Evaluating the risk of mosquito-borne diseases in non-endemic regions: A dynamic modeling approach

Nico Stollenwerk[1], Luís Mateus[1,2], Vanessa Steindorf[1], Bruno V. Guerrero[1], Rubén Blasco-Aguado[1,3], Aitor Cevidanes [4], Joseba Bidaurrazaga Van-Dierdonck[5] and Maíra Aguiar[1,6]

[1]Basque Center for Applied Mathematics (BCAM), Bilbao, Spain
[2]Center for Mathematical Studies (CEMS.UL) - Faculty of Sciences of the University of Lisbon (FCUL), Lisbon, Portugal
[3]Scuola Superiore Meridionale (SSM), Naples, Italy
[4]Animal Health Department, NEIKER - Basque Institute for Agricultural Research and Development, Basque Research and Technology Alliance (BRTA), Derio, Bizkaia, Spain
[5]Public Health, Basque Health Department, Rekalde Zumarkalea 39A, 48008 Bilbao, Spain
[6]Ikerbasque, Basque Foundation for Science, Bilbao, Spain

October 9, 2024

Abstract

Mosquito-borne diseases are spreading into temperate zones, raising concerns about local outbreaks driven by imported cases. Using stochastic methods, we developed a vector-host model to estimate the risk of import-driven autochthonous outbreaks in non-endemic regions. The model explores key factors such as imported cases and vector abundance. Our analysis shows that mosquito population abundance significantly affects the probability and timing of outbreaks. Even with moderate mosquito populations, isolated or clustered outbreaks can be triggered, highlighting the importance of monitoring vector abundance for effective public health planning and interventions.

1 Introduction

The presence of *Aedes* mosquitoes in non-endemic regions has become a significant public health concern [1, 2], as these areas now face the potential risk of local transmission of mosquito-borne diseases such as dengue fever, Zika virus disease, and chikungunya fever [3]. Historically confined to tropical and subtropical climates, these diseases are now spreading into temperate zones, driven by factors like climate change, increased global travel, and the adaptability of these mosquitoes [4], which are now distributed across several parts of the world [5].

The expansion of mosquito species into new regions creates an alarming epidemiological situation, referred to as an "invasion scenario of disease introduction". This scenario represents an increased risk of mosquito-borne disease cases transmitted by newly established mosquito species that serve as disease vectors, going beyond a small local outbreak or merely reaching a threshold for continuous transmission. Such a situation presents new challenges for public health systems in non-endemic areas, which may not be adequately equipped to manage these emerging disease outbreaks.

Understanding the concept of invasion scenarios of disease introduction is crucial for assessing the risk of autochthonous (locally transmitted) disease outbreaks in non-endemic regions [6, 7, 8, 9]. Even when a region is below the epidemiological threshold - indicating that the disease is not yet self-sustaining - significant outbreaks of autochthonous cases can still occur. Important examples include the Madeira Island outbreak in Portugal in 2012 [10] and more recent cases in France and Italy [11].

In these invasion scenarios, the risk of autochthonous cases is closely linked to both the abundance of mosquito vectors and the influx of viremic imported cases - individuals who are infected returning from travels in endemic areas who are capable of transmitting the virus [12, 13, 14, 15]. As mosquito populations increase, the likelihood of disease outbreaks also rises, often following power-law scaling in the distribution of confirmed cases [16, 17, 18]. This pattern, characteristic of systems approaching a critical threshold, indicates that small changes in key factors like mosquito population density or environmental conditions can lead to disproportionately large effects [19]. Recognizing this is essential for public health, emphasizing the need for early intervention, mosquito control, and continuous surveillance to prevent the system from reaching a tipping point where large-scale outbreaks become inevitable.

In the epidemiology of mosquito-borne diseases, classical compartmental models, both with and without explicit mosquito dynamics, have been

extensively used, primarily focusing on disease transmission in endemic regions [20, 21, 22, 23, 24]. These models provide essential parameters for understanding disease dynamics. However, assessing the potential impact of mosquito abundance and imported cases on invasion scenarios - where no autochthonous cases have yet occurred - presents additional challenges, such as data limitations, dynamic variability, and the stochastic nature of disease spread.

To evaluate and quantify the risk of imported driven autochthonous cases in non-endemic regions, we propose a dynamic modeling approach that incorporates factors such as the presence of *Aedes* mosquitoes and viremic imported cases. Using stochastic methods and their deterministic counterpart, this paper offers a comprehensive analysis of the risk landscape, including estimates of waiting times until the first autochthonous case is identified and statistics on the inter-event times of imported cases, with confidence intervals derived from likelihood functions. Using the Basque Country as a case study - where mosquito populations are established and imported cases are frequently reported - we analyze trends in imported cases and assess the impact of mosquito abundance on the risk of autochthonous disease outbreaks. Our study provides a timely framework for predicting imported driven autochthonous disease cases based on available data on mosquito presence and imported case notifications, which is crucial for preventing sustained transmission and mitigating the effects of isolated outbreaks in non-endemic regions.

2 The SIRUV modeling framework

To assess the risk of mosquito-borne disease transmission in non-endemic regions, we refined and extended the SIRUV modeling framework [25, 26]. Initially designed as a foundational model for simulating disease spread in endemic areas, the framework is adjusted to describe local transmission dynamics in non-endemic regions, potentially triggered by the introduction of viremic imported cases and the presence of competent mosquito vectors.

In this section, we start by outlining the structure and key assumptions of the deterministic SIRUV model, which captures the interaction between human and mosquito populations in an endemic scenario. Using parameters obtained from data from endemic countries [27, 28, 29], the model is then extended to incorporate imported cases - which have little impact on sustained transmission in endemic areas - lower vector abundance, and stochastic elements. This extension allows for a more comprehensive analysis of how

these factors influence the probability of autochthonous disease outbreaks in non-endemic regions.

2.1 The SIRUV model for endemic areas

The classical Susceptible-Infected-Recovered (SIR) model [30, 24], which describes the dynamics of a host-host transmission disease, is given by

$$\begin{aligned} \frac{d}{dt}S &= \alpha R - \frac{\beta_{SIR}}{N} SI \\ \frac{d}{dt}I &= \frac{\beta_{SIR}}{N} SI - \gamma I \\ \frac{d}{dt}R &= \gamma I - \alpha R \quad , \end{aligned} \tag{1}$$

where β_{SIR} represents the transmission rate, γ the recovery rate, N the human population size and α the waning immunity rate. These equations describe the flow of individuals between the compartments: S (susceptible), I (infected), and R (recovered).

To model mosquito-borne diseases, the SIR model is extended to include explicit vector dynamics [31, 32, 25, 26, 33, 34, 35]. In its basic concept to describe different mosquito-borne diseases such as dengue, Zika and chikungunya, this results in the SIRUV type model, which incorporates compartments for Uninfected mosquitoes U, and infected mosquitoes, i.e., the disease-Vector V. The equations for the SIRUV model are given by

$$\begin{aligned} \frac{d}{dt}S &= \alpha R - \frac{\beta}{mN} SV \\ \frac{d}{dt}I &= \frac{\beta}{mN} SV - \gamma I \\ \frac{d}{dt}R &= \gamma I - \alpha R \\ \frac{d}{dt}U &= \psi - \frac{\vartheta}{N} UI - \nu U \\ \frac{d}{dt}V &= \frac{\vartheta}{N} UI - \nu V \quad , \end{aligned} \tag{2}$$

where $\psi = \nu \cdot mN$ represents the mosquito supply rate, with m being the ratio of the mosquito population size to the human population size N, ν the natural mosquito mortality rate, and β and ϑ the transmission rates from mosquitoes to humans and humans to mosquitoes, respectively.

It is important to note that the average lifespan of an *Aedes* mosquito in nature is two weeks [36], whereas the protection conferred by human immunological responses persists for a much longer period. In the case of dengue fever, for example, protection against reinfection can last from 6 months to 2 years [28, 37, 38, 39, 40]. Given these differing timescales, a simplified model can be derived by using techniques such as singular perturbation or center manifold analysis [29, 26, 25] to describe time scale separation. As described in those studies, the mosquito dynamics are dominated by the slower dynamics of human infection on immunological time scales.

To rescale the mosquito parameters ν and ϑ to align with the human immunological timescales, we introduce a parameter ϵ, which represents the ratio between the fast mosquito timescale and the slow human timescale. This rescaling is done such that $\epsilon\nu = \hat{\nu} \approx \alpha$ and $\epsilon\vartheta = \hat{\vartheta}$. Hence, the last equation of system (2) becomes

$$\frac{d}{dt}V = \frac{1}{\epsilon}\left(\frac{\hat{\vartheta}}{N}UI - \hat{\nu}V\right) \quad . \tag{3}$$

In the limit as $\epsilon \to 0$, indicating that mosquito dynamics quickly reach a quasi-stationary state, the equation simplifies to

$$\epsilon\frac{d}{dt}V \to 0 = \frac{\hat{\vartheta}}{N}(M - V)I - \hat{\nu}V \quad . \tag{4}$$

When considering the quasi-stationary solution for $V(I)$ (i.e., $\frac{dV}{dt} \approx 0$), the number of infected mosquitoes V can be expressed as

$$V(I) = \frac{\frac{\vartheta}{\nu}\frac{I}{N}}{1 + \frac{\vartheta}{\nu}\frac{I}{N}}mN, \tag{5}$$

where $M(t) := U(t) + V(t)$ represents the total mosquito population, which quickly approaches $M^* = mN$. This result captures how the number of infected mosquitoes adjusts according to the infection dynamics within the human population.

When the infected population is much smaller compared to the total population size N, as is the case for typical SIR systems in endemic areas, we have, $\frac{\vartheta}{\nu}\frac{I}{N} \ll 1$, which simplifies equation (5) to

$$V(I) = \frac{\vartheta}{\nu}\frac{I}{N}mN \quad . \tag{6}$$

For more detailed calculations, see [29].

This simplified expression can be substituted for V in the human infection dynamics equations of system (2). Consequently, we obtain an effective SIR system with human-to-human transmission mediated by mosquitoes, given by

$$\begin{aligned} \frac{d}{dt}S &= \alpha R - \frac{\beta\frac{\vartheta}{\nu}}{N}SI \\ \frac{d}{dt}I &= \frac{\beta\frac{\vartheta}{\nu}}{N}SI - \gamma I \\ \frac{d}{dt}R &= \gamma I - \alpha R \quad . \end{aligned} \tag{7}$$

This relationship between models allows us to infer the basic infection parameters for mosquito-borne diseases from the host-host dynamic model given in system (1). Consequently, the infection rates from the SIRUV model and the effective SIR model are related by

$$\beta_{SIR} = \beta\frac{\vartheta}{\nu} \quad . \tag{8}$$

The parameter values used for our numerical simulations are listed in Table 1. Considering both primary and secondary infections, such as those observed in dengue, we set an approximate period of 10 years for the waning immunity α. The effective infectivity rate β_{SIR} is estimated from results in endemic regions, known to range from approximately $1.1 \cdot \gamma$ in low endemic areas to $2 \cdot \gamma$ in high endemic areas [27]. The selected value in our simulations reflects an effective infectivity rate, and it should be adjusted according to specific epidemiological information for the mosquito-borne disease and region under consideration. Any change in these values will be mentioned in the figure captions.

2.2 The SIRUV model for an invasion scenario

To model an invasion scenario of disease introduction, we consider an area where there is no sustained local transmission, but where mosquitoes are present - which are significantly less abundant than in endemic areas - and imported cases are frequently recorded.

To quantify this relative reduction in the mosquito abundance, we introduce the ratio $k \in [0, 1)$, defined as

$$k = \frac{\text{vector abundance in the non-endemic area}}{\text{vector abundance in a typical endemic area}} \quad . \tag{9}$$

Parameter	**Description**	**Value** ($y^{-}1$)
γ	Recovery rate (1 week)	52
α	Waning immunity rate (10 years)	0.1
β	Transmission rate from mosquitoes to humans	$1.2\,\gamma$
ϑ	Transmission rate from humans to mosquitoes	$1.2\,\nu$
ν	Mosquito mortality rate (10 days)	36.5
β_{SIR}	Effective transmission rate	$1.2^2\,\gamma$

Table 1: Baseline model parameters to describe mosquito-borne disease transmission dynamics in endemic regions [27, 26]. These parameters can be refined according to specific epidemiological data from endemic areas and diseases.

Since monitoring adult mosquitoes and effectively measuring their contribution to the infection process is challenging, the relative mosquito abundance k, could be measured using surrogate data such as ovitrap egg counts, positive ovitrap index, vector indices, composite index, passive larval and pupal collections, adult collections, and Stegomyia indices, for example [41, 42, 43].

The reduced supply rate of vectors ψ, in a non-endemic area can be expressed as a function of k

$$\psi(k) = k\,\nu m N \quad , \tag{10}$$

where $k = 1$ represents the vector supply rate typical of an endemic scenario.

On the other hand, the possibility of having imported infections Y into the population can be introduced in the model by assuming that there exists a steady input of new infections at constant ratio ϱ [28, 44]. In real life, those infections account for travelers arriving during their viremic period or residents returning from endemic regions with an active infection, which are often recorded in non-endemic areas when those individuals seek medical care. The number of imported cases is described by $Y := \varrho N$, with $\varrho > 0$, and it is included in the infection term of the uninfected mosquitoes equation.

After including these considerations in the System (2), the model for the

non-endemic area reads

$$\begin{aligned}
\frac{d}{dt}S &= \alpha R - \frac{\beta}{mN}SV \\
\frac{d}{dt}I &= \frac{\beta}{mN}SV - \gamma I \\
\frac{d}{dt}R &= \gamma I - \alpha R \\
\frac{d}{dt}U &= \psi(k) - \frac{\vartheta}{N}U(I+\varrho N) - \nu U \\
\frac{d}{dt}V &= \frac{\vartheta}{N}U(I+\varrho N) - \nu V \quad .
\end{aligned} \tag{11}$$

In this deterministic model, the expected number of imported cases is constant at ϱN, which is relatively small compared to the total human population size (usually on the order of a few individuals). However, because infectious disease transmission is inherently stochastic, this variability must be considered in the modeling process.

2.3 The stochastic SIRUV model for an invasion scenario

The stochastic version of the SIRUV model given in system (11) is described by the dynamics of the discrete state vector $\underline{X} := (S, I, R, U, V)^{\mathrm{tr}}$, representing the evolution of the state variables of the model. The probability of arriving to the state $\underline{X}$ at time t is modeled as a continuous time Markov process with the following master equation

$$\frac{d}{dt}\, p(\underline{X}, t) \;=\; \sum_{j=1}^{n} \Big(w_j(\underline{X} + \Delta\underline{X}_j) \cdot p(\underline{X} + \Delta\underline{X}_j, t) \;\; - w_j(\underline{X}) \cdot p(\underline{X}, t) \Big) \quad , \tag{12}$$

where $n = 7$ is the number of different deviations from state $\underline{X}$, which are given by $\Delta\underline{X}_j := \underline{r}_j$. For the SIRUV model for the non-endemic scenario, the transitions $w_j(\underline{X})$ and their shift vectors $\underline{r}_j$ are

$$\begin{aligned}
w_1(\underline{X}) &= \alpha R && , \quad \underline{r}_1 = (-1,0,1,0,0)^{tr} \\
w_2(\underline{X}) &= \tfrac{\beta}{mN}SV && , \quad \underline{r}_2 = (1,-1,0,0,0)^{tr} \\
w_3(\underline{X}) &= \gamma I && , \quad \underline{r}_3 = (0,1,-1,0,0)^{tr} \\
w_4(\underline{X}) &= \psi(k) && , \quad \underline{r}_4 = (0,0,0,-1,0)^{tr} \\
w_5(\underline{X}) &= \tfrac{\vartheta}{N}U(I+\varrho N) && , \quad \underline{r}_5 = (0,0,0,1,-1)^{tr} \\
w_6(\underline{X}) &= \nu U && , \quad \underline{r}_6 = (0,0,0,1,0)^{tr} \\
w_7(\underline{X}) &= \nu V && , \quad \underline{r}_7 = (0,0,0,0,1)^{tr} \quad .
\end{aligned} \tag{13}$$

With these $w_j(\underline{X})$ and $\underline{r}_j$ specified, we can obtain stochastic realizations of the process using the Gillespie algorithm [45].

In the case of a well-defined system size, we can express the stochastic process in terms of the densities of the state variables. From this, we can obtain the mean field ODE system and, using the Kramers-Moyal approximation of the master equation, derive a Fokker-Planck equation and from it a system of stochastic differential equations [46]. In this study, we investigate stochastic processes with small numbers of cases, which are close to extinction thresholds. Therefore, we will primarily use the version with absolute numbers, rather than the densities.

The expected number of autochthonous infected cases is denoted by $\langle I \rangle$ and is defined as

$$\langle I \rangle := \sum_{\underline{X}} I \; p(\underline{X}, t) = \langle I(t) \rangle \quad , \tag{14}$$

where $p(\underline{X}, t)$ represents the probability of the system being in state $\underline{X}$ at time t. The dynamics of $\langle I \rangle$ is governed by the evolution of these probabilities over time.

After performing the calculations, we obtain the following differential equation for the expected number of autochthonous infected cases

$$\frac{d}{dt}\langle I \rangle = \sum_{\underline{X}} I \; \frac{d}{dt} p(\underline{X}, t) = \frac{\beta}{mN} \langle SV \rangle - \gamma \langle I \rangle \quad . \tag{15}$$

In the mean field approximation, this simplifies to

$$\frac{d}{dt}\langle I \rangle = \frac{\beta}{mN} \langle S \rangle \langle V \rangle - \gamma \langle I \rangle \quad . \tag{16}$$

Thus, the ordinary differential equation (ODE) system captures the dynamics of the stochastic process in terms of expectation values [19].

2.4 Analytic results for the expected risk of autochthonous cases

After defining the SIRUV model for non-endemic areas and accounting for the stochastic nature of the process, we found that the expected number of autochthonous infected cases $\langle I \rangle$, depends on the number of imported cases $Y = \varrho N$, and the relative mosquito abundance k. We use the mean value of the quasi-stationary solution for I as a proxy to estimate the risk of autochthonous cases of arbovirus diseases in non-endemic regions.

From the invasion scenario model (see system (11)), which serves as a mean field approximation of the stochastic process, we derive an analytical expression for $\langle I \rangle$. For simplicity in notation, we will omit the expectation brackets in the following discussion.

Starting from the mosquito dynamics given by the last two equations in system (11), the dynamics of the total mosquito population $M = U + V$, is

$$\frac{d}{dt}M = \nu k m N - \nu M \quad . \tag{17}$$

At stationarity, the total mosquito population is $M^* = kmN$, and the uninfected mosquito population is $U = M^* - V = kmN - V$.

Assuming the time scale separation ($\frac{dV}{dt} \approx 0$) described in Section 2.1, the following expression for $V(I)$ can be derived from the last equation in system (11)

$$V(I) = \frac{\frac{\vartheta}{\nu}\frac{I+\varrho N}{N}}{1 + \frac{\vartheta}{\nu}\frac{I+\varrho N}{N}} k m N \quad . \tag{18}$$

The assumption that $\frac{\vartheta}{\nu}\frac{I+\varrho N}{N} \ll 1$ remains valid since the number of infected and imported cases is small compared to the total population. Consequently, the relationship between infected mosquitoes and infected humans is approximately linear and given by

$$V(I) = \frac{\vartheta}{\nu}\frac{I + \varrho N}{N} k m N \quad . \tag{19}$$

This expression can be substituted for V in the human disease dynamics equation in system (11)

$$\frac{d}{dt}I = \frac{\beta}{mN} S V(I) - \gamma I \quad . \tag{20}$$

Considering that in a non-endemic area almost the entire population is susceptible, hence $S(t) \approx S_0 \approx N$, the equation can be written as

$$\frac{d}{dt}I = (k\beta_{eff} - \gamma)I + k\beta_{eff}\varrho N \quad , \tag{21}$$

where $\beta_{eff} := \beta\frac{\vartheta}{\nu}\frac{S_0}{N} \approx \beta_{SIR}$ is the effective infection rate. Note that in the invasion scenario k can still modulate the infectivity.

For $k\beta_{eff} < \gamma$, i.e., below the epidemiological threshold for exponential growth upon index cases ($\beta_{eff} = \gamma$), we have, triggered by import, a sub-critical stationary state of the dynamics given by Equation (21)

$$I^* = k\beta_{eff}\frac{1}{\gamma - k\beta_{eff}}\varrho N \quad , \tag{22}$$

where $\gamma - k\beta_{eff} =: \varepsilon(k)$ is defined as the distance away from the exponential growth threshold with self-sustained spreading in the local population, depending on the relative mosquito abundance k. Hence, the expected number of autochthonous cases, given the underlying stochastic process is

$$\langle I \rangle^*(k, \varrho) = k\,\beta_{eff}\,\frac{1}{\varepsilon(k)}\varrho\,N \quad . \tag{23}$$

This expression depends on the mosquito abundance k, in the invasion region and the import of infection ϱ. Although $\langle I \rangle$ is typically much smaller than one case, stochastic simulations can reveal isolated autochthonous cases and occasionally rare clusters of outbreaks, where more than one case occurs simultaneously, as illustrated in Figure 1.

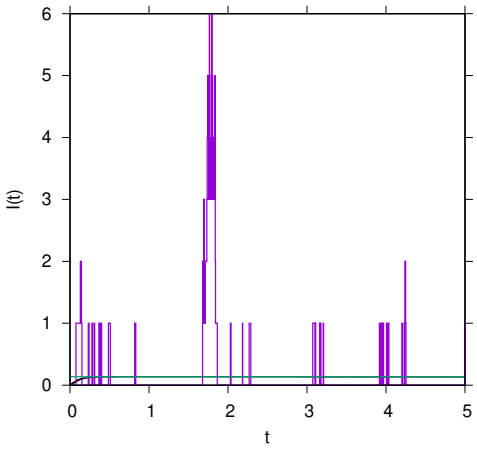


Figure 1: One stochastic realization (in magenta), the trajectory towards the stationary state of the SIRUV model (in black), and the approximation solution using the equation (in green) with $k = 40\%$ as the proportion of mosquitoes relative to endemic countries. For $N = 10^4$ and $\varrho = 10^{-5}$, the parameter values used in this simulation are listed in Table 1.

2.5 The SIRUV model with explicit dynamics for imported cases Y

Aimed at understanding the impact of imported cases on the dynamics of stochastic realizations, we extend the SIRUV framework by explicitly incorporating the dynamics of imported cases Y, to study their stochastic fluctuations.

The SIRUVY model can be expressed as

$$\begin{aligned}
\frac{d}{dt}S &= \alpha R - \frac{\beta}{mN}SV \\
\frac{d}{dt}I &= \frac{\beta}{mN}SV - \gamma I \\
\frac{d}{dt}R &= \gamma I - \alpha R \\
\frac{d}{dt}U &= \psi(k) - \frac{\vartheta}{N}U(I+Y) - \nu U \\
\frac{d}{dt}V &= \frac{\vartheta}{N}U(I+Y) - \nu V \\
\frac{d}{dt}Y &= \tilde{\varrho}N - \gamma Y \quad ,
\end{aligned} \tag{24}$$

where $\tilde{\varrho}$ represents the constant input rate of imported cases, which can recover at rate γ. Note that the last equation is decoupled from System (24), and consequently, the stationary value of the imported cases Y^*, is given by

$$Y^* = \frac{\tilde{\varrho}}{\gamma}N = \varrho N \quad , \tag{25}$$

consistent with the previous model in System (11).

When considering the cumulative imported cases C_Y, over a given time interval, we have

$$\frac{d}{dt}C_Y = \tilde{\varrho}N \quad . \tag{26}$$

The numerical solutions for both the number of imported cases $Y(t)$, and the cumulative imported cases $C_Y(t)$, are illustrated in Figure 2. With initial conditions set to zero imported cases $Y(t_0) = 0$, we observe that the mean field solution of the system described by Equation (24) converges quickly to its stationary state, as shown in Figure 2(a). After five years of simulation, the cumulative number of imported cases is expected to be approximately 26 cases, see Figure 2(b).

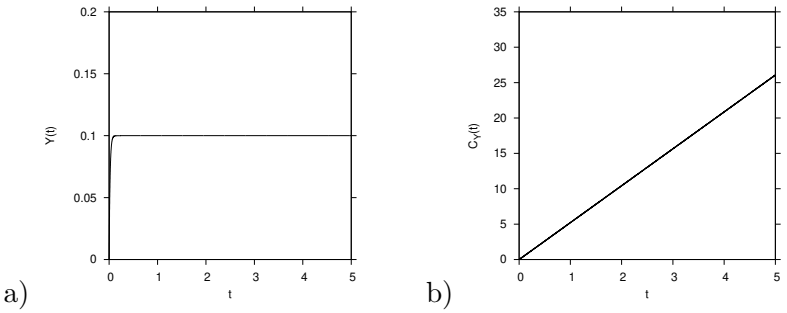


Figure 2: Expected values from the mean field solution over a 5-year simulation period. For $N = 10^4$ and $\varrho = 10^{-5}$, the parameter values used in this simulation are listed in Table 1. In (a), the imported cases and in (b), the cumulative cases $C_Y(t)$.

2.6 The stochastic SIRUVY model

Following the stochastic approach described in Section 2.3, we model the SIRUVY system using the master equation (see equation (12)). The stochastic model is characterized by the extended state vector $\underline{X} := (S, I, R, U, V, Y)^{\mathrm{tr}}$ and includes the following $n = 9$ transitions, each with its corresponding shift vectors

$$\begin{array}{lll}
w_1(\underline{X}) = \alpha R & , & \underline{r}_1 = (-1,0,1,0,0,0)^{tr} \\
w_2(\underline{X}) = \frac{\beta}{mN} SV & , & \underline{r}_2 = (1,-1,0,0,0,0)^{tr} \\
w_3(\underline{X}) = \gamma I & , & \underline{r}_3 = (0,1,-1,0,0,0)^{tr} \\
w_4(\underline{X}) = \psi & , & \underline{r}_4 = (0,0,0,-1,0,0)^{tr} \\
w_5(\underline{X}) = \frac{\vartheta}{N} U(I+Y) & , & \underline{r}_5 = (0,0,0,1,-1,0)^{tr} \\
w_6(\underline{X}) = \nu U & , & \underline{r}_6 = (0,0,0,1,0,0)^{tr} \\
w_7(\underline{X}) = \nu V & , & \underline{r}_7 = (0,0,0,0,1,0)^{tr} \\
w_8(\underline{X}) = \tilde{\varrho} N & , & \underline{r}_8 = (0,0,0,0,0,-1)^{tr} \\
w_9(\underline{X}) = \gamma Y & , & \underline{r}_9 = (0,0,0,0,0,1)^{tr} \quad .
\end{array} \tag{27}$$

Similarly to the deterministic approach, the dynamics of imported cases Y, can be described on its own by the following master equation

$$\frac{d}{dt}\, p(Y,t) = \tilde{\varrho} N\, p(Y-1,t) + \gamma(Y+1)\, p(Y+1,t) - (\tilde{\varrho} N + \gamma Y)\, p(Y,t) \quad . \tag{28}$$

For the cumulative number of imported cases C_Y, using the entry transitions of the stochastic dynamics of Y, we have the master equation given by

$$\frac{d}{dt}\, p(C_Y, t) = \tilde{\varrho} N\; p(C_Y - 1, t) - \tilde{\varrho} N\; p(C_Y, t) \quad . \tag{29}$$

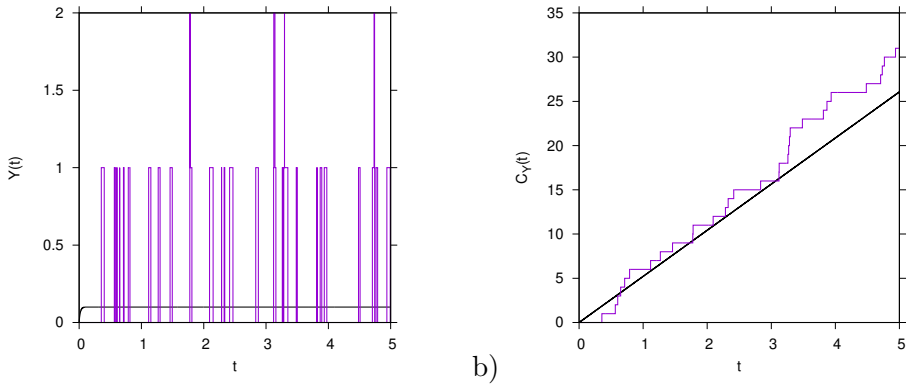
This represents a simple Poisson process with a known distribution in time evolution [47, 19].

The analytical solution to the master equation (29), given the initial condition $C_Y(t_0)$ at time t_0, is described by the conditional probability

$$p(C_Y, t | C_Y(t_0), t_0) = \frac{(\lambda(t - t_0))^{C_Y - C_Y(t_0)}}{(C_Y - C_Y(t_0))!}\; e^{-\lambda(t - t_0)} \quad , \tag{30}$$

where $\lambda := \tilde{\varrho} N = \varrho \gamma N$ is constant. This represents a Poisson distribution for each time point t.

Figure 3(a) illustrates the differences between the mean field approximation and a stochastic realization for imported cases. We observe that, in contrast to the mean field approximation (in black), where the solution converges to a constant infected population, on average less than one infected case, the stochastic realization of imported cases (in magenta) can exhibit time intervals with no cases and others with multiple cases. This pattern is also observed in the cumulative imported cases, as shown in Figure 3(b).



a) b)

Figure 3: One stochastic realization (in magenta) and the mean field solution (in black) of (a) imported cases $Y(t)$, and (b) cumulative imported cases $C_Y(t)$. For $N = 10^4$ and $\varrho = 10^{-5}$, the parameter values used in this simulation are listed in Table 1.

Motivated by stochastic modeling, we apply statistical methods to model imported cases and integrate these models with our risk estimator. Let τ

denote the waiting time between two consecutive imported cases, $C_Y(t_{n-1})$ to $C_Y(t_n)$. The probability distribution of τ is given by

$$p(\tau) = \lambda\, e^{-\lambda\tau} \quad , \tag{31}$$

where $\lambda := \varrho\gamma N$ is the rate parameter. This distribution allows us to analyze the statistical characteristics of imported case notifications based on the waiting times between their occurrences.

2.6.1 Parameter estimation from waiting time distribution

Given the recorded imported cases and the time intervals between consecutive cases, represented as the data vector $\underline{\tau} = (\tau_1, \tau_2, \ldots, \tau_n)$, we estimate the parameter λ using the log-likelihood method.

The likelihood function for the waiting times is given by

$$p(\tau_1, \tau_2, \ldots, \tau_n) = \prod_{i=1}^{n} \lambda\, e^{-\lambda\tau_i} = L(\lambda) \quad , \tag{32}$$

where the log-likelihood function is maximized to determine the maximum likelihood estimator

$$\hat{\lambda} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} \tau_i} \quad . \tag{33}$$

From the given data, we order the waiting times to produce the observed distribution function (shown in magenta in Figure 4). This is then compared to the cumulative distribution function

$$P(\tau) := \int_0^{\tau} p(\tilde{\tau})\, d\tilde{\tau} = 1 - e^{-\lambda\tau} \quad , \tag{34}$$

using both the known parameter $\lambda = \varrho\gamma N$ (represented by the black line in Figure 4) and the maximum likelihood estimator $\hat{\lambda}$ (middle green line in Figure 4).

We then quantify the confidence interval of the estimation, acknowledging that simulated data may deviate from the true parameters. Once the best estimate of λ, $\hat{\lambda}$, is determined, the expected number of imported cases can be expressed as

$$\langle Y^* \rangle = \frac{\hat{\lambda}}{\gamma} \quad . \tag{35}$$

Although the risk estimator refers to average values, the model accounts for the intrinsic stochasticity of the time series.

2.6.2 Confidence intervals from the likelihood function

Using the negative inverse Fisher matrix, which simplifies to a 1×1 matrix in the case of a single parameter, we can approximate the variance of a Gaussian distribution by examining the curvature of the log-likelihood function around its maximum. This variance is given by

$$\sigma^2 = -\frac{1}{\frac{d^2}{d\lambda^2} \ln L\Big|_{\lambda=\hat{\lambda}}} \quad . \tag{36}$$

For the exponential waiting time likelihood

$$L(\lambda) = p(\underline{\tau} \mid \lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} \tau_i} \quad , \tag{37}$$

the variance simplifies to

$$\sigma^2 = \frac{\hat{\lambda}^2}{n} \quad , \tag{38}$$

where n is the number of data points. This provides a 2σ confidence interval, which approximately covers 95% of the Gaussian distribution

$$\lambda = \hat{\lambda} \pm 2\sigma \quad , \tag{39}$$

as shown in Figure 4).

The estimate of λ can be applied to the cumulative distribution function $P(\tau)$. In Figure 4, $P(\tau_i)$ is obtained by ordering the τ_i values from smallest to largest (x-axis) and plotting $\frac{i}{n}$ on the y-axis (in magenta), where the τ_i values are derived from simulations with $\lambda = \varrho\gamma N$. The figure also includes the maximum likelihood estimate $P(\tau) = 1 - e^{-\hat{\lambda}\tau}$ (middle green line) along with its 95% confidence intervals (upper and lower green lines). Additionally, the theoretical cumulative distribution function $P(\tau) = 1 - e^{-\varrho\gamma N\tau}$ is shown (black line).

For a more thorough analysis of estimation uncertainties, Bayesian methods using conjugate priors, such as a Gamma distribution for the exponential case, could be used. The Bayesian posterior $p(\lambda \mid \underline{\tau})$ may provide asymmetric confidence intervals. For further details on these methods, see [47], as well as Appendices A and B for a more in-depth discussion in the present context.

Note that the current analysis of inter-event times is preliminary, providing a basic guide for stochastic simulations. As more data become available, the analysis can be refined and more sophisticated statistical methods can be applied.

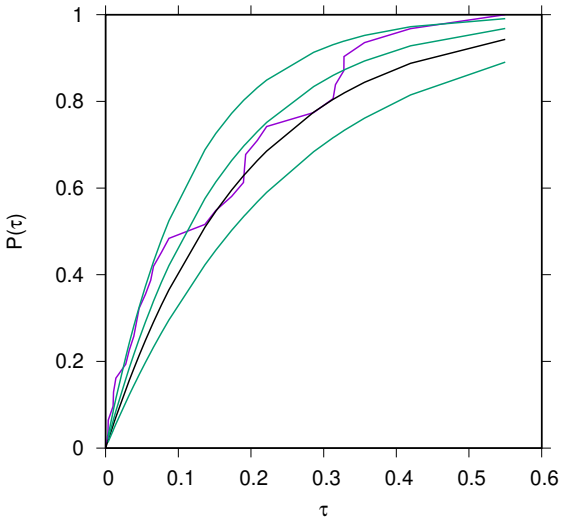


Figure 4: The empirical cumulative distribution function is shown in magenta, while the maximum likelihood estimate and its 95% confidence intervals are represented by the green lines. The theoretical cumulative distribution, represented by a black line, is known in simulations but not in empirical data analysis. The numerical values are $\hat{\lambda} = 6.276\text{y}^{-1}$ and $\varrho\gamma N = 5.214\ \text{y}^{-1}$.

2.6.3 Derivation of the maxima likelihood estimate variance and theoretical background

For large data sets, the likelihood function approximates a Gaussian distribution around the maximum log-likelihood estimator. Specifically, the likelihood function $L(\lambda)$ can be approximated by a Gaussian distribution $p_G(\lambda)$, given by

$$L(\lambda) \sim \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\lambda-\hat{\lambda})^2}{2\sigma^2}} =: p_G(\lambda) \quad , \tag{40}$$

where $\hat{\lambda}$ is the mean value and σ^2 is the variance, determined by the curvature of the log-likelihood function around the maximum. This curvature is quantified by the second derivative of the log-likelihood function at the maximum, expressed as

$$\frac{d^2}{d\lambda^2} \ln L \bigg|_{\lambda=\hat{\lambda}} \quad . \tag{41}$$

For the Gaussian approximation, the variance is

$$\sigma^2 = -\frac{1}{\frac{d^2}{d\lambda^2}\ln L\Big|_{\lambda=\hat{\lambda}}} \quad . \tag{42}$$

Applying this to the likelihood function of the exponential distribution

$$L(\lambda) = \lambda^n e^{-\lambda\sum_{i=1}^{n}\tau_i} \quad , \tag{43}$$

we obtain

$$\frac{d^2}{d\lambda^2}\ln L\Big|_{\lambda=\hat{\lambda}} \quad = \quad \frac{d^2}{d\lambda^2}\left(n\ln(\lambda) - \lambda\sum_{i=1}^{n}\tau_i\right)\Big|_{\lambda=\hat{\lambda}} = -\frac{n}{\hat{\lambda}^2} \quad . \tag{44}$$

Thus, the variance is

$$\sigma^2 = -\frac{1}{\frac{d^2}{d\lambda^2}\ln L\Big|_{\lambda=\hat{\lambda}}} = \frac{\hat{\lambda}^2}{n} \quad . \tag{45}$$

The standard deviation is $\sigma = \frac{\hat{\lambda}}{\sqrt{n}}$, which can be used to determine the confidence interval for λ as $\lambda = \hat{\lambda} \pm 2\sigma$. While this approach is practically useful, a more fundamental description and justification is provided in Appendix B through the Cramér-Rao consideration, which offers only an inequality as a lower bound for Equation (45).

3 Mosquito abundance and its impact on disease transmission

Analyzing mosquito abundance is crucial for predicting the transmission of mosquito-borne diseases, but it presents significant challenges. To address these complexities, a practical approach is to establish a threshold for mosquito abundance. Below this threshold, the likelihood of local disease transmission is minimal, while exceeding it significantly increases the risk of autochthonous cases.

3.1 Explicit calculation of Poisson process approximation for autochthonous cases

To model the dynamics of autochthonous infected cases, we start with the differential equation

$$\frac{d}{dt}I = k\beta\frac{\vartheta}{\nu}\frac{S_0}{N}(I + \varrho N) - \gamma I \quad , \tag{46}$$

where $I^* = k \cdot \beta \frac{\vartheta}{\nu} \frac{1}{\varepsilon} \cdot \varrho N$, see Equation (23), and $\varepsilon \to \gamma$. For small k, the equilibrium value I^* approaches zero, simplifying the dynamics.

The cumulative number of autochthonous cases C_I then follows

$$\frac{d}{dt} C_I = k\beta \frac{\vartheta}{\nu} \frac{S_0}{N} \varrho N =: \lambda_I \quad , \tag{47}$$

where λ_I represents the rate of new autochthonous cases.

Thus, the stochastic process for C_I can be approximated by a Poisson process. The master equation for $p(C_I, t)$ is

$$\frac{d}{dt} p(C_I, t) = \lambda_I \, p(C_I - 1, t) - \lambda_I \, p(C_I, t) \quad , \tag{48}$$

which has the Poisson distribution solution

$$p(C_I, t \mid C_I(t_0) = 0, t_0) = \frac{(\lambda_I (t - t_0))^{C_I}}{C_I!} \, e^{-\lambda_I (t - t_0)} \quad , \tag{49}$$

Where $C_I(t_0) = 0$ and $T = t - t_0$ represents the time interval, similar to the analysis of imported cases, denoted by C_Y.

The probability of having no autochthonous cases is

$$p(C_I = 0, T) = e^{-\lambda_I T} = e^{-k\beta \frac{\vartheta}{\nu} \frac{S_0}{N} \varrho N \, T} \quad , \tag{50}$$

which approaches 1 as k becomes small.

The probability of having one or more autochthonous cases is

$$p(C_I \geq 1, T) = 1 - p(C_I = 0, T) = 1 - e^{-k \cdot \beta \frac{\vartheta}{\nu} \frac{S_0}{N} \varrho N \cdot T} =: p_{th} \quad , \tag{51}$$

where p_{th} is a threshold probability. Solving for k yields

$$k_{th} = \frac{-\ln(1 - p_{th})}{\beta \frac{\vartheta}{\nu} \frac{S_0}{N} \cdot \varrho N \cdot T} \quad , \tag{52}$$

which gives the mosquito abundance k_{th} needed to achieve a certain probability p_{th} of having one or more autochthonous cases. This can be used for different values of p_{th}, such as $p_{th} = \frac{9}{10}$ or $p_{th} = \frac{1}{10}$, in subsequent analyses.

3.2 Analysing the mosquito abundance threshold

In this section, we analyze the bounds for relative mosquito abundance based on theoretical scenarios. We assume $S_0 = N$, $\varrho = 10^{-5}$, $N = 10^4$,

and $Y^* = 0.1$. The threshold values for mosquito abundance k, for a given probability p_{th} are determined using Equation (52).

The relationship between imported cases and local transmission, emphasizing the role of mosquito abundance in determining the probability and timing of autochthonous cases, is illustrated in Figure 5. In a real-world scenario, this is evident in the comparison between the Basque Country, which has reported zero non-travel-related cases so far, and Italy, which had over 80 cases notified in 2023 [11], highlighting the different risks potentially associated with varying mosquito densities and environmental conditions.

Considering a probability threshold of $p_{th} = \frac{1}{10}$, which means that in 1 out of 10 stochastic runs we expect at least one autochthonous case within the time interval $T = 1$ year, we obtain

$$k_{th} = \frac{-\ln(1-p_{th})}{\beta \frac{\vartheta}{\nu} \frac{S_0}{N} \cdot \varrho N \cdot T} = 0.01403 \approx 1.4\% \quad . \tag{53}$$

For a higher probability threshold of $p_{th} = \frac{9}{10}$, the value of k_{th} is approximately 31%. This indicates that we would nearly always observe at least one case in stochastic simulations within the first year.

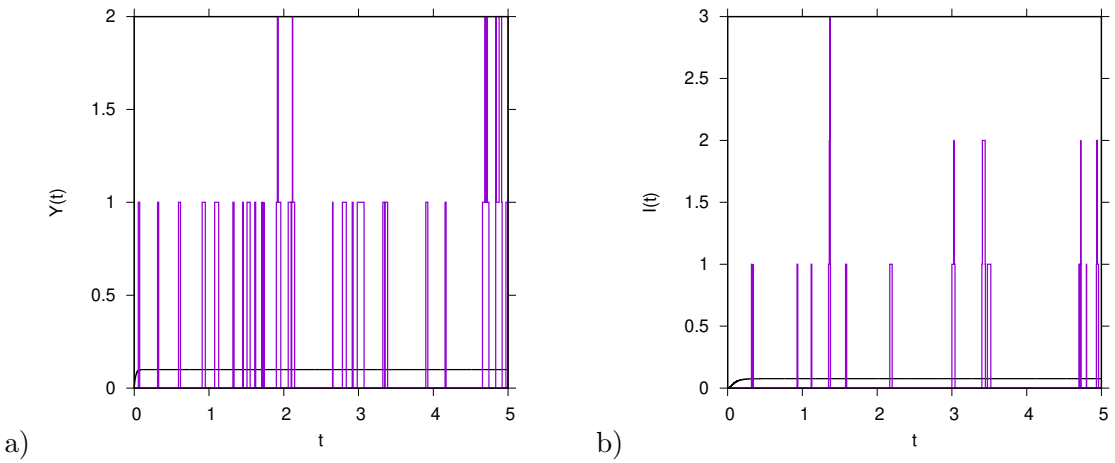


Figure 5: For mosquito abundance $k = 30\%$, $\varrho = 10^{-5}$, and $N = 10^4$, in a) imported cases $Y(t)$ and b) autochthonous cases $I(t)$. Stochastic simulation is shown in magenta and its mean-field solution in black.

Figure 5 illustrates the relationship between imported cases and local transmission, highlighting how mosquito abundance impacts the likelihood and timing of autochthonous cases to occur. For a mosquito abundance

of $k = 30\%$, Figure 5(a) shows the time series of imported cases $Y(t)$, while Figure 5(b) presents the time series of autochthonous cases $I(t)$. Two autochthonous cases occur in the first year (shown in magenta), but periods of up to a year with no autochthonous cases also occur, illustrating the stochastic nature of disease transmission even at relatively high mosquito abundance.

The Poisson distribution for the threshold value $k_{th} = 0.3067 \approx 30\%$ is presented in Figure 6, showing the probability of no autochthonous cases being $p(C_I = 0, T) = 0.1$ and the probability of one or more cases being $p_{th} = 0.9$.

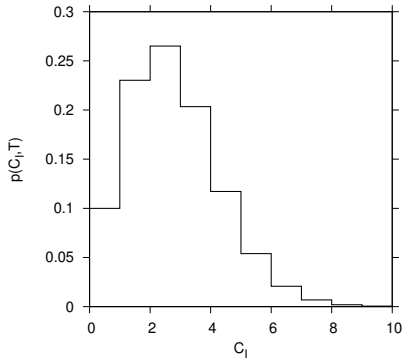


Figure 6: Poisson distribution for $k_{th} = 0.3067 \approx 30\ \%$.

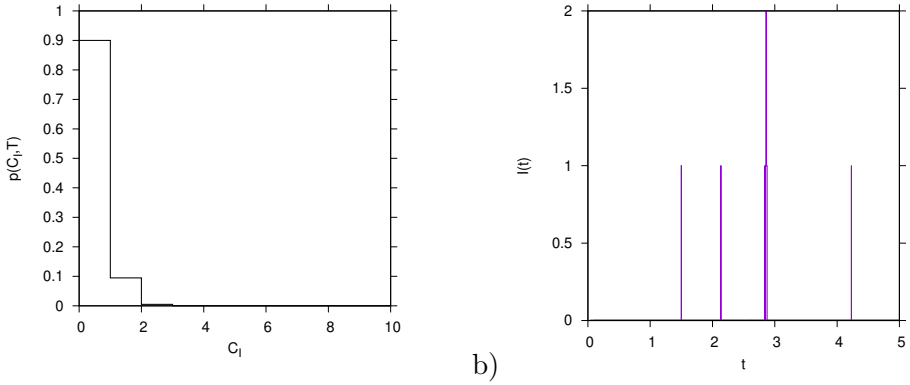


Figure 7: a) Poisson distribution for $k_{th} = 0.01403 \approx 1.4\%$. b) Time series of autochthonous cases $I(t)$, for a mosquito abundance of $k = 2\%$.

Conversely, for the threshold value $k_{th} = 0.01403 \approx 1.4\%$, Figure 7(a) demonstrates that the probability of observing zero autochthonous cases is $p(C_I = 0, T) = 0.9$, leading to $p_{th} = 0.1$ for one or more cases. Figure 7(b) indicates that with this threshold, we nearly always observe no autochthonous cases within the first year and rarely any cases in subsequent years.

While the Poisson process approximation for the autochthonous cases $I(t)$, especially for small mosquito abundance k, provides an indication of the expected frequency and duration without cases, the full stochastic model simulations reveal more clustering of cases. For example, Figure 5(b) shows up to three cases occurring simultaneously, and Figure 7(b) shows two cases at the same time. This clustering may be influenced by state-dependent factors in the stochastic variables w_j that are not captured by the basic Poisson approximation.

3.3 Analysis of the distance from exponential growth $\varepsilon(k)$

For small mosquito abundance k, the term $\varepsilon(k)$, defined in Section 2.4, can be approximated by the recovery rate γ. This approximation is illustrated in Figure 8, where k values are shown on a logarithmic scale (base 10). For instance, for $k = 1\%$, the approximation $\varepsilon \approx \gamma$ is already quite accurate, with $\gamma \approx 52$ y^{-1}.

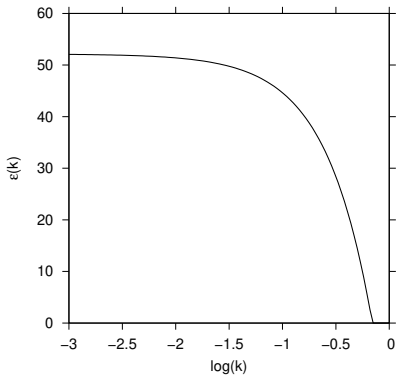


Figure 8: Values of $\varepsilon(k)$, the distance to the threshold of exponential growth, converge towards $\gamma = 52\text{y}^{-1}$ as the relative mosquito abundance k, decreases.

In an invasion scenario with relatively low mosquito abundance and no observed autochthonous cases, the expected number of autochthonous cases

$\langle I \rangle$ can be quantified using the simplified expression

$$\langle I \rangle(k, \varrho) = k\, \xi\, \varrho N \tag{54}$$

where $\xi = \frac{\beta}{\gamma}\frac{\vartheta}{\nu}\frac{S_0}{N}$ is constant. For a naive, completely susceptible population, $\frac{S_0}{N} = 1$. Thus, ξ can be compared to the basic reproduction number typically expected in vector-borne disease models. Consequently, our risk estimator is given by Equation (54).

It is important to note that, due to the presence of several imported cases over time, we might still expect occasional autochthonous cases, even if they occur infrequently. Additionally, observed increases in mosquito abundance over the years [11] could impact this risk.

The primary focus of the risk measure is to determine the probability of observing the first autochthonous case in a non-endemic area. Although the calculated expectation $\langle I \rangle$ is well below one (see Figure 1), stochastic simulations can provide insights into when the first cases might occur, depending on the levels of imported cases and mosquito abundance. We will analyze scenarios ranging from low to intermediate mosquito abundance and examine conditions where the approximation $\varepsilon \approx \gamma$ no longer holds. In such cases, the non-linearity of $\varepsilon(k)$ can lead to increasing clusters of outbreaks until the threshold for exponential growth and self-sustained transmissibility is reached.

3.4 The risk of outbreaks in increasing mosquito abundance scenarios

In this section, we analyze scenarios where mosquito abundance increases beyond the simplest approximation explored previously. To analyze deviations from the simplest linear relationship of the expected number of autochthonous infections $\langle I \rangle^*$ with respect to mosquito abundance k, we need to consider how the effective transmission rate $\varepsilon(k)$ changes as k increases.

3.4.1 Deviations from the simplest linear relation of expected autochthonous infected on mosquito abundance

The simplest risk estimator for the expected number of autochthonous infections $\langle I \rangle^*$ is given by

$$\langle I \rangle^* = k \cdot \beta_{eff} \cdot \frac{1}{\varepsilon} \cdot \varrho N \quad , \tag{55}$$

where ε is the distance to the threshold of exponential growth and is given by

$$\varepsilon(k) = \gamma - k \cdot \beta \frac{\vartheta}{\nu} \cdot \frac{S_0}{N} \quad . \tag{56}$$

In the initial phases of outbreaks, where $S_0 \approx N$, and as long as $\varepsilon > 0$, the expression $\varepsilon(k) \to \gamma$ as $k \to 0$ indicates that the distance to the threshold of exponential growth approaches the recovery rate γ.

However, deviations from the simplest linear relationship become significant as mosquito abundance k approaches the critical threshold k_c. In this regime, the linear approximation of $\langle I \rangle^* \sim \frac{k}{\varepsilon(k)} \approx \frac{k}{\gamma}$ fails to hold, and the expected number of autochthonous cases can increase very rapidly.

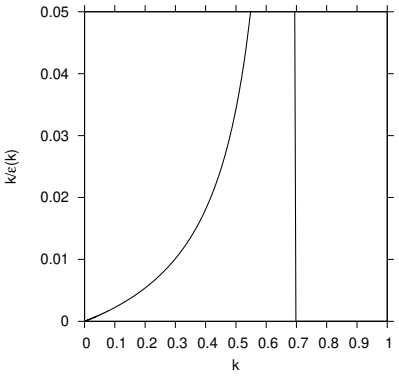


Figure 9: Graphical representation of the expression $\langle I \rangle^* \sim \frac{k}{\varepsilon(k)}$ as a function of the relative mosquito abundance k (bent curve). The vertical line represents the critical value k_c.

As shown in Figure 9, the relationship between $\langle I \rangle^* \sim k/\varepsilon(k)$ and mosquito abundance reveals that for small k the expression $k/\varepsilon(k)$ approximately behaves linearly as k/γ. This is expected since $\varepsilon(k) \approx \gamma$ for small values of k. Nevertheless, close to the critical threshold k_c there is a significant non-linear growth in $\langle I \rangle^*$. As k approaches k_c, $\varepsilon(k)$ decreases rapidly, causing $\langle I \rangle^*$ to rise sharply. This non-linearity is due to the rapid approach of $\varepsilon(k)$ towards zero, leading to a divergence in $\langle I \rangle^*$. These findings highlight the importance of accounting for the non-linearity of $\varepsilon(k)$ in disease transmission models and emphasize the need for careful empirical observation of mosquito abundances, as they may increase from year to year.

3.4.2 Temporal dynamics and scaling close to the epidemiological threshold

The dynamics described in Equation (21) for the sub-critical regime reads

$$\frac{d}{dt}I \quad = \quad -(\gamma - k\beta_{eff}) \cdot I + k\beta_{eff} \cdot \varrho N \tag{57}$$

and can be solved, yielding its time-dependent solution given by

$$I(t) = I(t_0) \cdot e^{-(\gamma - k\beta_{eff}) \cdot (t - t_0)} + \frac{k\beta_{eff} \cdot \varrho N}{\gamma - k\beta_{eff}} \left(1 - e^{-(\gamma - k\beta_{eff}) \cdot (t - t_0)}\right) \quad . \tag{58}$$

When the mosquito abundance k reaches its threshold value k_c, given by $\gamma - k_c \cdot \beta_{eff} = 0$, the time-dependent solution is

$$I(t) = I(t_0) + k\beta_{eff} \cdot \varrho N \cdot (t - t_0) \tag{59}$$

i.e. with the distance to threshold $\varepsilon = \gamma - k\beta_{eff} = 0$. This implies that close to the threshold, the total number of infected individuals becomes asymptotically self-similar, i.e., a homogeneous function of the general form

$$I(t) = t^{\hat{\theta}} \cdot F\left(\varepsilon t^{\frac{1}{\hat{\nu}_{\|}}}, \varrho t^{\frac{\hat{\mu}}{\hat{\nu}_{\|}}}\right) \quad . \tag{60}$$

The critical exponents $\hat{\theta}$, $\hat{\nu}_{\|}$, and $\hat{\mu}$ characterize the self-similar, scale-invariant behavior of the infection spread near the critical threshold, governing the growth, time evolution, and density fluctuations of the infected population [48].

Rescaling in terms of the parameter ε, we have

$$I(t) = \varepsilon^{-\hat{\theta}\hat{\nu}_{\|}} \cdot F(t\varepsilon^{\hat{\nu}_{\|}}, \varrho\varepsilon^{-\hat{\mu}}) \quad . \tag{61}$$

In the mean-field regime, i.e., no spatial effects, the exponents are $\hat{\theta} = 0$, $\hat{\nu}_{\|} = 1$, and $\hat{\mu} = 1$ (see [48] for further details on general scaling theory around critical thresholds).

For empirical epidemiological systems, this results in distributions of avalanches with power-law tails, leading to larger numbers of autochthonous cases more frequently than expected, for example, in classical Gaussian distributions.

Thus, from a management point of view, once mosquito abundance increases and leads to larger clustered outbreaks, even small additional increases in mosquito numbers will result in significantly larger outbreaks of autochthonous cases.

4 Applying the model to Basque Country data

The Basque Country is an autonomous region in northern Spain with approximately 2.2 million inhabitants. In this region, the Public Health Epidemiological Unit, in collaboration with NEIKER - the Basque Institute for Agricultural Research and Development - monitors records of both imported disease cases and the distribution and establishment of disease vectors. Although the Basque Country is considered a non-endemic area for tropical mosquito-borne diseases, it has established populations of *Aedes* mosquitoes (with increasing distribution) and frequently reports viremic imported cases of dengue, Zika, and Chikungunya. This makes it an ideal setting for this study.

4.1 Imported cases of *Aedes* mosquito-borne diseases in the Basque Country in 2019 (Pre-COVID) and 2022 (Post-COVID)

The raw empirical data of cumulative imported cases of dengue, Zika and chikungunya in the Basque Country for the years of 2019 and 2022, provided by the Health Basque Department, is illustrated in Figure 10.

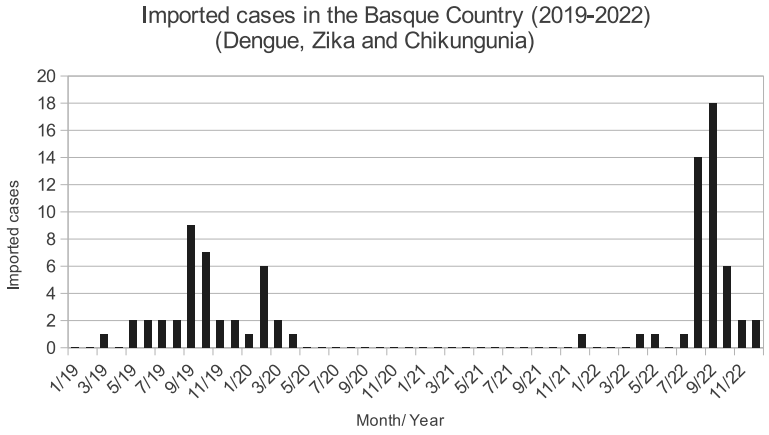


Figure 10: Epidemiological records of imported disease cases in the Basque Country from 2019 to 2022.

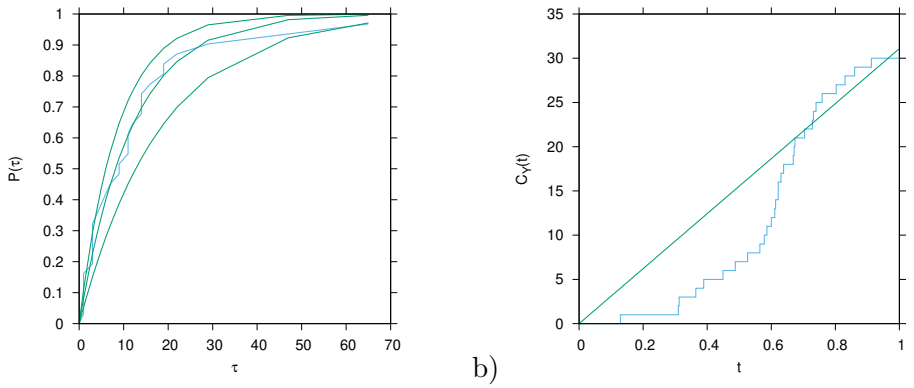
The data reveals a steady registration of imported cases throughout the year, with a notable increase during the summer months. This seasonal surge coincides with the period of peak mosquito activity, which can contribute to

higher disease transmission rates. Note that for the years of 2020 and 2021, very few imported cases were recorded since travel restrictions were in place due to the COVID-19 pandemic.

4.1.1 Estimation of imported cases

Applying the statistical analysis described in Section 2.6.2, we obtain the inter-event cumulative distribution of imported cases, and the cumulative counts of imported cases of dengue, chikungunya, and Zika, i.e. *Aedes* mosquito-transmitted diseases notified in the Basque Country for the years 2019. For 2022, there were no significant differences in the number of imported cases compared to 2019; the numbers remained relatively similar, although travel activity was higher in the latter half of the year. The primary difference was the increase in mosquito abundance, as illustrated in Figure 12.

For 2019, Figure 11 (a) shows the inter-event cumulative distribution of imported cases $P(\tau)$ for dengue, chikungunya, and Zika in the Basque Country (in blue), alongside the inter-event rate $\lambda = (0.085 \pm 0.031)\ \mathrm{d}^{-1}$ (in green). Figure 11 (b) shows the cumulative number of imported cases $C_Y(t)$ for dengue, chikungunya, and Zika (in blue) and the estimated mean number of cumulative imported cases $\langle C_Y \rangle = \hat{\lambda}(t - t_0)$ (in green).



a) b)

Figure 11: For the year of 2019, in (a) the inter-event cumulative distribution of imported cases $P(\tau)$ (in blue), alongside the inter-event statistics (in green). In (b) the cumulative number of imported cases $C_Y(t)$ (in blue), and the estimated mean number of cumulative imported cases $\langle C_Y \rangle = \hat{\lambda}(t - t_0)$ (in green). The inter-event times τ_i include the first and the last day of the year.

The estimation of imported cases in the Basque Country is derived from the data on imported cases using maximum likelihood estimators. The estimation of the parameter λ is given by Equation (33). The confidence interval for λ is calculated as $\sigma = \frac{\hat{\lambda}}{\sqrt{n}}$. For the Basque Country in 2019, this yields

$$\lambda = \hat{\lambda} \pm 2\sigma = 0.085 \pm 0.031 \text{ day}^{-1} \tag{62}$$

Additionally, data from the post-COVID-19 pandemic period in 2022 provides an estimated mean number of imported cases

$$\lambda = 0.128 \pm 0.037 \text{ day}^{-1} \quad . \tag{63}$$

This indicates some increase in the number of imported cases in 2022. Variations in the risk of autochthonous cases across different years are evident, with a clear upward trend. Differences in trends may also be observed at different spatial scales, such as by province, though these differences are generally not substantial, especially if confidence intervals overlap. Furthermore, variations over the years show lower risks before August and after September, with increased risks during the summer months, aligning with the period of higher mosquito activity.

The expected number of imported cases $\langle Y \rangle^*$ is related to the Poisson rate λ as follows

$$\langle Y \rangle^* = \varrho N = \frac{\lambda}{\gamma} \quad . \tag{64}$$

As an example, using the results for 2019 in the Basque Country, where $\lambda = 0.1$, we obtain

$$\langle Y \rangle^* = \frac{0.1 \text{ day}^{-1}}{\frac{1}{7 \text{ day}}} = 0.1 \cdot 7 = 0.7 \tag{65}$$

This initial statistical analysis provides insights into the trends of imported cases in the Basque Country, offering preliminary estimates of parameters critical for assessing the risk of mosquito-borne diseases in the region.

To further refine this analysis, a potential improvement would involve seasonal breakdowns, such as spring, summer, and autumn, to capture variations in case numbers. However, this approach faces challenges, including a smaller dataset for each season and the lack of robust patterns due to significant regional variations. Additionally, the sparse data for 2020 and 2021, due to travel restrictions from the COVID-19 pandemic, limits the analysis. Future data may reveal a return to more typical patterns and provide clearer trends.

4.1.2 Mosquito abundance data in the Basque Country

In the Basque Country, the Basque Institute for Agricultural Research and Development (NEIKER) has been monitoring mosquito breeding sites and recording egg count data across various localities for over a decade [49], providing increasingly detailed insights into mosquito population trends.

The data show a significant increase in egg counts from ovitraps in recent years. Notably, the counts in the most populated traps have reached levels comparable to those in endemic regions, with maximum counts nearing $1,000$ eggs. However, since the ovitraps are not yet standardized across temporal and spatial scales, comparing mosquito abundance between invasion scenarios and endemic areas remains challenging.

Given the increasing number of traps deployed each season, this analysis focuses on the sum of the highest 20 egg counts as a preliminary indicator of mosquito abundance in the Basque Country for 2019 and 2022.

Figure 12 shows the 20 highest egg counts in the Basque Country. There is a noticeable difference between 2019 (pre-COVID-19 pandemic) and 2022 (post-COVID-19 pandemic). In 2022, the maximum egg counts reached 1000 eggs, with a mean of approximately 530, while in 2019, the maximum count was only a quarter of that, with a mean around 140 eggs.

Since no autochthonous cases of *Aedes* mosquito-borne diseases have yet been detected in the region, these egg counts serve as initial indicators of mosquito abundance. These results can be applied to calculate the relative risk of disease transmission, considering the presence of imported cases and mosquito levels in the Basque Country.

4.1.3 Risk of autochthonous cases in the Basque Country

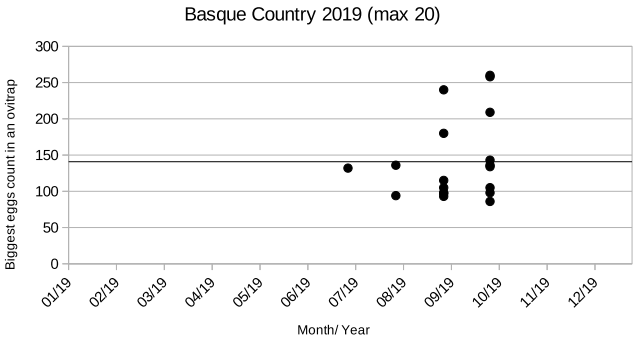
The expected risk of autochthonous cases $\langle I \rangle^*$ is given by

$$\langle I \rangle^* = k \cdot \beta_{eff} \frac{1}{\varepsilon} \cdot \langle Y \rangle^* = f(k, \langle Y \rangle^*), \tag{66}$$

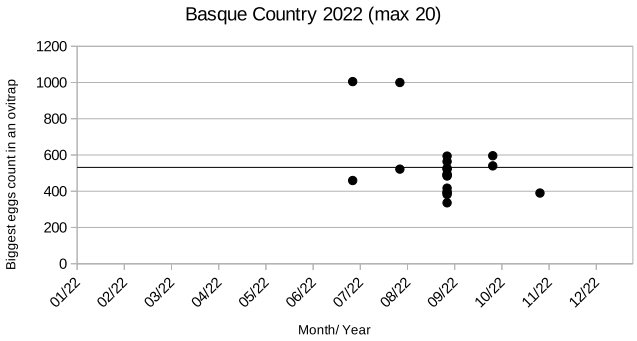
and for small mosquito abundance k, where $\varepsilon(k) \approx \gamma$, this simplifies to

$$\langle I \rangle^* = k \cdot \beta_{eff} \frac{1}{\gamma} \cdot \frac{\lambda}{\gamma} = k \cdot \frac{\beta}{\gamma} \frac{\vartheta}{\nu} \cdot \frac{\lambda}{\gamma} \quad . \tag{67}$$

Here, $\frac{\beta}{\gamma} \frac{\vartheta}{\nu} = \xi$ is a factor defined from basic information in endemic countries. As discussed in Section 2, we use $\frac{\beta}{\gamma} = 1.2$ and $\frac{\vartheta}{\nu} = 1.2$, giving $\frac{\beta}{\gamma} \frac{\vartheta}{\nu} = 1.44 \approx 1.5$. These values can be adjusted based on updated information for different vector-borne diseases.



a)



b)

Figure 12: The largest 20 eggs counts in Basque Country in a) 2019, and b) 2022.

The expected number of autochthonous cases in the Basque Country, based on mosquito abundance in 2019, can be estimated as $\langle I \rangle^* = 0.0671$, and in 2022 as $\langle I \rangle^* = 0.5602$, suggesting a potential rise in the number of cases. Although these estimates may indicate fewer than one person affected, stochastic simulations suggest that clusters of infected cases could emerge in the coming years.

Using this approach, the risk can be calculated on a smaller spatial scale, such as at the provincial and municipal levels, as more data become available. Additionally, risk maps with finer spatial resolution, such as at the municipality level, can be generated. These maps would consider imported cases, mosquito abundance, and the combined risk of autochthonous cases, providing a more detailed risk assessment.

5 Discussion and conclusion

Driven by climate change and increased global travel, mosquito-borne diseases like dengue, Zika, and chikungunya, once confined to subtropical and tropical regions, are now being reported in non-endemic areas. Although most cases originate from infected travelers - referred to as imported cases - they can trigger unexpected autochthonous outbreaks in regions with competent mosquito vectors. This has been evident recently, with non-travel-related dengue cases reported in France and Italy, as noted by the European Center for Disease Prevention and Control [11].

In endemic areas, mathematical models have been developed to estimate disease outbreak risks, with the Basic Reproduction Number (R_0) being widely used to assess the likelihood and intensity of outbreaks occurring throughout the year. This metric, which represents the average number of secondary cases produced by an infected individual in a completely susceptible population, is crucial for understanding the dynamics of disease spread in regions where transmission is ongoing. However, in non-endemic areas where no local cases have been recorded, R_0 is typically less than 1 ($R_0 < 1$), indicating that the disease is not currently spreading within the local population. As a result, R_0 is not a useful measure for assessing outbreak risks in these regions. Instead, alternative models and metrics must be used to account for the role of imported cases and the presence of competent mosquito vectors, which can introduce and potentially amplify disease transmission in previously unaffected areas.

In this study, a dynamic vector-host model is developed to assess the risk of autochthonous cases driven by imported infections in non-endemic regions. Using stochastic methods, the model analyzes waiting times for the first local cases and inter-event statistics of imported cases, with a specific application to the Basque Country. This approach incorporates imported cases and the presence of mosquito populations, relative to those in endemic regions, to estimate the risk of mosquito-borne disease outbreaks.

Stochastic simulations reveal that even in non-endemic regions with moderate mosquito abundance - such as 40% of that found in endemic areas - there is a potential for isolated autochthonous cases driven by imported infections, as well as the possibility of outbreak clusters. These simulations indicate that the dynamics of mosquito-borne disease transmission can be significantly influenced by the presence and density of mosquito populations, even at lower levels. By estimating the expected number of infected cases and assessing the risk of autochthonous infections based on imported case data and mosquito abundance, the model provides a robust tool for predicting

and managing mosquito-borne disease risks in non-endemic areas. This approach allows public health authorities to anticipate potential outbreaks and implement targeted interventions to mitigate the risk of disease spread.

By considering the Basque Country - a region in Spain with documented imported cases but no local cases to date - as a case study, we apply our method to assess the risk of an autochthonous disease outbreak. Essential data for this analysis include mosquito egg counts recorded by NEIKER across various localities and epidemiological information on viremic imported cases provided by the Epidemiology Unit of the Public Health Department of the Basque Country.

Using a Poisson process, we analytically computed the trend of imported cases and applied this method to data from the Basque Country for 2019 and 2022. We also calculated the confidence intervals for the cumulative curve of imported cases via likelihood functions. It is important to note that for the years 2020 and 2021, very few or no imported cases were recorded due to travel restrictions imposed by the COVID-19 pandemic, resulting in insufficient data to apply our method effectively. Our analysis demonstrates that mosquito population abundance significantly impacts the likelihood and timing of autochthonous cases. This explains why autochthonous cases have been identified in countries like France and Italy but not yet in other European regions, such as the Basque Country. The findings underscore the critical importance of considering non-linear dynamics in disease transmission models, particularly near critical thresholds of mosquito abundance, and call for continued research into seasonality and other factors influencing mosquito abundance.

By coupling mosquito abundance calculations with the estimated mean number of imported cases in 2019, we apply the risk estimator developed through the stochastic framework. The expected number of autochthonous cases in the Basque Country rose from $\langle I \rangle = 0.0671$ in 2019 to $\langle I \rangle = 0.5602$ in 2022, and this increasing trend is anticipated to continue in the coming years. Although these estimates suggest fewer than one case on average, stochastic simulations indicate that clusters of infected cases could emerge in the coming years.

Our approach allows for risk calculation on a smaller spatial scale, such as at provincial and municipal levels, as more data becomes available. The spatial distribution of imported cases versus the measurements points of ovitraps pose some challenges on spatial correlation of the risks, however it facilitates the creation of risk maps with finer spatial resolution, such as at the municipality level, which will be investigated in future research, providing a more detailed risk assessment for public health authorities.

Acknowledgments

This work is also supported by the ARBOSKADI project for monitoring vector-borne diseases in the Basque Country, Euskadi. We wish to extend our acknowledgments to Jesús Ángel Ocio Armentia, Oscar Goñi Laguardia and Ana Ramírez de La Peciña Pérez, Dirección de Salud Pública for their fruitful discussions, and to Madalen Oribe Amores, Unidad de Vigilancia Epidemiológica de Bizkaia, for her cooperation in providing the requested epidemiological data that was essential for carrying out this research.

M.A. acknowledges the financial support by the Ministerio de Ciência e Innovacion (MICINN) of the Spanish Government through the Ramon y Cajal grant RYC2021-031380-I. This research is also supported by the Basque Government through the "Mathematical Modeling Applied to Health" (BMTF) Project, BERC 2022-2025 program and by the Spanish Ministry of Sciences, Innovation and Universities: BCAM Severo Ochoa accreditation CEX2021-001142-S / MICIN / AEI / 10.13039/501100011033.

References

[1] Paz S. Climate change: A driver of increasing vector-borne disease transmission in non-endemic areas. 2024 PLOS Medicine;21(4):e1004382. Available from: https://dx.plos.org/10.1371/journal.pmed.1004382.

[2] Communicable disease threats report 10-16 August 2024 week 33. European Centre for Disease Prevention and Control; 2024. Available from: https://www.ecdc.europa.eu/en/publications-data/communicable-disease-threats-report-10-16-august-2024-week-33.

[3] Dengue - Global situation. World Health Organization; 2024. Available from: https://www.who.int/emergencies/disease-outbreak-news/item/2024-DON518.

[4] Aedes invasive mosquitoes - current known distribution: July 2024. European Centre for Disease Prevention and Control; 2024. Available from: https://www.ecdc.europa.eu/en/publications-data/aedes-invasive-mosquitoes-current-known-distribution-july-2024.

[5] Laporta GZ, Potter AM, Oliveira JFA, et al. Global distribution of Aedes aegypti and Aedes albopictus in a climate change scenario of regional rivalry. 2023 Insects;14(1):49. Available from: https://www.mdpi.com/2075-4450/14/1/49.

[6] Colón-González FJ, Sewe MO, Tompkins AM, et al. Projecting the risk of mosquito-borne diseases in a warmer and more populated world: A multi-model,

multi-scenario intercomparison modelling study. 2021 The Lancet Planetary Health;5(7):e404-14. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2542519621001327.

[7] Saldaña F, Stollenwerk N, Van Dierdonck JB, et al. Modeling spillover dynamics: Understanding emerging pathogens of public health concern. 2024 Scientific Reports;14(1):9823. Available from: https://www.nature.com/articles/s41598-024-60661-y.

[8] Young MJ, Fefferman NH. The dynamics of disease mediated invasions by hosts with immune reproductive tradeoff. 2022 Scientific Reports;12(1):4108. Available from: https://www.nature.com/articles/s41598-022-07962-2.

[9] Zardini A, Menegale F, Gobbi A, et al. Estimating the potential risk of transmission of arboviruses in the Americas and Europe: A modelling study. 2024 The Lancet Planetary Health;8(1):e30-40. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2542519623002528.

[10] Dengue outbreak in Madeira, Portugal, October – November 2012. European Centre for Disease Prevention and Control; 2013. Available from: https://www.ecdc.europa.eu/en/publications-data/dengue-outbreak-madeira-portugal-october-november-2012.

[11] Local transmission of dengue virus in mainland EU/EEA, 2010-present. European Centre for Disease Prevention and Control; 2024. Available from: https://www.ecdc.europa.eu/en/all-topics-z/dengue/surveillance-and-disease-data/autochthonous-transmission-dengue-virus-eueea.

[12] Fernández-Martínez B, Pampaka D, Suárez-Sánchez P, et al. Spatial analysis for risk assessment of dengue in Spain. 2023 Enfermedades Infecciosas y Microbiología Clínica:S0213005X23001830. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0213005X23001830.

[13] Gossner CM, Ducheyne E, Schaffner F. Increased risk for autochthonous vector-borne infections transmitted by Aedes albopictus in continental Europe. 2018 Eurosurveillance;23(24). Available from: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2018.23.24.1800268.

[14] Massad E, Amaku M, Coutinho FAB, et al. Estimating the probability of dengue virus introduction and secondary autochthonous cases in Europe. 2018 Scientific Reports;8(1):4629. Available from: https://www.nature.com/articles/s41598-018-22590-5.

[15] Sohail A, Anders KL, McGuinness SL, et al. The epidemiology of imported and locally acquired dengue in Australia, 2012–2022. 2024 Journal of Travel Medicine;31(2):taae014. Available from: https://academic.oup.com/jtm/article/doi/10.1093/jtm/taae014/7577676.

[16] Blasius B. Power-law distribution in the number of confirmed COVID-19 cases. 2020 Chaos: An Interdisciplinary Journal of Nonlinear Science;30(9):093123. Available from: https://pubs.aip.org/cha/article/30/9/093123/341972/Power-law-distribution-in-the-number-of-confirmed.

[17] Esquivel-Gómez JJ, Barajas-Ramírez JG. Rapid disease spread on dense networks with power-law topology. 2024 Eur Phys J B;97(53). Available from: https://doi.org/10.1140/epjb/s10051-024-00675-7.

[18] Meyer S, Held LH. Power-law models for infectious disease spread. 2014 The Annals of Applied Statistics;8(3):1612-39. Available from: http://www.jstor.org/stable/24522277.

[19] Stollenwerk N, Jansen V. Population Biology and Criticality: From Critical Birth-Death Processes to Self-Organized Criticality in Mutation Pathogen Systems. IMPERIAL COLLEGE PRESS; 2010. Available from: https://www.worldscientific.com/worldscibooks/10.1142/p645.

[20] Cai L, Li X. Analysis of a simple vector-host epidemic model with direct transmission. 2010 Discrete Dynamics in Nature and Society;2010(1):679613. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1155/2010/679613.

[21] Esteva L, Vargas C. Analysis of a dengue disease transmission model. 1998 Mathematical Biosciences;131-151(2):220–240. Available from: https://doi.org/10.1016/S0025-5564(98)10003-2.

[22] Martcheva M. In: Vector-borne diseases. Boston, MA: Springer US; 2015. p. 67-89. Available from: https://doi.org/10.1007/978-1-4899-7612-3_4.

[23] Pandey A, Mubayi A, Medlock J. Comparing vector–host and SIR models for dengue transmission. 2013 Mathematical Biosciences;246(2):252-9. Available from: https://www.sciencedirect.com/science/article/pii/S0025556413002435.

[24] Ross R. An application of the theory of probabilities to the study of a priori pathometry.—Part I. 1916 Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character;92(638):204-30. Available from: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1916.0007.

[25] Rashkov P, Venturino E, Aguiar M, et al. On the role of vector modeling in a minimalistic epidemic model. 2019 Mathematical Biosciences and Engineering;16(5):4314-38. Available from: http://www.aimspress.com/article/10.3934/mbe.2019215.

[26] Rocha F, Aguiar M, Souza M, et al. Time-scale separation and centre manifold analysis describing vector-borne disease dynamics. 2013 International Journal of Computer Mathematics;90(10):2105-25. Available from: http://www.tandfonline.com/doi/abs/10.1080/00207160.2013.783208.

[27] Aguiar M, Kooi BW, Martins J, et al. Scaling of stochasticity in dengue hemorrhagic fever epidemics. 2012 Mathematical Modelling of Natural Phenomena;7(3):1-11. Available from: http://eudml.org/doc/222190.

[28] Aguiar M, Ballesteros S, Kooi BW, et al. The role of seasonality and import in a minimalistic multi-strain dengue model capturing differences between primary and secondary infections: Complex dynamics and its implications for data analysis. 2011 Journal of Theoretical Biology;289:181-96. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0022519311004462.

[29] Aguiar M, Kooi B, Pugliese A, et al.. Time scale separation in the vector borne disease model SIRUV via center manifold analysis; 2021. Available from: http://medrxiv.org/lookup/doi/10.1101/2021.04.06.21254992.

[30] Kermack WO, McKendrick AG, Walker GT. A contribution to the mathematical theory of epidemics. 1927 Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character;115(772):700-21. Available from: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1927.0118.

[31] Esteva L, Vargas C. A model for dengue disease with variable human population. 1999 J Math Biol;38:220–240. Available from: https://doi.org/10.1007/s002850050147.

[32] Heidrich P, Jayathunga Y, Bock W, et al. Prediction of dengue cases based on human mobility and seasonality—An example for the city of Jakarta. 2021 Mathematical Methods in the Applied Sciences;44(17):13633-58. Available from: https://onlinelibrary.wiley.com/doi/10.1002/mma.7648.

[33] Rocha F, Mateus L, Skwara U, et al. Understanding dengue fever dynamics: A study of seasonality in vector-borne disease models. 2016 International Journal of Computer Mathematics;93(8):1405-22. Available from: http://www.tandfonline.com/doi/full/10.1080/00207160.2015.1050961.

[34] Skwara U, Mozyrska D, Aguiar M, et al. Applications of fractional calculus to mathematical modelling of vector-borne diseases. 2021 Available at SSRN. Available from: http://dx.doi.org/10.2139/ssrn.4610210.

[35] Skwara U, Mozyrska D, Aguiar M, et al. Dynamics of vector-borne diseases through the lens of systems incorporating fractional-order derivatives. 2024 Chaos, Solitons & Fractals;181:114643. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0960077924001942.

[36] Guidelines for the surveillance of invasive mosquitoes in Europe. European Centre for Disease Prevention and Control; 2012. Available from: https://www.ecdc.europa.eu/sites/default/files/media/en/publications/Publications/TER-Mosquito-surveillance-guidelines.pdf.

[37] Fung T, Clapham HE, Chisholm RA. Temporary Cross-Immunity as a Plausible Driver of Asynchronous Cycles of Dengue Serotypes. 2023 Bulletin of Mathematical Biology;85(12):124. Available from: https://link.springer.com/10.1007/s11538-023-01226-4.

[38] Sabin AB. Research on dengue during World War II. 1952 The American Journal of Tropical Medicine and Hygiene;1:30-50.

[39] Snow GE, Haaland B, Ooi EE, et al. Research on Dengue During World War II Revisited. 2014 The American Society of Tropical Medicine and Hygiene;91(6):1203-17. Available from: https://www.ajtmh.org/view/journals/tpmd/91/6/article-p1203.xml.

[40] Waggoner JJ, Balmaseda A, Gresh L, et al. Homotypic dengue virus reinfections in nicaraguan children. 2016 Journal of Infectious Diseases;214(7):986-93. Available from: https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/jiw099.

[41] Focks DA. A review of entomological sampling methods and indicators for dengue vectors. World Health Organization on behalf of the Special Programme for Research and Training in Tropical Diseases; 2023. Available from: https://iris.who.int/bitstream/handle/10665/68575/?sequence=1.

[42] Oliveira Noleto JV, Moura do Nascimento Moraes HL, De Moura Lima T, et al. Use of ovitraps for the seasonal and spatial monitoring of Aedes spp. in an area endemic for arboviruses in Northeast Brazil. 2020 The Journal of Infection in Developing Countries;14(04):387-93. Available from: https://jidc.org/index.php/journal/article/view/12245.

[43] Sasmita HI, Neoh KB, Yusmalinar S, et al. Ovitrap surveillance of dengue vector mosquitoes in Bandung City, West Java Province, Indonesia. 2021 PLOS Neglected Tropical Diseases;15(10):e0009896. Available from: https://dx.plos.org/10.1371/journal.pntd.0009896.

[44] Mateus L, Ghaffari P, Skwara U, et al. Semiclassical approximations of stochastic epidemiological processes towards parameter estimation using as prime example the SIS system with import. 2016 Ecological Complexity;27:63-73. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1476945X15000963.

[45] Gillespie DT. Stochastic Simulation of Chemical Kinetics. 2007 Annual Review of Physical Chemistry;58(1):35-55. Available from: https://www.annualreviews.org/doi/10.1146/annurev.physchem.58.032806.104637.

[46] Stollenwerk N, Sommer PF, Kooi B, et al. Hopf and torus bifurcations, torus destruction and chaos in population biology. 2017 Ecological Complexity;30:91-9. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1476945X16301143.

[47] Mateus L, Stollenwerk N, Zambrini JC. Stochastic models in population biology: From dynamic noise to Bayesian description and model comparison for given data sets. 2013 International Journal of Computer Mathematics;90(10):2161-73. Available from: http://www.tandfonline.com/doi/abs/10.1080/00207160.2013.792924.

[48] Lübeck S. Universal scaling behavior of non-equilibrium phase transitions. 2004 International Journal of Modern Physics B;18(31n32):3977-4118. Available from: https://www.worldscientific.com/doi/abs/10.1142/S0217979204027748.

[49] Cevidanes A, Goiri F, Barandika JF, et al. Invasive Aedes mosquitoes in an urban—peri-urban gradient in northern Spain: Evidence of the wide distribution of Aedes japonicus. 2023 Parasites & Vectors;16(1):234. Available from: https://parasitesandvectors.biomedcentral.com/articles/10.1186/s13071-023-05862-6.

[50] Basawa IV, Prakasa Rao BLS. Statistical Inference for Stochastic Processes. Probability and Mathematical Statistics. London New York: Academic Press; 1980.

[51] Merberg A, Miller SJ. Course Notes for Math 162: Mathematical Statistics The Cramér-Rao Inequality; 2008. Available from: https://web.williams.edu/Mathematics/sjmiller/public_html/BrownClasses/162/Handouts/CramerRaoHandout08.pdf.

Appendices:

A Analysis of mean and variance via hypothetical ensembles of data

In this section, we summarize both the frequentist and Bayesian approaches to parameter estimation, along with their theoretical foundations.

In the previous analysis, we constructed the maximum likelihood estimator (MLE) and assessed its variance using frequentist methods. The variance of the MLE was estimated as $\sigma^2 = \frac{\hat{\lambda}^2}{n}$, which was derived from the curvature of the likelihood function.

To theoretically justify this result, we consider the likelihood function $L(\lambda) = p(\underline{\tau} \mid \lambda)$, which describes the probability of observing data ensembles given a particular model. The Gaussian approximation of the likelihood function provides a good fit for the distribution of estimators $p(\hat{\lambda} \mid \lambda)$. This fit can be validated through simulations that evaluate multiple data sets and examine the resulting histogram of estimators $\hat{\lambda}_i$. Although an exact analytical solution for the ensemble variance $\langle(\hat{\lambda} - \langle\hat{\lambda}\rangle)^2\rangle$ is not feasible, it can be approximated using the variance expression obtained from the negative inverse Fisher matrix [50]

$$\sigma^2 = \frac{\hat{\lambda}^2}{n} \quad .$$

This expression is consistent with the Cramér-Rao inequality [51], which provides a lower bound for the variance of λ given by λ^2/n. The Cramér-Rao inequality involves the true but unknown parameter λ, while our heuristic method uses the estimator $\hat{\lambda}$. The complete form of the inequality also accounts for any bias in the estimator, resulting in

$$\langle(\hat{\lambda} - \langle\hat{\lambda}\rangle)^2\rangle \geq \frac{\lambda^2}{n}\left(\frac{n}{n-1}\right)^2 \quad ,$$

where the impact of bias diminishes quickly. Detailed calculations related to the Fisher information and the Cramér-Rao inequality can be found in Appendix B.

The likelihood function $L(\lambda) = p(\underline{\tau}_D \mid \lambda)$ from a single empirical data set $\underline{\tau}_D$ approximates the functional form of the distribution $p(\hat{\lambda} \mid \lambda)$ obtained by marginalizing over all possible hypothetical data sets $\underline{\tau}$, even if the estimator $\hat{\lambda}$ deviates from the original parameter λ.

In contrast, Bayesian analysis updates prior beliefs about the parameter λ using the empirical likelihood function from a single data set. The prior distribution $p(\lambda)$, specified before observing the data, is updated to form the posterior distribution $p(\lambda \mid \underline{\tau}_D)$. This posterior provides estimates and confidence intervals and can be applied to the empirical data of imported cases in the Basque Country, as demonstrated in the analysis for the Basque Country. Due to the very limited data, the maximum likelihood method of frequentists provides reasonable information, while the more elaborate Bayesian approach might not refine the results significantly in the present study. Hence, we focus on the frequentist framework here.

B Calculation of the lower bound of the ensemble variance of maximum likelihood estimators

B.1 Conditioned ensemble averages, the starting point of the analysis, to the variance inequality

Conditioned ensemble averages serve as the starting point for analyzing the variance inequality. For the ensemble mean of estimators, we have

$$\langle \hat{\lambda} \rangle = \int_0^\infty \hat{\lambda}\, p(\hat{\lambda}|\lambda)\, d\hat{\lambda} \quad , \tag{68}$$

where the probability $p(\hat{\lambda} \mid \lambda)$ of estimators $\hat{\lambda}$, given a stochastic model with parameter λ, can be expressed by marginalizing over ensembles of data sets $\underline{\tau}$ obtained from the likelihood function

$$L(\lambda) = p(\underline{\tau} \mid \lambda) = \lambda^n\, e^{-\lambda \sum_{i=1}^n \tau_i} \quad .$$

The likelihood function L gives the probability for the data given an underlying model with parameter λ. These data determine the estimators $\hat{\lambda} = f(\underline{\tau})$ via the probability

$$p(\hat{\lambda} \mid \underline{\tau}) = \delta(\hat{\lambda} - f(\underline{\tau})) \quad ,$$

where $\hat{\lambda} = f(\underline{\tau}) = \frac{1}{\frac{1}{n}\sum_{i=1}^n \tau_i}$, with the analytic functional form of the maximum likelihood estimator, giving

$$p(\hat{\lambda} \mid \lambda) = \int_0^\infty \cdots \int_0^\infty p(\hat{\lambda} \mid \underline{\tau}) \cdot p(\underline{\tau} \mid \lambda)\, d^n\underline{\tau} \quad , \tag{69}$$

which essentially marginalizes via

$$p(\hat{\lambda} \mid \underline{\tau}) \cdot p(\underline{\tau} \mid \lambda) = p(\hat{\lambda}, \underline{\tau} \mid \lambda)$$

the joint probability $p(\hat{\lambda}, \underline{\tau} \mid \lambda)$ over $\underline{\tau}$. Hence, we have

$$\begin{aligned} \langle \hat{\lambda} \rangle &= \int \int \cdots \int \hat{\lambda}\, p(\hat{\lambda} \mid \underline{\tau}) \cdot p(\underline{\tau} \mid \lambda)\, d^n\underline{\tau}\, d\hat{\lambda} \\ &= \int \cdots \int \int \hat{\lambda}\, \delta(\hat{\lambda} - f(\underline{\tau}))\, d\hat{\lambda}\, p(\underline{\tau} \mid \lambda)\, d^n\underline{\tau} \\ &= \int \cdots \int f(\underline{\tau})\, p(\underline{\tau} \mid \lambda)\, d^n\underline{\tau} \quad , \end{aligned} \tag{70}$$

which gives, for the exponential distribution, after some calculations, the explicit result

$$\langle \hat{\lambda} \rangle = \frac{n}{n-1}\lambda \quad ,$$

essentially via deriving and solving the ordinary differential equation

$$\frac{d}{d\lambda}\langle \hat{\lambda} \rangle = n\frac{\langle \hat{\lambda} \rangle}{\lambda} - n \quad ,$$

via the variable transformation $z := \langle \hat{\lambda} \rangle / \lambda$, hence

$$\langle \hat{\lambda} \rangle = z \cdot \lambda = \langle \hat{\lambda} \rangle(z, \lambda) \quad .$$

Then, the bias of the estimator is defined as the deviation of the ensemble mean of the estimator $\langle \hat{\lambda} \rangle$ and the original parameter λ, that is

$$b(\lambda) := \langle \hat{\lambda} \rangle - \lambda \quad ,$$

which is here in the case of the exponential distribution $b(\lambda) = \frac{1}{n-1}\lambda$, converging quickly to zero with increasing number n of data points.

An analogous calculation for the variance of the estimator gives

$$\begin{aligned} \langle (\hat{\lambda} - \langle \hat{\lambda} \rangle)^2 \rangle &= \int \int \ldots \int (\hat{\lambda} - \langle \hat{\lambda} \rangle)^2\, p(\hat{\lambda}|\underline{\tau}) \cdot p(\underline{\tau}|\lambda)\, d^n\underline{\tau} \quad d\hat{\lambda} \\ &= \int \ldots \int \int (\hat{\lambda} - \langle \hat{\lambda} \rangle)^2\, \delta(\hat{\lambda} - f(\underline{\tau}))\, d\hat{\lambda} \quad p(\underline{\tau}|\lambda)\, d^n\underline{\tau} \\ &= \int \ldots \int (f(\underline{\tau}) - \langle \hat{\lambda} \rangle)^2\, p(\underline{\tau}|\lambda)\, d^n\underline{\tau} \\ &= \int \ldots \int (f(\underline{\tau})^2 - \langle \hat{\lambda} \rangle^2)\, p(\underline{\tau}|\lambda)\, d^n\underline{\tau} \quad , \end{aligned} \tag{71}$$

where in the last step we used in expectation $\langle(\hat{\lambda}-\langle\hat{\lambda}\rangle)^2\rangle = \langle\hat{\lambda}\rangle^2 - \langle\hat{\lambda}\rangle^2$. This can only be treated via an inequality giving a lower bound of the estimator variance.

In order to apply the Cauchy-Schwarz inequality, (which is in ordinary vector spaces relating the scalar product of two vectors to the Euclidian norm of the vectors as $\underline{g}\cdot\underline{h} \leq ||\underline{g}||\cdot||\underline{h}||$), we first analyze

$$\begin{aligned}
\frac{d}{d\lambda}\langle\hat{\lambda}\rangle &= \int ... \int f(\underline{\tau})\ \frac{d}{d\lambda}p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \\
&= \int ... \int f(\underline{\tau})\ \left(\frac{d}{d\lambda} ln\ p(\underline{\tau}|\lambda)\right) p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \qquad (72)\\
&= \int ... \int f(\underline{\tau})\sqrt{p(\underline{\tau}|\lambda)}\cdot\left(\frac{d}{d\lambda} ln\ p(\underline{\tau}|\lambda)\right)\sqrt{p(\underline{\tau}|\lambda)}\ d^n\underline{\tau} \quad ,
\end{aligned}$$

where in the last step the integrand is written as a product of two functions $g(\underline{\tau}) := f(\underline{\tau})\sqrt{p(\underline{\tau}|\lambda)}$ and $h(\underline{\tau}) := \left(\frac{d}{d\lambda} ln\ p(\underline{\tau}|\lambda)\right)\sqrt{p(\underline{\tau}|\lambda)}$ to give in the Cauchy-Schwarz inequality, now for integrals ,the treatable expressions in form of expectation integrals.

After we considered $\frac{d}{d\lambda}\langle\hat{\lambda}\rangle$ evaluating the definition of $\langle\hat{\lambda}\rangle$ we now evaluate from the definition of the bias $b(\lambda) := \langle\hat{\lambda}\rangle - \lambda$, hence $\langle\hat{\lambda}\rangle = \lambda + b(\lambda)$ giving

$$\frac{d}{d\lambda}\langle\hat{\lambda}\rangle = 1 + \frac{d}{d\lambda}b(\lambda) \quad . \qquad (73)$$

Applying, thus, the Cauchy-Schwarz inequality replacing the sums by integrals and the finite dimensional vectors by functions, we get

$$\begin{aligned}
\left(\frac{d}{d\lambda}\langle\hat{\lambda}\rangle\right)^2 &= \left(1 + \frac{d}{d\lambda}b(\lambda)\right)^2 \qquad (74)\\
&= \left(\int ... \int f(\underline{\tau})\sqrt{p(\underline{\tau}|\lambda)}\cdot\left(\frac{d}{d\lambda} ln\ p(\underline{\tau}|\lambda)\right)\sqrt{p(\underline{\tau}|\lambda)}\ d^n\underline{\tau}\right)^2 \\
&= \left(\int ... \int g(\tau)\cdot h(\tau)\ d^n\underline{\tau}\right)^2 \quad ,
\end{aligned}$$

and applying the Cauchy-Schwarz inequality

$$\begin{aligned}\left(1+\frac{d}{d\lambda}b(\lambda)\right)^2 &= \left(\int ... \int g(\tau)\cdot h(\tau)\; d^n\underline{\tau}\right)^2 \\ &\leq \int ... \int g^2(\tau)\; d^n\underline{\tau} \quad \cdot \quad \int ... \int h^2(\tau)\; d^n\underline{\tau} \\ &= \int ... \int f^2(\underline{\tau})\; p(\underline{\tau}|\lambda)\; d^n\underline{\tau} \\ &\qquad \cdot \quad \int ... \int \left(\frac{d}{d\lambda} ln\; p(\underline{\tau}|\lambda)\right)^2 p(\underline{\tau}|\lambda)\; d^n\underline{\tau} \quad ,\end{aligned} \tag{75}$$

and hence, the inequality

$$\int ... \int f^2(\underline{\tau})\; p(\underline{\tau}|\lambda)\; d^n\underline{\tau} \quad \geq \quad \frac{\left(1+\frac{d}{d\lambda}b(\lambda)\right)^2}{\int ... \int \left(\frac{d}{d\lambda} ln\; p(\underline{\tau}|\lambda)\right)^2 p(\underline{\tau}|\lambda)\; d^n\underline{\tau}} \tag{76}$$

holds, which only needs little refinement for the final result.

For this, we consider now the normalization of the probability of the likelihood function $\int ... \int p(\underline{\tau}|\lambda)\; d^n\underline{\tau} = 1$ and its first two derivatives in respect to the parameter λ. We have

$$\frac{d}{d\lambda}\int ... \int p(\underline{\tau}|\lambda)\; d^n\underline{\tau} \quad = \quad \frac{d}{d\lambda}\, 1 = 0 \quad , \tag{77}$$

and hence

$$0 = \int ... \int \frac{d}{d\lambda}\; p(\underline{\tau}|\lambda)\; d^n\underline{\tau} = \int ... \int \left(\frac{d}{d\lambda}\; ln\; p(\underline{\tau}|\lambda)\right)\; p(\underline{\tau}|\lambda)\; d^n\underline{\tau} \quad . \tag{78}$$

The zero can be multiplied by any function to still be zero, and any function not depending on the data $\underline{\tau}$ can be taken into the integrals. Hence, for the function $\langle\hat{\lambda}\rangle$ we obtain

$$0 = \langle\hat{\lambda}\rangle \int ... \int \frac{d}{d\lambda}\; p(\underline{\tau}|\lambda)\; d^n\underline{\tau} = \int ... \int \langle\hat{\lambda}\rangle \frac{d}{d\lambda}\; p(\underline{\tau}|\lambda)\; d^n\underline{\tau} \quad , \tag{79}$$

and with this we can rewrite Eq. (72) as follows

$$\begin{aligned}
\frac{d}{d\lambda}\langle\hat{\lambda}\rangle &= \int ... \int f(\underline{\tau})\ \frac{d}{d\lambda}p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \quad - \quad 0 \\
&= \int ... \int f(\underline{\tau})\ \frac{d}{d\lambda}p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \quad - \quad \int ... \int \langle\hat{\lambda}\rangle\ \frac{d}{d\lambda}p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \\
&= \int ... \int (f(\underline{\tau}) - \langle\hat{\lambda}\rangle)\ \frac{d}{d\lambda}p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \\
&= \int ... \int (f(\underline{\tau}) - \langle\hat{\lambda}\rangle)\ \left(\frac{d}{d\lambda} ln\ p(\underline{\tau}|\lambda)\right) p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \qquad (80) \\
&= \int ... \int (f(\underline{\tau}) - \langle\hat{\lambda}\rangle)\sqrt{p(\underline{\tau}|\lambda)} \cdot \left(\frac{d}{d\lambda} ln\ p(\underline{\tau}|\lambda)\right) \sqrt{p(\underline{\tau}|\lambda)}\ d^n\underline{\tau} \quad ,
\end{aligned}$$

and redefining $g(\underline{\tau}) := (f(\underline{\tau}) - \langle\hat{\lambda}\rangle)\sqrt{p(\underline{\tau}|\lambda)}$ allows us, in all subsequent calculations above, to replace $f(\underline{\tau})$ by $(f(\underline{\tau}) - \langle\hat{\lambda}\rangle)$. Hence, in Equation (76) we have now the complete variance on the left hand side of the inequality

$$\int ... \int (f(\underline{\tau}) - \langle\hat{\lambda}\rangle)^2\ p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \quad \geq \quad \frac{\left(1 + \frac{d}{d\lambda}b(\lambda)\right)^2}{\int ... \int \left(\frac{d}{d\lambda} ln\ p(\underline{\tau}|\lambda)\right)^2 p(\underline{\tau}|\lambda)\ d^n\underline{\tau}} \qquad (81)$$

as desired.

The final step is to take the second derivative of the normalization of the likelihood function given by

$$\begin{aligned}
0 &= \frac{d^2}{d\lambda^2} \int ... \int p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \\
&= \int ... \int \frac{d}{d\lambda}\frac{d}{d\lambda}p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \\
&= \int ... \int \frac{d}{d\lambda}\left(\left(\frac{d}{d\lambda} ln\ p(\underline{\tau}|\lambda)\right) p(\underline{\tau}|\lambda)\right)\ d^n\underline{\tau} \qquad (82) \\
&= \int ... \int \left(\frac{d^2}{d\lambda^2} ln\ p(\underline{\tau}|\lambda)\right) p(\underline{\tau}|\lambda) + \left(\frac{d}{d\lambda} ln\ p(\underline{\tau}|\lambda)\right)\left(\frac{d}{d\lambda} p(\underline{\tau}|\lambda)\right)\ d^n\underline{\tau} \\
&= \int ... \int \left(\frac{d^2}{d\lambda^2} ln\ p(\underline{\tau}|\lambda)\right) p(\underline{\tau}|\lambda) + \left(\frac{d}{d\lambda} ln\ p(\underline{\tau}|\lambda)\right)\left(\frac{d}{d\lambda} ln\ p(\underline{\tau}|\lambda)\right) p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \quad ,
\end{aligned}$$

and hence

$$0 = \int ... \int \left(\frac{d^2}{d\lambda^2} ln\ p(\underline{\tau}|\lambda)\right) p(\underline{\tau}|\lambda)\ d^n\underline{\tau} + \int ... \int \left(\frac{d}{d\lambda} ln\ p(\underline{\tau}|\lambda)\right)^2 p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \qquad (83)$$

or

$$\int ... \int \left(\frac{d}{d\lambda} ln\ p(\underline{\tau}|\lambda) \right)^2 p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \ = \ - \int ... \int \left(\frac{d^2}{d\lambda^2} ln\ p(\underline{\tau}|\lambda) \right) p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \tag{84}$$

with the Cramér-Rao inequality now in its final form given by

$$\int ... \int (f(\underline{\tau}) - \langle \hat{\lambda} \rangle)^2\ p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \quad \geq \quad \frac{\left(1 + \frac{d}{d\lambda} b(\lambda)\right)^2}{\int ... \int \left(- \frac{d^2}{d\lambda^2} ln\ p(\underline{\tau}|\lambda) \right) p(\underline{\tau}|\lambda)\ d^n\underline{\tau}} \tag{85}$$

with the Fisher information

$$\langle - \frac{d^2}{d\lambda^2} ln\ p(\underline{\tau}|\lambda) \rangle = \int ... \int \left(- \frac{d^2}{d\lambda^2} ln\ p(\underline{\tau}|\lambda) \right) p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \tag{86}$$

in its expectation value form in the denominator of the inequality and the variance

$$\langle \left(\hat{\lambda} - \langle \hat{\lambda} \rangle \right)^2 \rangle = \int ... \int (f(\underline{\tau}) - \langle \hat{\lambda} \rangle)^2\ p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \quad . \tag{87}$$

So we can write in short

$$\langle \left(\hat{\lambda} - \langle \hat{\lambda} \rangle \right)^2 \rangle \quad \geq \quad \frac{\left(1 + \frac{db}{d\lambda}\right)^2}{\langle - \frac{d^2}{d\lambda^2} ln\ L(\lambda) \rangle} \tag{88}$$

as the easiest memorizable form of the Cramer-Rao inequality, using the likelihood notation $L(\lambda) = p(\underline{\tau}|\lambda)$, since the ensemble average is always taken over the data points $\underline{\tau}$ now.

B.2 Application to the exponential distribution and its maximum likelihood estimator

We first calculate explicitly the ensemble mean of the estimator by deriving an ODE and solving it. From Eq. (70) we have with $L(\lambda) = p(\underline{\tau}|\lambda) = \lambda^n\ e^{-\lambda \sum_{i=1}^n \tau_i}$ and $f(\underline{\tau}) = \frac{1}{\frac{1}{n}\sum_{i=1}^n \tau_i}$ for the exponential distribution

$$\begin{aligned} \langle \hat{\lambda} \rangle \ &= \ \int ... \int f(\underline{\tau}) \cdot p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \\ &= \ \int ... \int \frac{1}{\frac{1}{n}\sum_{i=1}^n \tau_i} \cdot \lambda^n\ e^{-\lambda \sum_{i=1}^n \tau_i}\ d^n\underline{\tau} \end{aligned} \tag{89}$$

and hence, for its derivative

$$\begin{aligned}\frac{d}{d\lambda}\langle\hat{\lambda}\rangle &= \int ... \int f(\underline{\tau}) \cdot \frac{d}{d\lambda} p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \\ &= \int ... \int \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\tau_i} \cdot \left(n\lambda^{n-1} \cdot e^{-\lambda\sum_{i=1} n\tau_i} + \lambda^n \cdot (-\sum_{i=1}^{n}\tau_i) e^{-\lambda\sum_{i=1}^{n}\tau_i} \right) d^n\underline{\tau} \\ &= \frac{n}{\lambda}\int ... \int f(\underline{\tau}) \cdot p(\underline{\tau}|\lambda)\ d^n\underline{\tau} - n\int ... \int p(\underline{\tau}|\lambda)\ d^n\underline{\tau} \\ &= \frac{n}{\lambda}\langle\hat{\lambda}\rangle - n \quad ,\end{aligned} \tag{90}$$

since the first integral $\int ... \int f(\underline{\tau}) \cdot p(\underline{\tau}|\lambda)\ d^n\underline{\tau} = \langle\hat{\lambda}\rangle$ gives the expectation value, and the second $\int ... \int p(\underline{\tau}|\lambda)\ d^n\underline{\tau} = 1$ gives the normalization of probability. Thus, we have the ODE for $\langle\hat{\lambda}\rangle$ given by

$$\frac{d}{d\lambda}\langle\hat{\lambda}\rangle = n\ \frac{\langle\hat{\lambda}\rangle}{\lambda} - n \tag{91}$$

to be solved. This can be done via the variable transformation $z := \langle\hat{\lambda}\rangle/\lambda$, hence $\langle\hat{\lambda}\rangle = z \cdot \lambda = \langle\hat{\lambda}\rangle(z, \lambda)$. We transform the original ODE into an ODE in new variable

$$\frac{d}{d\lambda}\langle\hat{\lambda}\rangle = \frac{d\langle\hat{\lambda}\rangle}{dz}\frac{dz}{d\lambda} + \frac{d\langle\hat{\lambda}\rangle}{d\lambda}\frac{d\lambda}{d\lambda} = \lambda\frac{dz}{d\lambda} + z \tag{92}$$

and on the other side of the original ODE

$$\frac{d}{d\lambda}\langle\hat{\lambda}\rangle = nz - n \tag{93}$$

to be solved

$$\lambda\frac{dz}{d\lambda} + z = nz - n \tag{94}$$

or

$$\frac{dz}{d\lambda} = \frac{1}{\lambda}(n-1)(z - \frac{n}{n-1}) \tag{95}$$

which can now be resolved easily via separation of variables. Then,

$$\int \frac{1}{z - \frac{n}{n-1}}\ dz = (n-1)\int \frac{1}{\lambda}\ d\lambda \quad , \tag{96}$$

giving as general solution

$$\ln\left(z - \frac{n}{n-1}\right) + c_z = (n-1)(\ln\lambda + c_\lambda) \tag{97}$$

and after some elementary steps, and replacing z with $\langle\hat{\lambda}\rangle/\lambda$ again, we have

$$\langle\hat{\lambda}\rangle = \frac{n}{n-1}\lambda + c\cdot\lambda^n \tag{98}$$

using integration constants c_z and c_λ to be combined to $c = e^{(n-1)c_\lambda - c_z}$. To leading order in λ we have the solution $\langle\hat{\lambda}\rangle = \frac{n}{n-1}\lambda$, since λ is in general a small number.

Hence, for the bias $b(\lambda) := \langle\hat{\lambda}\rangle - \lambda$ we have

$$b(\lambda) = \frac{1}{n-1}\lambda \tag{99}$$

and in the expression for the Cramér-Rao bound of the ensemble variance

$$\langle\left(\hat{\lambda} - \langle\hat{\lambda}\rangle\right)^2\rangle \quad \geq \quad \frac{\left(1+\frac{db}{d\lambda}\right)^2}{\langle -\frac{d^2}{d\lambda^2} ln\ L(\lambda)\rangle} = \frac{\lambda^2}{n}\cdot\left(\frac{n}{n-1}\right)^2 \tag{100}$$

since the expectation of the negative second derivative of the log-likelihood function is $-\frac{d^2}{d\lambda^2} ln\ L(\lambda) = \frac{n}{\lambda^2}$ independent of the data τ_i, as calculated previously. Therefore, the expectation $\langle -\frac{d^2}{d\lambda^2} ln\ L(\lambda)\rangle = \frac{n}{\lambda^2}$, also know as the Fisher information).

In conclusion, the analysis of the maximum and its vicinity of the likelihood using the empirical data gives the same information as the ensemble analysis of mean and variance of the estimator, just with small extra terms which vanish for large data sets. Already for 10 data points, (for instance using the Gaussian approximation), compared with $p(\hat{\lambda}|\lambda)$ it can be quite good, and gives qualitatively the correct result for confidence interval analysis (see for another example [47] in Figure 2 b)), using extended simulations for the conditioned probability of estimators from the original parameter.