribution. This study provides a novel contribution to the field of infectious disease modeling by introducing, for the first time, an extended Kalman filter framework for

tracking Zika virus transmission under partial observability. Unlike existing approaches, which predominantly rely on deterministic or simple stochastic compartmental models, the proposed methodology is rooted in stochastic epidemic models derived from large-population limits of continuous-time Markov chains. These models are specifically designed to incorporate unobserved epidemic states, enabling the systematic estimation of hidden infections. A central innovation is the formulation of a cascade of partially hidden compartments, where either the inflow or outflow is observed, but not both. This allows the model to leverage all available data without overestimating unobservable quantities. To capture nonlinear dynamics inherent in vector-borne disease transmission, we apply the extended Kalman filter (EKF), which facilitates real-time inference of latent states. Through comprehensive simulation studies based on a Zika transmission model, we demonstrate that although initial uncertainty in estimates is high, the EKF quickly assimilates information and accurately tracks the evolution of hidden states. These results highlight the potential of Kalman filtering techniques in improving public health decision making in the presence of incomplete data and under-reporting, particularly in diseases such as Zika, where traditional models fall short.

Notation. Let $x \in \mathbb{R}^d$ be a vector with components denoted by $x^1, \dots, x^d$, and let $\|x\|$ represent its Euclidean norm. For a matrix $A \in \mathbb{R}^{d \times d}$, the entry in the i -th row and j -th column is written as A^{ij} , and $\|A\|$ denotes the Frobenius norm. The $d \times d$ identity matrix is denoted by $\mathbb{I}_d$, and the zero vector in $\mathbb{R}^d$ is denoted by 0_d .

Paper organization. The rest of this paper is organized as follows. Section 2 presents the stochastic epidemic modeling framework underpinning the analysis. In Section 3, we develop the Zika virus transmission model, beginning with a compartmental representation, followed by a simplified version, a base model, and finally an extended model capturing more detailed dynamics. Section 4 addresses the estimation of unobservable states using Kalman filtering techniques for conditionally Gaussian state-space models. Section 5 discusses the numerical results, including parameter specification and initialization procedures for the extended model and Kalman filter. An appendix contains a list of notations and provides technical details and proofs that were removed from the main text.

2 Stochastic Epidemic Modeling Framework

Following the framework introduced in [33], we briefly recapitulate the core components of the stochastic compartmental model relevant to the current work. We consider a structured population model in which a closed population of constant size $N \in \mathbb{N}$ is subdivided into $d \in \mathbb{N}$ compartments, each representing a subpopulation with a distinct epidemiological status. The model incorporates $K \in \mathbb{N}$ distinct transition mechanisms between these compartments.

Let $T > 0$ denote a fixed time horizon. The stochastic processes describing the model are defined on a complete filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, where the filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in [0, T]}$ satisfies the usual conditions of right-continuity and completeness. Let $X = (X(t))_{t \in [0, T]}$ denote a $\mathbb{Z}^d$ -valued process, where each component $X^i(t) \in \{0, \dots, N\}$ represents the number of individuals in compartment $i \in \{1, \dots, d\}$ at time t . The filtration $\mathbb{F}$ is assumed to be generated by X , i.e., $\mathcal{F}_t = \sigma(X(s) : 0 \leq s \leq t)$.

Microscopic Dynamics. Let $\Theta_k(t)$ denote the cumulative number of transitions of type $k \in \{1, \dots, K\}$ that have occurred in the interval $[0, t]$. The evolution of the process $X(t)$ can then be expressed as

$$X(t) = X(0) + \sum_{k=1}^K \xi_k \Theta_k(t), \quad (2.1)$$

where each $\xi_k \in \mathbb{Z}^d$ is a transition vector encoding the change in state caused by a single transition of type k . Specifically, $\xi_k^i = +1$ if the transition causes an inflow into compartment i , $\xi_k^i = -1$ if it causes an outflow from compartment i , and $\xi_k^i = 0$ otherwise. This formulation presumes that at most one individual is transferred between compartments during a single transition event.

A standard assumption in many compartmental epidemic models is that the transitions are independent and memoryless. Accordingly, we model the transition counting processes $\Theta_1, \dots, \Theta_K$ as mutually independent, time-inhomogeneous Poisson processes with intensity functions $\lambda_k(t, X(t))$, which depend on the current time and system state. To express these processes in terms of homogeneous Poisson processes, let $\Pi_k(t)$ denote a standard unit-rate Poisson process for each $k = 1, \dots, K$. Then, define the stochastic time-change

$$\tau_k(t) := \int_0^t \lambda_k(s, X(s)) ds.$$

The non-homogeneous counting process can then be represented as $\Theta_k(t) = \Pi_k(\tau_k(t))$. Substituting into (2.1), the state dynamics of the system can be rewritten as

$$X(t) = X(0) + \sum_{k=1}^K \xi_k \Pi_k \left(\int_0^t \lambda_k(s, X(s)) ds \right), \quad X(0) = x_0. \quad (2.2)$$

This stochastic representation defines a continuous-time Markov chain (CTMC) over the discrete state space $\{0, \dots, N\}^d$, with time-dependent transition intensities governed by the system's current state.

Macroscopic Dynamics – Large Population Limits. For large population sizes N , the stochastic dynamics of the continuous-time Markov chain (CTMC) described in (2.2) can be approximated by a continuous-state diffusion process. This approach is based on the study of the asymptotic behavior of a properly rescaled CTMC for $N \rightarrow \infty$ and the application of a functional law of large numbers and central limit theorem, see Britton and Pardoux [7, Chapter 2, Section 2.2-2.3], Anderson and Kurtz [18, Chapter 1, Section 3.2], Ethier and Kurtz [22, Chapter 4, Section 7], Guy et al. [28], and [33, Section 3.4]. There it is shown that under suitable regularity conditions, the CTMC process converges weakly, as $N \rightarrow \infty$, to a continuous-state diffusion process $X^D = (X^D(t))_{t \in [0, T]}$ governed by the following system of stochastic differential equations (SDEs)

$$dX^D(t) = f_X(t, X^D(t)) dt + \sigma_X(t, X^D(t)) dW(t), \quad X^D(0) = x_0, \quad (2.3)$$

where $W(t)$ is a K -dimensional standard Brownian motion. The drift and diffusion terms are given by

$$f_X(t, x) = \sum_{k=1}^K \xi_k \lambda_k(t, x), \quad \text{and} \quad \sigma_X(t, x) = \left(\xi_1 \sqrt{\lambda_1(t, x)}, \dots, \xi_K \sqrt{\lambda_K(t, x)} \right).$$

This limiting system captures the macroscopic (aggregate) behavior of the epidemic under stochastic fluctuations and is commonly referred to as the diffusion approximation of the underlying CTMC.

In practical applications such as statistical inference or real-time filtering, it is often preferable to work with a time-discretized version of the dynamics, since observations (e.g., incidence data, case reports) are typically available at discrete time intervals (e.g., daily or weekly). To this end, we partition the time interval $[0, T]$ into $N_t \in \mathbb{N}$ equally spaced intervals of length $\Delta t = T/N_t$, and define the discrete time points $t_n = n\Delta t$ for $n = 0, \dots, N_t$. Applying the Euler–Maruyama discretization scheme to the diffusion system (2.3) yields the following stochastic recursion with the initial value $X_0^D = x_0$:

$$X_{n+1}^D = X_n^D + f_X(t_n, X_n^D) \Delta t + \sigma_X(t_n, X_n^D) \sqrt{\Delta t} \mathcal{E}_{n+1}, \quad (2.4)$$

where $X_n^D \approx X^D(t_n)$ denotes the discrete-time approximation of the state at time t_n , and $\{\mathcal{E}\}_{n=1}^{N_t}$ is a sequence of independent standard normally distributed vectors in $\mathbb{R}^K$, i.e., $\mathcal{E}_n \sim \mathcal{N}(0_K, \mathbb{I}_K)$.

This discrete-time representation will serve as the foundation for the development of inference procedures under partial observation in subsequent sections.

Models with Hidden States. In realistic Zika epidemic scenarios, not all components of the population can be directly observed. For instance, asymptomatic human infections and unmonitored mosquito infections are typically hidden from surveillance systems. To incorporate this partial observability, we partition the full epidemic state vector into hidden and observable components.

Suppose the total system is described by a d -dimensional state vector X_n^D at discrete time t_n , as in (2.4). We assume that $d_1 < d$ of these compartments are unobservable (latent, hidden), while the remaining $d_2 = d - d_1$ compartments are observable.

We decompose the state as

$$X_n^D = \begin{pmatrix} Y_n \\ Z_n \end{pmatrix},$$

where $Y_n \in \mathbb{R}^{d_1}$ contains the hidden states (e.g., asymptomatic human carriers), and $Z_n \in \mathbb{R}^{d_2}$ represents the observable states (e.g., confirmed symptomatic human infections).

The evolution of this system can be described by a coupled stochastic recursion with initial values $Y_0 = y_0, Z_0 = z_0$:

$$\begin{aligned} Y_{n+1} &= f(n, Y_n, Z_n) + \sigma(n, Y_n, Z_n) \mathcal{E}_{n+1}^1 + g(n, Y_n, Z_n) \mathcal{E}_{n+1}^2, \\ Z_{n+1} &= h(n, Y_n, Z_n) + \ell(n, Y_n, Z_n) \mathcal{E}_{n+1}^2, \end{aligned} \quad (2.5)$$

where $\mathcal{E}_n^1$ and $\mathcal{E}_n^2$ are independent sequences of standard normally distributed vectors:

$$\mathcal{E}_n^1 \sim \mathcal{N}(0_{k_1}, \mathbb{I}_{k_1}), \quad \mathcal{E}_n^2 \sim \mathcal{N}(0_{k_2}, \mathbb{I}_{k_2}),$$

with $k_1 + k_2 = K$, the total number of stochastic drivers in the system. The noise vector $\mathcal{E}_n^1$ captures randomness exclusively affecting hidden compartments, while $\mathcal{E}_n^2$ influences both hidden and observable states. The coefficient functions are defined on $\mathcal{G} = \{0, \dots, N_t\} \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ with $f: \mathcal{G} \rightarrow \mathbb{R}^{d_1}$, $h: \mathcal{G} \rightarrow \mathbb{R}^{d_2}$, $\sigma: \mathcal{G} \rightarrow \mathbb{R}^{d_1 \times k_1}$, $g: \mathcal{G} \rightarrow \mathbb{R}^{d_1 \times k_2}$, $\ell: \mathcal{G} \rightarrow \mathbb{R}^{d_2 \times k_2}$, and encode the drift and diffusion terms of the system. They will be specified in Appendix B based on the structure of the Zika transmission model presented in the next section.

3 Zika Models

3.1 Introduction

This section develops a class of stochastic models designed to describe the dynamics of Zika virus transmission. As a starting point, Subsection 3.2 discusses the development of a stochastic compartmental model that integrates both human and vector populations. The model extends classical epidemic formulations by explicitly accounting for asymptomatic and undocumented infections, waning immunity, and vector-host interactions, which are key features influencing the spread of ZIKV. The modelling assumptions are also captured in this subsection.

In Subsection 3.3, we introduce a foundational compartmental model capturing key features of Zika transmission dynamics. A notable challenge in Zika epidemiology is the high proportion of asymptomatic or mildly symptomatic cases, many of which remain undetected due to limited testing or clinical presentation. In addition, some people who experience symptoms may not seek a formal diagnosis or healthcare services, further contributing to under-reporting. This leads to a substantial discrepancy between reported and actual infection counts, which must be addressed within the model structure. The base model incorporates these hidden processes explicitly.

Among the hidden compartments are those representing individuals who have recovered from infection and are assumed to be temporarily immune. The transitions into these recovered compartments (e.g., from symptomatic infection) are often observable, as recovery is typically recorded. However, the transition out of these states, specifically, the gradual loss of immunity and return to susceptibility is typically unobserved in surveillance systems. For Zika virus, evidence suggests that recovery confers full immunity for a limited period (see, e.g., [27]), after which immunity wanes and susceptibility may return.

To incorporate this biological reality and leverage the available information on recovery timing, we refine the base model in Subsection 3.5. The extended model introduces cascades of recovered compartments, which stratify recovered individuals by the duration since recovery, often referred to as “recovery age”. These cascade compartments allow the model to track the gradual transition from full immunity to susceptibility in a structured manner, based on the time elapsed since infection resolution. By modeling this delay explicitly, the extended model supports more reliable estimation of underreported cases which are commonly referred to as “dark figures” by capturing both the observable inflow into recovery and the unobservable outflow back to susceptibility.

3.2 Compartmental Modeling of Zika Transmission

In this section, we introduce a stochastic compartmental model designed to capture the transmission dynamics of the ZIKV, with particular emphasis on unobserved epidemiological states. Motivation comes from two critical challenges associated with ZIKV surveillance and control. First, a substantial fraction of infections are asymptomatic or present with only mild symptoms, resulting in low test-seeking behavior and, consequently, under-reporting. Second, a subset of individuals can undergo an informal or unofficial diagnosis, for example, through self-testing or peer confirmation, without the cases being included in official health records. These factors contribute to significant uncertainty about the true burden of infection within the population.

Given the often benign course of ZIKV in symptomatic cases and the negligible associated mortality, a large pool of undetected cases can accumulate without triggering major public health responses. However, if this hidden infectious population grows unchecked, it can become a significant driver of transmission, undermining disease control efforts and possibly seeding new outbreaks. Therefore, accounting for these unobservable compartments is critical for effective modeling, forecasting, and policy formulation. Empirical studies have also shown that immunity to ZIKV is not necessarily life-long. Although initial infection induces a neutralizing antibody response, this immune protection tends to decrease over time, approximately within 24 months. [36, 27].

Model Structure. To model the above features, we start with a classic SIRS (Susceptible-Infected-Recovered-Susceptible) model, which is often used to study recurrent viral infections where immunity wanes over time. However, to accurately reflect the biological transmission mechanism of ZIKV, we extend the standard SIRS model to include the mosquito vector population to account for vector–host interactions. Furthermore, this model exclusively considers sexual transmission from men to women, as the number of transmissions from women to men is low [12, 19].

The human population is stratified into male and female subgroups, each subdivided into compartments reflecting disease progression and immunity status: susceptible, exposed, detected and undetected asymptomatic infected, symptomatic infected, and recovered. The mosquito vector population is modeled similarly using susceptible, exposed and infectious compartments, incorporating seasonality.

Model Assumptions. Our proposed Zika epidemic models operate under the following key assumptions with important practical justifications. The *human population size is held constant at N* , neglecting births and natural deaths - a reasonable simplification for short-term epidemic modeling where Zika-related mortality is negligible. The *mosquito birth rate varies seasonally* according to a time-dependent function $B_v(t)$ capturing realistic temporal fluctuations in vector abundance driven by climate and environmental conditions. Both the *human and vector populations are homogeneously mixed* within and between disease compartments, which means, in particular, that every susceptible individual has the same probability of encountering infectious individuals — a standard epidemiological assumption that simplifies the transmission dynamics. Finally, all model parameters are *time-homogeneous*, with the exception of mosquito birth rate which captures seasonal variation concentrated in vector population dynamics where it matters most. These assumptions create a computationally feasible framework suitable for epidemic forecasting and control optimization in resource-limited settings while capturing the essential biological features of Zika transmission.

3.3 Simplified Zika Model

Zika virus presents a unique epidemiological challenge due to its dual modes of transmission: vector and direct human-to-human transmission through sexual contact [35, 38]. Traditional vector host models are insufficient to capture the full scope of ZIKV dynamics, particularly in settings where sexual transmission plays a non-negligible role in sustaining outbreaks [55]. To address this, we construct a transmission framework that integrates vector-borne and sexual transmission pathways of the Zika virus. Here we start with a

simplified model capturing only the key properties. In subsequent subsections, we include a more refined compartmental structure and introduce explicit *exposed* compartments for male and female individuals and the vector to capture latency period between infection and infectiousness. Furthermore, in Section 3.5, we extend the model to include so-called cascade compartments in order to capture the transition from complete to waning immunity after recovery depending on the time elapsed since recovery.

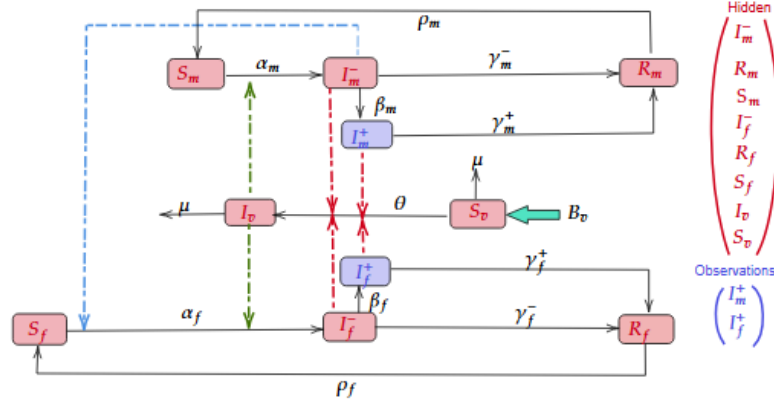


Fig. 3.1: Simplified Zika virus transmission dynamics considering gender (male human(m), female human(f)) and vector (v). The blue compartments indicate the observable states, while the red compartments are the unobserved states (dark figures). The green dashed lines indicate the vector-to-human transmission while red lines indicate the human-to-vector transmission and blue lines indicate the human-to-human transmission. A black solid arrow shows the transition from one compartment to the next.

The human compartments are subdivided into susceptible (S_m, S_f), infected but undetected (I_m^-, I_f^-), confirmed infected (I_m^+, I_f^+), and recovered classes, which are hidden (R_m, R_f) as illustrated in Figure 3.1. The mosquito population is modeled with susceptible (S_v) and infected (I_v) classes. The model tracks the transitions between these compartments driven by infection, recovery, sexual contact, and mosquito bites. Crucially, only a subset of infections are observed, reflecting the underreporting common in real-world surveillance data [42, 52]. This hybrid transmission structure and observability framework enable a more realistic representation of ZIKV transmission, supporting improved inference and evaluation of control strategies.

Susceptible males (S_m) and females (S_f) become infected by bites from infected mosquitoes (I_v). Using the placeholder notation $\dagger = m, f$ the infection rate is modeled by

$$\lambda_{\dagger}^{S \rightarrow I} = \lambda_{\dagger}^{S \rightarrow I}(S_{\dagger}, I_v) = \alpha_{\dagger} S_{\dagger} \frac{I_v}{\phi + I_v}, \quad (3.1)$$

with some positive constant parameters $\alpha_{\dagger}, \kappa_{\dagger}$. The reasoning behind this choice is as follows. In the CTMC modeling framework sketched in Section 2 the expected number of transitions from $S_{\dagger}$ to $I_{\dagger}$, i.e., the expected number of infections, within a small interval of time of length Δt is given by $\lambda_{\dagger}^{S \rightarrow I} \Delta t$. Denote by $A_{\dagger}$ the random event that a single susceptible person from $S_{\dagger}$ becomes infected within Δt time units, and its probability by $\mathbb{P}(A_{\dagger})$. Assuming a homogeneously mixed population and that bites and the subsequent development of infection among susceptible individuals are stochastically independent, we are within the framework of a Bernoulli scheme. Thus, the random number of infections

in Δt is binomially distributed with parameters $S_{\dagger}$ and $\mathbb{P}(A_{\dagger})$, and expectation $S_{\dagger}\mathbb{P}(A_{\dagger})$. Therefore, we find for the above considered expectation

$$\lambda_{\dagger}^{S \rightarrow I} \Delta t = S_{\dagger} \mathbb{P}(A_{\dagger}). \quad (3.2)$$

$\mathbb{P}(A_{\dagger})$ is proportional to Δt and the probability that a single susceptible is bitten by an infected mosquito. For the latter we use Holling's Type II function $\frac{I_v}{\phi + I_v}$, in which $\phi > 0$ is the so-called "half-saturation parameter". That means, for $I_v = \phi$ we have $\frac{I_v}{\phi + I_v} = \frac{1}{2}$. This function approaches 0 and 1 for $I_v \rightarrow 0$ and $I_v \rightarrow \infty$, respectively. For small numbers of infected mosquitoes I_v , it increases almost linearly with I_v , and for large I_v it captures saturation effects, since it tends to 1 for $I_v \rightarrow \infty$.

Summarizing, we have $\mathbb{P}(A) \sim \frac{I_v}{\phi + I_v} \Delta t$. Denoting the factor of proportionality by $\alpha_{\dagger}$ and substituting into (3.2) we obtain (3.1).

Following infection, individuals may remain undetected (I_m^-, I_f^-) , either due to an asymptomatic presentation or limited access to tests. Some asymptomatic infections are detected through surveillance or clinical diagnosis controlled by the parameters β_m, β_f and recorded as confirmed cases (I_m^+, I_f^+) . Recovery occurs in both states, such that both observed and hidden infected individuals move to the recovery compartment (R_m, R_f) , with transitions governed by rates $\gamma_m^-, \gamma_m^+, \gamma_f^-$, and γ_f^+ . Individuals who recover from infection (R_m, R_f) are assumed to acquire temporary immunity. However, over time, immunity can wane due to a natural decline in protective antibodies or lack of long-term immunity, thus returning individuals to the susceptible pool (S_m, S_f) . In the proposed model, the parameter ρ represents the rate at which individuals lose immunity and return to the susceptible class. This parameter applies separately to male (ρ_m) and female (ρ_f) populations, reflecting possible differences in immune waning dynamics between the sexes. A notable feature of this model is the inclusion of unidirectional sexual transmission from males to females. This mechanism reflects empirical findings that ZIKV can persist for a long time in semen [52, 38]. Vector dynamics involves recruitment of susceptible mosquitoes (S_v) at rate B_v which is time-dependent to capture seasonality effects. B_v is considered as the relative birth rate at which new mosquitoes are produced compared to their mortality rate.

Mosquitoes can be infected when biting infectious humans. The associated transmission rate is modeled by the usual contact term which is the product of S_v , the fraction of infectious humans and a scaling factor $\theta > 0$, and given by $\theta S_v (I_m^- + I_m^+ + I_f^- + I_f^+)/N$, where N is the total population size. Infected vectors (I_v) can then transmit the virus to new human hosts before dying at rate μ .

Table 3.1 outlines the structure and all possible transitions of the model. The system is composed of both observable and unobservable components, which are grouped into two state vectors. The vector of latent (hidden) states is given by $Y = (I_m^-, R_m, S_m, I_f^-, R_f, S_f, I_v, S_v)^{\top}$, while the vector of observable states is $Z = (I_m^+, I_f^+)^{\top}$. This results in a total of $d = 10$ distinct states and $K = 14$ possible transition pathways.

Within the hidden states, we have the *fully hidden* states $(S_m, S_f, I_v, \text{ and } S_v)$ which are not directly observable at either entry or exit. There are also the *partially hidden* states $(I_m^- \text{ and } I_f^-)$ where the inflow from S_m, S_f and the outflow to R_m, R_f cannot be tracked, but the outflow to I_m^+, I_f^+ is observed. R_m and R_f are also partially hidden, since the inflow from I_m^+, I_f^+ is observable whereas the outflow to S_m, S_f and the inflow from I_m^-, I_f^- cannot be observed.

The stochastic evolution of the system is described by a non-homogeneous continuous-time Markov chain (CTMC), whose transition directions and corresponding intensity func-

K	Transition	Transition vectors	Intensities
1	Infection to I_m^-	1, 0, -1, 0, 0, 0, 0, 0, 0, 0	$\alpha_m S_m \frac{I_v}{\phi + I_v} = \alpha_m Y^3 \frac{Y^7}{\phi + Y^7}$
2	Testing from I_m^-	-1, 0, 0, 0, 0, 0, 0, 0, 1, 0	$\beta_m^- I_m = \beta_m Y^1$
3	Recovery from I_m^-	-1, 1, 0, 0, 0, 0, 0, 0, 0, 0	$\gamma_m^- I_m = \gamma_m Y^1$
4	Recovery from I_m^+	0, 1, 0, 0, 0, 0, 0, 0, -1, 0	$\gamma_m^+ I_m = \gamma_m^+ Z^1$
5	Loss of immunity i R_m	0, -1, 1, 0, 0, 0, 0, 0, 0, 0	$\rho_m R_m = \rho_m Y^2$
6	Infection to I_f^-	0, 0, 0, 1, 0, -1, 0, 0, 0, 0	$(\alpha_f \frac{I_v}{\phi + I_v} + \omega \frac{I_m^- + I_m^+}{N_m}) S_f$ $= (\alpha_f \frac{Y^7}{\phi + Y^7} + \omega \frac{Y^1 + Z^1}{N_m}) Y^3$
7	Testing from I_f^-	0, 0, 0, -1, 0, 0, 0, 0, 0, 1	$\beta_f^- I_f = \beta_f Y^4$
8	Recovery from I_f^-	0, 0, 0, -1, 1, 0, 0, 0, 0, 0	$\gamma_f^- I_f = \gamma_f Y^4$
9	Recovery from I_f^+	0, 0, 0, 0, 1, 0, 0, 0, 0, -1	$\gamma_f^+ I_f = \gamma_f^+ Z^2$
10	Loss of immunity in R_f	0, 0, 0, 0, -1, 1, 0, 0, 0, 0	$\rho_m R_f = \rho_f Y^5$
11	Infection of vector	0, 0, 0, 0, 0, 0, 0, 1, -1, 0, 0	$\theta S_v (I_m^- + I_m^+ + I_f^- + I_f^+) / N$ $= \theta Y^8 (Y^1 + Z^1) + Y_f^4 + Z^2) / N$
12	Birth of vector	0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0	$B_v (I_v + S_v) = B_v (Y^9 + Y^{10})$
13	Death of susceptible vector	0, 0, 0, 0, 0, 0, 0, 0, -1, 0, 0	$\mu S_v = \mu Y^{10}$
14	Death of infected vector	0, 0, 0, 0, 0, 0, 0, 0, -1, 0, 0	$\mu I_v = \mu Y^9$

Table 3.1: Transition vectors and intensities of the simplified Zika model, with state processes $Y = (I_m^-, R_m, S_m, I_f^-, R_f, S_f, I_v, S_v)^\top$ and $Z = (I_m^+, I_f^+)^\top$; total number of compartments $d = 10$ and number of transitions, $K = 14$

tions are specified in Table 3.1. The CTMC is approximated by a diffusion process derived via a functional central limit theorem applied to the underlying Poisson processes driving the transitions. This results in a set of stochastic differential equations (SDE) system.

$$\begin{aligned}
dI_m^- &= \left(\alpha_m S_m \frac{I_v}{\phi + I_v} - \beta_m I_m^- - \gamma_m I_m^- \right) dt + \sqrt{\alpha_m S_m \frac{I_v}{\phi + I_v}} dW_1 - \sqrt{\beta_m I_m^-} dW_2 - \sqrt{\gamma_m I_m^-} dW_3, \\
dR_m &= (\gamma_m^- I_m^- + \gamma_m^+ I_m^+ - \rho_m R_m) dt + \sqrt{\gamma_m^- I_m^-} dW_3 + \sqrt{\gamma_m^+ I_m^+} dW_4 - \sqrt{\rho_m R_m} dW_5, \\
dS_m &= \left(-\alpha_m S_m \frac{I_v}{\phi + I_v} + \rho_m R_m \right) dt - \sqrt{\alpha_m S_m \frac{I_v}{\phi + I_v}} dW_1 + \sqrt{\rho_m R_m} dW_5, \\
dI_f^- &= \left(\alpha_f S_f \frac{I_v}{\phi + I_v} + \omega S_f I_m^- - \beta_f I_f^- - \gamma_f I_f^- \right) dt \\
&\quad + \sqrt{\alpha_f S_f \frac{I_v}{\phi + I_v}} dW_6 + \sqrt{\omega S_f I_m^-} dW_7 - \sqrt{\beta_f I_f^-} dW_8 - \sqrt{\gamma_f I_f^-} dW_9, \\
dR_f &= (\gamma_f^- I_f^- + \gamma_f^+ I_f^+ - \rho_f R_f) dt + \sqrt{\gamma_f^- I_f^-} dW_9 + \sqrt{\gamma_f^+ I_f^+} dW_{10} - \sqrt{\rho_f R_f} dW_{11}, \\
dS_f &= \left(-\alpha_f S_f \frac{I_v}{\phi + I_v} - \omega S_f I_m^- + \rho_f R_f \right) dt - \sqrt{\alpha_f S_f \frac{I_v}{\phi + I_v}} dW_6 - \sqrt{\omega S_f I_m^-} dW_7 + \sqrt{\rho_f R_f} dW_{11}, \\
dI_v &= \left(\theta S_v \frac{I_m^- + I_m^+ + I_f^- + I_f^+}{N} - \mu I_v \right) dt + \sqrt{\theta S_v \frac{I_m^- + I_m^+ + I_f^- + I_f^+}{N}} dW_{12} - \sqrt{\mu I_v} dW_{13}, \\
dS_v &= \left(B_v (S_v + I_v) - \theta S_v \frac{I_m^- + I_m^+ + I_f^- + I_f^+}{N} - \mu S_v \right) dt
\end{aligned}$$

$$\begin{aligned}
& + \sqrt{B_v} dW_{14} - \sqrt{\theta S_v \frac{I_m^- + I_m^+ + I_f^- + I_f^+}{N}} dW_{12} - \sqrt{\mu S_v} dW_{15}, \\
dI_m^+ &= (\beta_m I_m^- - \gamma_m^+ I_m^+) dt + \sqrt{\beta_m I_m^-} dW_2 - \sqrt{\gamma_m^+ I_m^+} dW_4, \\
dI_f^+ &= (\beta_f I_f^- - \gamma_f^+ I_f^+) dt + \sqrt{\beta_f I_f^-} dW_8 - \sqrt{\gamma_f^+ I_f^+} dW_{10}.
\end{aligned}$$

Subsequently, these diffusion approximations are discretized in time using the Euler–Maruyama scheme, as in Equations 2.4 leading to a discrete-time recursive formulation which governs the evolution of the state vectors Y and Z as described in Section 2. The functions and coefficients used in these formulations are provided in Appendix B.1.

3.4 Base Zika Model

In this section, we present an expanded version of the ZIKV transmission model. It includes a more refined compartmental structure to cater for the biological and epidemiological characteristics of the disease. In contrast to the simplified model in Section 2, this formulation introduces explicit *exposed* compartments for both male and female individuals (E_m, E_f) and the vector (E_v), capturing the latency period between infection and infectiousness. Additionally, the infected population is subdivided into symptomatic (I_m^s, I_f^s) and asymptomatic classes. The asymptomatic class is further stratified by whether or not the infection is detected (I_m^+, I_f^+) or not (I_m^-, I_f^-). This separation allows for a clearer distinction between observed and hidden dynamics. The recovery process is similarly disaggregated, with distinct compartments for individuals recovering from observed compartments (R_m^{2-}, R_f^{2-}) which are partially hidden) versus those recovering from hidden compartments (R_m^{1-}, R_f^{1-}). These additions enhance the model's ability to account for underreporting, sex-specific disease progression, and vector-host transmission dynamics.

In the base model, males and females are also represented in parallel structures, each beginning in the susceptible class (S_m, S_f). When an infectious mosquito comes into contact with susceptible human, the male and females, become exposed (E_m, E_f) at rates $\alpha_{\dagger} S_{\dagger} \frac{I_v}{\phi + I_v}$, $\dagger = m, f$. These exposed individuals at earlier stages are not yet infectious to both the human and vectors before viremia [13]. At later stages of exposure, shortly before the onset of the symptoms, the exposed individual becomes infectious to both human

and vector. The transition to the infected state after an intrinsic incubation period, ranges from 3 to 14 days, with an average estimate of approximately 7.5 days [35].

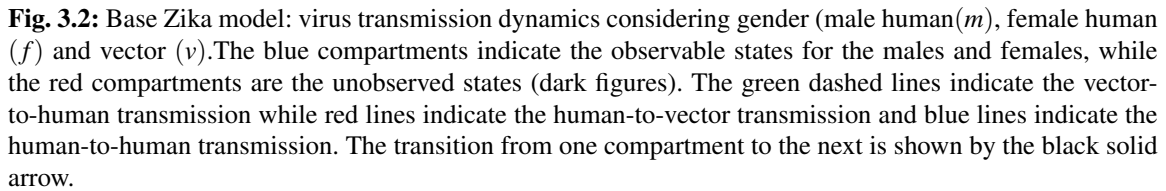


Table 3.3: Transition vectors and intensities of the base Zika model, with hidden state $Y = (I_m^-, E_m, R_m^{1-}, R_m^{2-} S_m, I_f^-, E_f, R_f^{1-}, R_f^{2-}, S_f, I_v, E_v, S_v)^\top$, and the observable state $Z = (I_m^+, I_m^s, I_f^+, I_f^s)^\top$, total number of compartments $d = 17$ and number of transitions $K = 20$.

As in the simplified model, there is also a possible transition from undetected I_m^- and I_f^- to detected I_m^+ and I_f^+ via testing rates β_m^+, β_f^+ . Cohort and surveillance studies indicate that Zika viremia is reliably detectable in serum shortly before or at the start of the symptom and is most reliably detected by Reverse Transcription Polymerase Chain Reaction (RT-PCR) tests [23]. For base model and later, the extended model, this dynamic is represented by the transition from E_m, E_f to I_m^+, I_f^+ via testing at the rate β_m^E, β_f^E . Recovery from infection also occurs through distinct pathways. Detected symptomatic or asymptomatic individuals transition into partially hidden recovery states R_m^{2-}, R_f^{2-} for men and women respectively. These compartments are considered partially hidden due to observed inflows into the compartment and unobserved outflows, which will be discussed in the next section on the extended Zika model. The undetected individuals enter fully hidden recovery compartments (R_m^{1-}, R_f^{1-}). These states reflect differing levels of immunity awareness and tracking. Individuals in all recovery compartments may eventually lose immunity and re-enter the susceptible state, due to waning immunity over time.

In the vector population, mosquitoes are classified into susceptible (S_v), exposed (E_v), and infectious (I_v) compartments. Susceptible mosquitoes become exposed after biting an infectious human in the compartments $I_{\dagger}^-, I_{\dagger}^+, I_{\dagger}^s$ or $E_{\dagger}$, and then progress to the infectious stage at the rate ν following an incubation period of approximately 7.5 days. It is worth noting that only a fraction ψ of the human population is infectious.

Figure 3.2 gives the structure of all possible transitions in the base model. For simplifying the notation we denote the sum of all infectious men and women by $\mathcal{I}_m$ and $\mathcal{I}_f$, respectively, and by $\mathcal{I} = \mathcal{I}_m + \mathcal{I}_f$ the total number of infectious in the human population. It holds $\mathcal{I}_{\dagger} = I_{\dagger}^- + I_{\dagger}^+ + I_{\dagger}^s + \psi E_{\dagger}$ for $\dagger = m, f$. Here, the scaling factor $\psi \in [0, 1]$ takes into account the fact that not all exposed individuals are infectious, but only those who are about to develop symptoms and transition to one of the infected compartments. Then the fraction $\mathcal{I}_m/N_m$ contributes to the rate of the transition $S_f \rightarrow E_f$ in the female population due to a contact with an infectious male, for which we have the rate $\omega S_f \mathcal{I}_m/N_m$. Note that the total male subpopulation size N_m and female population N_f are constant, since there is no birth and death in the human population.

We denote by β_m^+, β_f^+ and β_m^E, β_f^E the *testing rates*, which quantify the rates at which infected and exposed individuals respectively, are identified by diagnostic screening or clinical confirmation. This parameter governs the flow of individuals from undetected infection classes into observed compartments. The parameters β_m^s and β_f^s denote the *rates of symptomatic infection*, which describe the rate at which exposed individuals transition to the observed symptomatic infectious states. By contrast, β_m^-, β_f^- represents *rates of unobserved infection*, corresponding to infections that escape detection due to asymptomatic progression, limited health-seeking behavior, or surveillance inefficiencies. Recovery is likewise partitioned between observable and unobservable processes. The rates γ_m^s, γ_f^s and γ_m^+, γ_f^+ represent the *rates of observed recovery*, with γ_m^s, γ_f^s capturing recovery among symptomatic infections while γ_m^+, γ_f^+ representing recovery following confirmed positive tests. On the other hand, γ_m^-, γ_f^- corresponds to the *rates of unobserved recovery*, encompassing recoveries among individuals whose infections were never reported. The rates ρ_m^1, ρ_f^1 denotes the transition from the unobserved recovery classes R_m^{1-}, R_f^{1-} back to susceptibility, representing the waning immunity among those who recovered from undetected infections. Similarly, ρ_m^2, ρ_f^2 describe the *rate of waning immunity* from the partially

observed compartments R_m^{2-}, R_f^{2-} , which contains individuals who recovered from infections in the observed classes I^+ and I^s but their immunity eventually diminishes.

The recruitment of susceptible mosquitoes (S_v) is given by the rate B_v , which varies over time to account for seasonal fluctuations. The parameter B_v is the relative birth rate, at which new mosquitoes emerge in relation to their natural mortality rate. The rate for the transition $S_v \rightarrow E_v$ depends on the total proportion of infectious humans $\mathcal{I}/N$ and is given by $\theta S_v \mathcal{I}/N$ where θ is the rate of exposure of the vector. The exposed vector then becomes infectious and transition to I_v after approximately 7.5 days at the rate v .

Figure 3.2 is also composed of both observable and unobservable components. The vector of hidden states is given by $Y = (I_m^-, E_m, R_m^{1-}, R_m^{2-}, S_m, I_f^-, E_f, R_f^{1-}, R_f^{2-}, S_f, I_v, E_v, S_v)^\top$, while the vector of observable states is $Z = (I_m^+, I_m^s, I_f^+, I_f^s)^\top$. This results in a total of $d = 17$ distinct states and $K = 26$ possible transition pathways.

Summarizing, within the hidden states, we have the *fully hidden* states (S_m, S_f, I_v , and S_v) which are not directly observable at either inflow or outflow. There are also the *partially hidden* states (I_m^- and I_f^-) where the inflow is not visible, but the outflow is observed to the compartments I_m^+ and I_f^+ respectively. R_m and R_f also remains partially hidden with observed inflows and unobserved outflows.

As described earlier, Equations (2.1) and (2.2) are used to model the dynamics as a continuous-time Markov chain that is discretized as in (2.4). The specific transition directions and corresponding Poisson intensities are detailed in Table 3.3. As population sizes become large, the dynamics of CTMC can be approximated by a system of stochastic differential equations (SDEs), presented in Equation (2.3). For numerical analysis, the system is discretized, leading to the recursive formulation in Equation (2.5). The functions and coefficients used in these formulations (f, h, g, σ , and ℓ) are provided in Appendix B.2.

3.5 Extended Zika Model

Building upon the foundational structure of the base Zika transmission model developed in the preceding section, we introduce a refined and extended formulation that captures immunological memory through compartments organized by recovery age. Although the base model incorporated key epidemiological interactions between human (male and female) and mosquito populations, it assumed a homogeneous recovered class without accounting for the dynamics of periods during which an individual can have perfect immunity.

To incorporate the effect of temporary full immunity into the model and to improve the estimation of dark figures, we introduce the notion of *recovery age*, defined as the elapsed time since an individual has recovered from infection. Individuals are stratified into subcompartments according to their recovery age, so that each subcompartment contains only those with the same duration since recovery.

All previously defined variables and transitions retain their original meaning, and only the immunity cascade and its associated transitions represent new components in this extension. To incorporate information from observable random transitions into the estimation of unobserved (hidden) states, we adopt a cascade framework from [32, 33] and decompose compartments with observable inflows into a sequence of intermediate stages. The recovered population in R_m^{2-}, R_f^{2-} is stratified according to their *recovery age*, i.e., the time elapsed since recovery. For brevity, we use the placeholder notation $\dagger = m, f$ and simply write $R_{\dagger}^{2-}$.

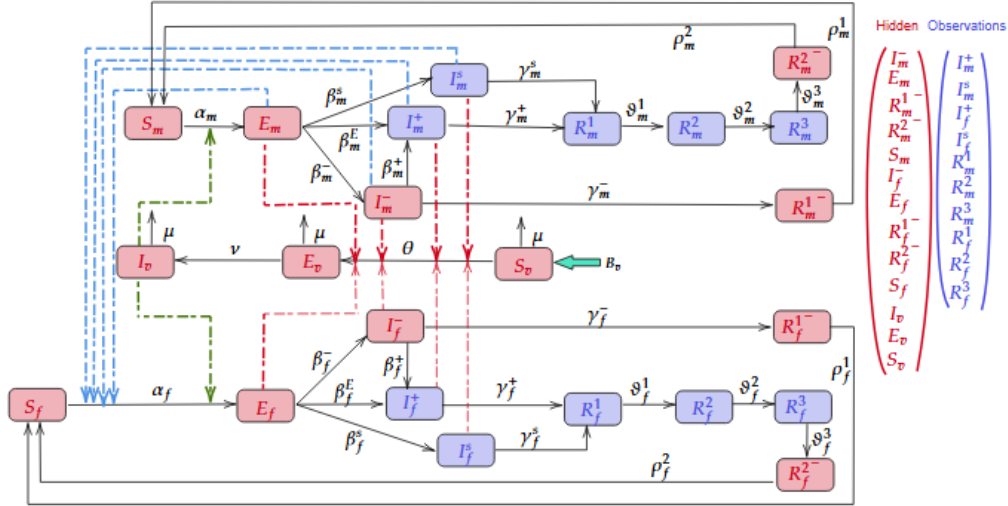


Fig. 3.3: Extended Zika virus transmission model highlighting the inclusion of $L = 3$ cascaded recovery compartments $R_m^1 \rightarrow R_m^2 \rightarrow R_m^3$ for males and $R_f^1 \rightarrow R_f^2 \rightarrow R_f^3$ for females. All these cascade compartments are included in the observed states and other compartmental transitions remain as defined in the base model.

Let $L \in \mathbb{N}$ denote the number of full-immunity periods of length Δt , the step size of the time discretization, such that complete immunity persists for the time $L\Delta t$ after recovery. The recovered class is then represented as a sequence of subcompartments $(R_{\dagger}^1, \dots, R_{\dagger}^L)$, followed by a residual class $R_{\dagger}^{2-}$, where $R_{\dagger}^j$ contains individuals with recovery age in $((j-1)\Delta t, j\Delta t]$. The last compartment $R_{\dagger}^{2-}$ includes those with a recovery age greater than $L\Delta t$ who have lost their full immunity and can transition to the susceptible compartments at a rate¹ of $\rho_{\dagger}^2$. Transitions within the cascade are deterministic, with $R_{\dagger,n+1}^j = R_{\dagger,n}^{j-1}$ for $j \geq 2$, while the first compartment receives a random inflow $R_{\dagger,n+1}^1 = R_{\dagger,n}^{\text{in}}$. Here, $R_{\dagger,n}^{\text{in}}$ counts newly recovered individuals who enter the first cascade compartment during the interval $(t_n, t_{n+1}]$, it is given by $R_{\dagger,n}^{\text{in}} = (\gamma_{\dagger}^s I_{\dagger,n}^s + \gamma_{\dagger}^+ I_{\dagger,n}^+) \Delta t$ and is observable since it depends on the observable states $I_{\dagger}^s, I_{\dagger}^+$. For the last compartment the dynamics reads $R_{\dagger,n+1}^{2-} = R_{\dagger,n}^{2-} + R_{\dagger,n}^{L,n} - R_{\dagger,n}^{\text{out}}$ with the outflow to $S_{\dagger}$ given by $R_{\dagger,n}^{\text{out}} = \rho_{\dagger}^2 R_{\dagger,n}^{2-} \Delta t$.

When L is large, this construction may introduce too many new compartments that add little statistical information because only $R_{\dagger}^1$ represents the new information from the observable inflow that was not yet been captured in the base model. To limit the number of compartments and reduce model complexity, we aggregate the L cascade stages into $d^R \leq L$ compartments $R_{\dagger}^1, \dots, R_{\dagger}^{d^R}$. Each $R_{\dagger}^j$ groups $P_j \in \mathbb{N}$ adjacent recovery-age compartments with $\sum_{j=1}^{d^R} P_j = L$, and represents recovery ages in $(\sum_{k < j} P_k, \sum_{k \leq j} P_k] \Delta t$. This reduction leads to some loss of resolution, but it is computationally more efficient and still captures the essential dynamics of immunity.

The transition between aggregated classes is approximated by assuming that, at each time step, a fraction $\vartheta_j = 1/P_j$ of the individuals in $R_{\dagger}^j$ progress to $R_{\dagger}^{j+1}$, which is an approximation that is accurate under an uniform distribution of recovery ages within $R_{\dagger}^j$.

¹ This rate is different to the one in the base model!

This leads to the following recursions for the new cascade states.

$$\begin{aligned} R_{\dagger,n+1}^1 &= (1 - \vartheta_1)R_{\dagger,n}^1 + R_{\dagger,n}^{\text{in}}, \\ R_{\dagger,n+1}^j &= (1 - \vartheta_j)R_{\dagger,n}^j + \vartheta_{j-1}R_{\dagger,n}^{j-1}, \quad \text{for } j = 2, \dots, d^R, \\ R_{\dagger,n+1}^{2-} &= R_{\dagger,n}^{2-} + \vartheta_{d^R}R_{\dagger,n}^{d^R} - R_{\dagger,n}^{\text{out}}. \end{aligned}$$

We refer to Appendix B.3 for a complete description of the extended Zika model dynamics.

4 Estimation of Unobservable States

Recall that in the context of Zika virus transmission modeling, not all compartments of the population are directly observable. For example, the number of asymptomatic infections or the vector infection level may be hidden from direct measurement. The goal of filtering is to estimate the hidden state variable Y_n at each discrete time point n , based on all available observations $Z_0, \dots, Z_n$ and prior knowledge of the initial hidden state distribution. These variables evolve according to the stochastic model defined in (2.5).

Formally, the filtering problem consists in computing the optimal mean-square estimate of Y_n given the filtration

$$\mathcal{F}_n^Z := \sigma\{Z_k : 0 \leq k \leq n\} \vee \mathcal{F}_0^I,$$

which captures all information available up to time n , including the observation history and prior information encoded in $\mathcal{F}_0^I$ such as estimates from past outbreaks, historical surveillance data, or expert knowledge. The optimal estimate is the conditional mean

$$M_n := \mathbb{E}[Y_n | \mathcal{F}_n^Z],$$

which minimizes the mean-square error $\mathbb{E}[\|Y_n - \hat{Y}_n\|^2]$ among all $\mathcal{F}_n^Z$ -measurable estimators $\hat{Y}_n \in L^2(\Omega, \mathcal{F}_n^Z)$. The uncertainty associated with the estimate M_n is quantified by the conditional covariance matrix

$$Q_n := \text{Var}(Y_n | \mathcal{F}_n^Z) = \mathbb{E}[(Y_n - M_n)(Y_n - M_n)^\top | \mathcal{F}_n^Z].$$

The recursive filtering process is initialized at time $n = 0$ by:

$$M_0 = m_0 := \mathbb{E}[Y_0 | \mathcal{F}_0^Z], \quad Q_0 = q_0 := \text{Var}(Y_0 | \mathcal{F}_0^Z),$$

A central challenge in the Zika model is the presence of nonlinear dynamics in both the hidden and observed components. The drift term $f(n, Y_n, Z_n)$ and the diffusion terms $\sigma(n, Y_n, Z_n), g(n, Y_n, Z_n), \ell(n, Y_n, Z_n)$ are nonlinear functions of the states. Consequently, classical Kalman filtering theory, which assumes linear-Gaussian dynamics, is not applicable in this setting.

To address this, appropriate filtering methods must be employed. In this work, we use the *Extended Kalman Filter* (EKF), which relies on local linearization of the nonlinear system around the current estimate. This approach provides a computationally efficient approximation to the optimal filter and is particularly suitable when the process noise remains approximately Gaussian.

The formulation and implementation of the EKF tailored to our epidemic model are developed in Subsections 4.2 and 4.3, following the general filtering framework introduced in the next subsection.

4.1 Kalman Filtering for Conditionally Gaussian State-Space Models

Consider a partially observed stochastic process $(\frac{Y_n}{Z_n})$ where $Y_n \in \mathbb{R}^{d_1}$ denotes the unobservable (hidden) state vector and $Z_n \in \mathbb{R}^{d_2}$ denotes the observed process, for discrete times $n = 0, \dots, N_t$, with $N_t \in \mathbb{N}$, $d_1, d_2 \in \mathbb{N}$. The evolution of this system is governed by the following stochastic recursions:

$$\begin{aligned} Y_{n+1} &= \tilde{f}_{n0}(\mathcal{Z}_n) + \tilde{f}_{n1}(\mathcal{Z}_n)Y_n + \tilde{\sigma}_n(\mathcal{Z}_n)\mathcal{E}_{n+1}^1 + \tilde{g}_n(\mathcal{Z}_n)\mathcal{E}_{n+1}^2, \\ Z_{n+1} &= \tilde{h}_{n0}(\mathcal{Z}_n) + \tilde{h}_{n1}(\mathcal{Z}_n)Y_n + \tilde{\ell}_n(\mathcal{Z}_n)\mathcal{E}_{n+1}^2, \end{aligned} \quad (4.1)$$

with given initial conditions $Y_0 \in \mathbb{R}^{d_1}$, $Z_0 \in \mathbb{R}^{d_2}$, and where $\mathcal{Z}_n := (Z_0, \dots, Z_n)$ denotes the trajectory of the observed process up to time n . The noise processes $\{\mathcal{E}_n^1\}_{n=1}^{N_t} \subset \mathbb{R}^{k_1}$ and $\{\mathcal{E}_n^2\}_{n=1}^{N_t} \subset \mathbb{R}^{k_2}$ are independent sequences of standard Gaussian random vectors, and are independent of the initial states Y_0, Z_0 . The coefficient functions $\tilde{f}_{n0}, \tilde{f}_{n1}, \tilde{\sigma}_n, \tilde{g}_n, \tilde{h}_{n0}, \tilde{h}_{n1}, \tilde{\ell}_n$ are measurable functions of the observation path $\mathcal{Z}_n$, ensuring that all terms in (4.1) are well-defined and dimensionally consistent.

We impose the following technical assumptions:

Assumption 4.1 *Let $\tilde{b}_n = \tilde{b}(\mathcal{Z}_n)$ denote any of the measurable coefficient functions $\tilde{f}_{n0}, \tilde{f}_{n1}, \tilde{\sigma}_n, \tilde{g}_n, \tilde{h}_{n0}, \tilde{h}_{n1}, \tilde{\ell}_n$. Then for each $n = 0, \dots, N_t - 1$, the following hold:*

(A1) **Square integrability:**

$$\mathbb{E} \left[\|\tilde{b}_n(\mathcal{Z}_n)\|^2 \right] < \infty.$$

(A2) **Uniform boundedness (almost surely):**

$$\|\tilde{f}_{n1}(\mathcal{Z}_n)\|, \|\tilde{h}_{n1}(\mathcal{Z}_n)\| \leq C < \infty \quad \mathbb{P}\text{-a.s.}$$

(A3) **Integrable initial state:**

$$\mathbb{E}[\|Y_0\|^2 + \|Z_0\|^2] < \infty.$$

(A4) **Gaussian prior:** *The conditional distribution of Y_0 given $\mathcal{F}_0^Z = \sigma(Z_0) \vee \mathcal{F}_0^I$ is Gaussian.*

Remark 4.2 In contrast to the nonlinear stochastic epidemic models such as (2.5), where the drift and diffusion terms are nonlinear functions of both the hidden and observed variables (e.g., due to nonlinear infection dynamics or seasonality in vector populations), the signal process in (4.1) evolves according to a *conditionally linear and Gaussian* structure. Specifically, the drift of Y_{n+1} is affine in the signal Y_n , and the diffusion coefficients $\tilde{\sigma}_n(\mathcal{Z}_n)$, $\tilde{g}_n(\mathcal{Z}_n)$, and $\tilde{\ell}_n(\mathcal{Z}_n)$ depend only on the observation history $\mathcal{Z}_n$, not on the current or past values of the hidden state Y_n . Such structure allows closed-form filtering recursions analogous to the classical Kalman filter, albeit in a non-autonomous setting where the coefficients vary with the observation history.

Given that only the observation sequence $\{Z_n\}$ is available, the filtering problem consists of computing, for each time $n = 0, \dots, N_t$, the conditional expectation $M_n := \mathbb{E}[Y_n | \mathcal{F}_n^Z]$, and the associated conditional covariance matrix $Q_n := \text{Var}(Y_n | \mathcal{F}_n^Z)$, which together

describe the *optimal mean-square estimate* and its uncertainty, based on the observation path $\mathcal{Z}_n$ and prior information $\mathcal{F}_0^I$.

Under Assumption 4.1, it is well known (cf. [47, Chapter 13]) that M_n is the unique mean-square optimal estimate of Y_n given $\mathcal{F}_n^Z$, and the total expected squared estimation error is given by $\text{tr}(\mathbb{E}[Q_n]) = \sum_{i=1}^{d_1} \mathbb{E}[(Y_n^i - M_n^i)^2]$.

While in general nonlinear filtering problems, recursive computation of (M_n, Q_n) requires approximations (e.g., particle filters, extended Kalman filters), the conditional Gaussian structure of (4.1) allows an *exact recursive solution* analogous to the classical Kalman filter. These recursions are derived below.

Theorem 4.3 (Liptser & Shiryaev (2001) [47, Theorem 13.3])

Let Assumption 4.1 hold. Consider the stochastic system defined by the recursive state-space model (4.1), where $\{Y_n\}$ is the hidden state sequence and $\{Z_n\}$ is the corresponding sequence of observations. Then, the joint process $\{(Y_n, Z_n)\}_{n=0}^{N_t}$ is conditionally Gaussian with respect to the filtration $\{\mathcal{F}_n^Z\}_{n \geq 0}$. That is, for every $n \in \{0, 1, \dots, N_t\}$, the conditional distribution of Y_n given $\mathcal{F}_n^Z$ is multivariate Gaussian.

Proof. A proof is provided in [47, Theorem 13.3] and also in [14]. The result follows by mathematical induction, relying on the recursive linear-Gaussian structure of the model (4.1). $\square$

The conditional Gaussianity established in Theorem 4.3 permits an explicit recursive computation of the conditional mean and conditional covariance of the hidden states $\{Y_n\}$, given observations $\{Z_n\}$. These recursive filtering equations, which generalize the Kalman filter to time-varying and nonlinear models, are stated next.

Theorem 4.4 (Liptser & Shiryaev (2001) [47, Theorem 13.4])

Let Assumption 4.1 be satisfied, and suppose the sequences $\{Y_n\}$ and $\{Z_n\}$ evolve according to the conditionally Gaussian model (4.1). Then the conditional distribution of Y_n given $\mathcal{F}_n^Z$ is the multivariate Gaussian distribution $\mathcal{N}(M_n, Q_n)$, with parameters computed recursively as follows:

$$M_{n+1} = \tilde{f}_{n0} + \tilde{f}_{n1}M_n + \left(\tilde{g}_n \tilde{\ell}_n^\top + \tilde{f}_{n1} Q_n \tilde{h}_{n1}^\top \right) \left(\tilde{\ell}_n \tilde{\ell}_n^\top + \tilde{h}_{n1} Q_n \tilde{h}_{n1}^\top \right)^+ \left(Z_{n+1} - \tilde{h}_{n0} - \tilde{h}_{n1} M_n \right), \quad (4.2)$$

$$Q_{n+1} = - \left(\tilde{g}_n \tilde{\ell}_n^\top + \tilde{f}_{n1} Q_n \tilde{h}_{n1}^\top \right) \left(\tilde{\ell}_n \tilde{\ell}_n^\top + \tilde{h}_{n1} Q_n \tilde{h}_{n1}^\top \right)^+ \left(\tilde{g}_n \tilde{\ell}_n^\top + \tilde{f}_{n1} Q_n \tilde{h}_{n1}^\top \right)^\top + \tilde{f}_{n1} Q_n \tilde{f}_{n1}^\top + \tilde{\sigma}_n \tilde{\sigma}_n^\top + \tilde{g}_n \tilde{g}_n^\top, \quad (4.3)$$

with initial conditions $M_0 = m_0$ and $Q_0 = q_0$, where all functions are evaluated at the observed path $\mathcal{Z}_n$.

Proof. See [47, Theorem 13.4] or [14] for a complete derivation. $\square$

Remark 4.5 (Use of the Moore–Penrose Pseudoinverse) The notation $[A]^+$ denotes the Moore–Penrose pseudoinverse of a matrix A . This generalized inverse ensures the recursive equations are well-defined even when A is singular or not full rank. It satisfies the following identities:

$$A[A]^+A = A, \quad [A]^+A[A]^+ = [A]^+, \quad (A[A]^+)^\top = A[A]^+, \quad ([A]^+A)^\top = [A]^+A.$$

The pseudoinverse provides the minimum-norm solution in least-squares problems and is particularly advantageous in state estimation for models with low-rank noise structure or ill-conditioned observation matrices. For example, in Zika transmission models where the observation matrix ℓ may be low-rank due to sparse or aggregated reporting, the pseudoinverse guarantees numerical stability of the filter.

Remark 4.6 (Online Covariance Updates in Nonlinear Systems) In contrast to the classical Kalman filter for linear time-invariant systems, where the covariance update equation (a Riccati equation) can be solved offline, the filtering equations (4.2)–(4.3) require real-time updates. This is because the model coefficients depend explicitly on the observation path $\mathcal{Z}_n$, and thus the conditional covariance Q_n is non-deterministic and observation-driven. This online updating is essential in real-time epidemic tracking of diseases such as Zika, where model parameters and surveillance data evolve dynamically over time.

4.2 Extended Kalman Filter

In the context of Zika virus transmission dynamics, the underlying epidemic processes are inherently nonlinear due to the complex interactions between human and mosquito populations, seasonally varying transmission rates, and the stochastic nature of disease spread. These nonlinearities manifest both in the deterministic drift and in the stochastic components of the system. Consequently, classical linear filtering techniques such as the standard Kalman filter are inadequate for accurately estimating the unobservable epidemic states.

To address this challenge, we employ the *Extended Kalman Filter (EKF)*, a recursive filtering technique tailored for nonlinear stochastic systems. The EKF generalizes the Kalman filter framework to accommodate nonlinear state dynamics by performing a first-order linearization of the system at each discrete time step. This approach allows for tractable, approximate inference of hidden states even when the underlying model deviates from linearity. Foundational references for this methodology are included in [8], and [3].

In the Zika epidemic setting, the drift function f captures the nonlinear transmission dynamics influenced by vector-host interactions, while the diffusion terms σ and ℓ may depend on Y_n to reflect state-dependent variability such as seasonally fluctuating mosquito populations or reporting delays. This model structure induces a filtering problem that is nonlinear both in the system dynamics and in the observation equation.

To implement the EKF, we proceed as follows: at each time step n , the drift function f is linearized via a first-order Taylor expansion about a reference point $\bar{Y}_n$, which is typically chosen to be the current state estimate $M_n = \mathbb{E}[Y_n | \mathcal{F}_n^Z]$. The linearized form of f is given by

$$f(Y_n) \approx f(n, \bar{y}, z) + \nabla_y f(n, \bar{y}, z)(y - \bar{y}),$$

where $\nabla_y f$ denotes the Jacobian matrix of f evaluated at $\bar{Y}_n$. Similarly, the diffusion matrices $\sigma(Y_n)$ and $\ell(Y_n)$ are approximated by their evaluations at $\bar{Y}_n$, i.e.,

$$\sigma(Y_n) \approx \sigma(\bar{Y}_n), \quad \ell(Y_n) \approx \ell(\bar{Y}_n),$$

as proposed in [46]. These approximations yield a locally linear Gaussian model, to which the Kalman filtering recursion for conditional Gaussian sequences (Theorem 4.4) can be applied.

4.3 Approximation by Conditional Gaussian Sequences

In this section, we construct an approximate formulation of the nonlinear state-observation system described by the recursions (2.5), with the aim of recasting it into a form suitable for applying Kalman filtering results for conditionally Gaussian models. Specifically, we seek to approximate the nonlinear dynamics of the state process (Y_n) and the observation process (Z_n) by a system of linear recursions of the form (4.1).

Our method is based on a local linearization of the drift function f with respect to the hidden signal Y_n , performed via a first-order Taylor expansion around a carefully chosen reference point $\bar{Y}_n$ at each discrete time step. In the Zika transmission models considered, the drift of the observation process is already linear with respect to Y_n , and thus does not require any further approximation. Accordingly, we restrict attention to models where the observation drift satisfies the affine structure:

$$h(n, y, z) = h_0(n, z) + h_1(n, z)y. \quad (4.4)$$

Denoting the reference point at time n by $\bar{Y}_n = \bar{y}$, the drift term $f(n, y, z)$ is approximated by its first-order Taylor expansion:

$$f(n, y, z) \approx f(n, \bar{y}, z) + \nabla_y f(n, \bar{y}, z)(y - \bar{y}), \quad (4.5)$$

where $\nabla_y f$ denotes the Jacobian matrix of f with respect to y . In addition, signal-dependent diffusion coefficients σ, g, ℓ are approximated by evaluating them at the reference point $\bar{Y}_n$, thus freezing their dependence on the unknown state.

This procedure leads to the following approximated state-observation system:

Lemma 4.7 *Let the system (2.5) satisfy the linear observation drift condition (4.4). Then the linearized approximation of the nonlinear dynamics, based on the expansion (4.5) and freezing of the diffusion coefficients at the reference point $\bar{Y}_n$, yields the following recursions:*

$$\begin{aligned} \tilde{Y}_{n+1} &= f_0(n, \bar{Y}_n, \tilde{Z}_n) + f_1(n, \bar{Y}_n, \tilde{Z}_n)\tilde{Y}_n + \sigma(n, \bar{Y}_n, \tilde{Z}_n)\mathcal{E}_{n+1}^1 + g(n, \bar{Y}_n, \tilde{Z}_n)\mathcal{E}_{n+1}^2, \\ \tilde{Z}_{n+1} &= h_0(n, \tilde{Z}_n) + h_1(n, \tilde{Z}_n)\tilde{Y}_n + \ell(n, \bar{Y}_n, \tilde{Z}_n)\mathcal{E}_{n+1}^2, \\ \tilde{Y}_0 &= y, \quad \tilde{Z}_0 = z. \end{aligned} \quad (4.6)$$

Here, the functions f_0 and f_1 are defined as:

$$f_0(n, \bar{y}, z) = f(n, \bar{y}, z) - \nabla_y f(n, \bar{y}, z)\bar{y}, \quad f_1(n, \bar{y}, z) = \nabla_y f(n, \bar{y}, z).$$

Proof. Substituting the linearized drift (4.5) and evaluating the diffusion coefficients at $\bar{Y}_n$, the dynamics of $\tilde{Y}_n$ become:

$$\tilde{Y}_{n+1} = f(n, \bar{Y}_n, \tilde{Z}_n) + \nabla_y f(n, \bar{Y}_n, \tilde{Z}_n)(\tilde{Y}_n - \bar{Y}_n) + \sigma(n, \bar{Y}_n, \tilde{Z}_n)\mathcal{E}_{n+1}^1 + g(n, \bar{Y}_n, \tilde{Z}_n)\mathcal{E}_{n+1}^2. \quad (4.7)$$

Similarly, the observation recursion becomes:

$$\tilde{Z}_{n+1} = h_0(n, \tilde{Z}_n) + h_1(n, \tilde{Z}_n)\tilde{Y}_n + \ell(n, \bar{Y}_n, \tilde{Z}_n)\mathcal{E}_{n+1}^2.$$

Rewriting (4.7) using the identities that define f_0 and f_1 produces the recursion (4.6). $\square$

The system (4.6) now takes the form of a linear state-space model with Gaussian innovations, conditional on the current approximating point $\bar{Y}_n$. As such, it is amenable to analysis via the Kalman filter for conditional Gaussian sequences, as described in Theorem 4.4.

In practice, the EKF method iteratively selects the reference point $\bar{Y}_n$ at each time step as the current estimated mean of the hidden state $\bar{Y}_n := \tilde{M}_n$. Moreover, the actual observed sequence (Z_n) is interpreted as being generated by the observation equation in (4.6), allowing the recursive computation of the filter approximations $(\tilde{M}_n), (\tilde{Q}_n)$. The full algorithmic implementation of this approximation procedure is given in Algorithm 4.1

Algorithm 4.1 EKF Algorithm

Input: $Z_0, \dots, Z_{N_t}$; model parameters, prior information $\mathcal{F}_0^I$

Output: Approximations $\tilde{M}_n, \tilde{Q}_n$ of $M_n := \mathbb{E}[Y_n | \mathcal{F}_n^Z]$ and $Q_n := \text{Var}(Y_n | \mathcal{F}_n^Z)$ for $n = 0, \dots, N_t$

Initialization : $n := 0, \quad \tilde{M}_0 := M_0 = \mathbb{E}[Y_0 | \mathcal{F}_0^Z], \quad \tilde{Q}_0 := Q_0 = \text{Var}(Y_0 | \mathcal{F}_0^Z)$

(i) State prediction

$$\tilde{M}_{n+1} = f_0 + f_1 \tilde{M}_n + (g\ell^\top + f_1 \tilde{Q}_n h_1^\top) [\ell\ell^\top + h_1 \tilde{Q}_n h_1^\top]^{-1} (\tilde{Z}_{n+1} - (h_0 + h_1 \tilde{M}_n))$$

(ii) Error measurement

$$\tilde{Q}_{n+1} = -(g\ell^\top + f_1 \tilde{Q}_n h_1^\top) [\ell\ell^\top + h_1 \tilde{Q}_n h_1^\top]^{-1} (g\ell^\top + f_1 \tilde{Q}_n h_1^\top)^\top + f_1 \tilde{Q}_n f_1^\top + \sigma\sigma^\top + gg^\top$$

All coefficient functions are evaluated at the point $(n, \tilde{M}_n, Z_n)$.

(iii) Repeat (i) and (ii) for the next time step until all samples are processed.

In the EKF approximation of the nonlinear filtering problem, the system coefficients $\tilde{f}_{n0}, \tilde{f}_{n1}, \tilde{g}_n, \tilde{\sigma}_n, \tilde{\ell}_n$ are defined as functions of both time and the observed data. At each time step n , they are evaluated using the current observation Z_n and the EKF-estimated signal mean $\tilde{M}_n$, according to the rule

$$\tilde{b}_n(\mathcal{Z}_n) = \tilde{b}_n((Z_0, \dots, Z_n)) = b(n, \tilde{M}_n, Z_n), \quad \text{for } \tilde{b}_n \in \{\tilde{f}_{n0}, \tilde{f}_{n1}, \tilde{g}_n, \tilde{\sigma}_n, \tilde{\ell}_n\}.$$

For the coefficients $\tilde{b}_n = \tilde{h}_{n0}, \tilde{h}_{n1}$ appearing in the observation equation, the dependence is restricted to the current observation:

$$\tilde{b}_n(\mathcal{Z}_n) = \tilde{b}_n((Z_0, \dots, Z_n)) = b(n, Z_n).$$

The EKF recursion is initialized at $n = 0$ with the conditional distribution of Y_0 , given the initial observation Z_0 and the prior information $\mathcal{F}_0^I$. This distribution is assumed to be Gaussian, $\mathcal{N}(M_0, Q_0)$, where M_0 and Q_0 represent the prior mean and covariance matrix, respectively. It is important to note that the EKF estimate $\tilde{M}_n$ is defined recursively and its computation requires access to the full observation path $\mathcal{Z}_n$.

Remark 4.8 Under suitable smoothness and regularity conditions on the model coefficients f and h , the EKF yields a valid first-order approximation of the true filter associated with the nonlinear system. For rigorous error analysis and convergence guarantees, we refer to the results in [45] and their extension in the continuous-time setting in [40].

5 Numerical Results

In this section, we present numerical experiments to illustrate the performance of the discrete-time stochastic Zika transmission model introduced earlier, particularly the version incorporating cascade compartments for temporary immunity and hidden states. The simulations are designed to demonstrate the impact of partial observability and immune waning on the system dynamics, as well as to assess the identifiability of key hidden compartments under noisy observations.

We recall that the extended Zika model consists of thirteen hidden compartments collected in the state vector

$$Y = (I_m^-, E_m, R_m^{1-}, R_m^{2-}, S_m, I_f^-, E_f, R_f^{1-}, R_f^{2-}, S_f, I_v, E_v, S_v)^\top,$$

and ten observable components forming the observation vector

$$Z = (I_m^+, I_m^s, I_f^+, I_f^s, R_m^1, R_m^2, R_m^3, R_f^1, R_f^2, R_f^3)^\top.$$

5.1 Model parameters

The reproductive dynamics of *Aedes aegypti* mosquitoes, the primary vectors for Zika virus transmission, are strongly shaped by environmental conditions, particularly rainfall and humidity. Rainfall creates standing water that serves as breeding sites, thereby increasing mosquito birth rates. To account for this ecological effect, we introduce a time-dependent mosquito birth rate $B_v(t)$ into the population dynamics.

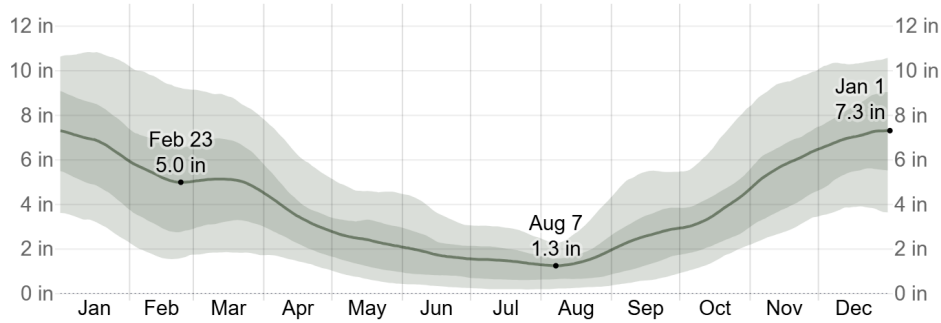


Fig. 5.1: Monthly rainfall pattern in Rio de Janeiro, Brazil, showing clear seasonality with peaks during the summer months (December–March) and lows in the winter (June–August)². The solid line represents the mean rainfall computed over a 31-day sliding window, while the shaded regions indicate empirical percentile bands: the darker band corresponds to the interquartile range (25th–75th percentiles), and the lighter band spans the 10th–90th percentiles.

Figure 5.1 illustrates the seasonal rainfall pattern in Rio de Janeiro, chosen here as a representative case study due to its tropical climate and its central role in the 2015–2016 Zika epidemic. The city exhibits strong seasonal fluctuations in temperature, humidity, and precipitation, which directly affect mosquito life cycles. Rainfall, in particular, plays a critical role by generating standing water that accelerates mosquito reproduction. Incorporating this rainfall-driven seasonality into the model is therefore essential to capture vector population dynamics and improve epidemic estimates under partial observations.

The rainfall profile peaks in January–February, averaging 7–8 inches per month, and reaches a minimum in July–August, with averages below 2 inches. Accordingly, we assume the mosquito birth rate $B_v(t)$ is strongly correlated with seasonal rainfall. Instead of using a purely parametric cosine-type function, we adopt a data-driven approach based on historical monthly rainfall records from Rio de Janeiro.² This provides a more realistic calibration of mosquito abundance.

To incorporate this information, we proceed as follows:

- Monthly average rainfall (in inches) is converted to millimeters (1 inch = 25.4 mm).
- The rainfall values are transformed into the interval $(a, \bar{B}_v) \subset (0, 1)$, with $a = 0.01$ as a positive lower bound and $\bar{B}_v = 0.1$ as the biologically (assumed) plausible upper bound for the mosquito birth rate. This ensures that the birth rate term $B_v(t)(S_v + E_v + I_v)$ remains stable and interpretable.
- The normalized monthly values are extended periodically and interpolated using a cubic spline to obtain a smooth daily function $B_v(t)$.

The normalization is given by

$$B_v(t_i) = a + (\bar{B}_v - a) \cdot \frac{R_i - \min(R)}{\max(R) - \min(R)},$$

where $R = \{R_1, R_2, \dots, R_{12}\}$ denotes the monthly average rainfall (mm), $a = 0.01$, and $\bar{B}_v = 0.1$. This rescaling ensures $B_v(t)$ remains within reasonable limits while reflecting seasonal variability.

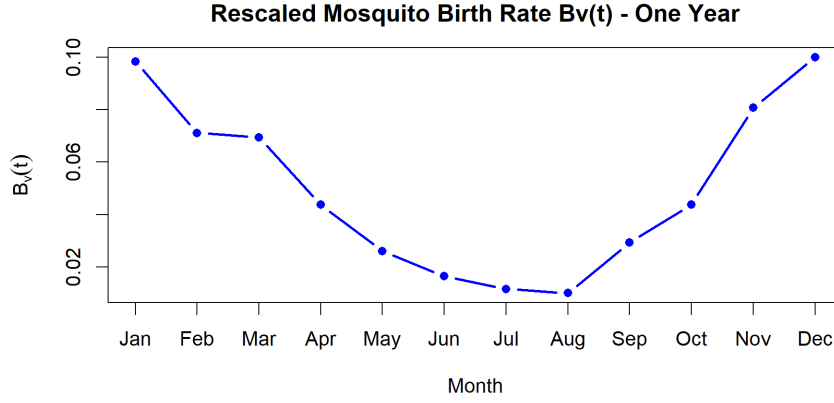
This calibrated function $B_v(t)$ is used in the numerical experiments that follow to simulate mosquito population dynamics and their effect on Zika transmission under partial information. Parameter values and their epidemiological interpretations are summarized in Table 5.1.

5.2 Initialization of the Extended Zika Model

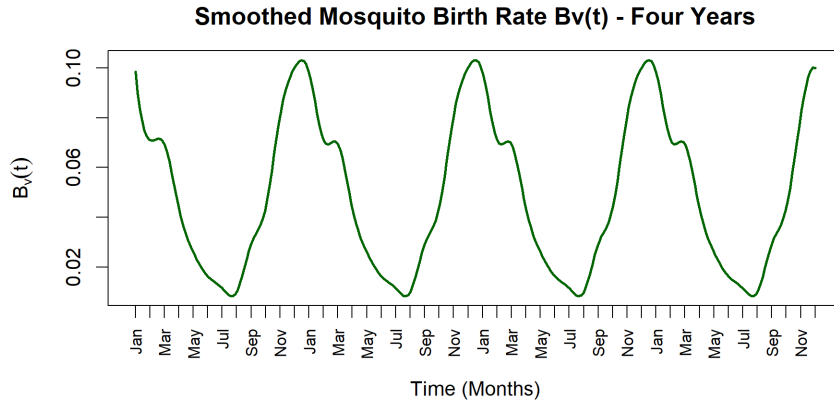
We present a strategy for initializing the extended Zika epidemic model under partial observation, tailored to the epidemiological context of the 2015–2016 Brazilian outbreak. The model distinguishes between a hidden state vector $Y = (Y^1, Y^2, \dots, Y^{13})^\top$, comprising unobserved compartments such as undetected infections, exposures, undetected recoveries, and susceptibles across male, female, and mosquito populations, and an observable vector $Z = (Z^1, Z^2, \dots, Z^{10})^\top$, which includes reported symptomatic and asymptomatic infections as well as confirmed recoveries.

A major challenge in modeling Zika dynamics is the substantial underreporting of cases, largely due to mild or asymptomatic infections and limited diagnostic capacity. To account for this, we introduce dark figure coefficients (DFCs), which constructs the initial filter estimates M_0, Q_0 . These coefficients are informed by epidemiological evidence and allow us to incorporate prior knowledge about the hidden epidemic burden into the initialization. The true initial values of the hidden state Y_0 are drawn from the Gaussian distribution $\mathcal{N}(M_0, Q_0)$.

² <https://weatherspark.com/y/30563/Average-Weather-in-Rio-de-Janeiro-Brazil-Year-Round>



(a) Transformed mosquito birth rate $B_v(t)$ based on monthly rainfall in Rio de Janeiro over one year.



(b) Smoothed mosquito birth rate $B_v(t)$ extended over four years using periodic spline interpolation.

Fig. 5.2: Calibration of the mosquito birth rate $B_v(t)$ from transformed rainfall data in Rio de Janeiro. The function reflects seasonal variability in mosquito reproduction driven by environmental conditions.

For each detected infected male (I_m^+), we assume approximately four undetected infections (I_m^-) and five exposures (E_m). Similarly, for each observed infected female (I_f^+), we assume three undetected infections (I_f^-) and four exposures (E_f). For recoveries, each observed case in stage one or two is associated with two undetected recoveries in the same stage. These assumptions provide a structured and epidemiologically plausible initialization of the hidden compartments.

Applying the dark figure coefficients to the initial observations provides consistent estimates for the hidden compartments, as summarized in Table 5.2. These estimates define the initial values of the hidden state vector Y at $t = 0$ and serve as the prior for simulating the Zika epidemic dynamics. The resulting trajectories are then used to assess the performance of the proposed filtering method.

This Table (5.2) reflects the initialization of the hidden state vector Y_0 used to simulate epidemic progression in a population. The dark-figure coefficient for compartments such as I_m^- , E_m , I_f^- , and E_f account for unreported cases. Recovered individuals are similarly adjusted to reflect potential underreporting. The susceptible compartments are calculated by removing all known and estimated compartments from the total. The vector compart-

Parameter	Description	Value
α_m	Male infection rate	1.5
α_f	Female infection rate	2.5
$\beta_m^-, \beta_m^+, \beta_m^s$	Transition from exposed to I^- , I^+ , I^s (male)	0.005, 0.005, 0.004
$\beta_f^-, \beta_f^+, \beta_f^s$	Transition from exposed to I^- , I^+ , I^s (female)	0.005, 0.005, 0.004
γ_m^-, γ_f^-	Recovery from I^-	0.1
γ_m^+, γ_f^+	Recovery from I^+	0.1
γ_m^s, γ_f^s	Recovery from symptomatic I^s	0.1
$\rho_m^1, \rho_m^2, \rho_f^3, \rho_f^4$	Immunity loss (recovered)	1/760, 1/30, 1/760, 1/30
ϕ	half-saturation parameter	200,000
θ	Exposure rate of susceptible vectors	0.02
ν	Infection rate of exposed vectors	0.10
μ	Vector death rate	0.05
$B_v(t)$	Vector birth rate (seasonal)	calibrated
ω	Male-to-female transmission probability	0.05
$N_m, N_f, N_{v,0}$	Population of males, females, and vectors	10,000, 10,000, 100,000
$\vartheta_m^{1,2,3}, \vartheta_f^{1,2,3}$	Transition between cascade states	3/730
$M_{U_{\dagger}}^I, M_{U_{\dagger}}^E, M_{U_{\dagger}}^R$	DFC Mean of I, E, R	7, 10, 7
$Q_{U_{\dagger}}^I, Q_{U_{\dagger}}^E, Q_{U_{\dagger}}^R$	DFC Variance of I, E, R	1, 1, 1
M_v^S, M_v^E, M_v^I	Mean of $S_{v,0}, E_{v,0}, I_{v,0}$	89,000, 6,000, 5,000
Q_v^S, Q_v^E, Q_v^I	Variance of $S_{v,0}, E_{v,0}, I_{v,0}$	9,500 ² , 800 ² , 700 ²

Table 5.1: Model Parameters for Zika Transmission Dynamics

ments (I_v, E_v, S_v) are set using biologically plausible proportions to maintain realistic initial transmission pressure.

5.3 Initialization of the Filter Estimate and Covariance Matrix

Computing the filter estimates requires the initialization of the filter processes $\tilde{M}, \tilde{Q}$ at time $n = 0$. In the following, we present for the extended stochastic Zika model with partial observations how $\tilde{M}_0$ and $\tilde{Q}_0$ can be constructed from the prior information contained in $\mathcal{F}_0^I$ and the first observation Z_0 . Analogously to the Covid-19 case described in [33] Section 2.3, the initialization is not yet affected by linearization errors that occur for $n > 0$ in the EKF approximation of the “true” filter processes M_n, Q_n . Therefore, we omit the tilde and simply write M_0, Q_0 in the sequel.

Dark Figure Coefficients (DFC). We employ DFC to integrate expert knowledge on the ratio of undetected to detected cases into the initialization. Using the placeholder notation $\dagger = m, f$ for male and female compartments, we define

$$U_{n,\dagger}^I = \frac{I_{\dagger,n}^-}{I_{\dagger,n}^+ + I_{\dagger,n}^s}, \quad U_{n,\dagger}^E = \frac{E_{\dagger,n}}{I_{\dagger,n}^+ + I_{\dagger,n}^s}, \quad U_{\dagger,n}^R = \frac{R_{\dagger,n}^{1-}}{R_{\dagger,n}^1 + R_{\dagger,n}^2 + R_{\dagger,n}^3}$$

so that $U_{\dagger,n}^I$ denotes the number of undetected infected per detected or symptomatic infected in the male ($\dagger = m$) and female ($\dagger = f$) population. Similarly, $U_{\dagger,n}^E$ represents the

Hidden States				
Index	State Variable	Initial value	$M_0^{(i)}$ (Initial Mean)	$(Q_0^{ii})^2$ (Initial Variance)
1	I_m^-	80	210	30^2
2	E_m	100	300	30^2
3	R_m^{1-}	36	126	18^2
4	R_m^{2-}	32	0	0
5	S_m	9,752	9,364	$30^2 + 18^2$
6	I_f^-	75	259	37^2
7	E_f	100	370	30^2
8	R_f^{1-}	36	126	18^2
9	R_f^{2-}	32	0	0
10	S_f	9,757	9,245	$37^2 + 17^2$
11	I_v	2,000	5,000	700^2
12	E_v	3,000	6,000	800^2
13	S_v	95,000	89,000	9500^2

Observable States					
i	Variable	Initial value	i	Variable	Initial value
	Z^i	Z_0^i		Z^i	Z_0^i
1	I_m^+	10	6	R_m^2	0
2	I_m^s	20	7	R_m^3	0
3	I_f^+	12	8	R_f^1	18
4	I_f^s	25	9	R_f^2	0
5	R_m^1	18	10	R_f^3	0

Table 5.2: Initial conditional expectations and variances for hidden state variables, consistent with Proposition 5.2, used to construct M_0 and the diagonal blocks of Q_0 . Top: Hidden states $Y_1, \dots, Y_{13}$ with their initial estimates used for filtering. Bottom: Observable states $Z_1, \dots, Z_4$

number of exposed per detected or symptomatic infected. The DFC $U_{\dagger,n}^R$ relates the recovered individuals with waning immunity to the observable recovered in the cascade states.

Throughout the following, the symbol # is used as a generic placeholder that may refer to any of the compartments I , E , or R , respectively.

Assumption 5.1 *At initial time $n = 0$ we assume:*

1. *Given the prior information $\mathcal{F}_0^I$, the initial DFCs $U_{0,\dagger}^\#$ are conditionally independent and Gaussian-distributed:*

$$U_{0,\dagger}^\# \sim \mathcal{N}(M_{U_\dagger}^\#, (Q_{U_\dagger}^\#)^2).$$

2. *$R_{\dagger,0}^{2-}$ start with zero initial values, i.e. $R_{\dagger,0}^{2-} = 0$.*

3. *The susceptible compartments $S_\dagger$ are determined by normalization within each subpopulation:*

$$S_{\dagger,0} = N_\dagger - (I_{\dagger,0}^- + E_{\dagger,0} + R_{\dagger,0}^{1-} + R_{\dagger,0}^{2-} + I_{\dagger,0}^+ + I_{\dagger,0}^s).$$

4. *Given the prior information $\mathcal{F}_0^I$, the initial vector compartments $(I_{v,0}, E_{v,0}, S_{v,0})$ are assumed to be conditionally independent, independent of Z_0 and Gaussian:*

$$I_{v,0} \sim \mathcal{N}(M_v^I, (Q_v^I)^2), \quad E_{v,0} \sim \mathcal{N}(M_v^E, (Q_v^E)^2), \quad S_{v,0} \sim \mathcal{N}(M_v^S, (Q_v^S)^2).$$

The second assumption, $R_{\dagger,0}^{2-} = 0$, is consistent with the epidemiological interpretation of these states. It only considers individuals with waning immunity after they have passed through the cascade states, in which complete immunity is assumed.

Based on this assumption we derive the following initial filter estimates M_0, Q_0 .

Proposition 5.2 *For the extended Zika model and under Assumption 5.1, the conditional distribution of the the initial state $Y_0 = (Y_0^1, \dots, Y_0^{13})$ given $\mathcal{F}_0^Z$, i.e., the prior information $\mathcal{F}_0^I$ and the initial observation Z_0 is Gaussian. The mean and the covariance matrix of this distribution are given by*

$$M_0 = \mathbb{E}[Y_0 | \mathcal{F}_0^Z] = M_0 = \begin{pmatrix} M_0^m \\ M_0^f \\ M_0^v \end{pmatrix}, \quad Q_0 = \text{Cov}(Y_0 | \mathcal{F}_0^Z) = \begin{pmatrix} \boxed{Q_0^m} & \mathbf{0}_{5 \times 5} & \mathbf{0}_{5 \times 3} \\ \mathbf{0}_{5 \times 5} & \boxed{Q_0^f} & \mathbf{0}_{5 \times 3} \\ \mathbf{0}_{3 \times 5} & \mathbf{0}_{3 \times 5} & \boxed{Q_0^v} \end{pmatrix}.$$

Here,

$$M_0^\dagger = \left(M_{U\dagger}^I A_\dagger, M_{U\dagger}^E A_\dagger, M_{U\dagger}^R B_\dagger, 0, N_\dagger - (M_{U\dagger}^I A_\dagger + M_{U\dagger}^E A_\dagger + M_{U\dagger}^R B_\dagger + A_\dagger) \right)^\top, M_0^v = (M_v^I, M_v^E, M_v^S)^\top,$$

$$Q_0^\dagger = \begin{pmatrix} a_I & 0 & 0 & 0 & -a_I \\ 0 & a_E & 0 & 0 & -a_E \\ 0 & 0 & b_R & 0 & -b_R \\ 0 & 0 & 0 & 0 & 0 \\ -a_I & -a_E & -b_R & 0 & a_I + a_E + b_R \end{pmatrix}, \quad Q_0^v = \begin{pmatrix} (Q_v^I)^2 & 0 & 0 \\ 0 & (Q_v^E)^2 & 0 \\ 0 & 0 & (Q_v^S)^2 \end{pmatrix},$$

$$A_\dagger := I_{\dagger,0}^+ + I_{\dagger,0}^s, \quad B_\dagger := R_{\dagger,0}^1 + R_{\dagger,0}^2 + R_{\dagger,0}^3, \quad C_\dagger := N_\dagger - (E_{\dagger,0} + I_{\dagger,0}^+ + I_{\dagger,0}^s)$$

$$a_I := A_\dagger^2 (Q_{U_\dagger}^I)^2, \quad a_E := A_\dagger^2 (Q_{U_\dagger}^E)^2, \quad b_R := B_\dagger^2 (Q_{U_\dagger}^R)^2.$$

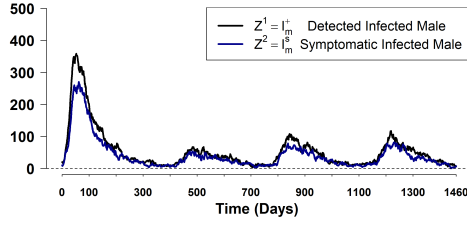
Proof. The result follows by writing each random component as an affine transformation of independent Gaussian variables and applying the standard properties of expectation, variance and covariance. A detailed derivation of the conditional mean and covariance is provided in Appendix C.

Initial Values for the Filter Estimate

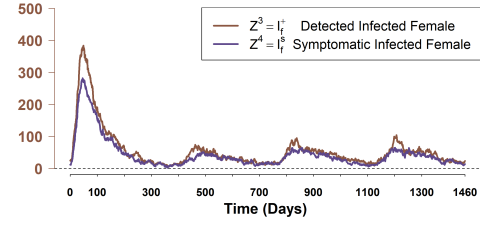
The construction of the initial mean vector M_0 and the associated conditional covariance matrix Q_0 follows Proposition 5.2.

The resulting conditional means and variances for each hidden state variable are summarized in Table 5.2. The entries are computed directly from the explicit formulas for M_0^m, M_0^f, M_0^v and Q_0^m, Q_0^f, Q_0^v given in Proposition 5.2. By construction, the covariance matrix Q_0 is block diagonal, encoding both the independence structure across male, female, and vector compartments and the correlations implied within each block through the population balance constraints.

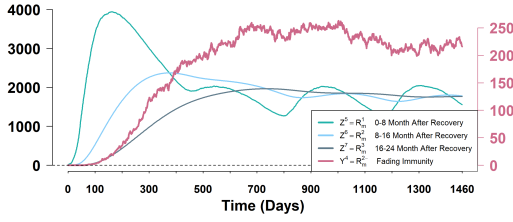
This initialization provides a structured and interpretable starting point for deploying the extended Kalman filter in the context of stochastic Zika epidemic modeling under partial observation.



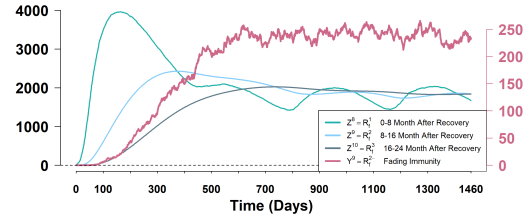
Detected infected male and detected symptomatic male.



Detected infected female and detected symptomatic female.



Recovery Cascade States for Male .



Recovery Cascade States for Female.

Fig. 5.3: Observations of the 10 observable compartments of the extended Zika model. Each figure shows the observable compartment.

5.4 Impact of Initial Estimates

We will now analyze the impact of initial estimates on the performance of the proposed filter method through a series of numerical experiments. To do this, we will vary the various initial estimates, namely the conditional mean $\tilde{M}_0 = M_0$ and the conditional variance $\tilde{Q}_0 = Q_0$, and observe the filter accuracy in the short and long term. Again, we focus on the results for the compartment I_m^- of undetected infected male individuals. For the other hidden compartments, we observed similar results.

	True value	Estimation	
		Conditional mean	Conditional variance
	$I_{m,0}^-$	$M_0^{I_m^-}$	$Q_0^{I_m^-}$
Scenario 1	80	210	30^2
Scenario 2	80	210	0
Scenario 3	80	80	0

Table 5.3: Different scenarios for the initials estimates $M_0^{I_m^-}$ and $Q_0^{I_m^-}$

5.5 Impact of Cascade States

In this section we investigate the impact of introducing cascade compartments on the performance of the filter, with particular emphasis on the accuracy of hidden-state estimation as measured by the corresponding standard deviations. We compare the base model with its extended counterparts that incorporate one, two, or three cascade compartments.

In both frameworks, individuals recovering from a confirmed infection are assumed to remain fully immune for a fixed period of $L = 730$ days. In the base model these individuals are assigned to the hidden compartments $R_{\dagger}^{2-}$, where they are aggregated with those whose recovery time already exceeds 730 days and who may subsequently undergo an unobservable transition back into the susceptible class. Consequently, the compartments $R_{\dagger}^{2-}$ must be treated as hidden. By contrast, the extended Zika models introduce observable cascade compartments, thereby allowing a more refined description of post-recovery and post-vaccination dynamics.

To ensure comparability, identical initial values are imposed across all models for the hidden states $S_{\dagger}, E_{\dagger}, I_{\dagger}^{-}$, and $R_{\dagger}^{1-}$. In the extended model, the first observable recovered cascade compartment is initialized as $R_{\dagger}^1 = 18$, with the remaining cascade compartments set to zero. In the baseline model, the corresponding hidden compartment $R_{\dagger}^{2-}$ is initialized at the same value, i.e. 18.

When cascade compartments are introduced, the immunity period of $L = 730$ days is subdivided into $d^R = 1, 2$, or 3 subintervals, corresponding respectively to one compartment of length 730, two compartments of 365 days each, or three compartments of 730/3 days. The loss rates of immunity are calibrated so that the expected duration of complete immunity remains identical across the baseline and extended models. Specifically, for immunity loss after recovery we set $\rho_{\dagger}^2 = 1/760$ in the baseline model and, analogously, $\rho_{\dagger}^2 = 1/30$ in the extended formulation, yielding in both cases an expected immune duration of $760 = L + 30$ days. Similarly, for immunity loss following recovery from $I_{\dagger}^{-}$ we fix $\rho_{\dagger}^1 = 1/760$ in both frameworks, ensuring again an expected duration of $760 = L + 30$ days.

Discussion on the convergence of standard deviation. An interesting feature of the results is that the standard deviations of the model without cascade states and the one with three cascade states converge in the long run, despite the expectation that additional cascade states should systematically reduce uncertainty. This behavior can be explained by the structure of the model: the only source of randomness enters through the inflow to the first cascade state, while subsequent cascade dynamics are purely deterministic. As a result, cascade states primarily act as filters that refine short- and medium-term uncertainty, which is particularly evident during outbreak and wave dynamics, when infections rise and decline sharply. In these phases, the system is more sensitive to randomness, and the models with cascade states display smaller variances by capturing more observable information. However, as the epidemic progresses toward the endemic equilibrium, the stochastic input naturally diminishes, leading to negligible long-run variability. Consequently, the additional information carried by cascade states loses its effect, and the variance across all models converges to the same baseline level. In other words, cascade states effectively reduce uncertainty during transient epidemic phases, but their impact fades once the driving noise vanishes and the system approaches equilibrium.

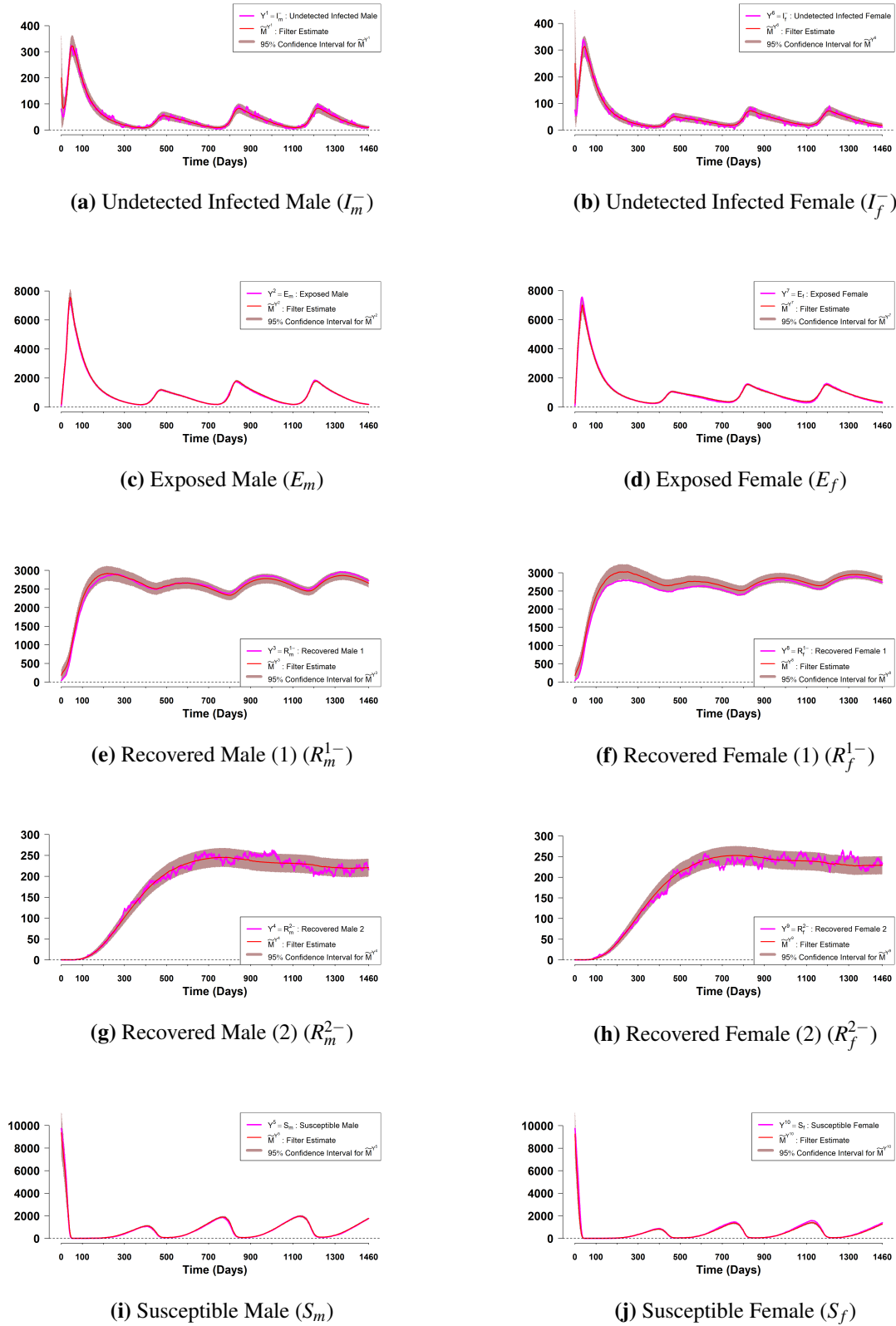


Fig. 5.4: Filter estimates and confidence bands for hidden human state variables over time. The left column shows male compartments, and the right column shows female compartments. The magenta curve shows the true simulated state, red the filter estimate, and the gray band represents the 95% confidence interval.

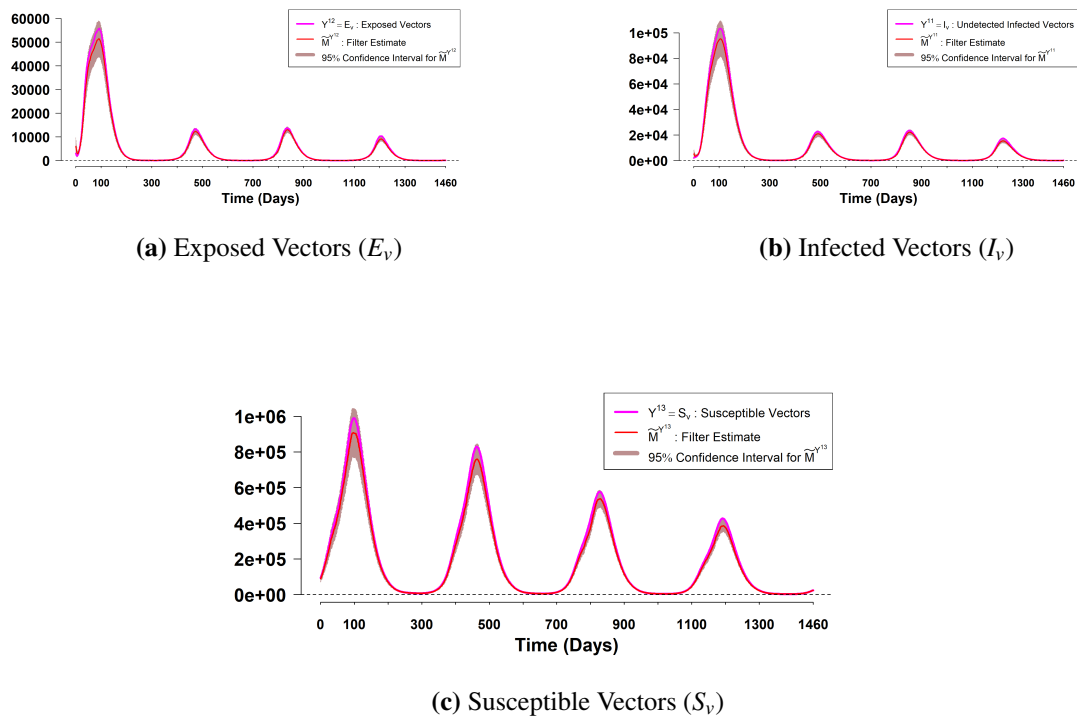
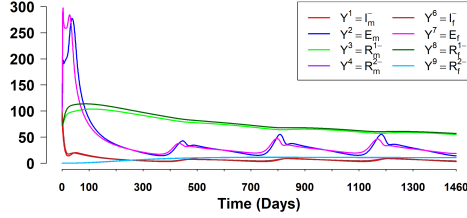
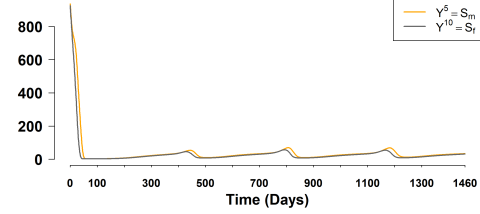


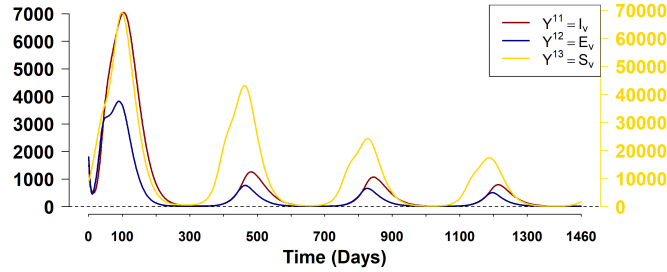
Fig. 5.5: Filter estimates and confidence bands for hidden mosquito vector state variables.



(a) Standard deviations $\sigma^{Y^i} = \sqrt{Q_{ii}}$ for the human compartments: $Y^1, Y^2, Y^3, Y^4, Y^6, Y^7, Y^8, Y^9$. These states correspond to latent and recovered classes for both males and females. The variance profiles reveal differences in uncertainty propagation across the infectious and exposed stages.

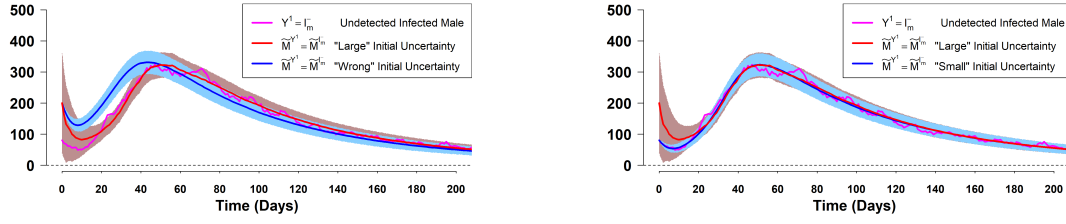


(b) Standard deviations $\sigma^{Y^i} = \sqrt{Q_{ii}}$ for the susceptible compartments Y^5 and Y^{10} (susceptible males and females). Their uncertainties remain relatively small and stable, consistent with population conservation and their indirect role in the observation process.



(c) Standard deviations $\sigma^{Y^i} = \sqrt{Q_{ii}}$ for vector-related compartments: Y^{11}, Y^{12}, Y^{13} . These show higher uncertainty levels, reflecting the indirect and unobserved dynamics of mosquito infection stages in the model.

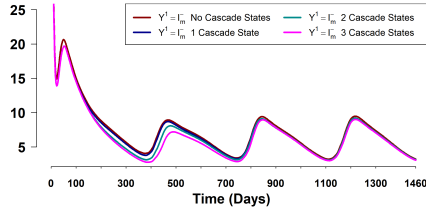
Fig. 5.6: Standard deviations of the estimated states $\sigma^{Y^i} = \sqrt{Q_{ii}}$ for different compartments in the Zika model under partial observation. Each panel highlights the uncertainty evolution for distinct groups: (a) human infectious and recovery classes, (b) human susceptible classes, and (c) vector populations.



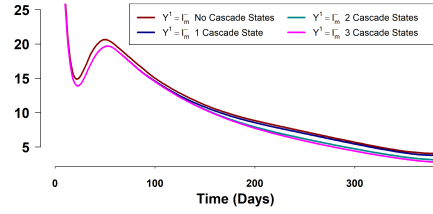
(a) Scenario 1 (red): Large initial uncertainty is reduced by learning from observations. Scenario 2 (blue): Incorrectly specified initial estimate with perfect accuracy needs long time to be corrected.

(b) Scenario 1 (red), scenario 3 (blue): Zero initial uncertainty is fading out by observation noise.

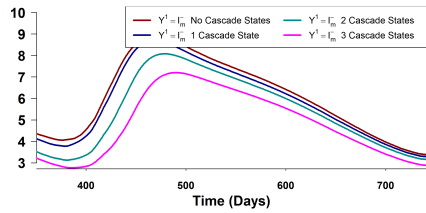
Fig. 5.7: Comparison of the effect of initial uncertainty on filtering performance. The figure shows for the three scenarios the true hidden state I_m^- , the filter estimate $\hat{M}_1 = \hat{M}^{I_m^-}$, and the associated 95% confidence band.



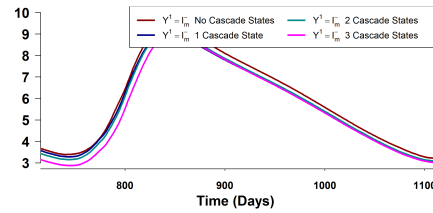
(a) Full trajectory over 4 years



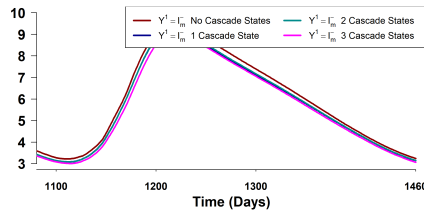
(b) Zoom: first year (Day 0–400)



(c) Zoom: epidemic peak around Day 400–700



(d) Zoom: epidemic peak around Day 800–1100



(e) Zoom: epidemic peak around Day 1100–1460

Fig. 5.8: Evolution of the standard deviation across different models depending on the number of cascade states. Panel (a) shows the full four-year trajectory, while Panels (b)–(e) zoom into different epidemic waves.

Appendix

A Notation

α_m, α_f	Force of infection rate from mosquitoes to susceptible males/females
β_m^+, β_f^+	Testing rates from exposed and asymptomatic male/female
β_m^s, β_f^s	Infection rates for symptomatic male/female
β_m^-, β_f^-	Infection rates for asymptomatic male/female
β_m, β_f	Infection rates for asymptomatic male/female in simplified model
γ_m^-, γ_f^-	Recovery rate from asymptomatic males/females
γ_m^+, γ_f^+	Recovery rate from detected males /females
γ_m^s, γ_f^s	Recovery rate from symptomatic males /females
$\rho_m^1, \rho_m^3, \rho_f^1, \rho_f^3$	Rate of immunity loss in asymptomatic males/females
$\rho_m^2, \rho_m^4, \rho_f^2, \rho_f^4$	Rate of immunity loss in symptomatic males/females
ρ_m, ρ_f	Rate of immunity loss in males/females in simplified model
θ, v, μ, B_v	Exposure rate, infection rate, natural mortality rate, and birth rate vectors
S_m, S_f	Susceptible male, female
E_m, E_f	Exposed male, female
R_m, R_f	Undetected/detected recovered in simplified model
R_m^{2-}, R_f^{2-}	Detected recovered with fading immunity in extended model
R_m^{1-}, R_f^{1-}	Undetected recovered with fading immunity in extended model
$I_m^-, I_m^+, I_m^s, I_f^-, I_f^+, I_f^s$	Undetected / detected, symptomatic infected
S_v, E_v, I_v	Susceptible, Exposed, Infected vectors
Δt	Time step
Θ_k	Counting process (for the number of transition k in $[0, t]$)
Y_n, Z_n	Hidden state and Observation vector at discrete time n
Π	Standard Poisson process with unit intensity
$\mathcal{E}^1, \mathcal{E}^2$	Independent $\mathcal{N}(0, \mathcal{I})$ random vectors
$\mathcal{F}_n^Z$	Filtration generated by observations up to time n
$\mathbb{E}[Y_n \mathcal{F}_n^Z]$	Conditional mean (filter estimate) of the hidden state Y_n up to time n
$\text{Var}(Y_n \mathcal{F}_n^Z)$	Conditional covariance matrix of Y_n given $\mathcal{F}_n^Z$
m_0, q_0	Initial prior mean and initial prior covariance of the hidden state Y_0
$\tilde{f}_{n0}, \tilde{f}_{n1}$	Constant term and linear operator in the drift of the state transition model
$\tilde{\sigma}_n$	Diffusion coefficient matrix for process noise in the state dynamics
$\tilde{g}_n$	Noise sensitivity matrix in the state equation
$\tilde{h}_{n0}, \tilde{h}_{n1}$	Constant term and linear operator mapping hidden states in the observation model
ℓ_n	Noise sensitivity matrix in the observation equation
$[A]^+$	Moore–Penrose pseudoinverse of matrix A
$\mathcal{Z}_n$	Observation trajectory $(Z_0, Z_1, \dots, Z_n)$ up to time n
$\hat{Y}_n$	Estimate of Y_n adapted to $\mathcal{F}_n^Z$ and square-integrable
Ω	Underlying probability space
$\mathcal{F}_0^I$	σ -algebra with prior information on the initial distribution of Y_0
f, f_0, f_1	Nonlinear drift function for deterministic dynamics of the hidden state
$\bar{Y}_n$	Reference point of linearized drift function f
$[Df](\bar{Y}_n)$	Jacobian matrix of f evaluated at the reference point $\bar{Y}_n$
R_n^{in}	newly recovered individuals entering the first cascade
R_n^{out}	The outflow from the final hidden compartment
d, d_1, d_2	Number of all, hidden and observable states
d^R	Number of cascade compartments
$\mathcal{I}_n$	$n \times n$ Identity matrix of order n
K	Total number of different transitions
L^R	Number of time steps with complete immunity
ℓ	Diffusion coefficient, observable state
$M, \tilde{M}$	Conditional mean/ EKF approximation
$N, \tilde{N}_t$	Total population size and total number of time steps
$Q, \tilde{Q}$	Conditional variance/ EKF approximation
P_k^R	Number of original cascade compartments grouped to one compartment
R_j^+	Cascade compartment with respect to observable recovered
W, W^1, W^2	multi-dimensional standard Brownian motions
$X, \bar{X}$	State vector for absolute and relative subpopulation size

$Y, \tilde{Y}$	Hidden state in original/linearized system
$\bar{Y}$	Reference point for Taylor expansion
$Z, \tilde{Z}$	Observable state in original/linearized system
ϕ	half-saturation parameter
$\mathcal{I}$	The total number of human infectious
ϑ	Transition parameter for cascade states

B Coefficients of Discrete-time Recursions of the State Process

B.1 Simplified Zika Model

In this section, we provide an explicit formulation of the model coefficients f, h, σ, g, ℓ that define the stochastic dynamics of the system introduced in Subsection 3.3. These coefficients appear in the state-space recursion equations (2.5), which govern the evolution of the hidden and observable states of the model given by $Y = (I_m^-, R_m, S_m, I_f^-, S_f, I_v, S_v)^\top$, and $Z = (I_m^+, I_f^+)^\top$, respectively. This results in a total of $d = 10$ distinct states and $K = 11$ possible transition pathways. In the coefficient f , the placeholder notation $x^\mathcal{I}$ is used in place of $\mathcal{I}$.

$$f(n, y, z) = y + \Delta t \cdot \begin{bmatrix} \alpha_m y^3 \frac{y^7}{\phi + y^7} - \beta_m y^1 - \gamma_m^- y^1 \\ \gamma_m^- y^1 + \gamma_m^+ z^1 - \rho_m y^2 \\ -\alpha_m y^3 \frac{y^7}{\phi + y^7} + \rho_m y^2 \\ \alpha_f y^7 \frac{y^7}{\phi + y^7} + \omega y^7 y^1 - \beta_f y^4 - \gamma_f^- y^4 \\ \gamma_f^- y^4 + \gamma_f^+ z^2 - \rho_f y^6 \\ -\alpha_f y^7 \frac{y^7}{\phi + y^7} - \omega y^7 y^1 + \rho_f y^6 \\ \theta y^9 \frac{x^\mathcal{I}}{N} - \mu y^8 \\ B_v(y^6 + y^7 + y^8) - \theta y^9 \frac{x^\mathcal{I}}{N} - \mu y^9 \end{bmatrix}$$

$$h(n, y, z) = h_0(n, z) + h_1(n, z)y$$

$$h_0(n, z) = z + \begin{bmatrix} -\gamma_m^+ z^1 \\ -\gamma_m^+ z^2 \end{bmatrix} \Delta t \quad h_1(n, z) = \begin{bmatrix} \beta_m & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta_f & 0 & 0 & 0 \end{bmatrix} \Delta t$$

$$\sigma(n, y, z) = \begin{bmatrix} \sqrt{\alpha_m y^3 \frac{y^7}{\phi + y^7}} & -\sqrt{\gamma_m^- y^1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{\gamma_m^- y^1} & -\sqrt{\rho_m y^2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\sqrt{\alpha_m y^3 \frac{y^7}{\phi + y^7}} & 0 & \sqrt{\rho_m y^2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sqrt{\alpha_f y^7 \frac{y^7}{\phi + y^7}} & \sqrt{\omega y^7 I_m^-} & -\sqrt{\gamma_f^- y^4} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sqrt{\gamma_f^- y^4} & -\sqrt{\rho_f y^6} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\sqrt{\alpha_f y^7 \frac{y^7}{\phi + y^7}} & -\sqrt{\omega y^7 I_m^-} & 0 & \sqrt{\rho_f y^6} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{\theta y^9 \frac{x^\mathcal{I}}{N}} & -\sqrt{\mu y^8} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\sqrt{\theta y^9 \frac{x^\mathcal{I}}{N}} & 0 & \sqrt{B_v(y^6 + y^7 + y^8)} & -\sqrt{\mu y^9} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$g(n, y, z) = \begin{bmatrix} -\sqrt{\beta_m y^1} & 0 & 0 & 0 \\ 0 & \sqrt{\gamma_m^+ z^1} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\sqrt{\beta_f y^4} & 0 \\ 0 & 0 & 0 & \sqrt{\gamma_f^+ z^2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \sqrt{\Delta t}, \quad \ell(n, y, z) = \begin{bmatrix} \sqrt{\beta_m y^1} & -\sqrt{\gamma_m^+ z^1} & 0 & 0 \\ 0 & 0 & \sqrt{\beta_f y^4} & -\sqrt{\gamma_f^+ z^2} \end{bmatrix} \sqrt{\Delta t}$$

The functions f_0 and f_1 appearing in Lemma 4.7 arise from the first-order linearization of the drift function f and are given as follows:

$$f_0(n, y, z) = \begin{bmatrix} -\alpha_m \frac{y^3 y^7}{N} \\ \gamma_m^+ z^1 \\ \alpha_m \frac{y^3 y^7}{N} \\ -\alpha_f \frac{y^6 y^7}{N} - \omega \frac{y^1 y^6}{N_m} \\ \gamma_f^+ z^2 \\ \alpha_f \frac{y^6 y^7}{N} + \omega \frac{y^1 y^6}{N_m} \\ -\theta y^8 (y^1 + y^4) \\ \theta y^8 (y^1 + y^4) \end{bmatrix}$$

$$f_1(n, y, z) = \begin{bmatrix} -\gamma_m^- - \beta_m^- & 0 & \alpha_m \frac{y^7}{N} & 0 & 0 & 0 & \alpha_m \frac{y^3}{N} & 0 \\ \gamma_m^- & -\rho_m & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \rho_m & -\alpha_m \frac{y^7}{N} & 0 & 0 & 0 & -\alpha_f \frac{y^6}{N} & 0 \\ \omega y^6 & 0 & 0 & -\gamma_f^- - \beta_f^- & 0 & \alpha_f \frac{y^7}{N} + \omega y^1 & \alpha_m \frac{y^3}{N} & 0 \\ 0 & 0 & 0 & \gamma_f^- & -\rho_f & 0 & 0 & 0 \\ -\omega y^6 & 0 & 0 & 0 & \rho_f & -\alpha_f \frac{y^7}{N} N_m - \omega y^1 & -\alpha_f \frac{y^6}{N} & 0 \\ \theta y^8 & 0 & 0 & \theta^8 & 0 & 0 & -\mu & \theta (y^1 + z^1 + y^4 + z^2) \\ -\theta y^8 & 0 & 0 & -\theta y^8 & 0 & 0 & 0 & B_v (y^6 + y^7 + y^8) - \theta (y^1 + z^1 + y^4 + z^2) - \mu \end{bmatrix}$$

B.2 The Base Zika Model

Here, we provide an explicit formulation of the model coefficients f, h, σ, g, ℓ that define the stochastic dynamics of the system introduced in Subsection ???. These coefficients appear in the state-space recursion equations (2.5), which govern the evolution of the hidden and observable states of the model are given by $Y = (I_m^-, E_m, R_m^{1-}, R_m^{2-}, S_m, I_f^-, E_f, R_f^{1-}, R_f^{2-}, S_f, I_v, E_v, S_v)^\top$, while the vector of observable states is $Z = (I_m^+, I_m^s, I_f^+, I_f^s)^\top$. This results in a total of $d = 17$ distinct states and $K = 20$ possible transition pathways.

$$f(n, y, z) = y + \Delta t \cdot \begin{bmatrix} \beta_m^- y^2 - \gamma_m^- y^1 - \beta_m^+ y^1 \\ \alpha_m y^5 \frac{y^{11}}{\phi + y^{11}} - (\beta_m^- + \beta_m^s + \beta_m^+) y^2 \\ \gamma_m^- y^1 - \rho_1 y^3 \\ \gamma_m^s z^2 + \gamma_m^+ z^1 - \rho_2 y^4 \\ -\alpha_m y^5 \frac{y^{11}}{\phi + y^{11}} + \rho_1 y^3 + \rho_2 y^4 \\ \beta_f^- y^7 - \gamma_f^- y^6 - \beta_f^+ y^6 \\ \alpha_f y^{10} \frac{y^{11}}{\phi + y^{11}} + \omega y^{10} \kappa_m - (\beta_f^- + \beta_f^s + \beta_f^+) y^7 \\ \gamma_f^- y^6 - \rho_3 y^8 \\ \gamma_f^s z^4 + \gamma_f^+ z^3 - \rho_4 y^9 \\ -\alpha_f y^{10} \frac{y^{11}}{\phi + y^{11}} - \omega y^{10} \kappa_m + \rho_3 y^8 + \rho_4 y^9 \\ v y^{12} - \mu y^{11} \\ -v y^{12} + \theta y^{13} (\kappa_m + \kappa_f) - \mu y^{12} \\ B_v (y^{11} + y^{12} + y^{13}) - \theta y^{13} (\kappa_m + \kappa_f) - \mu y^{13} \end{bmatrix}$$

$$h(n, y, z) = h_0(n, z) + h_1(n, z)y$$

$$h_0(n, z) = z + \begin{bmatrix} -\gamma_m^+ z^1 \\ -\gamma_m^s z^2 \\ -\gamma_f^+ z^3 \\ -\gamma_f^s z^4 \end{bmatrix} \Delta t, \quad h_1(n, z) = \begin{bmatrix} \beta_m^+ & \beta_m^+ & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta_m^s & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta_f^+ & \beta_f^+ & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \beta_f^s & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Delta t$$

[illegible]

The functions f_0 and f_1 appearing in Lemma 4.7 arise from the first-order linearization of the drift function f and are given as follows:

[illegible]

[illegible]

[illegible]

[illegible]

The functions f_0 and f_1 appearing in Lemma 4.7 arise from the first-order linearization of the drift function f and are given as follows:

[illegible]

$$f_0(n, y, z) = \begin{bmatrix} 0 \\ -\alpha_m \frac{y^5 y^{11}}{N} \\ 0 \\ \vartheta_m^3 z^7 \\ \alpha_m \frac{y^5 y^{11}}{N} \\ 0 \\ -\alpha_f \frac{y^{10} y^{11}}{N} - \omega \frac{(y^1 + y^2)}{N_m} \\ 0 \\ \vartheta_f^3 z^{10} \\ \alpha_f \frac{y^{10} y^{11}}{N} + \omega \frac{(y^1 + y^2)}{N_m} \\ 0 \\ -\theta y^{13} \left(\frac{y^1 + y^2}{N_m} + \frac{y^6 + y^7}{N_f} \right) \\ \theta y^{13} \left(\frac{y^1 + y^2}{N_m} + \frac{y^6 + y^7}{N_f} \right) \end{bmatrix}$$

C Proof of Proposition 5.2

Here, we provide the detailed derivations of the initial filter estimate. The computations rely on standard properties of expectation and covariance applied to the Gaussian random factors introduced in Assumption 5.1. In particular, we show how the conditional mean $M_0 = \mathbb{E}[Y_0 | \mathcal{F}_0^Z]$ and the conditional covariance $Q_0 = \text{Cov}(Y_0 | \mathcal{F}_0^Z)$ are obtained. The arguments consist mainly of using linearity of the expectation, linearity of the variance, and independence of the driving random coefficients.

We work conditionally on $\mathcal{F}_0^Z$, and use the placeholders $\dagger \in \{m, f\}$ (sex) and $\# \in \{I, E, R\}$ (compartment). Set

$$A_{\dagger} := I_{\dagger,0}^+ + I_{\dagger,0}^s, \quad B_{\dagger} := R_{\dagger,0}^1 + R_{\dagger,0}^2 + R_{\dagger,0}^3.$$

By Assumption 5.1, the initial dark figure coefficients (DFCs) are independent Gaussians,

$$U_{0,\dagger}^{\#} \sim \mathcal{N}(M_{U_{\dagger}}^{\#}, (Q_{U_{\dagger}}^{\#})^2), \quad \# \in \{I, E, R\}, \dagger \in \{m, f\},$$

and are independent of Z_0 . The vector components $(I_{v,0}, E_{v,0}, S_{v,0})$ are independent Gaussians with means (M_v^I, M_v^E, M_v^S) and variances $((Q_v^I)^2, (Q_v^E)^2, (Q_v^S)^2)$, and are independent of the human blocks.

Conditional mean $M_0 = \mathbb{E}[Y_0 | \mathcal{F}_0^Z]$. By linearity of expectation and the DFC definitions,

$$I_{\dagger,0}^- = U_{0,\dagger}^I A_{\dagger}, \quad E_{\dagger,0} = U_{0,\dagger}^E A_{\dagger}, \quad R_{\dagger,0}^{1-} = U_{0,\dagger}^R B_{\dagger}.$$

Using $R_{\dagger,0}^{2-} = 0$ and the population balance $S_{\dagger,0} = N_{\dagger} - (I_{\dagger,0}^- + E_{\dagger,0} + R_{\dagger,0}^{1-} + R_{\dagger,0}^{2-} + I_{\dagger,0}^+ + I_{\dagger,0}^s)$, we obtain, for each $\dagger \in \{m, f\}$, $M_0^{\dagger}$. For vectors, $M_0^v = (M_v^I, M_v^E, M_v^S)^{\top}$. Thus the full mean is given by $M_0 = (M_0^m, M_0^f, M_0^v)$.

Conditional covariance $Q_0 = \text{Cov}(Y_0 | \mathcal{F}_0^Z)$. Each random component is an affine function of mutually independent Gaussian factors $\{U_{0,\dagger}^{\#}\}$ (human) and $(I_{v,0}, E_{v,0}, S_{v,0})$ (vector). Therefore $Y_0 | \mathcal{F}_0^Z$ is Gaussian and Q_0 is block diagonal with human male/female blocks (size 5×5) and a vector block (size 3×3).

(a) *Human block for a fixed $\dagger \in \{m, f\}$.* With the ordering $(I_{\dagger,0}^-, E_{\dagger,0}, R_{\dagger,0}^{1-}, R_{\dagger,0}^{2-}, S_{\dagger,0})$, set

$$a_I := A_{\dagger}^2 (Q_{U_{\dagger}}^I)^2, \quad a_E := A_{\dagger}^2 (Q_{U_{\dagger}}^E)^2, \quad b_R := B_{\dagger}^2 (Q_{U_{\dagger}}^R)^2.$$

Independence across $\# \in \{I, E, R\}$ yields

$$\text{Var}(I_{\dagger,0}^-) = a_I, \quad \text{Var}(E_{\dagger,0}) = a_E, \quad \text{Var}(R_{\dagger,0}^{1-}) = b_R, \quad \text{Var}(R_{\dagger,0}^{2-}) = 0,$$

and all pairwise covariances among $I_{\dagger,0}^-, E_{\dagger,0}, R_{\dagger,0}^{1-}$ vanish. Since $S_{\dagger,0} = C_{\dagger} - (I_{\dagger,0}^- + E_{\dagger,0} + R_{\dagger,0}^{1-})$ with $C_{\dagger}$ constant given $(\mathcal{F}_0^Z)$, we have

$$\text{Var}(S_{\dagger,0}) = a_I + a_E + b_R, \quad \text{Cov}(S_{\dagger,0}, I_{\dagger,0}^-) = -a_I, \quad \text{Cov}(S_{\dagger,0}, E_{\dagger,0}) = -a_E, \quad \text{Cov}(S_{\dagger,0}, R_{\dagger,0}^{1-}) = -b_R.$$

Collecting the entries gives the explicit 5×5 block, $Q_0^\dagger$. By independence between the male and female DFCs, the cross-covariances between Q_0^m and Q_0^f are zero.

(b) *Vector block*. With ordering $(I_{v,0}, E_{v,0}, S_{v,0})$ and independence, we obtain Q_0^v .

(c) *Full covariance*. Therefore, based on (a) and (b) we obtain Q_0 .

Acknowledgment and Funding. L. Oluoch and R. Wunderlich gratefully acknowledge financial support from the cooperation program between the Deutsche Forschungsgemeinschaft (DFG) and The World Academy of Sciences (TWAS) under grant number 710382. F. Ouabo Kamkumo and R. Wunderlich gratefully acknowledge the support by the Deutsche Forschungsgemeinschaft (DFG), award number 458468407.

References

1. Alyami, L., & Das S. Extended skew Kalman filters for COVID-19 pandemic state estimation. In 2023 Sixth International Conference of Women in Data Science at Prince Sultan University (WiDS PSU) (2023), Institute of Electrical and Electronics Engineers (IEEE).
2. Baca-Carrasco, D., Velasco-Hernández, J.X. (2016) Sex, mosquitoes and epidemics: an evaluation of zika disease dynamics. *Bull Math Biol* 78(11):2228–2242. <https://doi.org/10.1007/s11538-016-0219-4>
3. Bain, A., & Crisan, D. (2009). *Fundamentals of stochastic filtering* (Vol. 3). Springer.
4. Blohm, G. M. et al. Evidence for mother-to-child transmission of Zika virus through breast milk. *Clin. Infect. Dis.* 66, 1120–1121(2018).
5. Brauer, F., Castillo-Chavez, C., Mubayi, A., & Towers, S (2016) Some models for epidemics of vector-transmitted diseases. *Inf Dis Model* 1(1):79–87. <https://doi.org/10.1016/j.idm.2016.08.001>
6. Britton, T., & O'Neill, P. D. Bayesian inference for stochastic epidemics in populations with random social structure. *Scandinavian Journal of Statistics* 29, 3 (2002), 375–390.
7. Britton, T., Pardoux, E., editors. (2019). *Stochastic Epidemic Models with Inference*. Springer. <https://doi.org/10.1007/978-3-030-30900-8>.
8. Burkholder, D. L., Pardoux, E., Sznitman, A.-S., & Pardoux, E. (1991). *Filtrage non linéaire et équations aux dérivées partielles stochastiques associées*. In *École d'Été de Probabilités de Saint-Flour XIX—1989* (pp. 68–163). Springer.
9. Calvetti, D., Hoover, A., Rose, J., & Somersalo, E. Bayesian particle filter algorithm for learning epidemic dynamics. *Inverse Problems* 37, 11 (2021), 115008.
10. Cappé, O., Moulines, E., & Rydén, T. (2007). *Inference in Hidden Markov Models*. Springer.
11. Centers for Disease Control and Prevention, *Clinical guidance for healthcare providers for prevention of sexual transmission of Zika virus*, <https://www.cdc.gov/zika/hc-providers/clinical-guidance/sexualtransmission.html> (accessed 27 June 2019).
12. Centers for Disease Control and Prevention, *First female-to-male sexual transmission of Zika virus infection reported in New York City*, <http://www.cdc.gov/media/releases/2016/s0715-zika-female-to-male.html>
13. Centers for Disease Control and Prevention. (2025). Sexual Transmission of Zika Virus. <https://www.cdc.gov/zika/hcp/sexual-transmission/index.html>
14. Chen, H.-F., Kumar, P., & Van Schuppen, J. On Kalman filtering for conditionally Gaussian systems with random matrices. *Systems & Control Letters* 13, 5 (1989), 397–404.
15. Chen, N., Li, Y., & Liu, H. (2022) Conditional Gaussian nonlinear system: A fast preconditioner and a cheap surrogate model for complex nonlinear systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 32, 5.
16. Chen, N., & Majda, A. J. (2018). Conditional Gaussian systems for multiscale nonlinear stochastic systems: Prediction, state estimation and uncertainty quantification. *Entropy* 20, 7, 509.
17. Colaneri, K., Damian, C., & Frez, R. Invisible infections: A partial information approach for estimating the transmission dynamics of the Covid-19 pandemic. *arXiv:2212.13443* (2022).
18. David, F.A., Thomas, G.K. Continuous-time Markov chain models for chemical reaction networks. In *Design and analysis of biomolecular circuits*, pages 3–42. Springer, 2011. <https://doi.org/10.1007/978-1-4419-6766-4-1>.

19. Davidson, A., Slavinski, S., Komoto, K., Rakeman, J. & Weiss, D.(2016) Suspected female-to-male sexual transmission of Zika virus – New York City, 2016. *MMWR Morb. Mortal. Wkly Rep.* 65, 716–717
20. Dénes, A., Ibrahim, M.A., Oluoch, L., Tekeli, M., Tekeli, T. (2019). Impact of weather seasonality and sexual transmission on the spread of Zika fever. *Sci Rep.* <https://doi.org/10.1038/s41598-019-53062-z>
21. Doucet, A., de.Freitas, N., & Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Springer.
22. Ethier, S. N., and Kurtz, T. G. *Markov processes: characterization and convergence*. John Wiley & Sons, 2009.
23. Fontaine. A., Franck, L., Didier, B. et al.(2018). Duration of Zika Viremia in Serum. *Clin Infect Dis.*67(7):1143–1149 <https://doi.org/10.1093/cid/ciy261>
24. Gao, D., Lou, Y., He, D., Porco, T.C.et al.(2016). Prevention and control of Zika as a mosquito-borne and sexually transmitted disease: a mathematical modeling analysis. *Sci Rep.* <https://doi.org/10.1038/srep28070>
25. Greenwood, P.E., Gordillo, L.F. (2009). Stochastic Epidemic Modeling. In: Chowell, G., Hyman, J.M., Bettencourt, L.M.A., Castillo-Chavez, C. (eds) *Mathematical and Statistical Estimation Approaches in Epidemiology*. Springer, Dordrecht. https://doi.org/10.1007/978-90-481-2313-1_2
26. Gregory, C. J. et al.(2017). Modes of transmission of Zika virus. *J. Infect. Dis.* 216, S875–S883 .
27. Griffin, I., Martin, S. W., Fischer, M., Chambers, T. V., Kosoy, O. L., Goldberg, C. C., et al. (2019). Zika virus IgM detection and neutralizing antibody profiles 12–19 months after illness onset, Miami-Dade County, Florida, USA. *Emerging Infectious Diseases*, 25(2), 299–303. <https://doi.org/10.3201/eid2502.181286>
28. Guy, R., Larédo, C., and Vergu, E. Approximation of epidemic models by diffusion processes and their statistical inference. *Journal of Mathematical Biology* 70, 3 (2015), 621–646.
29. Guy, R., Larédo, C., and Vergu, E. Approximation and inference of epidemic dynamics by diffusion processes. *Journal de la Société Française de Statistique* 157, 1 (2016), 71–100.
30. Hasan, A., Susanto, H., Tjahjono, V., Kusdiantara, R.et al.(2022). A new estimation method for COVID-19 time-varying reproduction number using active cases. *Scientific Reports* 12, 1 , 6675.
31. Indriani, R., & Hartono, B. (2021). Comparison of EnKF and EKF in forecasting dengue incidence in Indonesia. *Jurnal Sains dan Seni ITS*, 10(2), D194–D198. https://ejurnal.its.ac.id/index.php/sains_seni/article/view/75581
32. Kamkumo, F. O., Njiasse, I. M., & Wunderlich, R. (2025). Estimating Unobservable States in Stochastic Epidemic Models with Partial Information. *arXiv*. <https://arxiv.org/abs/2506.00906>
33. Kamkumo, F. O., Njiasse, I. M., & Wunderlich, R. (2025). Stochastic epidemic models with partial information. *arXiv*. <https://arxiv.org/abs/2503.07251>
34. Lal, R., Huang, W., & Li, Z. An application of the ensemble Kalman filter in epidemiological modelling. *Plos one* 16, 8 (2021), e0256227.
35. Lessler, J., Chaisson, L. H., Kucirka, L. M., & others. (2016). Assessing the global threat from Zika virus. *Science*, 353(6300), <https://doi.org/10.1126/science>.
36. Magalhaes, T., Morais, C.N.L., Azevedo, E.A.N., Jacques I.J.A.A. et al. (2022) Two-year decay of Zika virus neutralizing antibodies in people living in an endemic region in Brazil. *Am J Trop Med Hyg.* 6;107(1):186-189. doi:10.4269/ajtmh.21-1279.
37. Magalhaes, T., Foy, B. D., Marques, E. T. A., Ebel, G. D. & Weger-Lucarelli, J.(2018). Mosquito-borne and sexual transmission of Zika virus: recent developments and future directions. *Virus Research* 254, 1–9 .
38. Mead, P. S., Duggal, N. K., Hook, S. A., & others. (2018). Zika virus shedding in semen of symptomatic infected men. *The New England Journal of Medicine*, 378(15), 1377–1385. <https://doi.org/10.1056/NEJMoa1711038>
39. Njiasse, I. M., Kamkumo, F. O., & Wunderlich, R. (2025). Stochastic optimal control of an epidemic under partial information . *arXiv*. <https://arxiv.org/abs/2503.06804>
40. Njiasse, I. M., Kamkumo, F. O., & Wunderlich, R. (2025). Convergence analysis for the extended *Kalman* filter in continuous time. *Working Paper*.
41. O'Neill, P. D., & Roberts, G. O. Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society Series A: Statistics in Society* 162, 1 (1999), 121–129.
42. Oduyebo, T., Polen, K., Walke, H., & others. (2016). Update: Interim guidance for health care providers caring for women of reproductive age with possible Zika virus exposure—United States, 2016. *MMWR. Morbidity and Mortality Weekly Report*, 65(12), 315–322. <https://doi.org/10.15585/>

mmwr.mm6512e2

43. Padmanabhan, P., Seshaiyer, P., Castillo-Chavez, C. (2017) . Mathematical modeling, analysis and simulation of the spread of Zika with the influence of sexual transmission and preventive measures. *Lett Biomath*. <https://doi.org/10.30707/lib4.1padmanabhan>
44. Petersen, L. R., Jamieson, D. J., Powers, A. M. & Honein, M. A. (2016). Zika virus. *N. Engl. J. Med.* 375, 294–295 .
45. Picard, J. (1991). Efficiency of the extended *Kalman* filter for nonlinear systems with small noise. *SIAM Journal on Applied Mathematics*, 51(3), 843–885.
46. Picard, J. (1993). Estimation of the quadratic variation of nearly observed semimartingales with application to filtering. *SIAM Journal on Control and Optimization*, 31(2), 494–517.
47. Robert, S. L., Albert, N.S. (2001) *Statistics of Random Processes II: Applications*. Springer. <https://doi.org/10.1007/978-3-662-10028-8>.
48. Sarkka, S. (2013). *Bayesian Filtering and Smoothing*. Cambridge University Press. (Cambridge, UK)
49. Sasmal, S.K., Ghosh, I., Huppert, A., Chattopadhyay, J. (2018) . Modeling the spread of Zika virus in a stage-structured population: effect of sexual transmission. *Bull Math Biol* 80(11):3038–3067. <https://doi.org/10.1007/s11538-018-0510-7>
50. Song, B. H. et al.(2017). Zika virus: history, epidemiology, transmission, and clinical presentation. *J. Neuroimmunol.* 308, 50–64.
51. Suparit, P., Wiratsudakul, A.& Modchang, C.(2018). A mathematical model for Zika virus transmission dynamics with a time-dependent mosquito biting rate. *Theor. Biol. Med. Model.* 15, 11.
52. Turmel, J.M., Abgueuen, P., Hubert, B., & others. (2016). Late sexual transmission of Zika virus related to persistence in the semen. *The Lancet*, 387(10037), 2501.[https://doi.org/10.1016/S0140-6736\(16\)30775-9](https://doi.org/10.1016/S0140-6736(16)30775-9)
53. World Health Organization, (2015). *WHO Global Health Observatory data repository. Crude birth and death rate, Data by country*. <http://apps.who.int/gho/data/node.main.CBDR107?lang=en>
54. World Health Organization, (2018) *WHO list of blueprint priority diseases*, <https://www.who.int/blueprint/priority-diseases/en/> .
55. Yakob, L.,& Walker, T. (2016). Assessing the effectiveness of vector control for Zika virus: A modelling study. *The Lancet Infectious Diseases*, 16(2), 154–160.[https://doi.org/10.1016/S1473-3099\(15\)00433-8](https://doi.org/10.1016/S1473-3099(15)00433-8)
56. Yang, L., Zhao, Y., & Li, W. (2022). Metapopulation dengue forecast using EAKF in Guangdong, China. *PLoS Neglected Tropical Diseases*, 16(5), e011418. <https://doi.org/10.1371/journal.pntd.0011418>
57. Zeng, X., & Ghanem, R.(2020). Dynamics identification and forecasting of COVID-19 by switching Kalman filters. *Computational Mechanics* 66 , 1179–1193.
58. Zhou, Y., Liu, K., Wang, Y., et al. (2023). Forecasting dengue outbreaks using an ensemble adjustment Kalman filter model in Guangzhou, China. *BMC Public Health*, 23(1), 123. <https://doi.org/10.1186/s12889-025-22545-2>
59. Zhu, X., Gao, B., Zhong Y., Gu, C., & Choi, K.S.(2021). Extended Kalman filter based on stochastic epidemiological model for COVID-19 modelling. *Computers in Biology and Medicine* 137 , 104810.